# "الإحصاء الوصفى "

**لطلاب الفرقة الثانية**

**إعداد وتأليف**

**د. محمد عبدالحميد الربع**

**كلية التجارة**

**قسم الأساليب الكمية**

**العام الجامعى**

٢٠٢ ـ ٢٠٢٣

**مقدمة**

**مما لا شك فيه أن علم الإحصاء أصبح اليوم ركيزة أساسية فى جميع فروع المعرفة . فلقد بات هذا العلم لازمًا في مختلف مجالات العلوم الاجتماعية والاقتصادية والهندسية والطبيعية .**
**هذا ولقد تطور علم الإحصاء تطورًا كبيرًا في الآونة الأخيرة حيث تغيرت صورته القديمة والعالقة في أذهان الناس على أنه علم العد وجمع البيانات وعرضها بيانيًا إلى كونه الآن علم رسم السياسات واتخاذ القرارات في شتى مجالات المعرفة على أساس علمي سليم رغم عدم توافر المعلومات الكافية فاتسع لذلك نطاق استخدامه وازداد الاهتمام بأساسياته وأساليبة التحليلية . وأصبح لعلم الإحصاء في الوقت الحاضر دورًا هامًا في جميع أفرع العلوم المختلفة . ولقد أدى اتساع نطاق استخدامات الحاسبات الآلية في كافة أوجه الأنشطة إلى جانب التطوير المتسارع والموازي في البرمجيات الإحصائية الجاهزة مثل Minitab و Spss و Sass و Bmdp ...... وغيرها من البرامج قد سهل للإحصائي عملية التعامل مع البيانات والمجموعات الكبيرة من قواعد البيانات واستخدام الطرق والأساليب الإحصائية المعقدة بل والأكثر تعقيدًا إلى جانب السرعة والدقة العالية في الحصول على مخرجات عملية التشغيل عن ذي قبل .**

**أما عن مفهوم علم الإحصاء فهو ذلك العلم الذي يختص بالتعبير عن المبادئ والنظريات والقوانين المختلفة التي تنظم الطرق العلمية لجمع وتبويب وتلخيص وتحليل البيانات بالإضافة إلى استنباط واستخلاص النتائج والتوصل إلى قرارات معينة لحل المشكلات المختلفة على أساس من التحليل العلمي للبيانات .ومن خلال هذا المفهوم السابق لعلم الإحصاء فإنه يتم تقسيم الإحصاء إلى قسمين رئيسيين هما :**

- **الإحصاء الوصفي : Descriptive Statistics وهو ذلك الفرع الذي يختص بعملية جمع وعرض البيانات . كما يتم مشاهدة هذه البيانات في الواقع بغرض توفير المعلومات عن الاتجاهات المختلفة للظاهرة وذلك من خلال استخراج بعض المقاييس أو المؤشرات الإحصائية كمقاييس النزعة المركزية (المتوسطات) ومقاييس التشتت ومقاييس الارتباط والانحدار التي تفسر وتوضح طبيعة وشكل تلك**

البيانات والعلاقات فيما بينها . ويستعرض هذا الكتاب الجوانب المختلفة لهذا الفرع من فروع الإحصاء .

- <u>الإحصاء الرياضي أو الاستدلال الإحصائي</u> :

Mathematical Statistics or Statistical Inference

ويعتمد هذا الفرع على دراسة نظرية الاحتمالات والتوزيعات الاحتمالية لكي يختص بعملية اتخاذ قرارات عامة ودقيقة من خلال البيانات المتوفرة عن الظاهرة أو الظواهر محل الدراسة . كما أنه يساعد في عملية تقرير كيفية الحصول على البيانات المطلوبة بأكثر كفاءة ويسر . فيدرس هذا الفرع من فروع الإحصاء التوزيعات الاحتمالية وتطبيقاتها والعينات الإحصائية ونظرية التقدير الإحصائي واختبارات الفروض الإحصائية وتحليل التباين والاختبارات اللامعلمية والأرقام القياسية والطرق الإحصائية لمراقبة جودة الإنتاج والإحصاءات الحيوية . وهو ما سيتم دراسته بمشيئة الله في فصل دراسي لاحق بمشيئة الله تعالى .

هذا ولقد جاء هذا الكتاب باللغة العربية في كل محتوياته فيما عدا القوانين والرموز المستخدمة فقد جاءت باللغة الإنجليزية ( وبعضها حروفًا أو رموزًا لاتينية ) حتى يتم التواصل فيما بين ما ينشر باللغة العربية والإنجليزية مما يسهل على القارئ متابعة الموضوعات المختلفة في أي مؤلف عربي أو أجنبي .

ولقد راعينا في هذا الكتاب أن يكون العرض مبسطًا لموضوعاته المختلفة بالإضافة إلى وجود العديد من التطبيقات العملية التي توضح كيفية استخدام الأساليب والطرق المعروضة .ويرجو المؤلف أن يكون قد قدم لطلبة كلية التجارة ولدارسي علم الإحصاء في كافة فروع المعرفة مادة علمية تساعدهم في دراسة وتحليل الظواهر محل الاهتمام .

وأخيرًا نتمنى من الله أن يكون قد حالفنا التوفيق في إعداد هذا الكتاب
وفقنا الله وإياكم وجعلنا ممن يتعلمون العلم ويأخذون بأحسنه

دكتور

**محمد عبد الحميد الربع**

**( فى سبتمبر ٢٠٢٢)**

# الباب الأول

## المفاهيم الأساسية في الإحصاء

### مفهوم علم الإحصاء واستخداماته المختلفة : -

كثيراً ما تعتبر " الإحصاء " مرادفاً للبيانات الرقمية . إلا أن هذا التعريف لا يعبر عن الواقع . فالإحصاء أصبح الآن علماً مستقلاً بل أنه أسلوب علمي يستخدم في كافة مجالات الحياه والدراسات الهامة . ويمكن النظر إليه باعتباره أحد فروع الرياضة التطبيقية . كما يمكن تعريفه بأنه ذلك العلم الذي يبحث في مبادئ وطرق جمع البيانات وعرضها بطريقة يسهل معها استخراج المقاييس الملائمة ثم تحليل تلك المقاييس بهدف الوصول إلى نوع من المعرفة الرقمية عن المجتمعات محل الدراسة واتخاذ قرارات محددة بناءاً على هذه المعرفة .

وقد كان الاهتمام بالإحصاء قاصراً على جمع البيانات الرقمية العامة بهدف رسم سياسات الدولة مثل الاهتمام بالبيانات السكانية رغبة في تحديد القوة البشرية والعسكرية أو معرفة الخدمات الأساسية التي يجب أن تقوم بها وكيفية توزيعها على المناطق المختلفة للدولة . إلى أن جاء منتصف القرن السابع عشر بظهور نوع جديد من الدراسات الرياضية والتي تسمى "بنظرية الاحتمالات " . ولقد تطورت الدراسة في هذا الفرع خلال القرون السابقة على أيدي رياضيين مشهورين أمثال جالتون وباكون ولابلاس وجاوس وبرونولي ومع نهاية القرن التاسع عشر كانت نظرية الاحتمالات قد تطورت وأصبحت بمثابة الأساس العلمي للأسلوب الإحصائي مما أدى إلى ظهور الإحصاء الرياضي والتحليلي أو الاستدلالي وتطبيقاته في جميع مجالات أنواع المعرفة في الحياه مما ساهم في التعرف على خصائص المجتمعات أو التي لابد فيها من اتخاذ قرار محدد بالرغم من عدم توافر بيانات كاملة عن هذه المجتمعات . وذلك لأن الأسلوب الإحصائي أصبح قادراً على بحث وتقدير المخاطر التي تنجم عن اتخاذ قرار معين .

**هذا ونظراً لتطور علم الإحصاء فقد أصبح ذات أهمية كبرى في كثير من المجالات الحيوية المختلفة . ففي الدراسات الطبية والدوائية يستخدم الاسلوب الإحصائي لدراسة وتحديد أسباب انتشار الأمراض المختلفة ومدى فاعلية الأدوية المختلفة في علاجها . هذا بالإضافة إلى دراسة العوامل الوراثية وتأثيرها ...**

**وفي الصناعة يتم استخدام التحليل الإحصائي في عملية مستوى جودة المنتج والرقابة على هذا المستوى – مراقبة جودة المنتج Quality Control وكذلك في دراسة التنبؤ باتجاهات التغير في الطلب على بعض السلع وكذا مدى النجاح المتوقع للمنتجات الجديدة ومدى فاعلية الأساليب المختلفة للإعلان ...**

**وفي الأبحاث الزراعية يتم استخدام الإحصاء في عملية تصميم وتحليل نتائج التجارب التي قد تحدد أفضل أنواع السماد أو المبيدات الحشرية الواجب استخدامها أو أثر اختلاف نوع التربة وطريقة الري على كمية المحصول من أحد المنتجات ...**

**وكذلك نجد أن التحليل الإحصائي هو الأساس الكبير في الدراسات الطبيعية والكيميائية وأبحاث الفضاء والذرة .. إلى غير ذلك من كافة فروع أنواع المعرفة الحديثة . ويضاف إلى ذلك الزيادة الملحوظة لاستخدام الإحصاء في العلوم الاجتماعية كالدراسات الاقتصادية والسياسية واستطلاع الرأي العام وأبحاث الاجتماع وعلم النفس . كما يجب أن لا نتجاهل الاستخدام التقليدي للبيانات الإحصائية العامة في رسم السياسات الحكومية سواء المرتبطة بالتعليم أو الخدمات الصحية أو خدمات الأمن .... والآن وبعد تناولنا لمفهوم علم الإحصاء ومجالات استخدامه في فروع العلوم المختلفة دعنا نتسائل ما هو الهدف من دراسة علم الإحصاء ؟**

**والإجابة على هذا التساؤل يمكن تناولها في الآتي :-**

**الهدف الأول : " عملية جمع البيانات الإحصائية " :**

حيث يتم جمع البيانات الخاصة بالمشكلة موضع الدراسة بطريقة علمية صحيحة دقيقة بعيدة عن التحيز حتى يسهل استخدامها في عملية التحليل الإحصائي للظاهرة موضع الدراسة .

**الهدف الثاني :- " عرض البيانات الإحصائية "**

وهنا يقوم الإحصائي بعملية عرض البيانات الإحصائية التي تم جمعها عن الظاهرة موضع الدراسة إما جدولياً في صورة جداول يتم تبويبها وتصنيفها في صورة بسيطة ودقيقة كما سنرى فيما بعد يسهل فهمها للقارئ حتى الغير متخصص أو عرضها بيانياً في صورة رسوم بيانية يسهل قراءتها أو عرضها من خلال بعض المقاييس الإحصائية التي تستخدم لتفسير الظاهر محل الدراسة .

**الهدف الثالث : " تحليل البيانات الإحصائية : -**

فبعد عملية جمع وعرض البيانات الإحصائية سواء جدوليًا أو بيانيًا أو من خلال بعض المقاييس الإحصائية المستخلصة من البيانات يجب تحليل هذه البيانات تحليلاً علميًا دقيقًا يؤدي إلى كيفية التنبؤ المستقبلي واتخاذ القرارات الملائمة بصدد تلك الظواهر محل الدراسة ودراسة الطرق واختبارات الفروض الإحصائية حول المقاييس المستخلصة من بيانات تلك الظواهر .

هذا وهناك مجموعة من التعريفات الأساسية اللازمة لدراستنا في مبادئ الإحصاء الوصفي لعل من أهمها ما يلي :

**أولاً : المجتمع الإحصائي والعينة** : Statistical Population & Sample

تختلف معنى كلمة المجتمع عند الإحصائي عن المعنى الشائع الاستخدام . فهو لا يشير إلى تجمع الأفراد في منطقة معينة وإنما المقصود هو ذلك الكل

الذي نرغب في دراسته ومعرفة المعلومات التفصيلية عن خصائصه المميزة. وعلم الإحصاء يهتم بدراسة المجتمعات الإحصائية المختلفة .

والمجتمع الإحصائي عبارة عن مجموعة من المفردات المحددة والمعرفة بحدود زمنية أو مكانية تشترك في صفة أو مجموع خصائص محددة يمكن ملاحظتها أو قياسها كمياً .

فمثلاً إذا كان هدف الدراسة الإحصائية هو تحديد متوسط أجر ساعة العامل المصري . فإن المجتمع الإحصائي هو عبارة عن أجور جميع العاملين المصريين في الساعة . وإذا كان هدف الدراسة هو تحديد نسبة الوحدات المعيبة في هذا الخط الجديد من خطوط الإنتاج .

وكذلك إذا كان هدف الدراسة الإحصائية هو تحديد نسبة البطالة في مصر . فإن المجتمع الإحصائي هو جميع الأفراد من سكان ج . م . ع في سن العمل والقادرين على العمل ويبحثون عنه ولا يجدونه .

فالخلاصة أنه يمكن تعريف المجتمع الإحصائي بأنه عبارة عن مجموعة المفردات أو العناصر المحددة والمعرفة بحدود زمانية ومكانية معينة وتشترك في صفات وخصائص محددة يمكن ملاحظتها وقياسها كمياً .

ومهما تعددت المجتمعات الإحصائية حسب تنوع الظواهر والمتغيرات في مجالات كافة أنواع المعرفة فإن يمكن تقسيمها حسب تنوع الظواهر أو المتغيرات في مجالات كافة أنواع المعرفة إلى نوعين أساسين : –

**الأول : مجتمع إحصائي محدود** **Limited Population**

وهو المجتمع الذي يحتوي على عدد محدود من الوحدات أو الأحداث القابلة للعد أو الحصر ومن ثم القابلة لتحديد الحجم . ومثال للمجتمع المحدود مجموعة الطلاب المصريين للعام الدراسي ٢٠٠٨ / ٢٠٠٩ فالمجتمع

الإحصائي هنا مهما كبر حجمه ليصل لـلآلاف أو حتى الملايين من المفردات إلا أنه قابل للحصر وبالتالي للعد وتحديد الحجم . فما دام هناك إمكانية لتحديد إطار لمفردات أو عناصر المجتمع الإحصائي يتم تصنيفه لمجتمع إحصائي محدود .

**الثاني : مجتمع إحصائي غير محدود Unlimited Population**

وهو ذلك المجتمع الإحصائي الذي يحتوي على عدد غير محدود ( أي لا نهائي ) من الوحدات أو الأحداث الغير قابلة للحصر ومن ثم غير قابلة للعد أو لتحديد الحجم . مثال ذلك عدد كرات الدم الحمراء أو البيضاء في دم العنصر البشري . فهذا العدد غير قابل للحصر أو للعد لذا يصنف المجتمع في تلك الحالة على أنه مجتمع إحصائي غير محدود.

هذا وتسمى الخصائص الدالة على أي مجتمع إحصائي والتي يمكن ملاحظتها أو قياسها كمياً باسم معلمات المجتمع Parameters . وغالباً ما تكون القيم الحقيقية لمعلمات المجتمع الإحصائي مجهولة (غير معلومة). إلا أنه يمكن تقديرها من خلال الاستدلال عنها وذلك من خلال سحب عينة عشوائية بطريقة إحصائية سليمة لكي تكون ممثلة لهذا المجتمع تمثيلاً تاماً (جيداً) . وتجري الدراسة الإحصائية على تلك العينة مستخدمين في ذلك طرق أو أساليب الإحصاء التحليلي (الاستدلالي ) وهو ما سيرد دراسته في الفصل الدراسي الثاني بمشيئة الله تعالى .

**خطوات جمع البيانات الإحصائية : -**

إن عملية جمع البيانات الميدانية ليست بالسهولة وإنما تخضع لشروط وقيود يصنعها الباحث لضمان الحصول على بيانات جيدة لمجتمع المشكلة محل الدراسة فهي عملية طويلة تتطلب الآتي :

١- <u>تحديد هدف ومشكلة مجتمع الدراسة</u> : فإن تحديد هدف البحث هو الخطوة الأولى للقيام بأي بحث أو دراسة إحصائية فبناءًا على هذا الهدف يتحدد "<u>المجتمع</u>" محل البحث وكذلك تتحدد "<u>المفردة</u>" التي يجب جمع البيانات عنها ونوع ومصدر البيانات والأسلوب الذي يجب استخدامه في جمع تلك البيانات .

فمثلاً إذا كان الهدف من بحث ما هو المقارنة ما بين كيفية إنفاق دخل الأسرة في ريف ج.م.ع بكيفية إنفاق دخل الأسرة في حضر ج.م.ع خلال فترة زمنية معينة . فإن المجتمعات محل الدراسة تكون هي كلاً من ريف وحضر ج.م.ع خلال هذه الفترة الزمنية المحددة . كما أن المفردة التي تتكون منها هذه المجتمعات والتي يجب أن تُجمع عنها البيانات هي "الأسرة" والبيانات المطلوبة هي دخل الأسرة في تلك الفترة الزمنية المحددة وكيفية توزيعه على مصادر الإنفاق المختلفة .

فالمجتمع الإحصائي هو كما ذكرنا سابقًا عبارة عن مجموع المفردات التي يجب أن تُجمع عنها البيانات . كما أن المفردة محل البحث هي وحدة جمع البيانات والمفرده محل البحث قد تكون سلعة منتجة بطريقة معينة أو هي القرية ذات المساحة المحدودة أو هي الطفل في سن معين .... إلخ . فمن الواضح أن تحديد المجتمع وكذلك مفرده البحث يتوقف على نوع الدراسة أو الهدف من البحث .

٢- <u>تحديد مصادر جمع البيانات الإحصائية :</u>

فالخطوة التالية لتحديد المجتمع والمفردة محل الحث ونوع البيانات المطلوبة هي عبارة عن تحديد المصدر الذي يمكن استخدامه للحصول على هذه البيانات . فهناك نوعين أساسيين لمصادر البيانات الإحصائية هما :

**أ- <u>مصادر تاريخية</u> : ومنها يتم الحصول على البيانات المنشورة فعلاً والتي قامت بجمعها هيئة مختصة مثل التعدادات السكانية والإحصاءات المختلفة التي تقوم الدولة بجمعها أو نشرها . وإذا تم استخدام هذه المصادر التاريخية يجب مراعاة الحذر عند استخدام البيانات والتأكد من دقتها وتمثيلها للمجتمع تمثيلاً تامًا كما يجب دائمًا الإشارة إلى مصدر تلك البيانات .**

**ب- <u>مصادر ميدانية</u> : وهي عبارة عن جمع البيانات من المفردات المكونة لمجتمع الدراسة مباشرة سواء بالمقابلة الشخصية أو بالبريد أو بأحد وسائل الإعلام أو عن طريق المشاهدة (في حالة التجارب العملية مثلاً) . وحتى يمكن جمع البيانات من الميدان فإنه على الباحث أن يحدد أسلوب جمع البيانات بأن يقرر ما إذا كان سيحصل على البيانات من جميع أفراد المجتمع (أي باستخدام أسلوب الحصر الشامل) أم من بعض هذه المفردات فقط (أي باستخدام أسلوب العينة ) . وفي تلك الحالة الأخيرة أي في أسلوب العينة فإنه يجب أن يحدد أيضًا كيفية اختيار مفردات العينة أي تحديد نوع العينة (عشوائية بسيطة أم طبقية أم متعددة المراحل ... إلخ)**

**وكنبذة مبسطة فإن <u>العينة</u> Sample الإحصائية هي عبارة عن مجموعة من المفردات مسحوبة من مجتمع إحصائي حول ظاهرة معينة . ولكي يتم سحب عينة عشوائية بطريقة إحصائية سليمة تعطي كافة مفردات المجتمع الفرصة المتساوية لكي يتم تمثيلها في العينة يجب وضع قائمة تحتوي على كافة مفردات المجتمع محل الدراسة والتي تسمى <u>بالإطار</u> والذي يجب أن يتوافر فيه الدقة والشمول وعدم التكرار . ثم تجري عملية سحب متتالي لمفردات العينة من مجتمع الدراسة سواء يدويًا أو آليًا أو باستخدام جداول الأرقام العشوائية .**

**٣- اختيار طريقة جمع البيانات :**

**والمقصود هنا بطريقة جمع البيانات هو تحديد كيفية الاتصال بالمفردة محل البحث . فعلى الباحث أن يضع قائمة للأسئلة التي يريد الإجابة عنها ليحصل على ما يريد من بيانات . وهذه القائمة تسمى باستمارة البحث أو صحيفة الاستقصاء (Questionnaire) ويجب أن تراعي بعض القواعد عند وضع هذه الأسئلة . فمثلاً يجب أن تكون مختصرة بقدر الإمكان وأن لا تتضمن أسئلة محرجة . كما يجب أن تتجنب الاستمارة الأسئلة ذات الإجابات المعتمدة على الحكم الشخصي للفرد (فمثلاً بدلاً من سؤال الفرد المبحوث عن مستوى تعليمه يمكن أن يطلب منه وضع علامة أمام أحد الحالات التعليمية المعروفة : لا يقرأ ولا يكتب ، يقرأ فقط ، حاصل على شهادة متوسطة ، .... إلخ).كما يجب وضع الأسئلة في صورة يسهل الإجابة عليها بطريقة واضحة .**

**هذا وبعد التأكد من صلاحية استمارة البحث وبعد أن تتم عملية تهيئة المجتمع لعملية جمع البيانات (وذلك بالإعلان عن أهمية البحث وضرورة الدقة في الإجابة ومدى سرية البيانات) وبعد أن يتم تدريب الأفراد اللذين سيقومون بجمع البيانات يبدأ الجمع الفعلي للبيانات من عملية ميدانية تسمى عادة بالدراسة الاستكشافية Pilot Study والتي تجرى على عينة مماثلة إلى حد كبير للعينة أو المجتمع الأصلي . والهدف من تلك الدراسة المسبقة يتلخص في ثلاث أسباب رئيسية . فأولاً تساعد هذه المعاينة المصغرة على التعرف على الصعاب التي يمكن أن تواجه البحث الأصلي في الميدان مقدمًا فيمكن بذلك عمل أية تعديلات سواء في الاستمارة أو في طريقة المقابلة أو غير ذلك من التعديلات خصوصًا ما يتعلق منها بنظام سير العمل في الميدان. وثانيًا فإن هذه الدراسة الاستكشافية تساعد على التعرف على مدى فهم جامعي البيانات للمصطلحات الواردة بالاستمارة وتقدم بذلك فرصة طيبة لتدريبهم**

وزيادة خبرتهم في الميدان . وأخيرًا وهو ما يعتبر أهم فوائد الدراسة الاستكشافية ألا وهو التعرف على مدى فهم مفردات العينة للأسئلة والألفاظ المستخدمة ودرجة وضوحها ومدى سهولتها . كذلك تساعد تلك الدراسة على تغير لغة الاستمارة بما يتناسب مع بيئة المجتمع المستهدف . وتساعدنا أيضًا على التعرف على مدى وجود بدائل جديدة للإجابات يمكن أن تضاف لاختبار المستجيب . وعمومًا فالدراسة الاستكشافية ضرورية للاطمئنان على صحيفة الاستبيان من ناحية الشكل وحسن الصياغة والمضمون .

- **<u>وسائل جمع البيانات :</u>**

تتعدد وسائل جمع البيانات الميدانية . واختيار الوسيلة يعتمد على طبيعة البحث ودرجة الدقة المطلوبة وبطبيعة الحال على ميزانية البحث المالية والزمنية . كذلك فإن طبيعة المجتمع المستهدف قد تحتم أحيانًا وسيلة معينة دون أخرى . هذا ويمكننا تقسيم وسائل جمع البيانات إلى ما يلي :

* وسائل مباشرة      * وسائل غير مباشرة

أ- **<u>وسائل جمع البيانات المباشرة</u>** :

ويمكن تلخيص الوسائل المباشرة لجمع البيانات في الآتي :

- المشاهدة أو الملاحظة **Observation**
- البريد Mail
- الاتصال التليفوني Telephone
- المقابلة الشخصية Interview

فجمع البيانات يمكن أن يتم عن طريق المشاهدة أو الملاحظة ويتم ذلك في حالة الأبحاث العلمية ذات الطبيعة الخاصة كالأبحاث الطبية أو الكيميائية أو طبيعة اجتماعية معينة . فقد يقرر الباحث الاجتماعي أن يجمع بياناته عن

طريق مشاهدة سلوك أو انفعالات الفرد المبحوث دون أن يلجأ للسؤال مثل السلوك الاجتماعي لمتعاطي أو مدمني المخدرات .

أما عن تجميع البيانات من خلال وسيلة الخدمة البريدية فيتم ذلك عن طريق إرسال صحيفة البحث إلى الفرد المبحوث بريدياً أو حتى نشرها في صفحات الصحف أوالمجلات . ويكون على الفرد المبحوث أن يستوفي البيانات المطلوبة ويعيدها إلى الهيئة القائمة على البحث . وعادة ما تُرسل صحيفة البحث إلى الفرد المبحوث مرفقاً بها خطاباً عن أهمية البحث وضرورته . وكذلك يرفق به مظروف بعنوان الجهة أو الهيئة القائمة على البحث ملصقاً عليه طابع البريد حتى لا يتكلف الفرد المستجيب مالاً أو وقتاً . كما يجب أن يضمن الباحث للمستجيبين السرية الكاملة للبيانات كما يحق للفرد المستجيب عدم ذكر اسمه عند الرد .

أما وسيلة الاتصال التليفوني فهي تعتمد في جمعها للبيانات على الاتصال التليفوني بالمفردات محل البحث لاستيفاء البيانات المطلوبة . حيث يحصل الباحث على أرقام تليفونات المستجيبين ويتصل بهم لملأ صحيفة البحث من واقع إجاباتهم . وهذه الطريقة تتميز بالبساطة والسهولة والسرعة وقلة التكاليف . لكن أهم ما يعيبها أنه ربما تتضمن العينة محل البحث مجموعة من المفردات ليس في حوزتهم تليفونات .

أما عن وسيلة المقابلة الشخصية وهي أكثر الطرق شيوعاً واستخداماً في عملية جمع البيانات من الميدان . وهي تعتمد على أن يقوم جامع البيانات بالذهاب إلى حيث يلتقي بالفرد المبحوث للحصول على اجابات للأسئلة الموجودة في استمارة البحث .

ب- <u>وسائل جمع البيانات الغير مباشرة :</u>

وتعتمد هذه الوسائل عادة على الاستجابات التي يمكن تحليلها فيما بعد لاستنتاج نتائج معينة لا تعتمد على الاستجابة مباشرة . وتنتشر مثل هذه الوسائل خصوصاً في مجال الصحة والبحوث النفسية وبحوث الدوافع حيث يلجأ الباحثون إلى العمليات العقلية اللا شعورية . فعلى سبيل المثال طريقة التداعي الحر ووصف عينة من خلال السلوك أو الإسقاط وغيرها . والهدف منها التوصل إلى حقيقة نفسية ودوافع المبحوث عن طريق غير مباشر أي من غير طريقة السؤال المباشر فيما يختص بالسؤال المطلوب عن الموضوع المطلوب الإجابة عليه .

- <u>أساليب جمع البيانات :-</u>

عند جمع البيانات الاحصائية يمكن استخدام إما اسلوب الحصر الشامل أو اسلوب العينة ( المعاينة ) . فاستخدام اسلوب الحصر الشامل لجمع البيانات معناه جمع البيانات عن جميع مفردات المجتمع . أما اسلوب العينة فهو عبارة عن جمع البيانات من بعض مفردات المجتمع فقط . وفي بعض الأحيان تجد أن طبيعة المجتمع محل الدراسة وطبيعة البيانات المطلوبة قد يفرضان على الباحث استخدام أحد الاسلوبين دون الآخر فمثلاً إذا كان المجتمع مكون من عدد لا نهائي من المفردات – أي مجتمع غير محدود أو إذا كانت مفردات المجتمع غير موجودة بالكامل وقت جمع البيانات فإنه لابد من استخدام اسلوب العينة . وبلغة أخرى إذا كانت البيانات المطلوبة تؤدي إلى إهلاك المفردة محل البحث مثل حساب متوسط عدد ساعات إضاءة المصابيح الكهربائية التي ينتجها أحد المصانع أو متوسط عدد كرات الدم البيضاء أو الحمراء في المليمتر الواحد المكعب وكذلك البيانات الخاصة بالعمالة والأجور فإنه أيضاً لابد من استخدام اسلوب العينة. أما إذا كان المطلوب هو الحصول

على بيانات عن جميع مفردات مجتمع محدد مثل معرفة سن ومكان إقامة كل طالب في إحدى السنوات الدراسية بالكلية . فإنه لابد من استخدام اسلوب الحصر الشامل في تلك الحالة .

وجدير بالذكر فإنه في كثير من الأحيان قد نجد أن كل من طبيعة المجتمع وطبيعة البيانات تسمح باستخدام أي من الاسلوبين سواء الحصر الشامل أو العينة . وفي تلك الحالة فإن المقارنة بين الاسلوبين تعتمد أساساً على مدى الإمكانيات المادية والفنية المتاحة للبحث ومدى الدقة التي يمكن أن يحققها أحد الاسلوبين في عملية جمع البيانات . فجميع البيانات باستخدام اسلوب الحصر الشامل يحتاج إلى الكثير من النفقات والوقت بينما نجد أن اسلوب العينة قد يتطلب نفقات ووقت أقل . ولكن الحصر الشامل قد يتميز بأنه من الممكن إجراءه بمستوى عالي من الدقة في حالة المجتمعات المحدودة كما أنه يعطي نتائج مباشرة وكلية عن المجتمع بأكمله .

ويجب علينا أن نذكر أن استخدام اسلوب العينة بطريقة سليمة يتطلب توافر إطار يشمل جميع مفردات المجتمع حتى يمكن استخدامه عند اختيار تلك المفردات التي ستظهر في العينة . كما يجب مراعاه الدقة والشمول مع تجنب التكرار عند تكوين هذا الإطار حتى نضمن أن يكون لكل مفردة من مفردات المجتمع لها نفس الفرصة للظهور أو التمثيل في العينة . وبصفة عامة فإنه عند جمع البيانات قد يتعرض الباحث لنوعين من الأخطاء :-

- <u>خطأ التحيز</u> :- وينشأ هذا النوع من الخطأ نتيجة تحيز الأفراد – لا شعورياً – لبعض القيم أو البيانات . كما أنه قد ينشأ نتيجة الاهمال في جمع البيانات أو عدم دقة الرقابة أو عدم وضوح الأسئلة الموجودة باستمارة البحث أو عدم التعاون بين مفردات البحث والقائمين بعملية جمع البيانات وهي جميعاً عوامل تؤثر على البيانات التي يتم جمعها

سواء بأسلوب الحصر الشامل أو باسلوب العينة . إلا أنه في حالة استخدام العينة فإن استخدام إطار معيب وكذلك قيام جامعي البيانات بإحلال المفردات التي لا تتواجد أثناء جمع البيانات بمفردات أخرى تعتبر عوامل إضافية تؤدي إلى خطأ التحيز . ومن جهة أخرى فإنه نظراً لصغر عدد المفردات في العينة إذا ما قورن بعدد مفردات المجتمع فإن من الممكن تقليل خطأ التحيز بها .

ونظراً لعدم إمكانية قياس أثر هذا الخطأ على البيانات التي يتم جمعها فإنه يعتبر من العيوب التي يجب العمل على تجنبها بقدر الإمكان عند جمع البيانات .

- **خطأ الصدفة العشوائى ( الخطأ العشوائى ) :-** ويظهر هذا الخطأ عند استخدام اسلوب العينة فقط . فهو ينجم عن عملية اختيار بعض مفردات المجتمع دون البعض الآخر عند جمع البيانات منها وهذا النوع من الأخطاء ليس له اتجاه ثابت ويمكن التحكم فيه وتقليل آثاره وذلك إذا ما تم اختيار العينة بطريقة علمية أو إحصائية سليمة . وكذلك فإنه يمكن استخدام الأساليب الرياضية لتقدير مدى تأثير هذا النوع من الأخطاء على نتائج الدراسة .

وأخيراً فإنه يمكن القول أنه كون العشوائية ( الصدفة ) تحكم عملية اختيار وحدات المعاينة الإحصائية ( العينة) فإن نتائج عملية التحليل الإحصائي ستختلف باختلاف العينات الاحصائية ( حجماً أو نوعاً أو عدداً والتي يتم سحبها من نفس مجتمع المشكلة تحت الدراسة . هذا ويسمى الفرق ما بين القيمة الحقيقية لمعلمة المجتمع ( أو القيمة المشاهدة ) وبين القيمة المقدرة لتلك المعلمة من واقع بيانات العينة الاحصائية المختارة باسم خطأ

المعاينـة الإحصـائية Sampling Error . ويحكـم حجـم خطـأ المعاينـة مجموعة من العوامل وهي :-

١- نوع العينة : فيفضل العينات العشوائية الاحصائية حيث أن لكل مفردة في مجتمـع الدراسـة تكـون لهـا نفـس فرصـة الاختيـار والتمثيـل ضـمن مفـردات العينـة وفيهـا يمكـن الـتحكم وقيـاس أخطـاء المعاينـة العشوائية علـى خـلاف العينة العمدية Judgment Sample والتي يختار الباحث مفرداتها بشكل شخصي ( قصدي ) من مجتمع مشكلة البحث .

٢- حجم العينة :- حيث توجد علاقة عكسية فيما بين حجم أخطاء المعاينة وبلغـة أخـرى درجـة الدقـة الاحصـائية وبـين حجـم العينـة وحجـم الميزانيـة المرصودة للبحث .

٣- تباين المجتمع :- فكلما كان تباين المجتمع كبير فإن ذلك يؤدي إلى زيادة تمثيل عينة جيدة لمجتمع مشكلة البحث ومن ثم أخطاء المعاينة العشوائية. وإن كان نوع وحجم العينـة المسحوبة يؤديـان إلـى تحسين عمليـة الحصول على عينة احصائية جيدة .

وفيمـا يلـي الجـدول التـالي يبـين الفـرق بـين اسـلوبي جمـع البيانـات : الحصـر الشامل والعينة :

## الفرق ما بين أسلوب الحصر الشامل والعينة

| الأسلوب / بيان | الحصر الشامل | العينة |
|---|---|---|
| ١- المعنى | جمع البيانات من جميع مفردات المجتمع | جمع بيانات من بعض مفردات المجتمع |
| ٢- الاستخدام | يستخدم إذا كان الهدف يتطلب جمع بيانات من جميع مفردات المجتمع | يستخدم إذا كانت البيانات المطلوبة تؤدي إلى إهلاك المفردة محل البحث . |
| ٣- المزايا والعيوب ونوع الأخطاء الممكن الوقوع فيها | – الحصر الشامل يتميز بمستوى دقةعالي<br>– الحصر الشامل يحتاج إلى وقت وجهد وتكاليف أعلى<br>– الحصر الشامل يعطي نتائج مباشرة وكلية عن المجتمع محل الدراسة .<br>– الحصر الشامل يتعرض لخطأ الصدفة فقط | العينة تتميز بمستوى دقة أقل<br>العينة تحتاج إلى وقت وتكاليف وجهد أقل .<br>العينة لا تعطي نتائج مباشرة بل تقديرية عن المجتمع كله .<br>العينة تتعرض لخطأ التحيز كما تتعرض لخطأ الصدفة ولكن بدرجة أقل . |

- **<u>خطوات ( منهج ) البحث الاحصائي ( أو العلمي ) :–</u>**
- **كأي بحث علمي فإن التفكير الاحصائي يهدف إلى التوصل إلى الحقائق بطريقة أكثر دقة ومأمونية وذلك بقصد التعرف على العلاقات أو المتغيرات المختلفة التي تحكم ظاهرة ما أو مجموعة من الظواهر محل الدراسة عند محاولة الاجابة على العديد من التساؤلات في الحياه . وتختلف الطريقة الاحصائية عن باقي الطرق في أنها تعتمد على التعبير الكمي كلياً . حيث أن منطق الأرقام يعتبر من أقوى وسائل الإقناع والثبات وذلك لموضوعيته وحياديته وابتعاده عن التقدير الشخصي أو الآراء الفلسفية التي كانت تحكم منطق الأمور والتحليل في الماضي . هذا وترتكز خطوات البحث الاحصائي على مجموعة العناصر التالية :**

**<u>الأول : تعريف المشكلة : ـ</u>**

**فإن الشعور بوجود مشكلة معينة هو بداية عملية التفكير في عملية البحث . وتنشأ المشكلة أو يتم ملاحظتها أما بالصدفة أو قد تكون هناك حاجة إلى إيجاد تحليل أو تفسير الظاهرة أو سلوك أو حالة ما يراد علاجها . وفي تلك المرحلة من خطوات البحث لابد أن تتم عملية التحديد الواضح والدقيق للمشكلة موضع الدراسة كخطوة أولى في سبيل تحديد عناصر حل تلك المشكلة .**

**<u>الثاني : وضع الفروض الإحصائية :</u>**

**وهي تصاغ بمعرفة الإحصائي متوافقة مع الفروض النظرية والموضوعة من قبل الباحث كإقتراح مبدأي لتفسير أو حل المشكلة ، على أن تكون فروضًا موضوعية (بعيدة عن الميل الشخصي للباحث) ، وأن تصاغ بسهولة ووضوح ويتم ترتيبها حسب درجة أهميتها (أولويتها) .**

**<u>الثالث : جمع البيانات :</u> تأتي عملية جمع البيانات الإحصائية على ضوء التحديد المسبق لعناصر حل المشكلة . وعمومًا على الباحث أن يكون ملمًا بكل ما يستطيع جمعه من بيانات يمكن أن تلقي الضوء على المشكلة موضوع البحث من المصدرين الأوليين (التاريخي ، الداخلي) ، من قبل أن يلجأ إلى عملية جمع البيانات من المصدر الثالث (الميداني) بوسائله المتعددة . ويجب الحرص عند استخدام البيانات لتجنب الوقوع في مصادر أخطاء البيانات كون درجة الدقة المتحققة في البيانات الإحصائية ينعكس وبشكل مباشر على درجة الدقة أو المأمونية الإحصائية في مخرجات عمليات التحليل الإحصائي المختلفة فيما بعد .**

**<u>الرابع : تحليل البيانات</u>**

**من أساسيات تحليل البيانات هو تفهم الاختلاف فيها للوقوف على أسبابه ، ويمثل ذلك مفتاح تفهم نظام البيانات المتعلقة بالمشكلة تحت التحليل الإحصائي ، وفي التحليل الإحصائي نستخدم الطرق والأساليب الملائمة ، بمحددي نوع وخصائص البيانات ، وهدف الدراسة ، بغرض التوصيف (القياس) الإحصائي الجيد للعلاقات (النماذج) الإحصائية القادرة على تفسير سلوك الظواهر (المتغيرات) محل الدراسة .**

**ولقد ساعد التطوير والمتسارع في الحاسبات الآلية (خاصة على مستوى البرمجيات الإحصائية الجاهزة) في تحقيق السهولة واليسر ، والدقة العالية ، عند التعامل مع المجموعات الكبيرة من البيانات ، وإمكانية استخدام الطرق ، والأساليب الإحصائية المعقدة أو الأكثر تعقيدًا في عمليات التحليل الإحصائي عن ذي قبل .**

**<u>الخامس : تفسير النتائج :</u>اختبار صحة الفروض الإحصائية الموضوعة على ضوء نتائج مخرجات عمليات التحليل الإحصائي المستخدمة ، وهي احتمالية (أي إمكانية أن تكون صوابًا أو خطأ بمستويات ثقة محددة) ، وتتم عملية التفسير لمخرجات عمليات التحليل الإحصائي على مستويين ، وهما :**

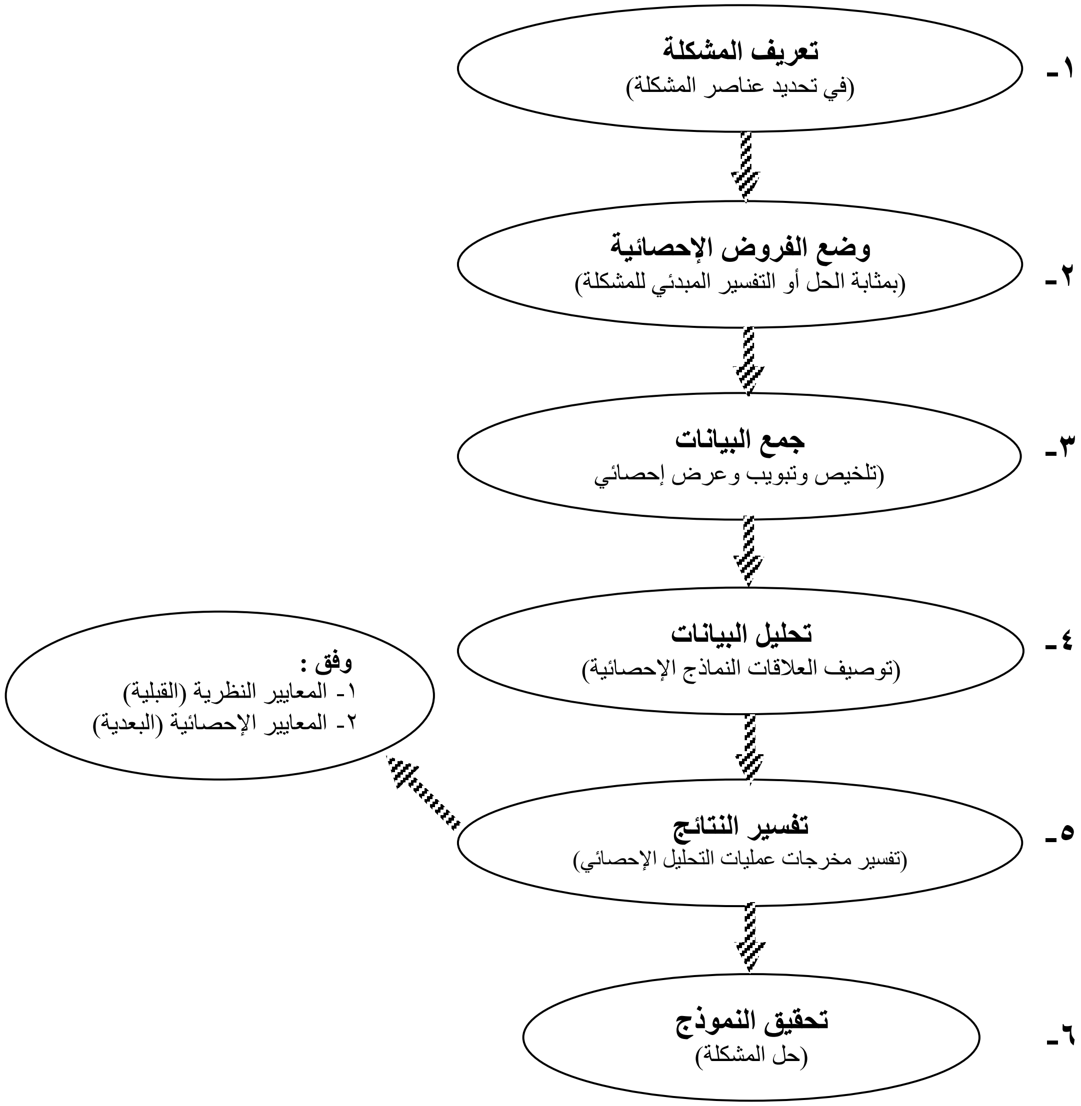
١- **<u>معيار قبلي : Prior</u> : مسترشدًا الإحصائي بالمعايير النظرية الحاكمة لعلاقات الظاهرة أو المتغيرة تحت المعالجات الإحصائية ، ويقع عبء هذا الدور أساسًا على المتخصص في مجال الدراسة ، فقد يكون : ديموجغرافي أو اجتماعي أو اقتصادي أو سياسي أو نفسي أو تربوي أو طبي ...... أو غيره ، في مجالات العلوم وأفرع المعرفة المختلفة .**

٢- **<u>معيار بعدي Posterior</u> : مسترشدًا الإحصائي بالمعايير النظرية الإحصائية الحاكمة للأساليب والطرق الإحصائية المستخدمة ، ويقع عبء هذا الدور على الإحصائي .**

**ويجب التنويه إلى أن المعايير الإحصائية (البعدية) هي داعم أو مؤكد للمعايير النظرية (القلبية) ولكن لا تجبها .**

**<u>السادس : تحقيق النموذج :-</u>**

**يأتي علاج المشكلة تحت الدراسة بتحقق صحة الفروض النظرية الموضوعة باحتمال على ضوء مخرجات عمليات التحليل الإحصائي المستخدمة ، وغالباً ما تصاغ في صورة علاقات إحصائية تتوافر فيها القدرة على تفسير سلوك الظاهرة تحت التفكير الاحصائي في الماضي والحاضر والمستقبل ، أما إذا لم يتحقق ذلك يبدأ الشك في النموذج الاحصائي المقترح ، ويجب على الباحث إعادة النظر في صياغة فروض نظرية جديدة ، والعودة إلى الخطوة الثالثة ، ويمكن إيجاز خطوات التفكير الاحصائي ، في الشكل التوضيحي الآتي :-**




**شكل (١-١)**

**توضيحي عام خطوات التفكير الإحصائي**

**<u>التحليل الاحصائى واستخدام الحاسب الآلى :-</u>**

**علم الإحصاء ، مع التطوير المتسارع في الحاسبات الآلية – خاصة على مستوى البرمجيات الاحصائية الجاهزة – أصبح معه استخدام الطرق ، والأساليب الاحصائية أكثر سهولة ويسراً ، ومتعة لدى مستخدمي الاحصاء في شتى مجالات العلوم الأخرى ، وهو ما امتد أثره لعملية التطوير الذاتي للاحصاء على المستويين النظري والتطبيقي .**

**غير أن الاحصائي المتخصص الذي يفتقر إلى الخلفية النظرية الاحصائية الكافية في المقام الأول يصبح معه استخدام الحاسب الآلي قليل المنفعة – إن لم تنعدم منفعته – فالحاسب الآلي لا يوفر لمستخدمه التفكير الاحصائي عند التعامل مع المشاكل تحت المعالجة الاحصائية ، فالحاسب الآلي لا يختار للاحصائي الاسلوب أو الطريقة الاحصائية واجبة الاستخدام أو يفسر له مخرجات عمليات التشغيل الاحصائي المختلفة ، وإنما يقع عبئه على الاحصائي المتخصص ، والأمر هنا يتوقف على مدى كفاءته الاحصائية ، وبما يتوافر لديه من مخزون معرفي بالنظرية الاحصائية ، ومستحدثاتها والتي تشتمل على فروع متنوعة صارت مجالات تخصص مختلفة تحت مظلة علم الاحصاء .**

**<u>خطوات التحليل الاحصائى باستخدام الحاسب الآلى :</u>**

**ويمكن تحقيق الاستفادة الكبيرة من استخدام الحاسب الآلي ، على مستوى عمليات التحليل الاحصائي للبيانات ، مستخدمين البرمجيات المختلفة ، منها : Minitab , Bmdp , Sas , Spss ... وغيرها ، في خدمة تنفيذ خطوات التفكير الاحصائي ، فيما يلي :**

**<u>الأول تفريغ البيانات :</u>وفيها تتم عملية تفريغ البيانات في استمارة خاصة ، يخصص فيها لكل متغيرة عدد من الحقول (يراعى أكبر قيمة ممكنة للمتغيرة) ، على أن يكون قد تمت المعالجات اللازمة لبعض مشاكل البيانات :**

**١- <u>عملية الترميز coding</u> : عند التعامل مع الظواهر أو المتغيرات الوصفية مثال : الإجابات ، وبمستوياتها على مستوى السؤال الواحد ، في استمارة الاستبيان .**

**٢- <u>القيم المفقودة مثال</u> : حالة النسيان أو الرفض أو الغير صحيحة لبعض إجابات استمارة الاستبيان ، وتعالج بالاستبعاد أو بإعطاءها رمز خاص ، والتعامل معها بطريقة ملائمة لموضوع الدراسة .**

**ويجب التأكيد على ضرورة المراجعة الدقيقة لعملية تفريغ البيانات ، كون درجة سلامتها تنعكس بشكل مباشر على درجة الدقة الاحصائية لمخرجات عمليات التشغيل الاحصائي المختلفة ، إن لم تكن خاطئة .**

**<u>الثاني : إنشاء ملف البيانات</u> :**

**وهي عملية إدخال البيانات إلى الحاسب في ملف يسمى بملف البيانات وفق قواعد البرنامج الاحصائي الجاهز المستخدم ، مع التأكيد على أهمية مراجعة عملية إدخال البيانات للتأكد من سلامتها ، خاصة والبرمجيات الاحصائية الجاهزة تعطي إمكانية وضع حدود لعدد الحقول عند إدخال البيانات ، فإذا تم إدخال بيانات أكبر من عدد الحقول المحددة مسبقاً يعطي مثلاً رمز ** بديلاً عن القيم في ملف البيانات .**

**<u>الثالث : وصف البيانات :</u>**

**تبدأ عملية التحليل الاحصائي للبيانات بخطوة وصف البيانات ، مستخدمين الجداول التكرارية والرسوم البيانية والمقاييس الاحصائية الأولية ( المتوسطات ، التشتت ، الالتواء ، التفلطح ) ، بهدف التعرف على الوصف الاحصائي**

**للبيانات ، لحسن التعامل معها ، واختيار الطرق والأساليب الاحصائية الملائمة .**

**<u>الرابع : توصيف العلاقات ( أو النماذج) الاحصائية :</u>**

**تحكم عملية اختيار الأساليب والطرق الاحصائية الملائمة ، كل من نوع وخصائص البيانات ، وهدف الدراسة تحت التحليل الاحصائي ، وفيها :**

**١- <u>اختبارات الفروض الاحصائية</u> :**

**يتم اختبار مدى صحة الفروص الاحصائية ، المقابلة للفروض النظرية ، والتوافق هنا ضروري مستخدماً الاحصائي الأساليب ، والطرق الاحصائية المعملية parametric أو اللا معملية Non- parametric الملائمة للبيانات وهدف الدراسة ، للحكم باحتمالية على مدى صحة الفروض الاحصائية من عدمه ( وفق المعايير الاحصائية ) .**

**٢- <u>قياس العلاقات ( أو النماذج ) الاحصائية</u> :**

**يهدف التحليل الاحصائي في نهاية المطاف ، التوصل إلى قياس العلاقات ( أو النماذج ) الاحصائية المختلفة ، والقادرة على تفسير التغيير الحادث في سلوك الظواهر أو المتغيرات ( الماضي أو الحاضر أو المستقبل ) تحت التحليل الاحصائي ، ومستخدمين من الأساليب والطرق الاحصائية المتنوعة ، مثال : تحليل الانحدار ( البسيط ، المتعدد ) ، تحليل التباين ( في اتجاه واحد أو أكثر ) ، تحليل التغاير ، التحليل العاملي ، تصميم التجارب ، .... وغيرها ، مع ملاحظة أنه يمكن عامة تقسيم مخرجات عمليات التحليل الاحصائي إلى ثلاث مجموعات ، وهي :**

**أ – وصف البيانات .**

**ب– اختبارات الفروض الاحصائية ( المعلمية – اللامعلمية ) .**

**جـ – قياس العلاقات ( أو النماذج ) الاحصائية .**

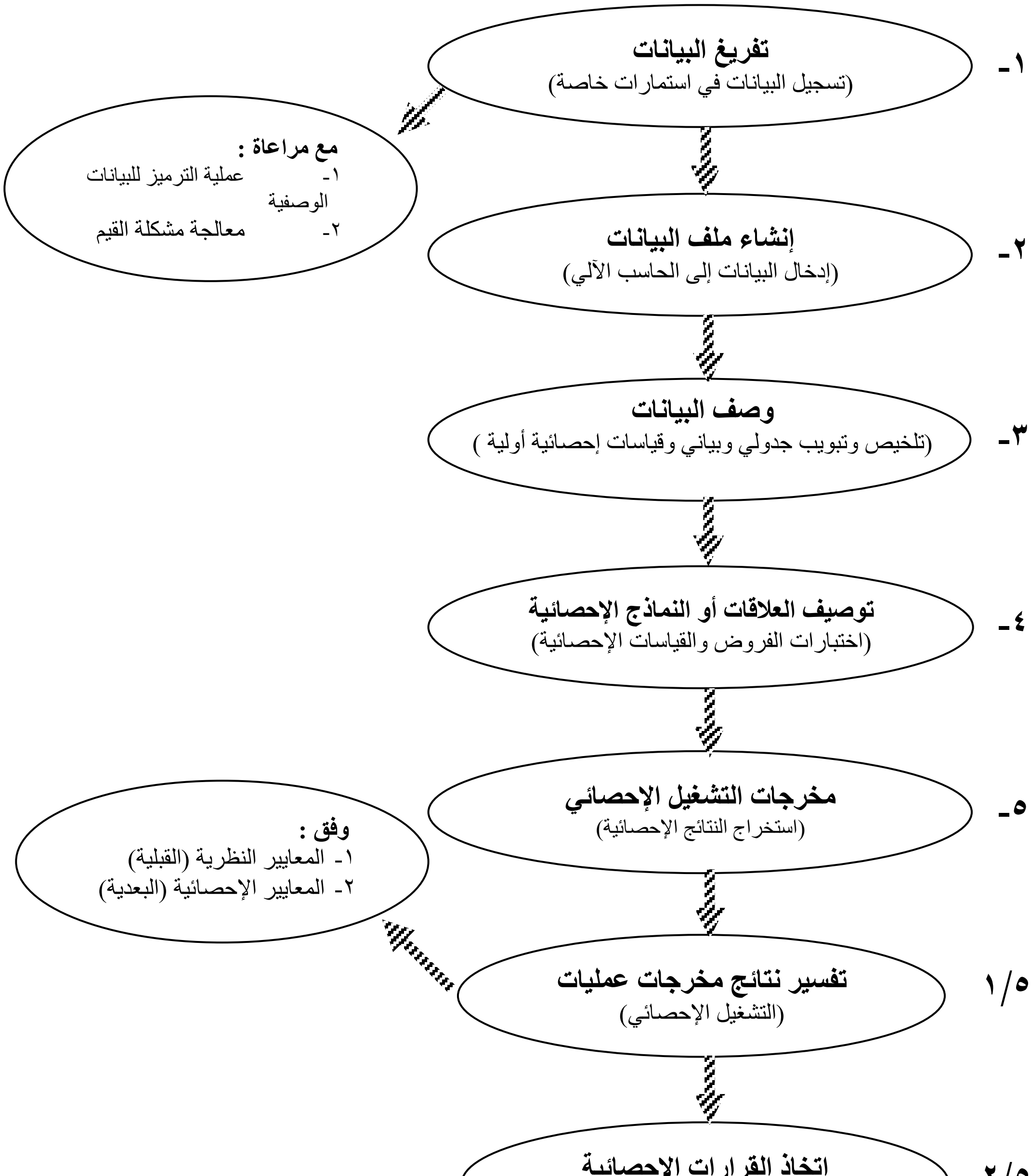
**<u>الخامس : مخرجات عملية التشغيل الاحصائى :</u>**

**ويتم التعامل مع مخرجات التشغيل الاحصائي المختلفة على مستويين :**

**١- تفسير نتائج مخرجات عمليات التشغيل الاحصائي ( وصف البيانات ، اختبارات الفروض ، قياس العلاقات ) ، على ضوء : المعايير النظرية ( القبلية ) الحاكمة لعلاقات النظرية للظاهرة أو المتغيرة محل التحليل الاحصائي ، وبدعم المعايير الاحصائية الحاكمة للطرق والأساليب الاحصائية المستخدمة .**

**٢- اتخاذ القرارات الاحصائية على ضوء التفسير السابق كمدخل لحل المشكلة تحت التحليل الاحصائي .**

**ويمكن إيجاز خطوات التحليل الاحصائي باستخدام الحاسب الآلي في الشكل التالي :**




**شكل (١-٢)**

**توضيحي عام خطوات التحليل الإحصائي باستخدام الحاسب الآلي**

**البرمجيات الاحصائية الجاهزة :**

**ويمكن تناول تعريف البرمجيات الاحصائية الجاهزة تحت نوعيين رئيسيين :**

**أ- البرمجيات الإحصائية البسيطة :**

**وهي برمجيات لا يتطلب تشغيلها كتابة أوامر معينة ، بل تعمل تحت مجموعة من الاختبارات ، والتي تشمل على العديد من الأساليب ، والطرق الإحصائية (الأقل تعقيدًا) ومنها البرنامج الإحصائي الجاهز Microstate .**

**ب- البرمجيات الإحصائية الأكثر تعقيدًا :**

**وهي برمجيات يتطلب تشغيلها كتابة مجموعة أوامر (حسب نوع العملية الإحصائية تحت الاستخدام) ، وذلك على مستوى البرمجيات التي كانت تعمل تحت نظام التشغيل Dos وبعد ظهور نظام التشغيل Windows سارعت شركات البرمجيات الجاهزة ، إلى إعادة توفيق برامجها لتعمل تحت نظام التشغيل الجديد ، مما جعلها أكثر سهولة ، ومتعة لدى المستخدمين ، حيث التعامل أصبح يتم من خلال الاختبارات بالتأشير باستخدام الماوس Mouse من مجموعة قوائم (وفق مسارات محددة ، ومن خلال مجموعة من النوافذ المتتابعة الظهور على سطح المكتب Desktop) ، والمسئولة عن تنفيذ العديد من الوسائل والطرق الإحصائية (المعلمية ، اللامعلمية) بدءًا من عمليات وصف البيانات ، إلى توصيف العلاقات (أو النماذج) الإحصائية والأكثر تعقيدًا ، ومنها الحزم الإحصائية الجاهزة Bmdp , Minitab , Sas , Spss , ..... وغيرها .**

**تعريف البرنامج الإحصائي الجاهز Spss :**

**وهو برنامج إحصائي متكامل يقوم بالعديد من عمليات التحليل الإحصائي البسيطة والمتقدمة ، وعلى الأخص في مجال العلوم الاجتماعية**

**(الظـواهر أو المتغيـرات الوصـفية) بدايــة مـن وصـف البيانــات إلــى التحليـل الإحصائي المتعدد .**

**<u>أهم العمليات الإحصائية تحت البرنامج الإحصائى Spss :</u>**

- Reliability Analysis **(لقياس ثبات ، ومصداقية البيانات) .**
- Store **(للترتيب التصاعدي أو التنازلي للبيانات) .**
- Frequencies **(لعمـل الجـداول التكراريـة ، العرض البيـاني ، حسـاب مقاييس إحصائية أولية [ الموضع ، التشتت ، الإلتواء والتفلطح ] ).**
- Cross tabs **(لعمل الجداول التكرارية المزدوجة)**
- Descriptive **(لحســاب العديــد مــن القياســات الإحصــائية الأوليــة للبيانات)**
- Tests of Hypotheses **(لإجـراء إختبـارات الفـروض الإحصـائية المعلمية) .**
- NPAR **(لإجراء اختبارات الفروض الإحصائية اللامعلمية) .**
- Graph **(لرسم أشكال الانتشار بين المتغيرات) .**
- Correlation **(لقيـاس معـاملات الارتبـاط بـين المتغيـرات الكميـة أو الوصفية) .**
- Regression **(لقيـاس الانحـدار الخطـي البسـيط ، المتعـدد ، غيـر الخطي بين المتغيرات) .**
- Cluster Analysis **(التحليل العنقودي) .**
- Discriminate Analysis **(تحليل التمايز) .**
- Factor Analysis **(التحليل العاملي) .**
- Logistic Regression **(الانحدار اللوجيستي) .**

- **(توفيق المنحنيات ، ومنها السلاسل الزمنية** Curve Estimation **.** (Time Series
- **وعمليات أخرى متنوعة .**

**<u>مع ملاحظة : أهم العمليات على البيانات :</u>**

**١- يمكن إجراء بعض التحويرات على البيانات ، سواء بخلق متغير جديد مكون من عدة متغيرات أصلية أو إعادة تعريف متغير أصلي أو إعادة تركيب متغير أصلي ... وهكذا .**

**٢- يمكن قصر التحليل الإحصائي على مجموعة جزئية من المتغيرات .**

**٣- يمكن فصل مجموعة من المتغيرات في ملف بيانات مستقل .**

**٤- ولمزيد من التفصيل حول كيفية التشغيل ، والتعامل مع البرنامج الأصلي الإحصائي الجاهز Spss فإنه يمكن الرجوع إلى أي من المراجع المتخصصة باللغة الإنجليزية أو العربية بالمكتبة .**

- **<u>أنواع العينات</u> :**

**سبق وأن ذكرنا أن العينة هي عبارة عن مجموعة مختارة من المجتمع وتمثل جزءاً صغيراً نسبياً من المجموع الكلي للمفردات المطلوب جمع بيانات عنها . وإذا ما تم اختيار مفردات العينة من المجتمع باسلوب إحصائي سليم فإن بياناتها تتسم بالدقة وقلة التكلفة والمشقة من بيانات الحصر الشامل .. ويمكن بدراسة خصائص العينة ( أو ما يسمى باحصائي العينة ) استنتاج الاحصائية المختلفة للعينة والتي يمكن تعميمها بالطرق الاحصائية لاستنتاج خصائص المجتمع المسحوبة منه تلك العينة . فيجب أن تتصف العينة بجميع خصائص المجتمع التي تمثله . لذا يلزم اختيار مفردات العينة بإحدى الطرق التي تضمن صحتها على قدر الإمكان . وتتلخص أهم العينات التي يمكن الحصول عليها بطريقة إحصائية جيدة من مجتمع الدراسة فيما يلي :-**

**١- العينة العشوائية البسيطة** Simple Random Sample

**وهي أبسط أنواع العينات وتتميز بأنها تعطي فرصاً متساوية لجميع مفردات المجتمع للظهور في العينة كما أن عامل الصدفة يكون هو العامل الوحيد المحدد للمفردات التي ستظهر بالعينة ولذلك فإن اختيار مفردات العينة يتم بأي طريقة عشوائية : ( أي بطريقة تضمن تساوي فرص الظهور في العينة أمام جميع مفردات المجتمع ) . فمثلاً عند اختيار طالبين من خمسة عشر طالباً بطريقة عشوائية يمكن استخدام ١٥ بطاقة متماثلة تماماً يكتب على كل منها اسم أحد الطلبة ثم تخلط البطاقات جيداً وتسحب إحدها فيكون الاسم المدون بها هو اسم أول طالب في العينة ثم تخلط البطاقات الباقية جيداً ويسحب منها بطاقة أخرى ويكون الاسم المدون بها هو اسم الاطالب الثاني في العينة ، ويمكن الاستعانة بجداول الإعداد العشوائية لتبسيط عملية الاختيار العشوائي من المجتمعات الكبيرة .**

**ويغلب على هذا النوع من العينات أن المفردات التي يتم اختيارها قد لا تمثل بعض فئات المجتمع كما أنها قد تكون منتشرة في أماكن متفرقة مما يؤدي إلى ارتفاع نفقات جميع البيانات . ويضاف إلى ذلك أنه إذا كان المجتمع محل البحث كبيراً فإن عملية الاختيار العشوائي تكون عملية مطولة ومملة وللتغلب على هذه العيوب تستخدم أنواع أخرى من العينات .**

**٢- العينة العشوائية المنتظمة** Systematic Random Sample

**والهدف من استخدام هذه العينة هو تجنب التكرار المطول الذي يظهر عند استخدام العينة العشوائية البسيطة . ويعتمد اختيار مفردات هذه العينة أساساً على تقسيم المجتمع إلى فئات متساوية ( عددها يساوي عدد مفردات العينة ) ثم اختيار مفردة واحدة من الفئة الأولى بطريقة عشوائية واختيار باقي المفردات من الفئات المتتالية بطريقة منتظمة .**

<u>فمثلاً</u> : إذا كان المجتمع مكون من ٤٨٠ مفردة وإذا كان المطلوب اختيار عينة حجمها ٦٠ مفردة ، فإننا نقسم المجتمع إلى ٦٠ فئة يتكون كل منها من ٨ مفردات : ( ٤٨٠/٦٠ = ٨ ) وتكون أرقام مفردات الفئة الأولى هي : ١ ، ٢ ٠٠٠٠ ، ٨ ، وأرقام مفردات الفئة الثانية هي : ٩، ١٠ ، ١٦٠٠٠٠٠ وهكذا . ثم نبدأ باختيار أحد مفردات الفئة الأولى بطريقة عشوائية ولنفرض أنها كانت المفردة رقم ٤ ، فتكون الخطوة التالية هي اختيار مفردة من الفئة الثانية بطريقة منتظمة فيكون رقمها هو ٤ + ٨ = ١٢ ثم نختار مفردة من الفئة الثالثة بنفس الطريقة فيكون رقمها ٤ + ١٦ = ٢٠ أو ١٢+ ٨ وهكذا ٠٠٠٠

فنجد أن المفردات التي تظهر في العينة هي المفردات المناظرة للأرقام :

٤ ، ١٢ ، ٢٠ ، ٢٨ ، ٢٦ ، ٠٠٠٠٠ ، ٤٦٨ ، ٤٧٦ .

ويجب أن لا يستخدم هذا النوع من العينات إذا كان هناك علاقة بين صفات المفردة وطريقة توزيع المفردات على الفئات المجتلفة للمجتمع ، وذلك لأن المفردات التي ستظهر في العينة في هذه الحالة سيكون لها صفات خاصة وبالتالي لن تكون ممثلة للمجتمع الذي سحبت منه .

<u>٣- العينة الطبقية</u> Stratified Sample

إذا كان المجتمع محل الدراسة مكون من مجموعات ( أو طبقات ) غير متجانسة وإذا أردنا التأكد من تمثيل كل مجموعة ( أو طبقة ) من هذه المجموعات في العينة فإننا نستخدم اسلوبٍ العينة الطبقية الذي يعتمد على اختيار عينة ( بالطريقة العشوائية البسيطة أو المنتظمة ) من كل طبقة . ومجموع هذه العينات يمثل العينة الكلية .

ويتوقف حجم العينة التي يتم اختيارها من كل طبقة على أحد العوامل التالية:

**أ – <u>حجم الطبقة</u> : فحجم العينمة من طبقة معينة يتناسب طردياً مع حجم هذه الطبقة .**

**ب– <u>مدى التجانس بين مفردات الطبقة</u> : فكلما ازداد التجانس بين مفردات الطبقة الواحدة كلما أدى ذلك إلى تخفيض حجم العينة المختارة من هذه الطبقة .**

**جـ – <u>تكاليف الحصول على بيانات المفردة من كل طبقة</u> : فكلما زادت نفقات جمع البيانات من المفردة في طبقة معينة كلما أدى ذلك إلى تخفيض حجم العينة من هذه الطبقة .**

**<u>٤ـ العينة المتعددة المراحل</u>** Multi-stage Sample

**إذا كان المجتمع محل الدراسة كبيراً وبه الكثير من الاختلافات والعوامل المؤثرة على المفردات . وإذا كان من المرغوب فيه اختيار عينة ممثلة لهذا المجتمع بما فيه من اختلافات دون أن تكون نفقاتها مرتفعة فإنه يمكن اختيار العينة على مرحلتين أو أكثر .**

**مثال ذلك إذا كان المطلوب هو استخدام اسلوب العينة لجمع بيانات عن كيفية اتفاق دخل الأسرة في الجمهورية العربية ، فإنه يمكن – كمرحلة أولى – تقسيم الجمهورية إلى مجموعات متشابهة من المحافظات ( مثلاً وجه قبلي ، وجه بحري ، الواحات ، ٠٠٠ )**

**واختيار عينة من كل مجموعة ، وتبدأ المرحلة الثانية باختيار مجموعة من مدن وقرى المحافظات التي ظهرت في المرحلة الأولى . ثم تأتي المرحلة الثالثة وهي اختيار مجموعة من أسر القرى والمدن التي ظهرت في المرحلة الثانية . وبذلك نكون قد استخدمنا عينة من ثلاث مراحل وتمتاز هذه العينة بأن الأسر التي تظهر في المرحلة الأخيرة تكون مركزة في مجموعة محدودة من المدن والقرى وبالتالي فإن نفقات جمع البيانات لن تكون مرتفعة . كما أنه**

نتيجة الطريقة المتبعة في الاختيار أصبحت العينة ممثلة لجميع أنحاء الجمهورية دون أن يكون حجمها مبالغاً فيه .

د- <u>الجداول العشوائية وكيفية استخدامها</u> :

الجداول العشوائية ما هي إلا الأرقام ٠ ، ١ ، ٢ ، ..... ، ٩ مرتبة في أعمدة متتالية تبعًا لترتيب الحصول عليها بإحدى الطرق العشوائية .

ولتوضيح كيفية استخدام هذه الجداول لنفرض أننا نريد اختيار عينة عشوائية بسيطة مكونة من أربع عمال شركة إنتاجية معينة وعددهم ٦٥٠ عاملاً . في هذه الحالة نبدأ بتكوين "الإطار" وهو قائمة تشمل أسماء جميع العمال بالشركة (دون حذف أو تكرار) وتعطى لكل منهم رقمًا مسلسلاً من ٠٠١ وحتى ٦٥٠ وبذلك نعتبر أن كل عامل يمثله عدد واحد (مكون من ثلاث خانات) فنختار ثلاث أعمدة من جداول الأعداد العشوائية ونقرأ أربعة أعداد متتالية وكل منهم يمثل أحد العمال اللذين يتم اختيارهم في العينة . فمثلاً إذا اخترنا الأعمدة الثلاث الأخيرة من الجدول المعطى في نهاية هذه المذكرة نجد أن الأربعة أعداد التي نحصل عليهم هي : ٥٠٣ ، ٥٢٩ ، ١٠٢ ، ١١٣ وكل منها يمثل أحد العمال في العينة (يلاحظ أننا تجاهلنا الرقم ٧٠٢ لأنه أكبر من ٦٥٠ ، أي لا يوجد عامل في القائمة التي لدينا يناظر هذا العدد) .

وإذا كان المطلوب هو اختيار عينة عشوائية بسيطة من مجتمع حجمه ١٠٠ مفرده فإنه يمكن توفير الكثير من الجهد المرتبط بعملية الاختيار العشوائي إذا تم إعداد القائمة باستخدام الأرقام ٠٠ ، ٠١ ، ٠٢ ، ... ، ٩٩ أو الأرقام ٠١ ، ٠٢ ، ... ، ٩٩ ، ٠٠ (أي .. بدلاً من ١٠٠) لأنه في هذه الحالات يتم استخدام عمودين فقط من أعمدة الجدول العشوائي بدلاً من استخدام ثلاثة

( إذا استخدمت الأعداد ١ ، ٢ ، ٠٠ ، ١٠٠ فإنه يجب أن نعتبر كل منها مكون من ثلاث خانات أي : ٠٠١ ، ٠٠٢ ، ٠٠٠ ، ٠٩٩ ، ١٠٠ حتى يكون لجميع الأعداد نفس فرصة الظهور في العينة ) .

كذلك إذا كان حجم المجتمع هو ١٧٠ مثلاً فيجب استخدام ثلاثة أعمدة ولكن بدلاً من إهمال جميع الأعداد من ١٧٠ أو حتى ٩٩٩ فإنه يمكن اعتبار الرقم المسلسل ١ مقابلاً للأعداد ٠٠١ ، ٢٠١ ، ٤٠١ ، ٦٠١ ، ٨٠١ . والرقم المسلسل ٢ مقابلاً للأعداد ٠٠٢ ، ٢٠٢ ، ٤٠٢ ، ٦٠٢ ، ٨٠٢ . وهكذا ٠٠ فالرقم المسلسل ١٧٠ يكون مقابلاً للأعداد ١٧٠ ، ٣٧٠ ، ٥٧٠ ، ٧٧٠ ، ٩٧٠ .

## الباب الثاني
## أساليب عرض البيانات الإحصائية

**بعد أن تتم عملية جمع البيانت الإحصائية يكون دور الباحث في المرحلة التالية هو عملية عرض البيانات في صورة واضحة ومكتملة بحيث يمكن الاستفادة منها . حيث يتم عرض تلك البيانات قبل إجراء الأساليب الخاصة بعملية التحليل الإحصائي وحتى يتسنى على الباحث استخدام الأسلوب الإحصائي الملائم ومن ثم استخلاص النتائج الإحصائية الصحيحة منها . هذا ويمكن عرض البيانات الإحصائية من خلال الطرق الثلاث التالية:**

- **<u>تبويب البيانات وعرضها في صورة جداول تكرارية</u> (العرض الجدولي للبيانات الإحصائية) . وهذه الطريقة تتيح لنا عملية عرض البيانات مهما كان حجمها في صورة مختصرة ومبسطة يسهل فهمها .**
- **<u>عرض البيانات الإحصائية بيانيًا</u> (العرض البياني) وذلك من خلال استخدام الرسوم أو الأشكال البيانية الملائمة للظاهرة موضع الدراسة والتي تساعد بمجرد النظر إليها فهم تلك الظاهرة حتى لغير المتخصصين .**
- **<u>إيجاد بعض المقاييس الإحصائية المعروفة لتلك البيانات</u> مثل مقاييس المتوسطات والتشتت وعلاقة الارتباط أو الانحدار بين مجموعة من الظواهر موضع الدراسة إلى غيرها من تلك المقاييس الإحصائية المختلفة.**
- **هذا وسوف يقتصر هذا الباب على دراسة طرق العرض الجدولي والبياني فقط أما الطريقة الثالثة – إيجاد المقاييس الإحصائية المختلفة – فسوف تكون موضع الدراسة للأبواب التالية بمشيئة الله .**

لكن قبل أن يتم تناول طريقة العرض الجدولي والبياني يجب أن ننوه إلى أن كلاً من هاتين الطريقتين تعتمد في البداية على تحديد نوع أو طبيعة المتغير أو الظاهرة موضع الدراسة .

فالمتغير الإحصائي هو عبارة عن ظاهرة تأخذ قيمًا – أو أوجهًا – مختلفة تسمى بعناصر نطاق المتغير . والمتغير الإحصائي يتم تقسيمه إلى قسمين رئيسيين هما :

**الأول : المتغير الوصفي**: Qualitative Variable

وهو ذلك المتغير الذي تكون فيه عناصر نطاق المتغير ممثله في صورة صفات ولا يأخذ قيمًا عددية . مثل ظاهرة الحالة الاجتماعية (أعزب – متزوج – مطلق – أرمل ) أوظاهرة تقدير الطالب (ضعيف جدًا – ضعيف – مقبول – جيد – جيد جدًا – ممتاز ) أوظاهرة النوع أو الجنس (ذكر – أنثى) أو الجنسية أو الديانة أو ...... إلخ .

هذا ويستلزم التحليل الإحصائي للبيانات أو الظواهر الوصفية القيام بعملية ترميز **Coding** لتلك البيانات وذلك من خلال تخصيص قيمًا رقمية لكل عنصر من عناصر نطاق المتغير أو الظاهرة موضع الدراسة أو لكل مستوىمن مستويات هذه الظاهرة لتحل محل هذا العنصر أو هذا المستوى دون أن تعكس قيمة كمية تتعلق بعناصر نطاق أو مستويات الظاهرة موضع الدراسة . ويعتبر الترميز أدنى مستويات القياس للمتغيرات أو الظواهر ويسمى بالقياس الاسمي Nominal

**الثاني : المتغير الكمي :** Qualitative Variable

وهو ذاك المتغير الذي تمثل عناصر نطاقه مجموعة الأعداد الحقيقية. ويمكن تقسيمه إلى قسمين رئيسيين وهما :

- **<u>متغير كمي منقطع (منفصل) :</u>** Discrete Variable **وهو متغير كمي عناصر نطاقه قيمًا عددية صحيحة فقط (أي قيمًا لا تقبل التقسيم أو التجزئة) . مثال ذلك المتغير أو الظاهرة ، التي تعبر عن حجم أو عدد أفراد الأسرة (١ ، ٢ ، ٣ ، ....... ) وكذلك المتغير الذي يعبر عن عدد الوحدات المنتجة من سلعة لا تقبل التقسيم أو التجزئة مثل عدد الوحدات المنتجة من الثلاجات أو الغسالات أو ..... إلخ .**
- **<u>متغير كمي متصل (مستمر) :</u>** Continuous Variable

**وهو متغير كمي عناصر نطاقه قيمًا عددية صحيحة أو كسرية . مثال ذلك المتغير الذي يعبر عن درجات الطلاب أو وزن مجموعة من الطلاب أو المتغير الذي يعبر عن الطول أو المساحة أو درجات الحرارة المسجلة في عواصم العالم أو الدخل أو الأجر اليومي أو الأسبوعي أو الشهري أو ...إلخ .**

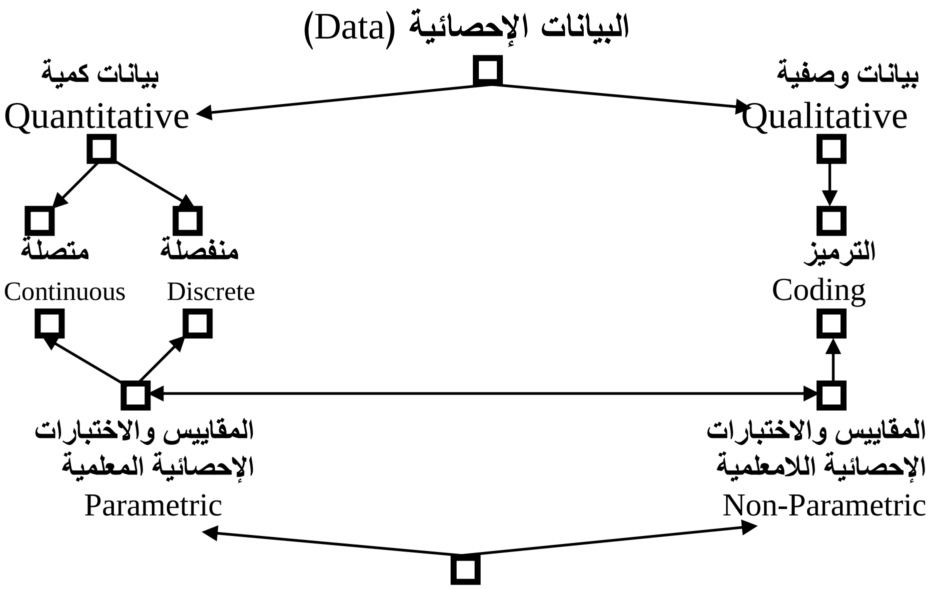
**هذا ويجب التنويه إلى أن البيانات الإحصائية غالبًا ما تكون غير كافية بحد ذاتها بل تستخدم للتوصل إلى استنتاجات معينة وهو ما يعرف بعملية التحليل أو المعالجة الإحصائية للبيانات . ويقوم علم الإحصاء بالدور الكبير في هذا الخصوص .**

**وعلى الدارس أن يكون على دراية كافية بالبيانات محل الظاهرة موضوع الدراسة عند اختيار الطرق والأساليب الإحصائية المستخدمة في عمليات التحليل الإحصائي للبيانات .**

**ويمكن عرض مجمل كافة المقاييس والاختبارات الإحصائية المستخدمة في عمليات التحليل الإحصائي للبيانات تحت فرعين أساسيين :**

**<u>المقاييس والاختبارات الإحصائية المعلمية</u>** <u>Parametric</u> **: ولها أولوية الاستخدام (ما دامت الشروط الإحصائية متحققة) ، وإلا استخدمت المقاييس والاختبارات اللامعلمية عند التعامل مع البيانات الكمية .**

والآخر <u>المقاييس والاختبارات الإحصائية اللامعلمية</u> Non-Parametric : ولها أولوية الاستخدام (ما دامت الشروط الإحصائية متحققة) عند التعامل مع البيانات الوصفية . غير أن اتساع دوائر استخدامات الحاسب الآلي ، ومع التطوير الموازي والمتسارع في الحزم الإحصائية الجاهزة ، قد سهل الأمر على كافة المستخدمين من غير الإحصائيين ، عند التعامل مع المجموعات الكبيرة من البيانات على مستوى العرض ، والتحليل الإحصائي بسرعة ودقة كبيرتين (وإن كان يفضل دائمًا عند التعامل مع التحليل الإحصائي للبيانات أن تتم بمعرفة أو تحت إشراف إحصائي) .ويمكن تقديم عرض تلخيصي لمجمل موضوع البيانات ومعالجتها الإحصائية في الشكل التوضيحي التالي .




شكل (٢) : الطرق والأساليب الإحصائية

<u>أولاً : العرض الجدولي للبيانات الإحصائية :</u>

سبق حالاً وأن ذكرنا أن البيانات الإحصائية تنقسم إلى قسمين رئيسيين هما البيانات الوصفية والبيانات الكمية . حيث يتعلق النوع الأول من

البيانات (الوصفية) بالصفات . بينما البيانات الكمية فهي التي يمكن التعبير عنها في شكل أرقام (منفصلة أو متصلة) أي أنها قابلة للقياس الكمي.

وبصفة عامة فأيًا كان نوع البيانات المستخدمة فإن العرض الجدولي للبيانات الإحصائية يتطلب المرور بالمراحل الثلاث التالية :

- تحديد أقسام (أو ما تسمى بفئات) الظاهرة أو المتغير محل الدراسة والمراد بتبويبها جدوليًا .
- عملية التفريغ في الجدول (تحديد العلاقات أو الحزم التكرارية) أمام كل قسم من أقسام الظاهرة أو المتغير .
- عملية العد (أي ترجمة العلامات أو الحزم التكرارية إلى أرقام) .

حيث تختلف عملية تحديد أقسام الظاهرة أو المتغير محل الدراسة وهي بمثابة المرحلة الأولى وذلك طبقًا لطبيعة الظاهرة أو المتغير من حيث كونه وصفيًا أو كميًا (منفصلاً أو متصلاً) . أما المرحلتين الثانية (عملية التفريغ) والثالثة (عملية العد) فهي لا تختلف في طبيعتها باختلاف طبيعة أو نوع الظاهرة أو المتغير محل الدراسة .

وفيما يلي نتعرض بالتفصيل لدراسة تلك المراحل الثلاث لكل نوع من أنواع المتغيرات أو الظواهر محل الدراسة مبينين ذلك من خلال تناول العديد من الأمثلة التوضيحية .

**أولاً : مراحل العرض الجدولي للبيانات الوصفية :**

حيث يتم تحديد كافة عناصر نطاق الظاهرة أو المتغير محل الدراسة التي تنقسم إليها البيانات . ويجب ألا تحتوي أقسام الظاهرة على أي قسم لا ينتمي إليه أي مفرده من مفردات الظاهرة محل الدراسة . ثم يلي عملية تحديد أقسام (أو فئات) الظاهرة عملية التفريغ في الجدول ويليها أخيرًا عملية العد كما توضحه الأمثلة التالية :

**<u>مثال (١)</u> :**

**فيما يلي تقديرات نجاح ٢٠ طالبًا في مادة الإحصاء :**

**جيد جدًا – جيد – مقبول – جيد – ممتاز – جيد – جيد جدًا – جيد – جيد جدًا – مقبول – جيد – جيد جدًا – جيد – جيد – جيد – مقبول – جيد جدًا – جيد جدًا – مقبول – ممتاز .**

**<u>الحل :</u>**

**من خلال استعراض البيانات الخام يتضح لنا أن الظاهرة موضع الدراسة (تقديرات نجاح الطلاب في مادة الإحصاء) تحتوي على أربعة أقسام وهي : مقبول ، جيد ، جيد جدًا ، ممتاز . لذا فإنه يتم رسم جدول من ثلاثة أعمدة ، تعكس هذه الأعمدة الثلاث مراحل تبويب هذه البيانات جدوليًا .**

**حيث يتم سرد أقسام الظاهرة الأبعة في العمود الأول تحت اسم فئات التقدير (أو تقديرات النجاح) وتفضل وصفها في صورة مرتبة سواء تنازليًا أو تصاعديًا . بعد ذلك يتم قراءة التقديرات الخاصة بالطلاب العشرين الواحد تلو الآخر ونضع علامة مناظرة لكل تقدير يتم قراءته في العمود الثاني . فإذا ما بدأنا في قراءة التقديرات كما هي معطاة في المثال الذي نحن بصدده فإننا نبدأ بوضع علامة مناظرة للتقدير جيد جدًا في العمود الثاني تعبر تلك العلامة عن تقدير نجاح الطالب الأول في هذه المجموعة ثم نبدأ في وضع علامة مناظرة أمام التقدير جيد تعبر عن تقدير نجاح الطالب الثاني ثم علامة مناظرة أمام التقدير مقبول ثم علامة مناظرة أمام القدير جيد وهكذا حتى يتم تفريغ كل المفردات المعبرة عن التقديرات والتي تنتهي بتقدير ممتاز . مع مراعاة أنه لتسهيل عملية العد فإنه اُتُفِق على أن تكون العلامة المعبرة عن المفردة عبارة عن شرطة مائلة ( / ) وأن تظهر العلامات في شكل حزمة تحتوي على خمسة علامات . ولتجنب الوقوع في أخطاء عند العد فإن العلامة الخامسة – إن**

وجدت – توضع بطريقة عكسية لوضع العلامات الأربعة التي تسبقها . وبذلك تأخذ الحزمة الصورة ////. وفي واقع الأمر فإن الحزم تساعد كثيرًا في تسهيل عملية العد النهائية حيث يكون معروفًا ان كل حزمة تحتوي على خمس مفردات .هذا ويجب أن يراعى عند تسطير الجدول الذي ستتم فيه عملية التبويب أن يكون العمود الثاني متسعًا إلى حد ما حتى يتسنى ظهور الحزم بشكل واضح وذلك من خلال ترك فراغ ملائم فيما بين كل حزمة وأخرى . ومما لا شك فيه أن درجة اتساع هذا العمود تتوقف على عدد المفردات أو حجم البيانات المطلوب تبويبها .

والجدول التالي يلخص المراحل الثلاث لعملية تبويب بيانات تقديرات نجاح الطلاب جدوليًا :

جدول (٢–١)

عملية تفريغ تقديرات نجاح عشرون طالبًا في مادة الإحصاء

| تقديرات النجاح (الفئات) | العلامات | عدد الطلاب (التكرار) |
|---|---|---|
| مقبول | //// | ٤ |
| جيد | //// /// | ٨ |
| جيد جدًا | //// / | ٦ |
| ممتاز | // | ٢ |
| المجموع | | ٢٠ |

ملحوظة : يجب ألا تحتوي أقسام الظاهرة (الفئات) على أي قسم لا تنتمي إليه أي من مفردات الظاهرة أو المتغير محل البحث والدراسة سواء كانت البيانات وصفية أو كمية . بمعنى أنه في المثال السابق لا يوجد في عمود الفئات مجموعة تقديرات الطلاب الراسبين (ضعيف جدًا ، ضعيف) حيث أن البيانات

المعطاة هي تقديرات نجاح مجموعة من الطلاب لا تحتوي على هذين التقديرين (ضعيف جدًا ، ضعيف) .

**ثانيًا : مراحل العرض الجدولى لبيانات الظواهر الكمية :**

في تصنيفنا للبيانات الإحصائية سبق وأن ذكرنا أن البيانات الإحصائية الكمية هي تلك البيانات التي يمكن قياسها رقميًا . وذكرنا أن البيانات الكمية يمكن تصنيفها إلى متغير كمي متقطع (منفصل) ومتغير كمي متصل (مستمر) .

- وبصفة عامة فإن العرض الجدولي للبيانات الإحصائية الكمية المتقطعة لا يختلف عن العرض الجدولي للظواهر الوصفية سالفة الذكر . حيث يتم سرد كافة القيم التي تنتمي لعناصر نطاق هذا المتغير في عمود أقسام الظاهرة (الفئات) . ثم يتم تحديد عدد العلامات (الحزم) التي تناظر كل قسم من أقسام المتغير الكمي المتقطع ثم ترجمة هذه العلامات إلى أرقام (مرحلة العد) . والمثال التالي يوضح ذلك . هذا ويمكن أن نلجأ إلى تجميع القيم التي تزيد عن قيمة معينة كبيرة أو القيم التي تقل عن قيمة معينة صغيرة في قسم واحد من أقسام الظاهرة . وذلك في حالة ما إذا كانت هذه القيم قليلة من حيث التكرار .

**مثال (٢) :**

فيما يلي البيانات التالية تعبر عن أحجام ٣٠ أسرة :

٥ ، ٤ ، ٣ ، ٥ ، ٣ ، ٢ ، ٤ ، ٦ ، ١ ، ٣ ، ٥ ، ٤ ، ٣ ، ٢ ، ٥ ، ٤ ، 
٢ ، ٣ ، ٦ ، ٥ ، ٢ ، ٣ ، ٤ ، ١ ، ٢ ، ٣ ، ٤ ، ٢ ، ١ ، ٤

والمطلوب تبويب هذه الظاهرة (تفريغها) جدوليًا .

**الحل** :لاحظ أن الظاهرة محل الدراسة في هذا المثال هي حجم الأسرة وهذا الحجم لا شك أنه يعبر عن متغير كمي متقطع (لأنه يأخذ قيمًا منفصلة ولا

يمكن أن يقبل التقسيم أو التجزئة) لذا يتم تفريغ هذه البيانات من خلال استعراض الأحجام المختلفة لمجموعة الأسر المعطاة . أي أن أقسام الظاهرة عبارة عن مجموعة الأرقام التالية ١ ، ٢ ، ٣ ، ٤ ، ٥ ، ٦ . فإذا بدأنا بقراءة أحجام الأسر واحدًا تلو الآخر ملتزمين بترتيب المفردات كما هي واردة بالمثال (٢) فأبنا نبدأ بوضع علامة مناظرة للحجم (٥) في العمود الثاني بالجدول (٢ – ٢) ثم علامة مناظرة للحجم (٤) ثم علامة مناظرة للحجم (٣) .... وهكذا حتى يتم تفريغ الجدول حسب أحجام الأسر والذي يوضحه الجدول التالي :

**جدول (٢–٢)**

**تفريغ أحجام ٣٠ أسرة**

| حجم الأسرة | العلامات | عدد الأسر |
|---|---|---|
| ١ | /// | ٣ |
| ٢ | ~~////~~ / | ٦ |
| ٣ | ~~////~~ // | ٧ |
| ٤ | ~~////~~ // | ٧ |
| ٥ | ~~////~~ | ٥ |
| ٦ | // | ٢ |
| المجموع | | ٣٠ |

- **أما عن مراحل العرض الجدولى للبيانات الإحصائية للظواهر الكمية المتصلة:** وهي تلك الظواهر أو المتغيرات التي تأخذ قيمًا عددية صحيحة أو كسرية . مثل أوزان أو أطوال أو درجات مجموعة من الطلاب أو مرتبات أو أجور أو أعمار مجموعة من العاملين وغير ذلك من المتغيرات التي يمكن

أن يحتوي مداها على عددًا لا نهائيًا من القيم . فإنه في المرحلة الأولى وهي تحديد أقسام (فئات) الظاهرة محل الدراسة يتم إجراء عملية تقسيم لقيم الظاهرة التي يمثلها المتغير المتصل إلى فئات أو فترات بحيث تضم كل فئة أو فترة قيمًا متتالية تنحصر ما بين حدي كل فئة والتي يطلق على أحدهما بالحد الأدنى Lower Bound (Limit) والآخر اسم الحد الأعلى لأي فئة يعتبر بمثابة الحد الأدنى Upper Bound (Limit) للفئة . وذلك مع اعتبار أن الحد الأعلى لأي فئة يعتبر بمثابة الحد الأدنى للفئة التالية لها . وفي هذه الحالة لكي يتم تحديد عدد الفئات والحد الأدنى والأعلى لكل فئة فإنه يتم اتباع الخطوات التالية :

١- <u>تحديد المدى</u> Range

والمدى إحصائيًا ما هو إلا الفرق ما بين أكبر وأصغر قيمة في مفردات المتغير أي أن :

المدى = أكبر قيمة – أصغر قيمة + ١

أي أن :

Range (R) = Maximum value – Minimum value + 1

٢- <u>تحديد عدد الفئات</u> The Number of Intervals or Classes (NI)

وهناك عددًا من القواعد المختلفة التي تساعد في حساب عدد الفئات المناسب . إلا أن هذه القواعد تتفق في النهاية على أساس واحد وهو وجوب أن يتناسب عدد فئات الجدول التكراري طرديًا مع عدد المفردات ولتكن(n) مفرده . وأشهر هذه القواعد في تحديد عدد الفئات قاعدة ستورجس Sturges والتي تفيد بأن عدد الفئات الذي يتناسب مع عدد (n) من المفردات يستنتج من القانون التالي :

NI = 1 + 3.3 Log (n)

**وهذه القاعدة تعطي نتائج مناسبة حينما يتراوح إجمالي عدد المفردات ما بين ١٠٠ ، ١٠٠٠ مفرده .**

**وبصفة عامة يجب مراعاة ألا يكون عدد الفئات كبيرًا حتى لا يتنافى مع أحد أهداف عملية التفريغ وتكوين الجدول التكراري وهو عرض البيانات بصورة مختصرة يسهل معها فهم البيانات . كما يجب ألا يكون عدد الفئات صغيرًا لأن ذلك يؤدي إلى طمس معالم التوزيع . وفي النهاية فإن تحديد عدد الفئات المناسب قد يكون اختياريا متروك للتقدير الشخصي لمن يقوم بتفريغ وأعداد الجداول التكرارية . وعمومًا فإنه من المفضل أن يتراوح عدد الفئات ما بين ٦ ، ١٠ فئات ويجب ألا يزيد عن ٢٠ فئة .**

٣- **تحديد طول الفئة** (Interval Length (I L (or) Width (W))

**بعد تحديد عدد الفئات السابقة يتم تحديد طول الفئة (I L) من خلال قسمة المدى (R) على عدد الفئات (NI) أي أن :**

$$\text{طول الفئة} = \frac{\text{المدى}}{\text{عدد الفئات}}$$

**أي أن :** $I\,L(or)\,W = R / NI$

**وينتج في تلك الحالة طول فئة ثابت لكل فئات التوزيع أي يعطي جدولاً تكراريًا متساويًا من حيث أطوال فئاته . هذا ويفضل أن تكون أطوال فئات الجدول التكراري متساوية حتى يسهل معها عرض البيانات سواء جدوليًا أو بيانيًا كما سنرى فيما بعد . مع مراعاة في حالة ما إذا كان ناتج عملية حساب طول الفئة رقمًا كسريًا فيفضل تقريب الناتج إلى أقرب عدد صحيح . فمثلاً إذا كان المدى المحسوب في الخطوة الأولى وليكن ٢٠٠ وكان عدد الفئات هو ٧ فئات فإن ناتج القسمة هو ٢٨.٥٧١ يتم تقريبه ليصبح طول الفئة ٢٩ وحدة قياس تابعة لنفس وحدة قياس الظاهرة أو المتغير محل الدراسة .**

هذا ويفضل اختيار عدد الفئات رقمًا بحيث يكون ناتج قسمة المدى عليه يعطي طول الفئة مساويًا للرقم (٥) أو (١٠) أو مضاعفاتهم .

هذا وتجدر الإشارة إلى أنه في بعض الحالات تستلزم طبيعة البيانات استخدام فئات غير متساوية من حيث أطوالها وذلك نظرًا لقلة التكرارات المقابلة لهذه الفئات . أو قد تكون فئات الجدول أو التوزيع التكراري محددة سلفًا في صورة فئات غير متساوية من حيث أطوالها .

وبصفة عامة يطلق على الجدول أو التوزيع التكراري المتساوي من حيث أطوال فئاته بالتوزيع المنتظم . أما الجدول التكراري الذي تشذ فيه أحد فئاته على الأقل عن باقي فئات التوزيع فيسمى بالجدول التكراري الغير منتظم.

٤- <u>تحديد الحد الأدنى والحد الأعلى لكل فئة من فئات التوزيع :</u>

عند تحديد الحد الأدنى للفئة الأولى من فئات التوزيع التكراري يراعى ألا يزيد هذا الحد عن أقل قيمة من قيم المشاهدات أو الظاهرة محل الدراسة وإلا فإن المفردة التي تأخذ قيمة أقل من هذا الحد الأدنى للفئة الأولى لن يتم تمثيلها في الجدول . كما يجب ألا يقل الحد الأعلى للفئة الأخيرة من فئات الجدول أو التوزيع التكراري عن أكبر قيمة في قيم المشاهدات وإلا فإن ذلك يعني ان المفردة التي تأخذ قيمة أكبر من الحد الأعلى للفئة الأخيرة لن يتم تمثيلها في الجدول التكراري .

وخلاصة ما سبق هو أن الحد الأدنى للفئة الأولى يأخذ قيمة تقل عن أو تساوي أقل قيمة في قيم المفردات . ثم يضاف لهذا الحد الأدنى طول الفئة الناتج في الخطوة السابقة ليعطي الحد الأدنى للفئة الثانية . ويضاف للحد الأدنى للفئة الثانية طول الفئة لينتج الحد الأدنى للفئة الثالثة و ..... وهكذا حتى نصل للحد الأدنى للفئة الأخيرة .

أما عن الحد الأعلى لكل فئة من فئات التوزيع فلتجنب أي لبس فقد أُتُفِقَ على أن تنتمي كل فئة بحد أعلى هو عبارة عن القيمة الأقل مباشرة من الحد الأدنى للفئة التالية . بمعنى إذا تم تحديد فئات الأجر اليومي لمجموعة من العاملين وكانت على الصورة :

٢٠- وتقرأ : من ٢٠ جنيه حتى أقل من ٣٠ جنيه .

٣٠- وتقرأ : من ٣٠ جنيه حتى أقل من ٤٠ جنيه .

٤٠- وتقرأ : من ٤٠ جنيه حتى أقل من ٥٠ جنيه .

٥٠- وهكذا ......

أي أنه وفقًا لهذه الطريقة فإننا نكتب الحد الأدنى لكل فئة فقط تاركين الحد الأدنى ليتحدد ضمنيًا بقيمة أقل من الحد الأدنى للفئة التالية لها . وعلى ذلك فالعامل الذي يتقاضى أجرًا يوميًا وليكن ٢٩.٥ جنيه يقع في الفئة الأولى ( ٢٠- ) .

وبعد تحديد الحد الأدنى والحد الأعلى لكل فئة من فئات التوزيع تتم عملية تفريغ البيانات من خلال قراءة المفردات الواحدة تلو الأخرى والتعبير عن كل مفردة بشرطة مائلة ( / ) أمام الفئة المقابلة إلى أن تنتهي عملية تفريغ كل المفردات .

٥- <u>مرحلة العد</u>

وفي تلك المرحلة يتم الحصول على الجدول التكراري من خلال استبعاد عمود العلامات وذلك بعد ترجمة العلامات إلى أرقام التي تعبر عن التكرارات والمثال التالي يوضح كيفية تكون جدول تكراري لبيانات متغير متصل(مستمر).

<u>مثال (٣) :</u>

فيما يلي البيانات التالية تمثل الأجور الأسبوعية بالجنيه التي حصل عليها ٤٠ عامل في أحد المصانع .

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| ٤٠ | ٤٩ | ٤٣ | ٣٥ | ٤٢ | ٤٣ | ٤٦ | ٣٦ |
| ٤٢ | ٣٦ | ٣٧ | ٤٤ | ٣٩ | ٤١ | ٣١ | ٣٥ |
| ٣٨ | ٤٨ | ٤٤ | ٥١ | ٣٨ | ٥٣ | ٣٥ | ٣٢ |
| ٣٠ | ٤٣ | ٤١ | ٥٢ | ٤٦ | ٤٣ | ٥٠ | ٤٠ |
| ٣٩ | ٤١ | ٤٨ | ٤٧ | ٣٢ | ٥٢ | ٤٧ | ٤٢ |

**والمطلوب تفريغ هذه البيانات في صورة جدول تكراري .**

**الحل :**

**لتفريغ بيانات أجور العاملين المعطاه يتم الآتي :**

**١- حساب المدى :**

**حيث أن**

$$R = Max - Min + 1$$
$$= 53 - 30 + 1 = 24 \text{ L.E}$$

**٢- تحديد عدد الفئات :**

**أما أن يتم تحديده عددًا اختياريًا وليكن ستة فئات مثلاً أو باستخدام قاعدة ستورجس فيصبح :**

$$NI = 1 + 3.3 \text{ Log } n$$
$$= 1 + 3.3 \text{ Log } 40$$
$$= 1 + 3.3\ (1.6021)$$
$$= 6.2868$$
$$= 6 \quad \text{(Classes)}$$

**٣- تحديد طول الفئة :**

**فحيث أن**

$$I\,L = \frac{R}{NI}$$

$$I\,L \text{ (or) } W = R / NI = 24/6 = 4 \quad \text{L.E}$$

**ومن ثم فإن فئات التوزيع تأخذ الصورة التالية :**

٣٠- ، ٣٤- ، ٣٨- ، ٤٢- ، ٤٦- ، ٥٠- ٥٤

وهي عبارة عن ستة فئات كما سبق وإن حددناه في الخطوة السابقة .

٤- <u>تحديد الحد الأدنى والأعلى لفئات الأجور الأسبوعية</u> : وتفريغ الجدول كما هو موضح بالجدول التالي حيث يتم تمثيل فئات الأجور في العمود الأول ثم تحديد العلامات في العمود الثاني وذلك من خلال توزيع بيانات الأجور المعطاة على فئات الأجور التي تنتمي إليها . فنحصل على الجدول التكراري (٣-٢) . فبقراءة المفردات المعبرة عن أجور العمال الأسبوعية المفردة تلو الأخرى وذلك بنفس الترتيب الوارد في بيانات مثال (٣) فأننا نجد أن قيمة الأجر ٤٠ جنيه تنتمي إلى الفئة (٣٨-) لذا يتم وضع علامة ( / ) أمام تلك الفئة . بينما نجد أن قيمة الأجر الثاني وهو ٤٩ جنيه تنتمي إلى الفئة (٤٦-) لذا يتم وضع علامة أمام تلك الفئة . أما المفردة الثالثة وهي قيمة الأجر ٤٣ جنيه فهي تنتمي للفئة (٤٢-) فيتم وضع علامة أمام تلك الفئة و ..... و هكذا حتى نصل للمفردة الأخيرة وهي قيمة الأجر ٤٢ جنيه وهي تنتمي للفئة (٤٢-) فيتم وضع علامة أمامها . ومن ثم تنتهي عملية التفريغ . فيتم ترجمة تلك العلامات إلى أرقام تعبر عن التكرارات المقابلة لكل فئة من فئات الأجور الأسبوعية .

هذا ولكي نتأكد من أن عملية التفريغ شملت كافة المفردات الموضحة بالمثال يجب أن يكون مجموع التكرارات الموضحة في العمود الثالث مساويًا لعدد المفردات الإجمالي المطلوب تبويب بياناتها وهي (٤٠) مفرده في مثالنا

**جدول (٣-٢)**
**تفريغ الأجر الأسبوعي لـ ٤٠ عاملاً ( بالجنيه )**

| فئات الأجر الأسبوعي | العلامات | عدد العمال (التكرار) |
|---|---|---|
| -٣٠ | //// | ٤ |
| -٣٤ | ~~////~~ / | ٦ |
| -٣٨ | ~~////~~ //// | ٩ |
| -٤٢ | ~~////~~ //// | ٩ |
| -٤٦ | ~~////~~ // | ٧ |
| ٥٠-٥٤ | ~~////~~ | ٥ |
| المجموع | | ٤٠ |

**وفي حالة عدم تساوي مجموع التكرارات لعدد المفردات المطلوب تبويب بياناتها فيكون هناك ثمة خطأ تم الوقوع فيه . بالإضافة إلى أن هذا لا يمنع من أن نكون قد أخطأنا في وضع علامة تخص فئة معينة من فئات التوزيع في غير موضعها . لذا يجب توخي الحذر أثناء عملية التفريغ حتى يعكس الجدول التكراري أو جدول التفريغ البيانات الفعلية للظاهرة موضع الدراسة .**

**٥- مرحلة العد**

**وفي تلك المرحلة يتم الحصول على التوزيع التكراري للأجور الأسبوعية للعمال وهو عبارة عن نفس جدول التفريغ وذلك بعد استبعاد العمود الأوسط والذي يمثل العلامات . وبالطبع فإنه يتم استبعاد هذا العمود وذلك بعد ترجمة العلامات إلى أعداد تظهر في العمود الثالث من جدول التفريغ . وعليه فإن الجدول التكراري لأجور العاملين يأخذ الصورة المبسطة التالية :**

**جدول (٤-٢)**

| فئات الأجر الأسبوعي | ٣٠- | ٣٤- | ٣٨- | ٤٢- | ٤٦- | ٥٠-٥٤ | المجموع |
|---|---|---|---|---|---|---|---|
| عدد العمال (التكرار) | ٤ | ٦ | ٩ | ٩ | ٧ | ٥ | ٤٠ |

وفي هذا التوزيع التكراري الوارد بالجدول (٤-٢) نجد أن الحد الأدنى للفئة الأولى في التوزيع معلوم (٣٠) وأن الحد الأعلى للفئة الأخيرة (٥٤) كذلك معلوم لذا يسمى هذا التوزيع التكراري بأنه توزيع مغلق (مقفل) من الطرفين. وفي حالة غياب أحد هذين الحدين يسمى التوزيع التكراري مفتوح سواء من طرف واحد أو من الطرفين . فمثلاً إذا كان التوزيع غير معلوم الحد الأدنى للفئة الأولى ومعلوم الحد الأعلى للفئة الأخيرة فيسمى التوزيع بأنه مفتوح من أعلى والعكس صحيح إذا كان التوزيع التكراري معلوم الحد الأدنى للفئة الأولى وغير معلوم الحد الأعلى للفئة الأخيرة يسمى التوزيع التكراري بأنه مفتوح من أسفل . أما في ظل غياب أو عدم معلومية الحدين الأدنى للفئة الأولى والأعلى للفئة الأخيرة فكما سبق حالاً يسمى التوزيع التكراري مفتوح الطرفين .

**الجداول أو التوزيعات التكراري المفتوحة :**

في بعض الأحيان قد نضطر إلى اللجوء إلى الجداول التكرارية المفتوحة وذلك في حالة تعذر تحديد الحد الأدنى للفئة الأولى أو الحد الأعلى للفئة الأخيرة أو كليهما معًا . والجدول التالي يبين توزيعًا تكراريًا مفتوح الطرفين .

**جدول (٥-٢)**

**التوزيع التكراري لدرجات ٥٠ طالب**

| فئات الدرجات | أقل من ١٠ | ١٠- | ٢٠- | ٣٠- | ٤٠- | ٥٠- | ٦٠ فأكثر |
|---|---|---|---|---|---|---|---|
| عدد الطلاب (التكرار) | ٧ | ٩ | ١٥ | ٦ | ٤ | ٧ | ٢ |

ففي هذا الجدول فإن الحد الأدنى للفئة الأولى (أقل من ١٠) والحد الأعلى للفئة الأخيرة (٦٠ فأكثر) غير معلومين . لذا فإن هذا التوزيع التكراري يسمى مفتوح الطرفين .

ولا تنسى عزيزي القارئ أنه ليس بالضرورة تساوي أطوال فئات التوزيع بل يمكن أن يختلف طول أحد فئات التوزيع على الأقل فيسمى التوزيع في هذه الحالة بأنه توزيع تكراري غير متساوي أطوال الفئات (أو ما يسمى بالتوزيع التكراري الغير منتظم ) والجدول التالي يعطي صورة واضحة عن الجداول التكرارية الغير منتظمة .

جدول (٦-٢)

التوزيع التكراري للدخل الشهري (بالجنيه) لمائة أسرة

| فئات الدخل | ٢٣٠- | ٢٤٠- | ٢٥٠- | ٢٦٠- | ٢٧٠- | ٢٨٠- | ٢٩٠- | ٣١٠-٣٤٠ |
|---|---|---|---|---|---|---|---|---|
| عدد الأسر (التكرار) | ٤ | ٦ | ١٠ | ١٢ | ٢٧ | ٢٨ | ١٤ | ٩ |

فيلاحظ على بيانات الجدول (٦-٢) أن أطوال فئاته الستة الأولى هي ١٠ جنيه أما الفئتين الأخيرتين (٢٩٠-) ، (٣١٠-٣٤٠) فإن أطوالهما هما ٢٠ ، ٣٠ جنيه على الترتيب . لذا فإن هذا التوزيع التكراري توزيعًا غير متساوي أطوال فئاته (توزيعًا غير منتظمًا) .

هذا وتجدر الإشارة إلى أنه في حالة الجداول التكرارية المتساوية من حيث أطوال فئاتها (الجداول المنتظمة) فإنه ليس بالضرورة أن يكتب صراحة الحد الأعلى للفئة الأخيرة إذا يمكن بسهولة تحديده في حالة غيابه وذلك من خلال إضافة طول الفئة إلى الحد الأدنى للفئة الأخيرة . هذا على عكس الجداول الغير منتظمة حيث يلزم كتابة الحد الأعلى للفئة الأخيرة صراحة .

<u>ملحوظة</u> : بعد إتمام عملية تفريغ البيانات والحصول على الجداول التكرارية كما هو وارد عل سبيل المثال في جدول (٥-٢) أو (٦-٢) فإن الجدول التكراري البسيط (الذي يحوي ظاهرة واحدة أو متغير واحد محل الدراسة ) يفيد في تحديد عدد التكرارات التي تنتمي لكل فئة من فئات التوزيع التكراري أي عدد التكرارات التي تقع ما بين الحدين الأدنى والأعلى لكل فئة من فئات التوزيع بصورة صريحة فمثلاً في الجدول (٥-٢) فإن هناك سبعة طلاب درجاتهم تقل عن عشرة درجات وكذلك فإن هناك تسعة طلاب تتراوح درجاتهم ما بين عشرة درجات وأقل من ٢٠ درجة وهناك ١٥ طالبًا تتراوح درجاتهم ما بين عشرة درجات وأقل من ٣٠ درجة وهكذا .

هذا ويمكن من تلك الجداول التكرارية البسيطة الحصول على نسبة التكرارات التي تنحصر ما بين حدي كل فئة من فئات التوزيع التكراري وهو ما يسمى بالجداول التكرارية النسبية والتي سيتم تناولها في الجزء التالي . وننوه بأنه تصح مقارنة توزيعين تكراريين بسيطين متى تساوت مجموع التكرارات الأصلية لهما .

**<u>التوزيعات التكرارية النسبية</u> : Relative Frequency Distribution**

في كثير من الأحيان قد يكون الباحث في حاجة إلى معرفة التكرارات النسبية لكل فئة من فئات الجدول أو التوزيع التكراري . والتوزيع التكراري النسبي ما هو إلى صورة أخرى من صور التوزيعات التكرارية البسيطة مع استبدال تكرار كل فئة من فئات التوزيع بقيمة تكرارها النسبي . فإذا كان هناك فئة ولتكن الفئة ( $i$ ) ذات التكرار الأصلي ولتكن ( $F_i$ ) فإن التكرار النسبي Relative Frequency للفئة ( $i$ ) وليكن ($R_i$) ما هو إلا النسبة ما بين التكرار الأصلي للفئة($F_i$) وبين مجموع تكرارات التوزيع التكراري( $\Sigma F_i$). أي أن:

$$R_i = \frac{F_i}{\sum_{i=1}^{r} F_i}, \quad \text{where} \quad i = 1, 2, \ldots, r$$

$$R_i = \frac{F_i}{\sum F_i} \times 100$$

**والفرق ما بين الصيغتين أن الأولى تعطي التكرار النسبي كسرًا عشريًا وفي هذه الحالة يكون مجموع عمود التكرارات النسبية مساويًا للواحد الصحيح أما الصيغة الثانية فهي نسبة مئوية وباستخدامها يكون مجموع عمود التكرارات النسبية مساويًا ١٠٠% .**

**حيث ( r ) تمثل عدد فئات التوزيع التكراري .**

**هذا ولقد سبق وأن ذكرنا في التوزيعات التكرارية البسيطة أنه تصح المقارنة ما بين توزيعين تكراريين متى تساوت مجموع التكرارات لهما . أما في حالة اختلاف مجموع تكرارات التوزيعين فلا تصح المقارنة مباشرة بل يجب التخلص من أثر عدم تساوي مجموع التكرارات ويتم ذلك من خلال حساب التوزيع التكراري النسبي لكلاً من التوزيعين ثم تتم المقارنة من خلال التكرارات النسبية. أي أنه إذا كان الباحث يهدف إلى مقارنة توزيعين مختلفين في مجموع التكرارات الأصلية فيصبح من غير الملائم المقارنة باستخدام التكرارات المطلقة (الأصلية) بل يجب حساب التكرارات النسبية حتى يتم التخلص من أثر عدم تساوي مجموع التكرارات المطلقة . فبحساب التكرارات النسبية نجد أن كل عمود من عمودي التكرارات النسبية يكون مساويًا للواحد الصحيح ومن هنا تم التخلص من أثر عدم التساوي في تلك الحالة ومن ثم تصح عملية المقارنة**

**مثال (٤) : الجدول التالي يوضح التوزيع التكراري لدرجات مجموعة الطلاب والطالبات.**

**جدول (٧-٢)**

**التوزيع التكراري لدرجات مجموع من الذكور والإناث**

| فئات الدخل | أقل من ١٠- | ١٠- | ٢٠- | ٣٠- | ٤٠- | ٥٠- | ٦٠ فأكثر | مجـ |
|---|---|---|---|---|---|---|---|---|
| عدد الطلاب | ٧ | ١٢ | ١٧ | ٩ | ٨ | ٥ | ٢ | ٦٠ |
| عدد الطالبات | ١٥ | ١٧ | ١٨ | ١٨ | ١٠ | ٩ | ١٣ | ١٠٠ |

**والمطلوب إجراء المقارنة بين التوزيعين باستخدام التوزيع التكراري الملائم.**

**الحل :بالنظر لمجموع تكرارات كلاً من الذكور والإناث لنفس فئات التوزيع نجد أن هناك اختلاف في مجموع التكرارات الأصلية للذكور عن الإناث . لذا لا تصح المقارنة من خلال التكرارات المطلقة ( الأصلية المعطاة في المثال ) ويجب للتخلص من أثر عدم تساوي هذا المجموع (٦٠ ، ١٠٠) حساب التكرارات النسبية لكل من الذكور ولتكن ($R_iM$) والإناث ولتكن ($R_iF$) كما يوضحه الجدول التالي :**

**جدول (٨–٢)**

**التوزيع التكراري الأصلي والنسبي من الذكور (M) والإناث (F)**

| Classes | التكرارات الأصلية | | التكرارات النسبية (%) | | Remarks |
|---|---|---|---|---|---|
| | $F_iM$ | $F_iF$ | $R_iM$ | $R_iF$ | |
| Less than 10 | 6 | 15 | $^{6}/_{60} \times 100 = 10$ | $^{15}/_{100} \times 100 = 15$ | $R_1M < R_1F$ |
| 10 - | 12 | 17 | 20 | 17 | $R_2M > R_2F$ |
| 20 - | 12 | 18 | 20 | 18 | $R_3M > R_3F$ |
| 30 - | 9 | 18 | 15 | 18 | $R_4M < R_4F$ |
| 40 - | 12 | 10 | 20 | 10 | $R_5M > R_5F$ |
| 50 - | 6 | 9 | 10 | 9 | $R_6M < R_6F$ |
| 60 and more | 3 | 13 | 5 | 13 | $R_7M < R_7F$ |
| Σ | 60 | 100 | 100 % | 100 % | |

**لاحظ أنه مجموع عمودي التكرارات النسبية سواء للذكور أو الإناث مساويًا ١٠٠% ومن خلال هذين العمودين نبدأ بعملية المقارنة على النحو التالي :**

- **في فئة الدرجات الأولى (أقل من عشرة درجات) فإن نسبة التكرارات للذكور تقل عن نسبة التكرارات للإناث . أي أن نسبة الطلاب الذكور الحاصلين على أقل من عشرة درجات تقل عن نسبة الإناث الحاصلين على أقل من عشرة درجات .**
- **بالنسبة للفئة الثانية (من الدرجة ١٠ حتى أقل من ٢٠ درجة) فإن نسبة الطلاب الذكور (٢٠%) تزيد عن نسبة الطالبات (١٧%) . لاحظ أن مقارنة التكرارات المطلقة لهذه الفئة تعتبر مقارنة مضللة فهي عكس ما يحدده التكرار النسبي .**
- **وهكذا إلى أن نصل للفئة الأخيرة (٦٠ درجة فأكثر) فإن نسبة الطلاب الذكور (٥%) تقل عن نسبة الطالبات (١٣%) .**

**<u>التوزيعات أو الجداول التكرارية المزدوجة :</u>** Crosstabs

**في بعض الأحيان قد يكون لدينا مجموعتان من البيانات تقيس كل واحدة منهما ظاهرة ما وهناك ثمة علاقة تربط بينهما مثل أطوال وأوزان مجموعة من الأفراد أو السن والدخل أو درجات مجموعة من الطلاب في مادتين مختلفتين أو ....... إلخ .**

**فإذا رغبنا في عمل دراسة لمعرفة ما إذا كان هناك علاقة من عدمه فيما بين هاتين الظاهرتين فإنه يصعب إجراء مثل تلك الدراسة إذا ما وضعنا بيانات كل ظاهرة في جدول تكراري بسيط . لكن يمكن إجراء مثل تلك الدراسات إذا ما تم وضع بيانات الظاهرتين (أيًا كانت الظاهرتين وصفتين أو كميتين أو احدهما وصفية والأخرى كمية) معًا في جدول واحد يسمى بالجدول التكراري المزدوج (أي ذو اتجاهين) . هذا الجدول يكون مقسمًا تقسيمًا أفقيًا ورأسيًا إلى مجموعة من الصفوف والأعمدة . حيث يتم تمثيل فئات أحد الظاهرتين في الصفوف والأخرى في الأعمدة . فالتوزيعات التكرارية المزدوجة تستخدم في تلخيص بيانات أزواج من المشاهدات أو القيم بغرض دراسة نوع ودرجة قوة العلاقة فيما بين الظاهرتين أو المتغيرين محل الدراسة أو ما يسمى بعلاقة الارتباط وهو ما سيرد دراسته فيما بعد في هذا الكتاب .**

**هذا ولتكوين الجدول التكراري المزدوج فإننا نتبع نفس الخطوات السابق شرحها في حالة الجداول التكرارية البسيطة وذلك من خلال تطبيقها على الظاهرتين أو المتغيرين محل الدراسة كل على حده . فبالنسبة لكل ظاهرة أو متغير يتم تحديد أقسام الظاهرة أو المتغير محل الدراسة (الفئات) ثم يتم تصوير جدول تكراري مزدوج حيث يتم تمثيل فئات الظاهرة الأولى في الصفوف والثانية في الأعمدة . ثم يلي ذلك عملية تفريغ بيانات الظاهرتين معاً في هذا الجدول التكراري المزدوج (مرحلة التفريغ) . ثم يتم ترجمة العلامات الناتجة من**

الخطـوة السـابقة إلـى أرقـام أو تكـرارات (عمليـة العـد) . أي أن مراحـل إنشـاء الجدول التكراري المزدوج هي :

١- تحديد أقسام الظاهرتين كلاً على حدة ( فئات كل ظاهرة على حدة ) . حيث يتم تحديد أقسام كل ظاهرة على حدة وذلك باتباع نفس خطوات الأسلوب السابق شرحه في حالة الجدول التكراري البسيط .

٢- عمليـة التفريـغ : ( العلامـات ) بعد تحديـد أقسـام أو فئـات الظـاهرتين يـتم تصميم الجدول التكراري المزدوج بحيث تخصص الصفوف لأحد المتغيرين ويكـون عـدد الصـفوف مسـاوياً لعـدد فئـات هـذا المتغيـر . بينمـا تخصـص الأعمدة لتوزيع أو الفئات المتغير الآخر بحيث يكون عدد الأعمدة مساوياً لعدد فئات المتغير الثاني . وبهذا فإن الجدول التكراري المزدوج يتكون من عدد من الخلايا يبلـغ عددها ( عدد الصفوف × عدد الأعمدة ) أو بلغة أخرى ( عدد فئات المتغير الأول × عدد فئات المتغير الثاني ) .

أما عن عملية التفريغ ذاتها فإنه يتم تحديد الخلية التي ينتمي إليها البيـان أو زوج المشاهدات حيث يتم تحديد الصف والعمـود الـذي ينتمـي إليـه هذا البيـان ومن ثم يتم تحديد الخلية وهي تقاطع صف وعمود هذا البيان ثم توضع شرطة ( / ) في تلك الخلية . وتتم عملية تفريغ بقراءة البيانـات أو أزواج المشـاهدات للبيانات أو الزوج تلو الآخر وفي كل مرة توضع شرطة ( / ) في الخليـة التي ينتمي إليها هذا البيان مع استخدام طريقة الحزم السابق الإشارة إليها .

٣- عملية العد : وهنا تـتم عمليـة ترجمـة العلامـات داخل كل خليـة مـن خلايـا الجدول المختلفة إلى أرقام فنحصل بذلك على الجدول التكراري المزدوج.

هـذا ويمكـن مـن خـلال الجـدول التكـراري المـزدوج اسـتنتاج التوزيعـات التكراريـة الهامشـية Marginal Distribution لكـل متغيـر علـى حـدة .

**التوزيع التكراري الهامشي للظاهرة أو المتغير الأول وليكن ذلك المتغير الذي تم التعبير عن فئاته في الصفوف وما يناظرها من مجاميع تكرارات صفوف الجدول التكراري المزدوج . والتوزيع الهامشي للمتغير الثاني المعبر عن فئات في الأعمدة وما يناظرها من مجاميع تكرارات الأعمدة في التكراري المزدوج . ومن هذه التوزيعات التكرارية الهامشية يمكن الحصول على التكرارات النسبية كما سبق وأن أوضحناها . والمثال التالي يوضح كيفية تكوين الجدول التكراري المزدوج وكيفية اشتقاق التوزيعات الهامشية للظاهرتين محل الدراسة .**

**<u>مثال (٥) :</u>**

**البيانات التالية تمثل الدخل الأسبوعي (x) والإنفاق الأسبوعي (y) بالجنيهات وذلك لعينة مكونة من خمسين مفردة :**

| X | Y | X | Y | X | Y | X | Y |
|---|---|---|---|---|---|---|---|
| ٦٠ | ٤٣ | ٩٠ | ٥٩ | ٨٤ | ٧٨ | ١٠٠ | ٨٩ |
| ٨٣ | ٧٥ | ١٠٠ | ٨٩ | ٧٣ | ٤١ | ٨٠ | ٦٩ |
| ٧٩ | ٤٣ | ٨٦ | ٥١ | ٩١ | ٦٥ | ١٠ | ٩٠ |
| ١٠١ | ٨٤ | ١٠٥ | ٨٨ | ٦٨ | ٥٤ | ٧٨ | ٧١ |
| ٨٠ | ٧٢ | ٧١ | ٥٢ | ٧٤ | ٥٦ | ٩٦ | ٧٩ |
| ٩٥ | ٨٤ | ٩٣ | ٧٥ | ٩٠ | ٦٩ | ١١٢ | ١٠٦ |
| ٧٢ | ٦٢ | ٨٨ | ٥٦ | ٦٥ | ٤٤ | ٩١ | ٧٢ |
| ٧٧ | ٥٣ | ٧٣ | ٦٩ | ٨٢ | ٦٨ | ٨١ | ٧٥ |
| ١٠٠ | ٧٨ | ١٠٩ | ٨٧ | ٨٩ | ٦٧ | ١٠٣ | ٩٤ |
| ٨٥ | ٧٦ | ١١٥ | ٩٨ | ٩٢ | ٧٩ | ٨٤ | ٧٩ |
| ١٠٤ | ٩١ | ٩٧ | ٨٥ | ٩٠ | ٦٨ | ٩٨ | ٩٠ |
| ٩٥ | ٧٥ | ١٢٥ | ١٠٥ | ٨١ | ٦٩ | ١١٩ | ١٠٨ |
| ٨٩ | ٨٠ | ٨١ | ٦٨ | | | | |

**<u>والمطلوب :-</u>**

**١- إنشاء جدول تكراري مزدوج لبيانات الدخل (x) والإنفاق (y) .**

**٢- من خلال نتائجك في (١) استنتج التوزيع الهامشي للدخل الأسبوعي (x) والتوزيع الهامشي للإنفاق الأسبوعي (y) .**

**٣- من خلال نتائجك في (٢) اوجد التوزيع التكراري النسبي لكل من الدخل الأسبوعي (x) والإنفاق الأسبوعي (y) .**

**الحل :**

**١- لتكوين جدول تكراري مزدوج فإننا نبدأ بعملية تحديد فئات كل من الدخل (x) والإنفاق (y) كلاً على حدة وذلك على النحو التالي :**

**أولاً : بالنسبة لبيانات الدخل (x) :-**

- Rang (x) = Max – Min + 1
  = 125 – 60 + 1 = 66 (L.E)
- NI = 1 + 3.3 Log (n)
  = 1 + 3.3 Log (so )
  = 6.6066 = 7 Intervals (classes )
- LI = Range ÷ NI
  = 66 ÷ 7 = 9.43 (L.E)

**ويتم تقريب طول الفئة بالزيادة فتصبح طول الفئة مساوياً عشرة جنيهات . وحيث أن أصغر قيمة في الدخل هي ستون جنيهاً لذا فإن من الملائم أن تبدأ فئات الدخل بالحد الأدنى للفئة الأولى وهو ستون جنيهاً . وحيث أن أكبر دخل هو مائة وخمس وعشرون جنيهاً لذا فإن فئات الدخل (x) تكون عبارة عن :**

60- , 70- , 80- , 90- , 100- , 110- , 120-130

**ثانياً: بالنسبة لبيانات الإنفاق (y) :**

- Range (y) = 108 – 41 + 1 = 68 (L.E)
- NI = 7 Intervals (classes )

**لاحظ أن عدد الفئات للمتغير (y) باستخدام قاعدة ستورجس سيكون نفس عدد الفئات المتغير (x) نظراً لثبات قيمة (n) .**

- LI = 68 ÷ 7 = 9.71 $\approxeq$ 10 (L.E)

**وحيث أن أصغر قيمة في الإنفاق هي إحدى وأربعون جنيهاً . لذا فإن من المناسب أن يكون الحد الأدنى للفئة الأولى هو أربعون جنيهاً .**

**وحيث أن أكبر قيمة للإنفاق هي مائة وثمانية جنيهات . لذا فإن فئات الإنفاق (y) تأخذ الشكل التالي :** 40- , 50- , 60- , 70- , 80- , 90- , 100- 110

**ومن خلال نتائجنا عن فئات كل من المتغيرين y,x يتم تصوير جدول تكراري مزدوج حيث يتم تمثيل فئات الدخل (x) وليكن في الصفوف وفئات الإنفاق (y) في الأعمدة فيعطي الصورة الموضحة في الجدول(٩ – ٢)على النحو التالي :**

**جدول (٩–٢)**

**تصميم جدول تكراري مزدوج يوضح العلاقة بين الدخل الأسبوعي (x) والإنفاق الاسبوعي (y) لخمسين أسرة**

| y / x | 40- | 50- | 60- | 70- | 80- | 90- | 100-110 | ∑ |
|---|---|---|---|---|---|---|---|---|
| 60- | | | | | | | | |
| 70- | | | | | | | | |
| 80- | | | | | | | | |
| 90- | | | | | | | | |
| 100- | | | | | | | | |
| 110- | | | | | | | | |
| 120-130 | | | | | | | | |
| ∑ | | | | | | | | **50** |

**أما عن مرحلة تفريغ البيانات (وضع العلامات) فإننا نبدأ بقراءة أزواج المشاهدات (y,x) البيان تلو الآخر . وفي كل مرة نضع شرطة (/) في الخلية التي ينتمي إليها هذا البيان . فمثلاً لو بدأنا بيانات الدخل والإنفاق متبعين الترتيب الوارد في هذا المثال فإننا نبدأ بالبيان الأول[ وهو x=60 تقابلها y=43 ] . فحيث أن قيمة الدخل ٦٠ تنتمي للفئة الأولى من فئات الدخل وهي (-60) وأن قيمة الإنفاق ٤٣ تنتمي للفئة الأولى من فئات الإنفاق وهي**

(40-) فإن الخلية التي ينتمي إليها هذا الزوج من المشاهدات أو هذا البيان هي تقاطع الصف الأول مع العمود الأول . لذا يتم وضع شرطة ( / ) في هذه الخلية . أما البيان الثاني [ وهو x = 83 ، y = 75 ] فإن ينتمي إلى الخلية الناشئة من تقاطع الصف الثالث الممثل للفئة (80-) مع العمود الرابع الممثل للفئة (70-) . لذا يتم وضع شرطة في الخلية الناشئة عن تقاطع الصف الثالث مع العمود الرابع .

وهكذا بالنسبة لبقية البيانات حتى نصل إلى البيان الأخير [ وهو119 =x تقابلها y=108 ] والذي ينتمي للخلية الناشئة عن تقاطع الصف السادس الممثل للفئة (110-) مع العمود السابع الممثل للفئة (110-) . لذا يتم وضع شرطة في الخلية الناشئة عن تقاطع الصف السادس مع العمود السابع كما هو موضح في جدول التفريغ (١٠– ٢) والذي يبين العلامات التي تعكس ازدواج المشاهدات .

**جدول (١٠–٢)**
**تفريغ مزدوج للدخل الأسبوعي (x)**
**والإنفاق الأسبوعي (y) لخمسين أسرة .**

| y / x | 40- | 50- | 60- | 70- | 80- | 90- | 100-110 | ∑ |
|---|---|---|---|---|---|---|---|---|
| 60- | // | / | | | | | | 3 |
| 70- | // | /// | // | / | | | | 8 |
| 80- | | // | $\cancel{////}$ | $\cancel{////}$ / | / | | | 14 |
| 90- | | / | /// | //// | /// | / | | 12 |
| 100- | | | | / | $\cancel{////}$ | // | | 8 |
| 110- | | | | | | // | // | 4 |
| 120-130 | | | | | | | / | 1 |
| ∑ | 4 | 7 | 10 | 12 | 9 | 5 | 3 | 50 |

وبترجمة العلاقات الواردة بخلايا جدول التفريغ المزدوج المبين في الجدول (١٠-٢) إلى أرقام نحصل علي الجدول التكراري المزدوج الذي تحتوي خلاياه على التكرارات المختلفة كما هو موضح في الجدول (١١-٢) .

**جدول (١١-٢)**

**جدول تكراري مزدوج يوضح العلاقة بين الدخل الأسبوعي (x) والإنفاق الأسبوعي (y) لخمسين أسرة**

| y / x | 40- | 50- | 60- | 70- | 80- | 90- | 100-110 | ∑ |
|---|---|---|---|---|---|---|---|---|
| 60- | 2 | 1 | | | | | | 3 |
| 70- | 2 | 3 | 2 | 1 | | | | 8 |
| 80- | | 2 | 5 | 6 | 1 | | | 14 |
| 90- | | 1 | 3 | 4 | 3 | 1 | | 12 |
| 100- | | | | 1 | 5 | 2 | | 8 |
| 110- | | | | | | 2 | 2 | 4 |
| 120-130 | | | | | | | 1 | 1 |
| ∑ | 4 | 7 | 10 | 12 | 9 | 5 | 3 | 50 |

هذا ويمكن من الجدول (١١-٢) القول أن تكرار فئة معينة من فئات الدخل (x) هو عبارة عن مجموع التكرارات المشتركة بين هذه الفئة . فمثلاً يمكن القول من خلال بيانات الصف الأول في هذا الجدول ( وهو صف فئة الدخل ٦٠ جنيه ) أن هناك ثلاث أسر يتراوح دخلها ما بين ستون وأقل من سبعون جنيهاً . وكذلك هناك ثماني أسر يتراوح دخلها ما بين سبعون جنيهاً وأقل من ثمانون جنيهاً . وكذلك الحال بالنسبة لفئات الإنفاق . حيث يمكن القول أن هناك أربعة أسر يتراوح إنفاقها الأسبوعي ما بين أربعون وأقل من خمسون جنيهاً وكذلك هناك سبعة أسر يتراوح إنفاقها الأسبوعي ما بين خمسون وأقل من ستون جنيهاً وهكذا .

هذا ويجب أن ننوه إلى أنه من المألوف أن نجد عدداً من الخلايا في الجدول التكراري المزدوج لا يحتوي على تكرارات . ويكون ذلك بصفة خاصة إذا كانت العلاقة قوية ما بين المتغيرين أو الظاهرتين موضع الدراسة والممثلين في الجدول التكراري المزدوج . وتزداد قوة العلاقة ما بين الظاهرتين أو المتغيرين حينما تتركز التكرارات على جانبي أحد أقطار الجدول التكراري المزدوج . وهذه العلاقة تسمى بعلاقة الارتباط ما بين الظاهرتين وهو ما سوف نتناوله فيما بعد بالتفصيل إن شاء الله .

٢- من خلال التوزيع التكراري المزدوج الوارد بالجدول (١١-٢) يمكن استنتاج التوزيع التكراري الهامشي للدخل الاسبوعي (x) وهو عبارة عن فئات الدخل (x) المختلفة وما يقابلها من مجاميع تكرارات الصفوف الواردة بالعمود الأخير في الجدول (١١-٢) دون أخذ الإنفاق الاسبوعي (y) في الاعتبار . والجدول (١٢-٢) يبين التوزيع التكراري الهامشي للدخل الاسبوعي (x) .

جدول (١٢-٢)

التوزيع التكراري الهامشي للدخل الأسبوعي

(x) لعينة من خمسين أسرة

| **Weekly Incom classes (x)** | 60- | 70- | 80- | 90- | 100- | 110- | 120-130 | ∑ |
|---|---|---|---|---|---|---|---|---|
| Frequency | 3 | 8 | 14 | 12 | 8 | 4 | 1 | 50 |

ومن خلال هذا التوزيع الهامشي للدخل يتبين لنا على سبيل المثال أن هناك أربعة عشرة أسرة يتراوح دخلها ما بين ثمانون وأقل من تسعون جنيهاً وهكذا .

أما عن التوزيع التكراري الهامشي للإنفاق الأسبوعي (y) فهو يتكون من فئات الإنفاق المختلفة الممثلة في الأعمدة الواردة بالجدول التكراري

**المزدوج (١١–٢) وما يقابلها من مجاميع تكرارات لهذه الأعمدة دون الأخذ في الاعتبار فئات الدخل الأسبوعي (x) .**

**والجدول (١٣–٢) يبين التوزيع التكراري الهامشي للإنفاق الأسبوعي (y)**

**جدول (١٣–٢)**

**التوزيع التكراري الهامشي للإنفاق الأسبوعي (y)**
**لعينة من خمسين أسرة**

| **Weekly Expenditure classes (y)** | 40- | 50- | 60- | 70- | 80- | 90- | 100-110 | ∑ |
|---|---|---|---|---|---|---|---|---|
| Frequency | 4 | 7 | 10 | 12 | 9 | 5 | 3 | **50** |

**ومن خلال هذا الجدول يتبين لنا على سبيل المثال أن هناك سبعة أسر يتراوح إنفاقها الأسبوعي ما بين خمسون جنيهاً وأقل من ستون جنيهاً وهكذا .**

**٣– بعد الحصول على التوزيعات الهامشية للمتغير (x) وكذلك للمتغير (y) فإنه يمكن قسمة قيم تكرارات تلك التوزيعات التكرارية الهامشية على إجمالي عدد المفردات (خمسون عينة) فنحصل على التوزيع التكراري النسبي للمتغير (x) وكذلك للمتغير (y) كما هو موضح في الجدولين (١٤– ٢ ) ، (١٥ –٢) على الترتيب .**

**جدول (١٤–٢)**

**التوزيع التكراري النسبي للدخل الأسبوعي (x)**
**لعينة من خمسين أسرة**

| Classes of (x) | 60- | 70- | 80- | 90- | 100- | 110- | 120-130 | ∑ |
|---|---|---|---|---|---|---|---|---|
| R.F(%) | 6 | 16 | 28 | 24 | 16 | 8 | 2 | 100% |

**ومن خلال هذا الجدول يتبين لنا على سبيل المثال أن ٢٤% من إجمالي عدد الأسر يتراوح دخلهم ما بين تسعون وأقل من مائة جنيه . وهكذا ومن خلال**

هذا التوزيع التكراري النسبي للإنفاق يتضح لنا على سبيل المثال أن هناك ١٨% من إجمالي عدد الأسر يتراوح إنفاقها ما بين ثمانون وأقل من تسعون جنيهًا وهكذا .

جدول (١٥-٢)
التوزيع التكراري النسبي للإنفاق الأسبوعي (y)
لعينة من خمسين أسرة

| Classes of (y) | 40- | 50- | 60- | 70- | 80- | 90- | 100-110 | ∑ |
|---|---|---|---|---|---|---|---|---|
| R.F(%) | 8 | 14 | 20 | 24 | 18 | 10 | 6 | 100% |

هذا وفي نهاية دراستنا للجدول التكرارية المزدوجة تجدر الإشارة إلى ما يلي :

١- ليس بالضرورة أن يتساوى عدد الفئات أو طول الفئة للمتغيرين في الجدول التكراري المزدوج . إذ أن عدد الفئات أو طول الفئة لكل متغير يتوقف على المدى المطلق (Range) لهذا المتغير .

٢- قد يكون المتغيران الممثلان في الجدول التكراري المزدوج متغيران كميان ( أياً كان نوعهما كمي متقطع أم كمي منفصل ) أو وصفيان أو أحدهما على الأقل وصفياً . فمثلاً يمكن تكوين جدول تكراري مزدوج يمثل حجم الأسرة وعدد حجرات المسكن (متغيران كميان منفصلان ) أو جدول تكراري يمثل حجم الأسرة (متغير كمي منفصل ) والإنفاق الشهري (متغير كمي متصل ) أو جدول تكراري يمثل العلاقة ما بين الحالة الاجتماعية للفرد وظاهرة التدخين ( متغيران وصفيان ) أو جدول تكراري يمثل العلاقة ما بين النوع والدخل الشهري لعينة من الأسر و ..... وهكذا .

**٣- التوزيع التكراري الهامشي لمتغيرًا ما هو إلا التوزيع التكراري البسيط لهذا المتغير . والتوزيع التكراري البسيط يبين عدد التكرارات التي تقع ما بين الحد الأدنى والأعلى لكل فئة من فئات التوزيع .**

**<u>الجداول( التوزيعات ) التكرارية المتجمعة :</u>** Cumulative Frequency

**سبق وأن ذكرنا حالاً أن التوزيع التكراري البسيط يعطي عدد التكرارات التي تقع ما بين الحد الأدنى والأعلى لكل فئة من فئات التوزيع بصورة صريحة . لكن في كثير من الأحيان قد يهتم الباحث بمعرفة عدد التكرارات أو المفردات التي تأخذ قيمة معينة سواء كانت تلك القيمة المعنية تمثل الحد الأدنى أو الأعلى لأي فئة من فئات التوزيع أو تقع ما بينهما . فمثلاً قد يكون الباحث في حاجة لأن يعرف عدد الطلاب اللذين تقل درجاتهم عن ٩٠ درجة لتوزيع تكراري بسيط معطي . وفي أحيان أخرى قد يهتم الباحث بمعرفة عدد المفردات أو التكرارات التي تأخذ قيمة تزيد على أو تساوي قيمة معينة سواء كانت تلك القيمة تمثل الحد الأدنى أو الأعلى لأي فئة من فئات التوزيع أو أي قيمة تقع ما بينهما . وليكن على سبيل المثال عدد الموظفين اللذين تبلغ مرتباتهم الأسبوعية ٩٥ جنيهاً على الأقل ( أي ٩٥ جنيهاً فأكثر ) .**

**ففي واقع الأمر فإن الجداول التكرارية البسيطة لا تلبي – بطريقة مباشرة – هذه الحاجات . ولكي يصل الباحث إلى ما يريد فإنه يقوم بتجميع التكرارات تصاعدياً في الحالة الأولى بينما يقوم بتجميعها تنازلياً في الحالة الثانية . ويسمى الجدول في الحالة الأولى بالجدول التكراري المتجمع الصاعد**

Ascending Cumulative Frequency ( Asc.C.F.T).

**بينما يسمى الجدول في الحالة الثانية بالجدول التكراري المتجمع الهابط أو النازل**

Descending Cumulative Frequency ( Dsc.C.F.T ) .

**ويتكون كلاً من الجدول المتجمع الصاعد أو الهابط باستخدام الجدول التكراري البسيط للظاهرة موضع الدراسة . ويستخدم الجدول التكراري المتجمع الصاعد (Asc.C.F.T) في تحديد عدد المفردات التي تأخذ قيماً أقل من قيمة معينة . كما يستخدم في تحديد القيمة التي يقل عنها عدداً معيناً أو نسبة من المفردات . أما الجدول التكراري المجتمع الهابط (Dsc.C.F.T) فإنه يستخدم في تحديد عدد المفردات التي تأخذ قيماً تساوي أو تزيد عن قيمة معينة كما يستخدم في إيجاد القيمة التي يزيد عنها أو يساويهما عدداً أو نسبة من مفردات التوزيع التكراري .**

**هذا وتجدر الإشارة عزيزي القارئ لبعض مفردات اللغة العربية التي تعادل تماماً في مضمونها كلمة أقل من وأيضاً كلمة فأكثر وهي :-**

**١- إذا كان المطلوب تحديد عدد التكرارات أو نسبة التكرارات التي تقل عن قيمة معينة سواء كانت هذه القيمة هي الحد الأدنى أو الحد الأعلى لفئة ما بين حدي فئة معينة أو ما بينهما فلاحظ أن هناك مرادفات لغوية مكافئة لكلمة <u>أقل من</u> حيث أن : -**

**أقل من ≡ يقل عن ≡ لا تزيد عن ≡ على الأكثر ≡ الحد الأقصى**

**فكل هذه المرادفات أو ما يعادلها لغوياً يفيد ضرورة تكوين الجدول التكراري المتجمع الصاعد ( Asc.C.F.T ) واستنتاج المطلوب منه .**

**٢- عكس ما سبق صحيح تماماً . بمعنى إذا كان المطلوب هو تحديد عدد التكرارات أو نسبة التكرارات التي تزيد عن ( أو تبلغ قيمة ما فأكثر ) سواء كانت هذه القيمة تمثل الحد الأدنى أو الحد العلى لفئة معينة أو قيمة معينة ما بين حدي فئة من فئات التوزيع التكراري الأصلي . فأنه يجب ملاحظة أن هناك مترادفات لغوية مكافئة لكلمة <u>تزيد عن</u> وهي على سبيل المثال : -**

**تزيد عن ≡ فأكثر ≡ لا تقل عن ≡ على الأقل ≡ الحد الأدنى .**

فكل هذه المرادفات وما يعادلها لغوياً يفيد ضرورة تكوين التكراري الهابط ( Dsc.C.F.T ) واستنتاج المطلوب منه .

٣- أما إذا كان المطلوب هو إيجاد عدد التكرارات ( أو نسبة من التكرارات التي تقع ما بين قيمتين لا تشكلان حدي أي فئة من فئات التوزيع الأصلي على الأقل ( أو على الأكثر ) فإنه يمكن استنتاج المطلوبين معاً أما من خلال التوزيع التكراري المتجمع الصاعد أو من خلال التوزيع التكراري المتجمع الهابط كما سنرى في الأمثلة فيما بعد .

**مثال (٦) :-**

فيما يلي لديك التوزيع التكراري للأجور الأسبوعية لعينة من خمسين عاملاً (بالجنيه) كما يوضحها الجدول التالي :

جدول (١٦ - ٢)

| Weekly wages classes | 30- | 40- | 50- | 60- | 70- | 80- | 90- | $\sum$ |
|---|---|---|---|---|---|---|---|---|
| Frequency ($F_i$) | 1 | 4 | 7 | 14 | 11 | 8 | 5 | 50 |

والمطلوب :-

١- إيجاد التوزيع التكراري المتجمع الصاعد وكذا التوزيع التكراري النسبي المتجمع الصاعد .

٢- إيجاد التوزيع التكراري المتجمع الهابط وكذا التوزيع التكراري النسبي المتجمع الهابط .

٣- باستخدام التوزيع التكراري الملائم حدد ما يلي :

أ- عدد العمال اللذين تتراوح أجورهم ما بين ٨٠ وأقل من ٩٠ وكذا النسبة .

ب- عدد العمال اللذين تبلغ أجورهم ٧٣ جنيهاً على الأكثر وكذا النسبة .

ج- ما هي قيمة الأجر الذي يتقاضاه ٦٥ % من إجمالي عدد العمال على الأقل .

د- ما هو الحد الأدنى للأجر الذي يبلغه ٤٧ عاملاً كحد أدنى .

ه- ما هي نسبة العمال اللذين تتراوح أجورهم ما بين ٥٥ ، ٨٧ جنيه .

و- إذا كانت نسبة العمال اللذين تتراوح أجورهم ما بين ٥٣ ، **X** جنيه تبلغ ٤٠% من إجمالي عدد العمال . فما هي قيمة الأجر **x** .

**الحل : -**

١- وفقاً لتعريف التوزيع التكراري المجتمع الصاعد . فإن التكرار المتجمع الصاعد (Asc.C.F.T) المناظر لأي فئة من فئات التوزيع الأصلي ما هو إلا عبارة عن تكرار هذه الفئة مضافاً إليه تكرارات الفئات السابقة . ويتكون هذا الجدول من عمودين يتم كتابة الفئات التكرارية المتجمعة الصاعدة ( أقل من الحد الأعلى لكل فئة في فئات التوزيع ) في العمود الأول أما التكرارات المتجمعة الصاعدة (Asc.C.F.T) فيتم تحديدها في العمود الثاني من هذا الجدول . أما التكرارات النسبية المتجمعة الصاعدة (R.Asc.C.F) للفئة ( **i** ) فهي عبارة عن التكرارات المتجمعة الصاعدة للفئة ( **i** ) مقسومة على مجموع التكرارات كنسبة مئوية . أي أن :-

$$R.Asc.C.F ( i ) = Asc.C.F ( i ) / \sum F_i \ , \ where \ i = 1,2,\ldots,r$$

و **r** تمثل عدد فئات التوزيع

حيث Asc.C.F ( i ) هو التكرار المتجمع الصاعد للفئة ( **i** )

والجدول (٢-١٧) يبين حسابات التوزيع التكراري المتجمع الصاعد والمتجمع النسبي الصاعد .

**جدول (١٧-٢)**
**التوزيع التكراري المتجمع الصاعد والنسبي المتجمع الصاعد**
**لأجور عينة من خمسين عاملاً في إحدى الشركات**

| Classes | $F_i$ | less than the Class Upper Limit | ACFD | RACFD(%) |
|---|---|---|---|---|
| 30- | 1 | less than 30 | 0 | 0 |
| 40- | 4 | less than 40 | 1 | 2 |
| 50- | 7 | less than 50 | 5 | 10 |
| 60- | 14 | less than 60 | 12 | 24 |
| 70- | 11 | less than 70 | 26 | 52 |
| 80- | 8 | less than 80 | 37 | 74 |
| 90- | 5 | less than 90 | 45 | 90 |
| $\sum$ | 50 | less than 100 | 50 | 100 |

**لاحظ في الجدول (١٧-٢) ما يلي : -**

- **بافتراض أنه توجد فئة وهمية (افتراضية ) تسبق الفئة الأولى فيكون الحد الأعلى لتلك الفئة الوهمية ما هو إلا الحد الأدنى للفئة الأولى (٣٠جنيه) فيكون التكرار المتجمع الصاعد المقابل للحد الأدنى للفئة الأولى ( وهو الحد الأعلى لتلك الفئة الوهمية يكون دائماً أبداً مساوياً للصفر حيث لا توجد هناك تكرارات سابقة لتلك الفئة الأولى . هذا ويمكن الاستغناء عن السطر الأول في التوزيع المتجمع الصاعد . وفي هذه الحالة نبدأ حسب تمييز العمود الأول من التوزيع المتجمع الصاعد وهو أقل من الحد الأعلى للفئة الأولى والذي يبلغ ٤٠ جنيه . فيكون عدد العاملين اللذين يقل أجرهم عن أربعون جنيهاً هو عامل وحيد وهو بمثابة تكرار الفئة الأولى . ويلي ذلك السطر الثاني في الجدول المتجمع الصاعد فإن عدد العاملين اللذين يقل أجرهم عن الحد الأعلى للفئة الثانية ( والذي يبلغ**

٥٠ جنيه ) وهو عبارة عن ٥ عمال وهذا التكرار المتجمع بمثابة تكرار الفئة الأولى مضافاً إليه تكرار الفئة الثانية ( أي أن : ( ٤ +١ = ٥ ) وبذلك يكون التكرار المتجمع الصاعد المقابل للحد الأعلى للفئة الثانية ما هو إلا تكرار الفئة الأولى مضافاً إليه تكرار الفئة الثانية . وهكذا يتم إيجاد التكرار المتجمع الصاعد لبقية الفئات فئة تلو الأخرى حتى نصل إلى التكرار المتجمع الصاعد للحد الأدنى للفئة الأخيرة (وهو بمثابة الحد الأدنى لفئة وهمية تلي الفئة الأخيرة والذي يعتبر بمثابة الحد الأعلى للفئة الأخيرة بالجدول التكراري الأصلي) والذي يجب أن يكون مساوياً للمجموع الكلي للتكرارات أي خمسون عاملاً في هذا المثال .

- يمكن ملاحظة أن التكرار المتجمع الصاعد المناظر لفئة معينة من فئات التوزيع الأصلي ما هو إلا عبارة عن التكرار المتجمع الصاعد للفئة السابقة لها مضافاً إليه تكرار هذه الفئة الأصلية من التوزيع التكاري الأصلي . فمثلاً للتأكد من ذلك فإن التكرار المتجمع الصاعد المقابل لأقل من ٨٠ جنيه وهو ٣٧ عاملاً ما هو إلا التكرار المتجمع الصاعد السابق لتلك الفئة ( أقل من ٧٠ جنيه ) مضافاً إليه تكرار الفئة من ٧٠ حتى أقل من ٨٠ جنيه أي أن ( ٢٦ + ١١ = ٣٧ ) .
- من التوزيع التكراري المتجمع الصاعد يمكن تحديد قيمة الأجر الذي يقل عنه عدداً معيناً من العمال . فمثلاً قيمة الأجر الذي يقل عنه ٤٥ عاملاً ما هو إلا الأجر ٩٠ جنيه . كذلك يمكن من الجدول المتجمع الصاعد تحديد أو حساب عدد العمال اللذين تقل أجورهم عن قيمة معينة لا توجد في عمود الفئات وليكن مثلاً الأجر ٦٤ جنيه وهي قيمة غير موجودة في الجدول وهو ما سيتم استنتاجه جبرياً أو بيانياً كما سنرى فيما بعد . وأيضاً العكس صحيح يمكن تحديد ما هو الأجر الذي يقل عنه عدداً معيناً

من التكرارات المتجمعة التي لا توجد في التوزيع التكراري المتجمع الصاعد بنفس الآلية الجبرية ( طريقة الاستكمال العددي أو النسبة والتناسب ) أو بيانياً كما سنرى فيما بعد في موضع لاحق إن شاء الله .

- الجدول التكراري المتجمع الصاعد النسبي ما هو إلا صورة من الجدول التكراري المتجمع الصاعد لكن بعد قسمة عمود التكرارات المتجمعة الصاعدة بالكامل على مجموع تكرارات التوزيع الأصلي وهي القيمة ٥٠ عامل في مثالنا محل الدراسة .
- التكرار المتجمع الصاعد أما أن يبدأ بالصفر ( وذلك في حالة افتراض أن هناك فئة وهمية تسبق الفئة الأولى فيكون الحد الأعلى لتلك الفئة الوهمية بمثابة الحد الأدنى للفئة الأولى ) أو يبدأ بتكرار الفئة الأولى

( وذلك في حالة بداية الفئات المتجمعة بالحد الأعلى للفئة الأولى ) وينتهي بالمجموع الكلي لتكرارت التوزيع الأصلي .

وكذلك فإن الجدول التكراري المتجمع الصاعد النسبي أما أن يبدأ بالصفر أو بالتكرار النسبي للفئة الأولى وينتهي بالواحد الصحيح أو النسبة ١٠٠ % وهذا المعيار يؤكد صحة حسابات الجدول التكراري المتجمع الصاعد وكذا الجدول التكراري المتجمع الصاعد النسبي .

٢- إيجاد التوزيع أو الجدول التكراري المتجمع الهابط ( Dsc.C.F.T ) وكذا التوزيع التكراري المتجمع الهابط النسبي ( R. Dsc.C.F.T ) .

فلتكوين الجدول التكراري المتجمع الهابط فإنه يعاد تعريف الفئات كما هو موضح في الجدول ( ١٨ – ٢ ) حيث يتم كتابة فئات التوزيع التكراري المتجمع الهابط على صورة " الحد الأدنى للفئة الأصلية فأكثر ) . ومن ثم فبالاستعانة ببيانات الجدول ( ١٦ – ٢ ) سوف نوضح كيفية تكوين الجدول التكراري المتجمع الهابط وكذا الهابط النسبي .

**جدول ( ١٨ – ٢ )**
**التوزيع التكراري المتجمع الهابط (Dsc.C.F.T) وكذا التوزيع التكراري المتجمع الهابط ( R.Dsc.C.F.T) لأجو عينة من ٥٠ عامل**

| Classes | $F_i$ | More than Lower Limit | DCFD | RDCFD(%) |
|---|---|---|---|---|
| 30- | 1 | More than 30 | 50 | 100 |
| 40- | 4 | More than 40 | 49 | 98 |
| 50- | 7 | More than 50 | 45 | 90 |
| 60- | 14 | More than 60 | 38 | 76 |
| 70- | 11 | More than 70 | 24 | 48 |
| 80- | 8 | More than 80 | 13 | 26 |
| 90- | 5 | More than 90 | 5 | 10 |
| $\sum$ | 50 | More than 100 | 0 | 0 |

**فوفقاً للتعريف الذي سبق وأن أوضحناه لفئات التوزيع التكراري المتجمع الهابط فإن :–**

- **التكرار المتجمع الهابط المقابل لفئة معينة ( للحد الأدنى لفئة معينة ) من فئات التوزيع الأصلي ما هو إلا عبارة عن مجموع التكرارات مطروحاً منه تكرارات الفئات السابقة لتلك الفئة .**

**فعلى سبيل المثال فإن التكرار المتجمع الهابط للفئة الأولى ما هو إلا عبارة عن المجموع الكلي للتكرارات ( ٥٠عاملاً ) حيث أنه لا توجد تكرارات سابقة لهذه الفئة . بينما التكرار المتجمع الهابط للفئة الثانية ( ٤٠ فأكثر ) فهو عبارة عن المجموع الكلي للتكرارات مطروحاً منه تكرار الفئة السابقة ( أي تكرار الفئة الأولى وهو ١ عاملاً ) . أي أن التكرار المتجمع الهابط للفئة الثانية هو عبارة عن ٥٠ – ١ = ٤٩ . وهكذا بالنسبة لبقية الفئات .**

- يمكن تكوين الجدول التكراري المتجمع الهابط بطريقة أسهل عن طريق اتباع نفس الأسلوب المستخدم في حالة الجدول التكراري المتجمع الصاعد. وذلك من خلال بداية عملية التجميع المتتالية من أسفل لأعلى على عكس التوزيع التكراري المتجمع الصاعد . فمثلاً التكرار المتجمع الهابط المقابل للحد الأدنى للفئة الأخيرة ( وهو ٩٠ جنيه ) يقابله التكرار المتجمع الهابط ( ٥ عامل ) وهو نفس تكرار الفئة الأخيرة في التوزيع التكراري الأصلي . بينما يكون التكرار المتجمع الهابط للفئة قبل الأخيرة أو الحد الأدنى للفئة قبل الأخيرة فهو عبارة عن تكرار الفئة الأخيرة مضافاً إليه تكرار الفئة قبل الأخيرة وذلك من التوزيع التكراري الأصلي أي ٥ + ٨ = ١٣ وهكذا إلى أن نصل إلى التكرار المتجمع الهابط المقابل للحد الأدنى للفئة الأولى والذي يجب أن يكون مساوياً المجموع الكلي للتكرارات .

أو يمكن القول أن التكرار المتجمع الهابط لفئة ما هو إلا التكرار المتجمع الهابط للفئة السابقة مطروحاً منه تكرار الفئة السابقة من التوزيع الأصلي . فمثلاً التكرار المتجمع الهابط للفئة الأولى هو (٥٠) نظراً لعدم وجود فئات سابقة للفئة الأولى . أما التكرار المتجمع الهابط للفئة الثانية فما هو إلا التكرار المتجمع الهابط للفئة الأولى من التوزيع التكراري الأصلي (١ عامل ) . كذلك فإن التكرار المتجمع الهابط للفئة الثالثة ما هو إلا التكرار المتجمع الهابط للفئة الثانية ( ٤٩ عاملاً ) مطروحاً منه التكرار الأصلي للفئة الثانية (٤ عامل ) أي أن ٤٩ – ٤ = ٤٥ . وهكذا إلا أن نصل إلى تكرا الفئة الأخيرة ( ٩٠ فأكثر ) المتجمع الهابط فللفئة قبل الأخيرة ( ٨٠ فأكثر ) أي ( ١٣ عاملاً ) مطروحاً منه تكرار الفئة (٨٠– ) من التوزيع الأصلي ( أي ٨ عامل ) أي أن ١٣ – ٨ = ٥ . أما عن السطر الأخير في الجدول التكراري المتجمع الهابط ( ١٠٠ فأكثر ) فتعتبر كتابته اختيارية . وفي هذا الخصوص

يفترض أن هناك فئة وهمية تلي الفئة الأخيرة فيكون الحد الأدنى لتلك الفئة الوهمية ما هو إلا الحد الأعلى للفئة الأخيرة في التوزيع ( ٩٠ – ١٠٠ ) أي القيمة ١٠٠ جنيه . فلا شك أن ليس هناك أي عامل يزيد أجره عن ١٠٠ على حسب بيانات هذا المثال

- التكرارات المتجمعة الهابطة تبدأ بمجموع التكرارات وتنتهي بتكرار الفئة الأخيرة ( وفي حالة الوقوف بفئات التوزيع التكراري المتجمع الهابط عند الحد ( ٩٠ فأكثر في مثالنا محل الدراسة ) أو تنتهي بالتكرار المتجمع الهابط صفراً وذلك في حالة افتراض أن هناك فئة وهمية تلي الفئة الأخيرة وأن الحد الأدنى لتلك الفئة ما هو إلا الحد الأعلى للفئة الأخيرة .
- الجدول التكراري المتجمع الهابط النسبي ما هو إلا صورة من الجدول التكراري المتجمع الهابط مع الاختلاف الوحيد وهو أن التكرارات النسبية المتجمعة الهابطة النسبية ( R.Dsc.C.F.T ) ما هي إلا التكرارات المتجمعة الهابطة منسوبة إلى أو مقسومة على مجموع التكرارات .
- من الجدول التكراري المتجمع الهابط يمكن الحصول على كثيراً من المعلومات والتي لا يمكن الحصول عليها بطريقة مباشرة من خلال استخدام التوزيع التكراري الأصلي . فمثلاً :

– هناك ٣٨ عاملاً تبلغ أجورهم ٦٠ جنيه فأكثر أسبوعيًا .

– أن قيمة الأجر الذي يحصل عليه أو أكثر منه ٤٩ عاملاً هو ٤٠ جنيه .

– أنه يمكن تحديد عدد العاملين الذين يحصلون على أجوراً تزيد على أو تساوي قيمة معينة لا توجد بصورة صريحة في الجدول المتجمع . وكذلك يمكن تحديد قيمة الأجر الذي يحصل عليه أو أكثر منه عدداً معيناً من العمال حتى لو كانت هذه القيم غير موضحة صراحة في الجدول . وهو ما

سنراه باستخدام طريقة الاستكمال العددي أو النسبة والتناسب كما سنراه في نهاية هذا المثال أو أثناء تناولنا لدراسة الوسيط كأحد مقاييس المتوسطات أو مقاييس النزعة المركزية وهو ما سيرد فيما بعد في أبواب لاحقة بمشيئة الله .

- من خلال قراءة الجدولين ( ١٧ – ٢ ) ، ( ١٨ – ٢ ) لاحظ الآتي :-

( التكرار المتجمع الصاعد + التكرار المتجمع الهابط ) لأي فئة من فئات التوزيع مساوياً لمجموع تكرارات التوزيع التكراري الأصلي فمثلاً عدد العاملين الحاصلين على أقل من ٣٠ جنيه + عدد العاملين الحاصلين على ٣٠ جنيه فأكثر = صفر + ٥٠ = ٥٠ عامل وهو نفس المجموع الكلي للتكرارات . كذلك عدد العاملين الحاصلين على أقل من ٤٠ جنيه + عدد العاملين الحاصلين على ٤٠ جنيه فأكثر = ١ + ٤٩ = ٥٠ عامل وهو نفس المجموع الكلي للتكرارات . وهكذا لكل فئات التوزيعين المتجمع الصاعد والهابط . وكذلك على مستوى التوزيعات التكرارية النسبية المتجمعة الصاعدة والهابط فإن مجموع التكرارين المتجمعين الصاعد والهابط النسبيان يكون مساوياً للواحد الصحيح أو ١٠٠ %

أي أن :-

( التكرار المتجمع الصاعد + التكرار المتجمع الهابط ) = مجموع تكرارات التوزيع الأصلي لأي فئة من فئات التوزيع

وكذلك فإن :

التكرار المتجمع الصاعد النسبي + التكرار المتجمع الهابط النسبي = ١ أو ١٠٠% لأي فئة من فئات التوزيع .

**٣- <u>باستخدام التوزيع التكراري الملائم تحديد</u> :-**

**(أ) – عدد العمال اللذين تتراوح أجورهم ما بين ٨٠ جنيه حتى أقل من ٩٠ جنيه وكذلك نسبة هذا العدد من إجمالي التكرارات .**

**بداية مثل هذه التساؤلات يتم الإجابة عليها أولاً من خلال البحث عن ماهية وجود فئة في التوزيع التكراري الأصلي كما هو معطى في المثال عدمه . فإذا كانت هناك فئة صريحة في التوزيع الأصلي فإن الإجابة تكون من خلال التوزيع التكراري الأصلي وإلا فإن الإجابة تكون من خلال أحد التوزيعات التكرارية المتجمعة كما سنرى حالاً .**

**فبالنسبة للفئة ( ٨٠ – ٩٠ جنيه ) فهناك فئة صريحة في التوزيع التكراري الأصلي تفيد بوجود ٨ عمال يتراوح أجورهم ما بين ٨٠ ، أقل من ٩٠ جنيه . ومن ثم فإن النسبة المطلوبة وهي : –**

**نسبة العمال اللذين تتراوح أجورهم ما بين ٨٠ أقل من ٩٠ جنيه**

**= ( $^{٨}/_{٥٠}$ ) × ١٠٠ = ١٦%**

**أي أن هناك ١٦% من إجمالي عدد العاملين تتراوح أجورهم ما بين ٨٠ وأقل من ٩٠ جنيه اسبوعياً .**

**(ب) – عدد العمال اللذين تبلغ أجورهم الاسبوعين ٧٣ جنيه على الأكثر وكذلك النسبة .**

**<u>لاحظ أن كلمة على الأكثر</u> = أقل من كما سبق أن أوضحنا . لذا فالمطلوب يمكن استنتاجه من الجدول التكراري المتجمع الصاعد ( أو من المنحنى التكراري المتجمع الصاعد كما سنرى فيما بعد ) .**

**في مثل هذه الحالات لاحظ تمييز العمودين اللذين يشكلان الجدول التكراري المتجمع الصاعد حيث تمثل فئات الأجور بالجنيه أما التكرارات المتجمعة الهابطة فوحدة قياسها هي العامل . والمعطى ٧٣ جنيه تقع في عمود الفئات**

المتجمعة . وهنا يجب التنويه أن فكرة إنشاء الجداول التكرارية البسيطة وكذا المتجمعة تبنى على أساس أن التكرارات تتوزع بانتظام
( بالنسبة والتناسب ) على مدار طول الفئة الواحدة . لذا فإن قيمة الأجر الأسبوعي ٧٣ جنيه تقع ما بين الأجر ٧٠ جنيه ، ٨٠ جنيه اللذان يقابلهم تكرارات متجمعة صاعدة كما يوضحها الجدول (١٧–٢) هي ٢٦ ، ٣٧ عامل. لذا نفرض أن عدد العمال اللذين تقل أجورهم عن ٧٣ جنيه وليكن (A) عامل ولاستنتاج القيمة A يتم تصوير الشكل الكروكي التالي :–

| التكرارات المتجمعة الصاعدة | | الفئات المتجمعة الصاعدة |
|---|---|---|
| 26 | ← | Less than 70 |
| A | ← | Less than 73 |
| 37 | ← | Less than 80 |

وباستخدام فكرة النسبة والتناسب فإن :

الحد الأخير – الحد الأول / الحد الأخير – الحد الأوسط (للفئات المتجمعة) = الحد الأخير – الحد الأول / الحد الأخير – الحد الأوسط (للتكرارات المتجمعة)

أي أن :

$$\frac{80-70}{80-73} \qquad \frac{37-26}{37-A}$$

i.e.,

$$\frac{10}{7} = \frac{11}{37-A}$$

i.e., $10 (37 - A) = 77$
i.e., $370 - 77 = 10A$

فتكون قيمة $A = 293 \div 10$

$= 29.3 \quad 29$ (Persons)

ومن ثم فإن عدد العمال اللذين تقل أجورهم عن ٧٣ جنيه ما هو إلا ٢٩ عاملاً . لاحظ أن الناتج لابد أن يقع ما بين التكرارين المتجمعين الصاعدين أي ما بين ٢٦ ، ٣٧ عاملاً وهو معياراً لدقة الحسابات . ولاحظ أيضاً أن الناتج ( ٢٩ عاملاً ) يقترب من التكرار المتجمع الصاعد (٢٦) حيث أن المقابل على مستوى الفئات المتجمعة نجد أن الأجر ٧٣ جنيه يقترب أيضاً من ٧٠ جنيه وليس ٨٠ جنيه . أما عن النسبة فإن نسبة العمال اللذين تقل أجورهم عن ٧٣ جنيه ( أو تبلغ أجورهم ٧٣ جنيه على الأكثر ) ما هي إلا النسبة ما بين عدد العمال الذين تقل أجورهم عن ٧٣ جنيه وبين إجمالي التكرارات أي أن :

$$R.Asc.C.F_i = ( Asc.C.F_i \div \sum F_i ) \times 100 = ( 29 \div 50 ) \times 100$$
$$= 58\%$$

أي أن هناك ٥٨% من إجمالي عدد العمال تبلغ أجورهم ٧٣ جنيه على الأكثر.

(جـ) لتحديد قيمة الأجر الذي يتقاضاه ٦٥% من إجمالي عدد العمال على الأقل . لاحظ أولاً أن اللفظ على الأقل = فأكثر . لذا فإن الإجابة عن هذا التساؤل تكون من التوزيع التكراري المتجمع الهابط أو من التوزيع التكراري المتجمع الهابط النسبي .

فحيث أن ٦٥% من إجمالي عدد العمال ما هو إلا :

$$50 \times 65\,\% = 32.5 \approx 33 \qquad \text{(Persons)}$$

وهذا العدد الناتج ( ٣٣ شخص ) يقع ما بين التكرارين المتجمعين الهابطين ٣٨،٢٤ كما هو موضح في الجدول (١٨-٢) أو أن النسبة ٦٥% من الإجمالي تقع ما بين ٧٦%، ٤٨% الواردة بنفس الجدول .

لذا نفرض أن قيمة الأجر الذي يتقاضاه ٣٣ عاملاً على الأقل (٦٥%) هو الأجر (x) . ولاستنتاج قيمة (x) يتم استخدام قاعدة النسبة والتناسب كما يلي :

$$\text{More than } 60 \longrightarrow 38$$
$$\text{More than} \times \longrightarrow 33$$
$$\text{More than } 70 \longrightarrow 24$$

فيكون

$$\frac{70-60}{70-x} = \frac{24-38}{24-33}$$

$$\frac{10}{70-x} = \frac{-14}{-9}$$

$$\text{i.e.,} \quad 90 = 14\,(70 - x)$$

$$14x = 890$$

$$\text{i.e.,} \quad x = 63.571 \quad (\text{L.E})$$

أي أن الدخل الذي يتقاضاه على الأقل ٦٥% من إجمالي عدد العمال هو ٦٣.٥٧١ جنيهاً .

لاحظ أن الدخل الناتج يقع ما بين ٦٠ ، ٧٠ جنيه وهو معيار لدقة الحسابات فيجب أن لا يخرج عن الحدود المتجمعة المقابلة .

**وعلى الطالب استنتاج نفس المطلوب باستخدام التكرارات النسبية المتجمعة الهابطة ويتحقق من الوصول لنفس النتائج .**

**(د) لتحديد قيمة الأجر الذي يبلغه كحد أدنى ٤٧ عاملاً . لاحظ أن اللفظ كحد أدنى = فأكثر.لذا يتم استنتاج المطلوب من الجدول التكراري المتجمع الهابط . فبملاحظة العدد ٤٧ عاملاً نجد أنه يقع في عمود التكرارات المتجمعة الهابطة ( وليس الفئات ) ما بين التكرارين المتجمعين الهابطين ٤٩،٤٥ . وبافتراض أن الأجر المطلوب وليكن ( y). فلاستنتاج قيمة (y) يتم تكوين الرسم الكروكي التالي :**

**استخدام قاعدة النسبة والتناسب كما يلي :**

More than 40 ⟶ 49
More than (y) ⟶ 47
More than 50 ⟶ 45

**فيكون :-**

$$\frac{50 - 40}{50 - Y} = \frac{45 - 49}{45 - 47}$$

i.e., $\frac{10}{50 - Y} = \frac{-4}{-2} = 2$

i.e. ,

$$10 = 2\,(50 - y)$$
$$2y = 100 - 10$$

i.e. ,

$$y = 90 \div 2 = 45 \qquad \text{(L.E)}$$

**أي أن قيمة الأجر الذي يتقاضاه ٤٧ عاملاً كحد أدنى ما هو إلا ٤٥ جنيه**

**ملحوظة : لاحظ أنه حينما كان عدد العمال ٤٧ يقع في منتصف التكرارات المتجمعة الهابطة ما بين ٤٩ ، ٤٥ عاملاً بالضبط فإن الناتج وهو قيمة الدخل (Y) وهي القيمة ٤٥ جنيه تقع في منتصف الفئة بالضبط أي في نقطة المنتصف ما بين ٤٠ ، ٥٠ جنيه .**

**وبلغة أخرى : فحيث أن**

$$47 = \frac{49 + 45}{2} \qquad \text{(Persons)}$$

$$Y = \frac{40 + 50}{2} = 45 \qquad \text{(L.E)}$$

**هـ – لتحديد نسبة العمال الذين تتراوح أجورهم ما بين ٥٥ ، ٨٧ جنيه فحيث أنه لا توجد فئة صريحة من فئات التوزيع التكراري الأصلي الواردة بالجدول ( ١٦–٢) تعطي صورة مباشرة عن تلك الفترة أو هذه الفئة من ( ٥٥–٨٧ ) لذا يجب استخدام أحد الجداول المتجمعة أما الصاعد أو الهابط. والآن دعنا نستنتج المطلوب من التوزيع التكراري المتجمع الصاعد .**

**فحيث أن الدخل ٥٥ جنيه يمثل نقطة المنتصف للفئتين ٥٠ ، ٦٠ جنيه . لذا فإن :**

**عدد العمال اللذين يقل أجرهم الأسبوعي عن ٥٥ جنيه = (٥+١٢) ÷ ٢ أي ٧.٥ ≈ ٨ عاملاً**

**هذا وبافتراض أن عدد العمال اللذين يقل أجرهم الأسبوعي عن ٨٧ جنيه وليكن B عامل فإن لاستنتاج قيمة (B) من الجدول التكراري المتجمع الصاعد الوارد في الجدول ( ١٧–٢ ) فإن :**

Less than 80 $\longrightarrow$ 37
Less than 87 $\longrightarrow$ B
Less than 90 $\longrightarrow$ 45

فيكون :

$$\frac{90-80}{90-87} = \frac{45-37}{45-B}$$

**i. e. ,**

$$\frac{10}{3} = \frac{8}{45-B}$$

$24 = 10\,(45 - B)$

$10\,B = 450 - 24 = 426$

i.e. , $B = 42.6 \approx 43$ ( Persons )

**والآن دعنا نلخص النتائج المستخلصة حتى الآن :-**

**هناك ٨ عمال يقل أجرهم عن ٥٥ جنيه**

**، وهناك ٤٣ عامل يقل أجرهم عن ٨٧ جنيه .**

**أي يمكن تصور الشكل الكروكي التالي**

Less than 30 L.E $\longrightarrow$ 0 Persons
Less than 55 L.E $\longrightarrow$ 5 Persons
L.E Less than 87 $\longrightarrow$ 43 persons

**فيكون عدد العمال اللذين تتراوح أجورهم الأسبوعية ما بين ٥٥ جنيه ، ٨٧ جنيه هو بمثابة الفرق ما بين ٤٣ ، ٥ عمال أي ٤٣ – ٥ = ٣٨ عاملاً . وعليه فإن نسبة العمال اللذين تتراوح أجورهم الأسبوعية ما بين ٥٥ ، ٨٧ جنيه هي :**

$$R_i = \frac{F_i}{\sum F_i} \times 100 = \frac{38}{50} \times 100 = 76\ \%$$

**أي أن هناك ٧٦% من إجمالي عدد العمال يتراوح أجرهم الأسبوعي ما بين ٥٥ ، ٨٧ جنيهاً .**

**(و) – حيث أنه لا توجد فئة من فئات التوزيع التكراري الأصلي الوارد بالجدول (١٦–٢) تبدأ بالقيمة ٥٣ جنيه . لذا فإن هذا المطلوب لا يمكن استنتاجه من التوزيع التكراري الأصلي . ويمكن استنتاج هذا المطلوب من أحد التوزيعين التكرارين المتجمعين أما الصاعد وأما الهابط .**

**والآن دعنا نلخص معطيات هذا المطلوب واستنتاجه وليكن من التوزيع التكراري المتجمع الصاعد .**

**فحيث أن ٤٠% من إجمالي عدد العمال هو ٥٠ × ٤٠% = ٢٠ عامل .**

**وحيث أن عدد العمال اللذين يقل أجرهم عن ٥٣ جنيه يمكن استنتاجه من خلال افتراض أن هذا العدد وليكن (z) عامل . فلاستنتاج قيم (z) يتم الآتي :**

Less than 50 L.E $\longrightarrow$ 5
Less than 53 L.E $\longrightarrow$ 7
Less than 60 L.E $\longrightarrow$ 12

**فيكون :**

$$\frac{60-50}{60-53} = \frac{12-5}{12-z} \qquad \text{i. e. ,}$$

$$\frac{10}{7} = \frac{7}{12-z} \longrightarrow 49 = 10\ (12 - z)$$

i.e., $10\,z = 120 - 49 = 71$

أي أن $z = 7.1 \approx 7$ (persons)

**لاحظ أن الناتج ٧ أشخاص يقع ما بين التكرارين المتجمعين الصاعدين ٥ ، ١٢ المقابلين للفئات المتجمعة ٥٠ ، ٦٠ جنيه .**

**والآن لاستنتاج قيمة الدخل (x) فدعنا نلخص المعطيات ونتائجنا حتى الآن وهي :-**

- **هناك ٧ عامل يقل أجرهم الاسبوعي عن ٥٣ جنيه .**
- **هناك ٢٠ عاملاً يتراوح دخلهم ما بين ٥٣ ، x جنيه**

**ولاستنتاج قيمة الدخل x دعنا نفرض الفرضين التاليين :-**

*** X < ٥٣ جنيه     * x > ٥٣ جنيه**

**ولذا مدى صحة هذين الفرضين من خلال الرسم الكروكي التالي :**

***I*** **– في حالة افتراض أن**

**x < ٥٣ جنيه فإن :**

30 L . E

7 عامل ↓ 20 ↓ X L . E

53 L . E

**وهـذا مرفـوض لأن القيمــة (٢٠) عامل لا يمكن أن تكون جزءاً من الكل وهو ( ٧) عمال .**

***II*** **– في حالة افتراض أن**

**x > ٥٣ جنيه فإن :**

7 ↓ 30 L . E

53 L . E

**20** عامل ↓ X L . E

**ونظراً لاستقلال السهمين فإن هذا الاتجـــاه مقبـــول رياضـــياً .**

ومن خلال الشكلين السابقين يتضح لنا أن الفرض الثاني هو الاتجاه الوحيد المقبول لدينا في هذا المثال . ومن ثم فإن :

عدد العمال اللذين يقل أجرهم عن (x) جنيه = ٧ + ٢٠ = ٢٧ عامل.

ومن ثم فالمطلوب الآن هو تحديد قيمة الأجر (x) الذي يقل عنه ٢٧ عاملاً. وبالرجوع للجدول التكراري المتجمع الصاعد الوارد بالجدول ( ١٧ – ٢ ) نجد أن التكرار المتجمع الصاعد ٢٧ يقع ما بين التكرارين المتجمعين الصاعدين ٢٦ ، ٣٧ المقابلين للفئتين المتجمعتين ٧٠ ، ٨٠ جنيه على الترتيب ومن ثم يصبح لدينا الشكل التالي :

$$\text{Less than } 70 \longrightarrow 26$$
$$\text{Less than } x \longrightarrow 27$$
$$\text{Less than } 80 \longrightarrow 37$$

فيكون :

$$\frac{80-70}{80-x} = \frac{37-26}{37-27}$$

i. e. ,

$$\frac{10}{80-x} = \frac{11}{10} \longrightarrow 100 = 11\,(80-x)$$

i.e. , $11x = 88 - 100 = 780$

فتكون قيمة x هي :

$X = 780 \div 11 = 70.909$ ( L.E )

أي أن قيمة الدخل (x) هو ٧٠.٩٠٩ جنيه .

<u>ملحوظة</u> :- في المطلوب الأخير من هذا المثال لاحظ أنـه تم رفض الفرص الأول الـذي افتـرض أن قيمـة x تقـل عـن ٥٣ جنيـه نظـراً لأن هـذا الفـرض يتنـافى مـع المنطـق الرياضـي الـذي يحكم بضرورة أن يكـون الجـزء أقـل مـن الكل . أي أن هذا الفرض سيتم رفضـه فـي الحـالات التـي يتنـافى فيهـا مـع المنطق الرياضي . لكن إذا كانـت معطيـات التمرين ومـا تـم استنتاجه أفـادت أن الجزء أقل من الكل فسوف تكـون النتيجـة هـي قبـول هذا الفرض أيضـاً . وفي هذه الحالـة يتم طرح التكرارين المتجمعين اللذين تم استنتاجهم بـدلاً مـن الجمع كما هو في الفرض الثاني من هذا المثـال وخلاصـة مـا سبق هو أنـه في حالة قبول الفرض الأول فستكون هناك إجابتين صـحيحتين لقيمـة الأجر (x) . حيث أن الاتجاه الثاني سيكون مقبول دائماً نظراً للاستقلال دائماً فيما بين الجزئين اللذين تم استخلاصهم من المعطيات ونتائـج الحل .

**ثانياً : ـ العرض البياني للبيانات الاحصائية .**

رأينا فيما سبق أن الجداول التكرارية بالرغم من دقتها في تلخيص الظواهر أو المتغيرات وعرض ما بينهما من علاقات ( الجداول التكرارية المزدوجة ) إلا أنه قد يصعب معها في كثير من الأحيان على القارئ العادي استقراء مدلولات تلك الأرقام أو تفهم العلاقة فيما بين المتغيرات المختلفة خصوصاً في ظل حالات البيانات الضخمة . لذلك يأتي دور العرض البياني للظواهر المختلفة مكملاً لدور العرض الجدولي للبيانات . حيث تتيح طرق العرض البياني إعطاء نبذة أو فكرة سريعة وصحيحة عن كيفية تغيير تلك الظواهر محل الدراسة وبيان نوع ومدى هذا التغيير وذلك من خلال رسم بياني في حيز معقول وفي مده وحيزه .

وخلاصة الأمر هو أنه رغم أن عملية عرض البيانات جدولياً تيسر على القارئ فهم هذه البيانات كما تساعد على استخلاص النتائج والعلاقات التي تربط بين الظواهر المختلفة إلا أن هذه العملية ( العرض الجدولي ) قد لا تلقي قبولاً من القارئ العادي الغير متخصص حيث يجد صعوبة في عملية استنباط النتائج من واقع بيانات الجداول التكرارية . ولعل العرض البياني للبيانات يكون أكثر جاذبية للقارئ وذلك لما يعكسه من سرعة في فهم الظاهرة موضع الدراسة . إذ أنه من اليسير على القارئ حتى غير المتخصص – فهم طبيعة تطور الظاهرة بمجرد النظر للشكل البياني . أضف إلى ذلك أن العرض البياني للبيانات كثيراً ما يستخدم في إجراء عمليات المقارنة والتي يمكن إدراك نتائجها بشكل أيسر وأسرع مما لو استخدمنا العرض الجدولي لهذه البيانات . هذا بالإضافة إلى أن العرض البياني يمكننا من تحديد أشكال التوزيعات التكرارية المختلفة . وبصفة عامة وكما هو الحال في العرض الجدولي فإن العرض البياني للظواهر يعتمد على طبيعة البيانات أي طبيعة المتغيرات محل الدراسة من حيث كونها وصفية أم كمية

بنوعيها المنفصل أو المتصل . أي أننا سنتناول بشيء من الإيجاز والتوضيح العرض البياني للظواهر أو المتغيرات :-

أ – الوصفية

ب- الكمية بنوعيها المنفصل والمتصل والمبوبة تكرارياً

هذا بالإضافة إلى أنه سيتم التعرض لكيفية عرض بيانات السلاسل الزمنية وهي عبارة عن تناول دراسة لظاهرة ما خلال مجموعة من الفترات الزمنية المتساوية . هذا وسوف تقتصر دراستنا في تناولنا لأساليب العرض البياني للبيانات على أهم وأكثر هذه الأساليب استخداماً .

**أولاً :- العرض البياني للبيانات الوصفية :-**

وهناك مجموعة من الأساليب البيانية التي تستخدم في عرض جداول الظواهر الوصفية بيانياً منها :

**١- الأعمدة البيانية :-** Bar Charts

وهي عبارة عن مجموعة من الأعمدة الرأسية ( مستطيلات ) قواعدها متساوية ( سواء منفصلة أو متصلة أو مجزأة ) تتناسب في ارتفاعاتها مع قيم الظاهرة التي تعكسها تلك الأعمدة بيانياً . وفي هذه الحالة يتم تمثيل أقسام أو فئات الظاهرة على المحور الأفقي بينما يتم تخصيص المحور الرأسي كقياس الظاهرة محل الدراسة حسب وحدة قياسها حيث يتم تقسيم هذا المحور الرأسي بمقياس رسم مناسب إلى أقسام متساوية .

وتستخدم الأعمدة البيانية في مقارنة مجموعتين أو أكثر عند النقاط الزمنية المختلفة ( سنة أو شهر أو يوم أو ..... إلخ ). كما يمكن أن تستخدم هذه الأعمدة فى التعرف على التطور التاريخى لظاهرة ما أو أكثر خلال فترات زمنية مختلفة. ولرسم الشكل البيانى باستخدام الأعمدة البيانية يتم الآتى :

بعد تقسيم المحور الأفقي حسب أقسام الظاهرة يتم رسم مجموعة من المستطيلات المنفصلة ( الأعمدة ) ذات قواعد متساوية ويتم تخصيص كل مستطيل أو عمود لقسم من أقسام الظاهرة مع ترك مسافة مناسبة فيما بين كل عمودين . وهناك رأى بترك مسافة مساوية لنصف أو ثلثى قاعدة تلك المستطيلات المعبرة عن أقسام الظاهرة .

- يتم تحديد ارتفاع كل عمود بقيمة كل قسم من أقسام الظاهرة أو قيمة الظاهرة عند السنة التي يمثلها هذا العمود .
- يتم كتابة اسم كل قسم من أقسام الظاهرة أسفل كل عمود .

والأمثلة التالية توضح ذلك .

**مثال (٧) :**

فيما يلي الجدول التالي يوضح تقديرات نجاح ١٥ طالباً في مادة الاقتصاد :

| Grade | pass | Good | Very good | Excellent | $\sum$ |
|---|---|---|---|---|---|
| Frequency | 6 | 5 | 3 | 1 | 15 |

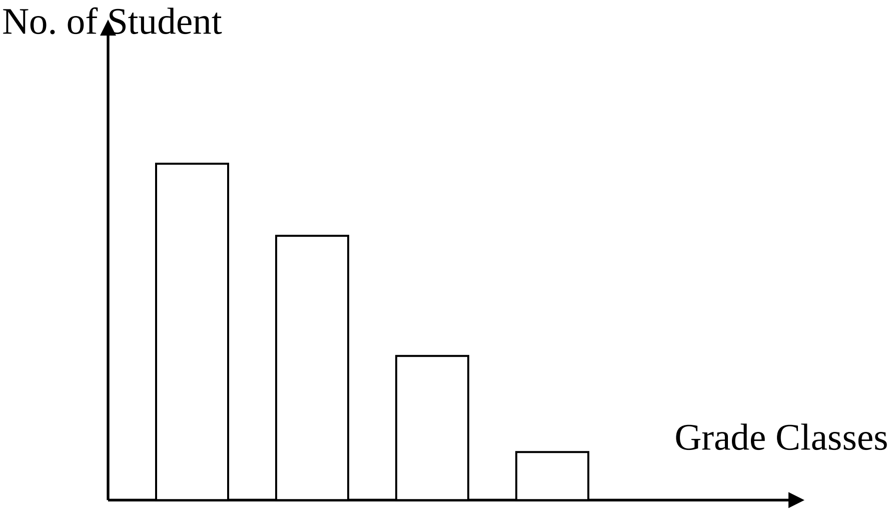
والمطلوب تمثيل بيانات هذا الجدول باستخدام الشكل البياني الملائم .

**الحل :-**

حيث أن الظاهرة محل الدراسة – تقدير الطالب – هي ظاهرة وصفية لذا يمكن عرضها بيانياً باستخدام الأعمدة البيانية ( المستطيلات ) كما يوضحها الشكل (٢–١) .

**شكل (٢-١ )**

**أعمدة بيانية توضح أعداد الطلاب وفقاً لتقديرات نجاحهم في مادة الاقتصاد.**




**هذا وتجدر الإشارة إلى أن الأعمدة في تلك الحالة تسمى باسم الأعمدة البيانية البسيطة** Simple Bar Charts **وذلك للتفرقة بينهما وفيما بين الأعمدة الملتصقة** Clustered Bar Charts **والأعمدة المجزأة** Stacked Bar Charts

**مثال (٨) :- فيما يلي الجدول التالي يبين نتيجة طلاب إحدى كليات التجارة في إحدى السنوات :**

| النتيجة / الفرقة | طلاب ناجحون pass | طلاب منقولون بمواد Pass with | طلاب راسبون Failed | Σ |
|---|---|---|---|---|
| الأولى | ٢٠٠٠ | ١٢٠٠ | ٨٠٠ | ٤٠٠٠ |
| الثانية | ١٣٠٠ | ٤٠٠ | ٣٠٠ | ٢٠٠٠ |
| الثالثة | ٧٠٠ | ٥٠٠ | ٦٠٠ | ١٨٠٠ |
| الرابعة | ١٢٠٠ | ٨٠٠ | ١٠٠٠ | ٣٠٠٠ |

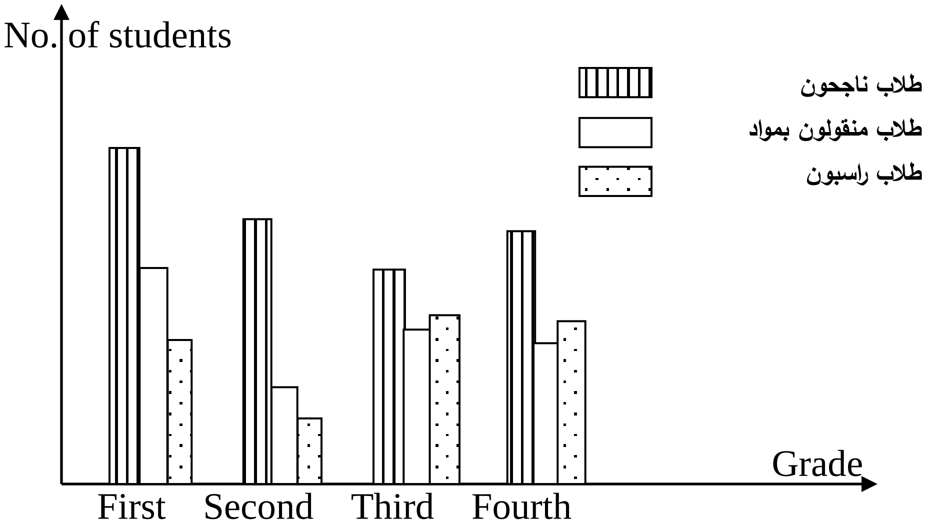
**والمطلوب استخدام الشكل البياني الملائم لعرض بيانات هذا الجدول .**

<u>الحـل</u> :- يمكـن تمثيـل البيانـات الـثلاث ( نـاجحون - منقولـون بمـواد - راسبون)

لكـل فرقـة علـى حـدة إمـا باسـتخدام الأعمـدة أو المسـتطيلات الملتصـقة أو المجزأة السابق الإشارة إليها فيما قبل المثال السابق وهو ما يوضحه الشكل البياني( ٢- ٢-أ ) ، ( ٢-٢- ب ) .

شكل ( ٢ - ٢ - أ )

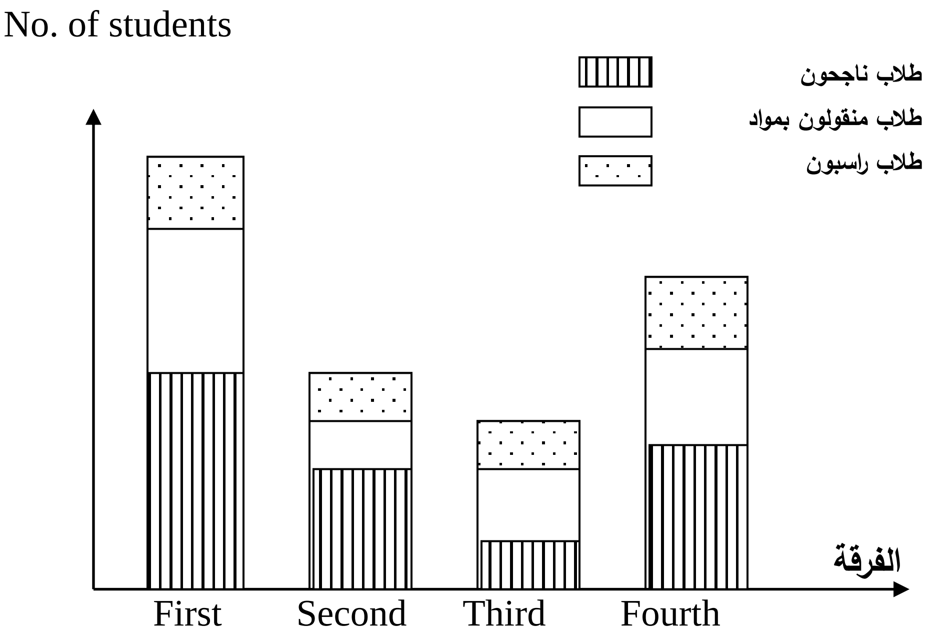
تمثيل بيانات نتيجة الفرق الدراسية الأربعة لإحدى كليات التجارة باستخدام الأعمدة الملتصقة .




حيث يتم تمثيل كل فرقة دراسية بثلاثة أعمدة ملتصقة تبين بالترتيب الطلاب الناجحون والمنقولون بمواد والراسبون أما بارتفاع تلك الأعمدة فهو يتناسب مع عدد الطلاب المقابل لكل فرقة من الفرق الأربعة .

أما عن تمثيل كل فرقة بعمود واحد يتم تقسيمه إلـى ثلاثـة أجزاء تبين علـى الترتيـب الطـلاب النـاجحين والمنقولـون بمـواد والراسـبون وهـو مـا يوضـحه الشكل ( ٢ - ٢ - ب ) .

**الشكل ( ٢ – ٢ – ب )**
**تمثيل بيانات نتيجة الفرق الدراسية الأربعة لإحدى كليات التجارة باستخدام الأعمدة المجزأة**




**هذا ويمكن تمثيل البيانات السابقة باستخدام سواء الأعمدة الملتصقة أو المجزأة وذلك على أساس نسبي متبعين في ذلك نفس الأسلوب السابق . حيث يكون في تلك الحالة ارتفاعات الأعمدة تتناسب في ارتفاعاتها مع نسبة تكرارات كل ظاهرة أو كل قسم من أقسام الظاهرة . ومن أهم مزايا هذا الأسلوب أنه يسهل عمليات المقارنة فيما بين أقسام الظاهرة في السنوات المختلفة وذلك على أساس نسبي وليس مطلق . ويكون ذلك نصفه خاصة إذا ما اختلفت المجاميع كما سبق ذكره .**

**ملاحظات :**

**عند استخدام الأعمدة البيانية كأسلوب من أساليب العرض البيانى فإنه يجب مراعاة ما يلى :–**

**١- أن تكون قواعد الأعمدة متساوية**

٢- حيث إن مساحة المستطيل (العمود)= القاعدة × الارتفاع.
فلما كانت قواعد الأعمدة متساوية فإن مساحة الأعمدة تتناسب مع ارتفاعها والتى تعكس بدورها قيم أقسام الظاهرة موضع الدراسة.
لذلك فإنه لا يجوز كسر المحور لأن ارتفاع كل عمود يمثل قيمة البيان.
وعندما تكون قيمة البيان كبيرة (شاذة)عن باقى القيم فإنه يمكن كسر العمود نفسه فى هذه الحالة مع كتابة قيمة البيان أعلى العمود كما سيرد فى المثال التالى :

٣- ان تكون المسافات فيما بين الاعمدة متساوية.

٤- فى حالة زيادة عدد الاعمدة فى الرسم زيادة كبيرة يمكن رسم محور رأسى أخر على يمين الرسم (مماثل للمحور الرأسى الاصلى) وذلك بغرض تسهيل قراءة وتمثيل قيمة البيان .

**مثال رقم (٨) :**

فيما يلى الجدول التالى يبين قيمة المبيعات بآلاف الجنيهات فى احدى السنوات وذلك لخمسة مصانع تنتج المنسوجات القطنية.

جدول(٢١-٢)

| المصنع | أ | ب | ج | د | هـ |
|---|---|---|---|---|---|
| قيمة المبيعات | ٢٤ | ٢٨ | ٢٠ | ٣٢ | ٨٦ |

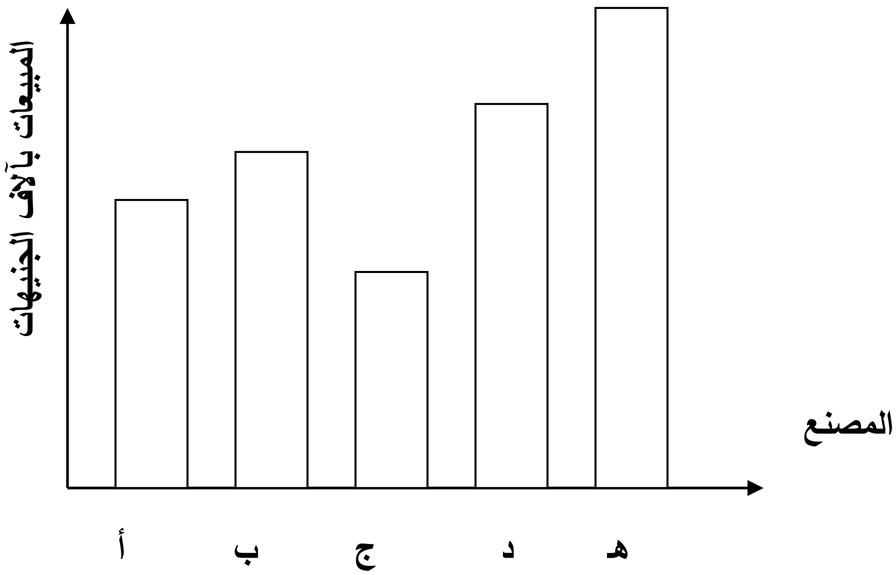
والمطلوب : عرض هذه البيانات بإستخدام الاعمدة البيانيه .

**الحل :**

لاحظ ان قيمة المبيعات للمصنع(هـ)كبيرة جدا بالمقارنة بباقى قيم المبيعات . لذا حتى يتم استخدام مقياس رسم مناسب لكافة بيانات الجدول يمكن كسر العمود الممثل لقيم مبيعات المصنع (هـ) وكتابة قيم مبيعاته اعلى العمود كما هو مبين فى الشكل(٢-٣) .

**شكل(٢-٣)**

**مبيعات خمسة مصانع للمنسوجات القطنية(بآلاف الجنيهات)**




**٢- الرسوم الدائرية :-** Pie charts

**تعتبر الدوائر من الاشكال الهندسية الشائعة لتمثيل الظاهرة محل الدراسة بيانيا وبصفة خاصة إذا كانت تلك الظاهرة مقسمة الى مجاميع جزئية . فإذا كانت لدينا ظاهرة واحدة فإنها تمثل بدائرة مناسبة دون أية قيود لتحديد نصف قطر تلك الدائرة . ثم يتم تقسيم تلك الدائرة او بمعنى زاوية مركز الدائرة الى قطاعات دائرية تتناسب مساحتها مع النسب المئوية للمكونات الجزئية للظاهرة.**

**ولما كانت الزاوية المركزية للدائرة مساوية ٣٦٠° فأن الزاوية المركزية لكل قطاع جزئى تتحدد بضرب تلك النسبة المئوية المقابلة لكل جزء او قسم من اقسام الظاهرة × ٣٦٠° . ثم يتم تظليل كل قطاع بشكل مخالف.**

**أما اذا كنا نود اجراء المقارنة فيما بين ظاهرتين كل واحدة تحتوى على مجموعة من الأوجهه ، ففى هذه الحالة يتم تخصيص دائرة لتمثيل كل**

ظاهرة على حده. ولكن يجب فى تلك الحالة رسم مساحة دائرتين متناسبتين مع قيم الظاهرتين ، وحيث ان :

$$\frac{\text{مساحة الدائرة الاولى}}{\text{مساحة الدائرة الثانية}} = \frac{\text{القيمة الاجمالية للظاهرة الاولى (ق١)}}{\text{القيمة الاجمالية للظاهرة الثانية(ق٢)}}$$

اى ان :

$$\frac{\text{ق}_{١}}{\text{ق}_{٢}} = \frac{\text{ط نق}^{٢}_{١}}{\text{ط نق}^{٢}_{٢}} = \frac{\text{نق}^{٢}_{١}}{\text{نق}^{٢}_{٢}}$$

حيث $\text{نق}_{١}$ ، $\text{نق}_{٢}$ هما نصفى قطر الدائرة الاولى والثانية على الترتيب وبإعتبار الجذر التربيعى للعلاقة السابقة فيكون :-

$$\frac{\text{نق}_{١}}{\text{نق}_{٢}} = \sqrt{\frac{\text{القيمة الإجمالية للظاهرة الأولى (ق}_{١}\text{)}}{\text{القيمة الإجمالية للظاهرة الثانية(ق}_{٢}\text{)}}}$$

فمثلا اذا كانت اجمالى قيم اقسام الظاهرة الاولى هو ١٦٠٠ واجمالى اقسام الظاهرة الثانية هو ٩٠٠ فأن:-

$$\frac{\text{نق}_{١}}{\text{نق}_{٢}} = \sqrt{\frac{١٦٠٠}{٩٠٠}} = \frac{٤}{٣}$$

فإذا افترضنا ان الدائرة الاولى نصف قطرها ٦ سم فأن:

نصف قطر الدائرة الثانية = ٦× ( $\text{نق}_{١} \div \text{نق}_{٢}$ ) = ٦ × ( ٣ ÷ ٤ )

= ٤,٥ سم

وبعد تحديد انصاف اقطار الدوائر المعبرة عن الظاهرتين يتم تقسيم كل دائرة الى قطاعات دائرية تتناسب مساحتها مع قيم مكونات كل ظاهرة على حده أى ان:-

$$\text{زاوية اى قطاع} = \frac{\text{قيمة الظاهرة فى هذا القطاع}}{\text{مجموع قيم الظاهرة}} \times ٣٦٠^\circ$$

<u>مثال (٩)</u> : فيما يلى البيانات التالية تمثل المبالغ المنفقه (بالمليار جنيه) على القطاعات المختلفة فى احدى الدول عام ١٩٩٤م :

| القطاع | المبالغ المنفقة بالمليار جنيه |
|---|---|
| القطاع الصناعى | ٢٠ |
| القطاع الزراعى | ١٦ |
| القطاع السياحى | ٤ |
| قطاع الخدمات | ٨ |

والمطلوب استخدام الدوائر لتمثيل هذه البيانات بيانيا.

<u>الحل:-</u>

لعرض هذه البيانات باستخدام الدائرة فإننا نقوم برسم دائرة بنصف قطر مناسب. ثم يتم تقسيم المساحة الكلية للدائرة الى قطاعات دائرية حيث ان :

$$\text{الزاوية التى تمثل اى قطاع} = \frac{\text{قيمة القطاع}}{\text{إجمالي القطاعات}} \times ٣٦٠^\circ$$

اى ان :

الزاوية التى تمثل القطاع الصناعى = (٢٠ ÷ ٤٨ ) × ٣٦٠ = ١٥٠°

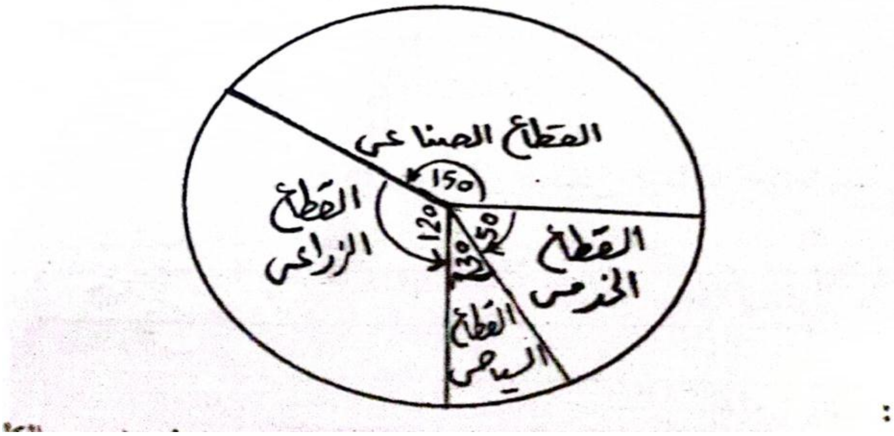
الزاوية التى تمثل القطاع الزراعي = (١٦ ÷ ٤٨ ) × ٣٦٠ = ١٢٠°

الزاوية التى تمثل القطاع السياحى = (٤ ÷ ٤٨ ) × ٣٦٠ = ٣٠°

**الزاوية التى تمثل القطاع الخدمى = (٨ ÷ ٤٨ ) × ٣٦٠ = ٦٠°**

**ومن ثم يكون شكل الدائرة على النحو الاتى :-**

شكل(2-4)




**مثال(١٠) :**

**الجدول التالى يوضح اعداد الطلاب (ذكورا واناثا) فى إحدى الكليات خلال السنوات(١٩٦٠-١٩٦٥) .**

**جدول(٢٢-٢)**
**اعداد الطلاب ذكورا واناثا فى احدى الكليات**
**خلال السنوات١٩٦٠-١٩٦٥**

| السنة | ١٩٦٠ | ١٩٦١ | ١٩٦٢ | ١٩٦٣ | ١٩٦٤ | ١٩٦٥ |
|---|---|---|---|---|---|---|
| ذكور | ١٢٠ | ١٥٠ | ١٧٥ | ٢١٠ | ٢٨٠ | ٣٢٠ |
| إناث | ٩٠ | ١٠٠ | ١٢٠ | ١٥٠ | ١٦٠ | ١٧٥ |

**المطلوب مقارنة عدد الطلاب بنوعيهم بيانيا باستخدام الدوائر**

**الحل :**

**للمقارنة فيما بين عدد الطلاب بنوعيهم فى عامى ١٩٦٠، ١٩٦٥ يتم الاتى :-**

**١- تحديد نصف قطر كلا من الدائرتين ورسم دائرة لكل نوع على حده:-**

**فحيث ان :**

$$\frac{\text{نق}_{١}}{\text{نق}_{٢}} = \sqrt{\frac{\text{مجموع الظاهرة خلال عام ١٩٦٠}}{\text{مجموع الظاهرة خلال عام ١٩٦٥}}}$$

$$\sqrt{\frac{٩٠+١٢٠}{١٧٥+٣٢٠}} = \sqrt{\frac{٢١٠}{٤٩٥}} = ٠.٦٥$$

**لذلك يمكن اختيار نصف قطر الدائرة الأولى = ٠,٦٥**

**ونصف قطر الدائرة الثانية = ١**

**او اختيار اى قيم اخرى لأنصاف الأقطار للدائرتين على ان تكون النسبة فيما بينهما مساوية دائما للقيمة٠.٦٥ فمثلا يمكن ضرب كلا من نق$_{١}$ ، نق$_{٢}$ فى اى مقدار ثابت وليكن مثلا(٥) فنجد ان:**

**نق$_{١}$=٠,٦٥ × ٥ =٣,٢٥ سم**

**نق$_{٢}$= ١ ×٥ = ٥ سم**

**٢- تحديد زوايا كل قطاع : وهى ما يوضحها الجدول التالى:**

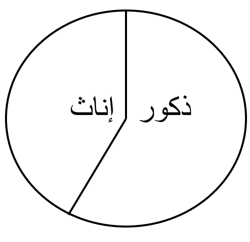
| بيان / النوع | ١٩٦٠ | | ١٩٦٥ | | زاوية القطاع | |
|---|---|---|---|---|---|---|
| | عدد الطلاب | النسبة % | عدد الطلاب | النسبة % | ١٩٦٠ | ١٩٦٥ |
| ذكور | ١٢٠ | ٥٧% | ٣٢٠ | ٦٥% | (١٢٠÷ ٢١٠) ×٣٦٠= ٢٠٦ | (٣٢٠÷ ٤٩٥) ×٣٦٠= ٢٣٣ |
| إناث | ٩٠ | ٤٣% | ١٧٥ | ٣٥% | (٩٠÷٢١٠) ×٣٦٠= ١٥٤ | (١٧٥÷ ٤٩٥) ×٣٦٠= ١٢٧ |
| مجـ | ٢١٠ | ١٠٠% | ٤٩٥ | ١٠٠% | ٣٦٠ | ٣٦٠ |

**وبذلك يمكن رسم الدائرتين مع تحديد القطاعات داخل كل دائرة كما يوضحه الشكل(٢-٥).**

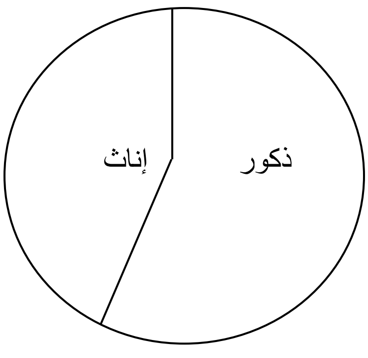
**شكل(٢-٥)**

**توزيع الطلاب(ذكورا ، واناثا) فى احدى الكليات فى عامى ١٩٦٥,١٩٦٠**

**عام ١٩٦٠**




**عام ١٩٦٥**



واخيرا فإن عيوب استخدام الدوائر فى التمثيل البيانى يمكن تحديده فيما يلى:

- تحتاج لحسابات ليست بالبسيطة
- المقارنة فيما بين قيم الظواهر فى الاقسام المختلفة يكون اصعب فى حالة الدوائر منه فى حالة الاعمدة البيانية.
- إذا كان عدد اقسام الظاهرة كبيرا فإنه من الافضل عدم استخدام الدائرة وذلك لصعوبة تمييز كل قطاع فى هذه الحالة ويفضل هنا عدم استخدام الأعمدة البيانية .

**ثانيا : العرض البيانى للظواهر الكمية المبوبة تكراريا :-**

**أ- العرض البيانى للمتغير الكمى المتقطع (المنفصل):-**

بصفة عامة فإنه في حالة البيانات المبوبة(الجداول التكرارية) للظاهرة التى تم التعبير عنها بمتغير كمى متقطع(منفصل) يتم تمثيلها بيانيا وذلك من خلال تحديد أقسام الظاهرة او المتغير على المحور الافقى ويتم رسم او إسقاط أعمدة تتناسب فى ارتفاعها مع التكرار المقابل لكل قسم من أقسام الظاهرة .اى ان ارتفاع كل عمود يكون مساويا لتكرار قيمة المتغير المتقطع.

**مثال(١١):** فيما يلى الجدول التالى بين توزيع ٢٠اسرة حسب حجم الاسرة:

جدول(٢٣-٢)

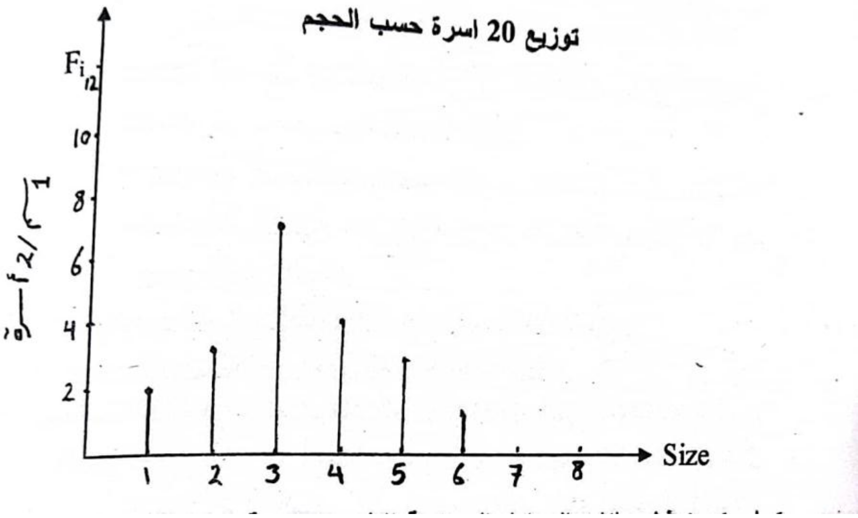
توزيع ٢٠ أسرة حسب الحجم

| Family size | 1 | 2 | 3 | 4 | 5 | 6 | $\sum$ |
|---|---|---|---|---|---|---|---|
| Frequency($F_i$) | 2 | 3 | 7 | 4 | 3 | 1 | 20 |

**الحل:-** يتم تمثيل حجم الاسرة على المحور الافقى والتكرارات على المحور الرأسى بمقياس رسم مناسب(١سم/وحدة تكرار). والشكل(٢-٦) يوضح التمثيل البيانى للجدول(٢٣-٢).

الشكل(2-6)

توزيع 20 اسرة حسب الحجم




**كمـا يمكـن تمثيـل بيانـات الجـداول التكراريـة للظـواهر الكميـة المنفصـلة باستخدام التكرارات النسبية بدلا من التكرارات المطلقة وفـى تلـك الحالـة فإنـه يتم تمثيل التكرارات النسبية على المحور الرأسى**

**ب _ التمثيل البيانى لجداول الظواهر الكمية المتصلة(المستمرة):**

**هنـاك ثلاثـة طـرق رئيسـية للتمثيـل البيـانى للجـداول التكراريـة التـى تحـوى متغيرات متصلة وهذه الطرق هى :**

- **المدرج التكرارى .**
- **المضلع التكرارى .**
- **المنحنى التكرارى .**

**هذا وعند رسم اى من الاشكال الثلاثة يجب التأكد من ماهية انتظام ( تساوى أطوال فئات) الجدول التكرارى من عدمه .فإذا كان الجدول التكرارى منتظما فلا توجد ادنى مشكلة فيتم الرسم مباشرة كما سنرى حالا . اما اذا كان الجدول التكرارى غير منتظم فإنه يجب تعديل التكرارات للتخلص من اثر**

**عدم الانتظام وذلك من خلال حساب ما يسمى بالتكرارات المعدلة** Modified Frequency**)**

**التكرار المعدل للفئة**($MF_i$) **= تكرار الفئة ( i ) ÷ طول الفئة ( i )**

**اى ان :**

$$MF_i = \frac{F_i}{L_i}$$

where $i = 1, 2, \ldots\ldots, r$

**وسوف نتعرض لدراسة الإشكال بشيء من التفصيل كما يلى :-**

**<u>١-المدرج التكرارى :</u>** Histogram

**والمدرج التكراراى ما هو إلا مجموعة من المستطيلات المتلاصقة كل مستطيل يعبر عن فئة من فئات التوزيع ويتناسب ارتفاعه مع تكرار تلك الفئة.**

**وفى تناولنا لرسم المدرج التكرارى كما سبق وإن ذكرنا يجب بداية ان نفرق فيما بين الجداول التكرارية المنتظمة والجداول التكرارية الغير منتظمة.**

**أ- <u>رسم المدرج التكرارى للجداول المنتظمة(متساوية اطوال الفئات):</u>**

**الخطوات :-**

- **رسم المحورين الافقى والرأسى . بحيث يتم تمثيل فئات الجدول على المحور الافقى وتكرارات تلك الفئات على المحور الرأسى بمقياس رسم مناسب . هذا ويراعى عند رسم المحور الافقى انه ليس من الضرورى ان يبدأ هذا المحور فى تدريجه من الصفر ولكن يبدأ من فئة سابقة عن الفئة الاولى فى الجدول وبنفس طول هذه الفئة الاولى . كما يراعى ان يكون هناك مسافة بعد الفئة الاخيرة فى الجدول، اما فيما يتعلق بالمحور الرأسى فأن تدريجه يجب ان يبدأ من الصفر على ان يتم اختيار مقياس الرسم المناسب للتكرارات الأصلية .**

- يتم تمثيل كل فئة بمستطيل تمثل قاعدته طول الفئة اما ارتفاعه فيمثل تكرار الفئة. وحيث ان الجدول التكرارى الذى نحن بصدده متساوى من حيث اطوال الفئات فإن مساحة المستطيلات المختلفة سوف تتناسب فى ارتفاعاتها مع التكرارات المقابلة التى تمثلها هذه المستطيلات .ومن ثم فإن إجمالى مساحات هذه المستطيلات المختلفة سوف يكون متناسبا مع المجموع الكلى للتكرارات.

والمثال التالى يوضح كيفية رسم المدرج التكرارى لجدول تكرارى منتظم .

**<u>مثال(١٢ ) :</u>فيما يلى لديك التوزيع التكرارى التالى لأجور ٥٠عاملا :**

**جدول(٢٣-٢)**

| Weakly Wages | 30- | 40- | 50- | 60- | 70- | 80- | 90- | ∑ |
|---|---|---|---|---|---|---|---|---|
| Frequencies | 1 | 4 | 7 | 14 | 11 | 8 | 5 | 50 |

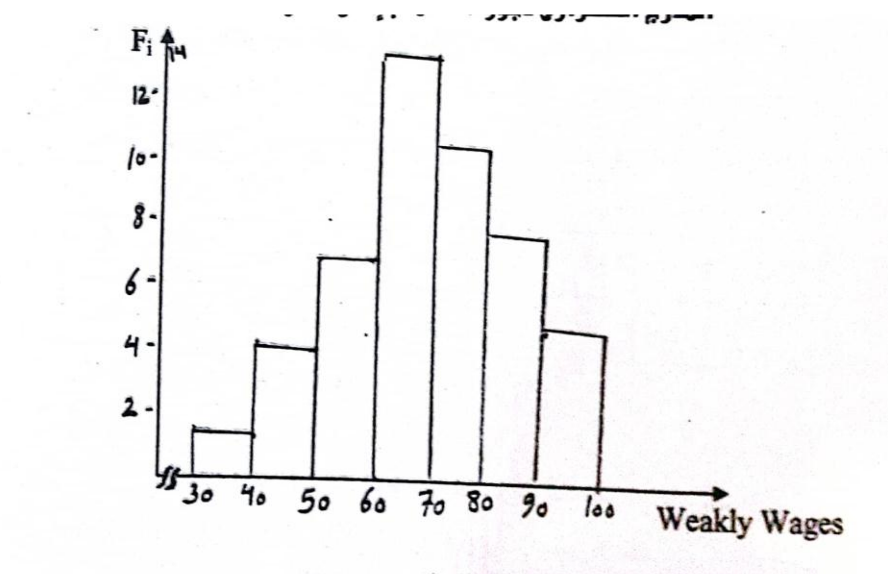
والمطلوب تمثيل هذا التوزيع بيانيا باستخدام الشكل البيانى الملائم

**<u>الحل:-</u>**

بالنظر لطبيعة الظاهرة الموضح عنها التوزيع التكرارى فهى عبارة عن اجور مجموعة من العمال.وهى ظاهرة كمية متصلة لذا فإن الشكل البيانى الملائم هو اما المدرج أو المضلع او المنحنى التكرارى .فدعنا نقوم بالتمثيل البيانى للجدول السابق من خلال رسم المدرج التكرارى.

وحيث ان : الجدول التكرارى فئاته متساوية من حيث الطول (جدول منتظم) لذا يتم رسم المدرج التكرارى مباشرة .حيث يتم تمثيل فئات الاجر على المحور الافقى وعدد العمال على المحور الرأسى بمقياس رسم وليكن ١سم/٢عامل كما هو موضح فى الشكل(٢-٧)

**شكل(٢–٧)**
**المدرج التكرارى لأجور ٥٠عاملا بإحدى الشركات**




**هذا ويمكن أيضا تمثيل تلك البيانات باستخدام التكرارات النسبية بدلا من التكرارات المطلقة.وفى تلك الحالة فإن المحور الرأسي يمثل التكرار النسبي للعمال .**

**ب- <u>رسم المدرج التكرارى فى حالة الجداول الغير منتظمة:</u>**

**فى حالة الجداول التكرارية الغير منتظمة أى الغير متساوية من حيث اطوال الفئات يجب حساب التكرار المعدل اولا للتخلص من أثر عدم تساوى اطوال الفئات بالقانون السابق الاشارة له وهو**

**اى ان :**

$$MF_i = \frac{F_i}{L_i}$$

where $i = 1,2,\ldots\ldots,r$

**وبعد حساب التكرارات المعدلة $MF_i$ يتم اتباع نفس خطوات الرسم البيانى مع فارق وحيد وهو انة المحور الرأسى يتم تمثيل التكرارات المعدلة عليه بدلا من التكرارات الاصلية والمثال التالى يوضح ذلك.**

**<u>مثال(١٣) :</u>**

**الجدول التالى يبين توزيع درجات مجموعة من الطلاب فى إحدى المواد.**

**جدول(٢٤-٢)**

| Classes | 5- | 10- | 20- | 30- | 50- | 60- | 80-100 | $\sum$ |
|---|---|---|---|---|---|---|---|---|
| No.of Students | 15 | 70 | 100 | 120 | 200 | 300 | 160 | 965 |

**والمطلوب رسم المدرج التكراري لتوزيع الدرجات**

**<u>الحل :-</u>**

**لاحظ ان فئات هذا التوزيع غير متساوية فى اطوالها . لذا يجب حساب التكرارات المعدلة لهذا التوزيع قبل رسم المدرج التكرارى وذلك من خلال قسمة التكرار الاصلى لكل فئة على طول الفئة كما يوضحه الجدول (٢٥-٢) الاتى :**

**جدول (٢٥-٢)**

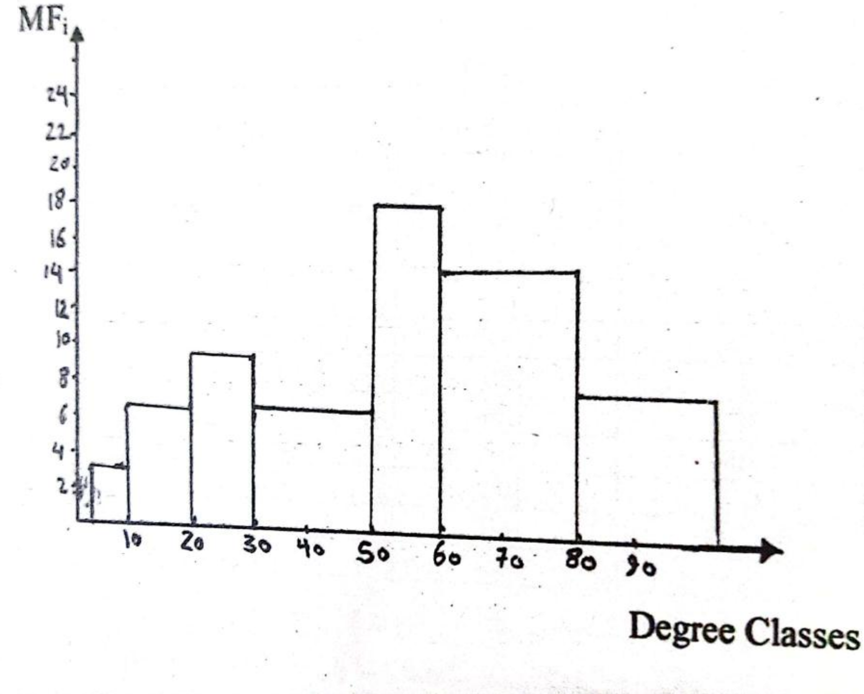
**حساب التكرار المعدل ($MF_i$) لتوزيع الدرجات**

| Classes | $F_i$ | $L_i$ | $MF_i = \frac{F_i}{L_i}$ |
|---|---|---|---|
| 5- | 15 | 5 | 3 |
| 10- | 70 | 10 | 7 |
| 20- | 100 | 10 | 10 |
| 30- | 120 | 20 | 6 |
| 50- | 200 | 10 | 20 |
| 60- | 300 | 20 | 15 |
| 80-100 | 160 | 10 | 8 |
| $\sum$ | 965 | | |

**والهدف من حساب التكرارات المعدلة هو تحقيق التناسب بين مساحات المستطيلات المكونة للمدرج التكرارى وارتفاعاتها . أى تحقيق التناسب مابين مساحة المستطيل والتكرار الذى يمثله هذا المستطيل هذا ويرسم الفئات كما سبق على المحور الافقى والتكرارات المعدلة بمقياس رسم مناسب على المحور الرأسى ١سم/٢وحدة تكرار معدل نحصل على الرسم البيانى التالى الموضح فى الشكل (٢-٨)**

شكل(٢-٨)

المدرج التكرارى لتوزيع درجات الطلاب




ويلاحظ على الشكل(٢-٨) مايلى:-

- ان قواعد المستطيلات المكونة للمدرج التكرارى غير متساوية وذلك نظرا لعدم تساوى أطوال فئات التوزيع.
- حيث ان مساحة المستطيل هى حاصل ضرب الطول(وهو هنا بمثابة التكرار المعدل) فى العرض (قاعدة المستطيل وهى بمثابة طول الفئة)فإن ناتج مساحة كل مستطيل ماهو إلا التكرار الاصلى للفئة التى تمثلها هذا المستطيل. ومن ثم فإنه إذا ما تم إيجاد مجموع مساحات المستطيلات المكونة للشكل البيانى السابق فإن ناتج المجموع سيكون بمثابة مجموع التكرارات الاصلية للتوزيع وهى ٩٦٥ طالب فى هذا المثال.

**١- <u>المضلع التكرارى :</u>** Frequency Polygon

**وهو وسيلة اخرى لتمثيل بيانات التوزيعات التكرارية لمتغيرات او ظواهر كمية متصلة. وهو عبارة عن خط منكسر يصل مابين النقاط التى يتم تمثيلها بيانيا من نقاط تلاقى الفئات المقابلة للتكرارت الاصلية (فى حالة الجداول التكرارية المنتظمة)أو التكرارات المعدلة(فى حالة الجداول الغير منتظمة) وبصفة عامة فإن مركز الفئة(**Midpoint-Mp or xi**) يمكن استنتاجها من خلال احد الزوايا التالية:-**

$$\text{مركز الفئة} = \frac{\text{الحد الادنى للفئة} + \text{الحد الأعلى للفئة}}{٢}$$

$$Xi \text{ or } MP = \frac{\text{Class Lower Limit} + \text{Class Upper Limit}}{\mathbf{2}}$$

i.e.,

$$x_i = \frac{C.L.L + C.U.L}{2}$$

**كذلك يمكن إيجاد مركز الفئة من خلال اضافة نصف طول الفئة للحد الادنى لتلك الفئة أى أن:**

$$X_i \text{ ( or ) } MP = C.L.L + 0.5\, L_i \qquad \text{for} \quad i = 1,2,\ldots\ldots\ldots\ldots,r$$

**حيث(**r**) تمثل عدد فئات التوزيع.**

**هذا ويجب التنويه الى انه عند رسم المدرج وكذلك المنحنى التكراري الذى سيتم دراسته حالا فإنه يفترض ان كافة التكرارات داخل كل فئة تأخذ قيمة**

مساوية لمركز الفئة ، وهذا الافتراض تبنى على اساسه حسابات المقاييس الاحصائية التى سيرد دراستها (مثال الوسط الحسابى والانحراف المعيارى).
ومن ثم لرسم المضلع التكرارى يجب بداية التأكد من مدى انتظام التوزيع التكراري من عدمه أولا. فإذا كان التوزيع التكرارى منتظما يتم حساب مراكز الفئات والرسم مباشرة حيث يتم تمثيل مراكز الفئات على المحور الافقى والتكرارات الأصلية على المحور الرأسى. وإذا سبق رسم المدرج لرسم المضلع فإنه يمكن تحديد نقاط منتصفات قمم مستطيلات المدرج وتوصيلها بخط منكسر بالمسطرة ليعطى المضلع التكرارى . اما اذا كان رسم المضلع مباشرة فيتم تمثيل نقاط تلاقى منتصف كل فئة وما يقابلها من تكرار اصلى بخط منكسر بالمسطرة نحصل على المضلع التكرارى ويجب فى تلك الحالة إقفال الشكل البيانى ، وذلك من خلال افتراض وجود فئة وهمية أو افتراضية تسبق الفئة الاولى وكذلك فئة وهمية تلى الفئة الاخيرة ويتم تحديد مراكز تلك الفئات الوهمية بنقاط على المحور الافقى والمثال التالى يوضح كيفية رسم المضلع التكرارى .

**مثال(١٤):**

**الجدول التالى يوضح توزيع اجور مجموعة من العاملين الشهرية بإحدى الشركات.**

**جدول (٢٥-٢)**

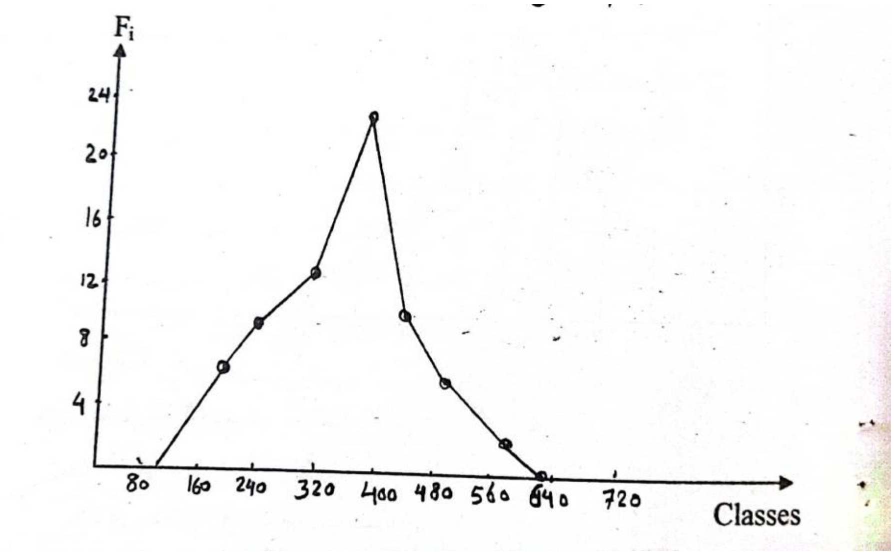
| classes | 150- | 220- | 270- | 360- | 430- | 500- | 570-640 |
|---|---|---|---|---|---|---|---|
| $F_i$ | 6 | 10 | 18 | 25 | 9 | 6 | 2 |

**والمطلوب تمثيل هذا التوزيع باستخدام الشكل البياني الملائم ؟**

**<u>الحـل :</u>** لاحظ ان بيانـات التوزيـع التكـرارى الموضـحة فـى الجـدول(٢٣-٢) عبارة عن جدول تكرارى منتظم لذا يجب رسم المدرج مباشرة .حيث يتم تمثيل الفئـات علـى المحـور الافقـى .كمـا يـتم بيـان التكـرارات الاصليـة علـى المحـور الراسى بمقياس رسم مناسب كما هو موضحه بالشكل(٢-٩).

**شكل(٢-٩)**

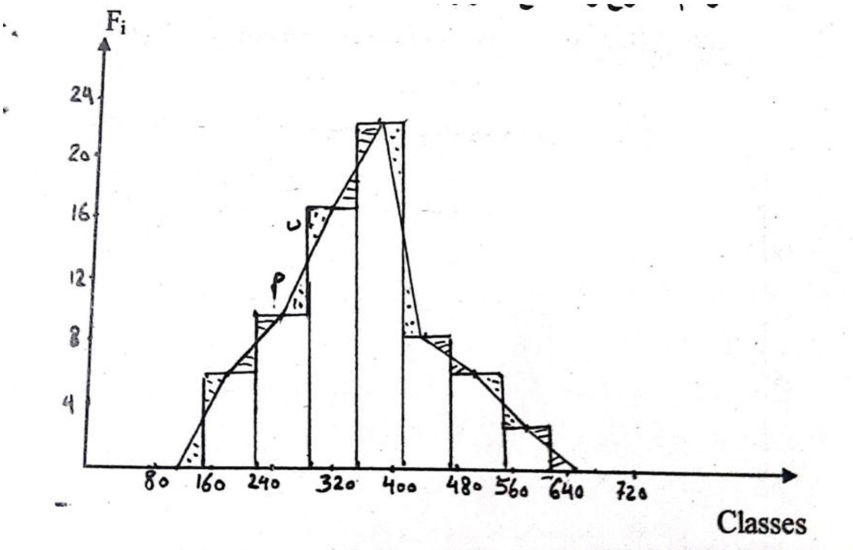
**رسم المضلع التكرارى للأجر الشهرى للعاملين**




**هذا ويمكن الحصول علـى المضـلع التكـرارى مـن المـدرج التكـرارى كمـا سـبق وان ذكرنا من خلال تنصيف قمـم المستطيلات (القواعد العليـا للمستطيلات) وتوصـيل هـذه المنتصـفات بخطـوط مسـتقيمة مـع إقفـال الشـكل مـن البدايـة والنـهاية وذلك لأن منتصفات قمم المستطيلات ما هى إلا مراكز فئات التوزيع التكرارى . ويمكن توضيح ذلك من خلال الرسم الموضـح فـى شكل (٢-١٠) حيث تـم رسـم المـدرج التكـرارى أولا ثـم تنصيف قمـم مسـتطيلاته وتوصـيلـها بـالخطوط المستقيمة .**

**شكل(٢-١٠)**

**رسم المدرج والمضلع التكرارى للاجر الشهرى للعاملين**




**ومن الشكل(٢-١٠) يمكن ملاحظة مايلى:-**

- **ان مركز الفئتين الافتراضيتين وقعا على المحور الافقى وذلك لأن تكرار كلا من هاتين الفئتين مساويا للصفر.**
- **المساحة تحت المضلع التكرارى مساوية للمساحة تحت المدرج التكرارى (أى مجموع مساحات المستطيلات المكونة للمدرج) وذلك لأن رسم المضلع التكرارى ادى الى استبعاد مثلثات من المدرج وأضاف ايضا مثلثات أخرى بدلا منها وتساويها فى المساحة . فعلى سبيل المثال مساحة المثلث المشار اليه بالرمز ( أ ) فى الشكل (٢-١٠) تساوى تماما مساحة المثلث المشار اليه بالرمز(ب) فى نفس الشكل.**
- **ان الملاحظة السابقة لا تتحقق اى لا يتحقق تساوى المساحة تحت المضلع مع المساحة تحت المدرج إلا اذا تم اقفال شكل المضلع التكرارى وذلك من خلال اضافة الفئتين قبل الاولى وبعد الاخيرة ولعل ذلك يبرر اضافة هاتين الفئتين .**

- وأخيرا ليس بالضرورة رسم المدرج التكرارى قبل المضلع فى حالة الجداول المنتظمة حيث يمكن رسم المضلع مباشرة دون رسم المدرج كما سبق وان اوضحناه فى الشكل (٢-٩) .

اما فى حالة الجداول التكرارية غير المنتظمة فإنه لرسم المضلع التكرارى يجب أولا رسم المدرج التكرارى وذلك من خلال حساب التكرارات المعدلة . ثم رسم المدرج التكرارى . ثم تنصيف مقدار الزيادة أو النقص فى التكرارات المعدلة المتتالية التى تعكسها الزيادة أو النقص فى ارتفاعات المستطيلات عن بعضها البعض فبهذه الطريقة نضمن تساوى المساحة تحت المضلع مع المساحة تحت المدرج التكرارى ، والمثال التالى يوضح ذلك.

**مثال(١٥):**

**الجدول التالى يوضح توزيع الدخل الشهرى لمائة اسرة (بالجنيه)**

**جدول(٢٦-٢)**

| classes | 230- | 240- | 250- | 260- | 270- | 280- | 290- | 310-340 |
|---|---|---|---|---|---|---|---|---|
| $F_i$ | 4 | 6 | 10 | 12 | 27 | 18 | 14 | 9 |

والمطلوب رسم المضلع التكرارى لهذا التوزيع.

**الحل:-**

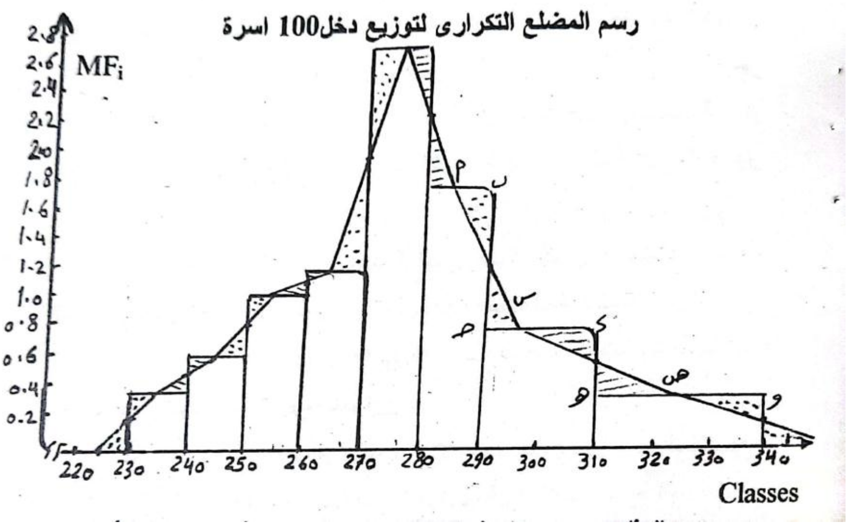
لاحظ أن التوزيع التكرارى لدخل هؤلاء الأسر غير منتظم لذا فإن لرسم المضلع التكرارى يجب الحصول على التكرارات المعدلة اولا ثم رسم المدرج النكرارى ثانيا ويلى ذلك رسم المضلع التكرارى وذلك على النحو التالى :

| Classes | $F_i$ | $L_i$ | $MF_i = Fi \div L_i$ |
| --- | --- | --- | --- |
| 230- | 4 | 10 | 0.4 |
| 240- | 6 | 10 | 0.6 |
| 250- | 10 | 10 | 1 |
| 260- | 12 | 10 | 1.2 |
| 270- | 27 | 10 | 2.7 |
| 280- | 18 | 10 | 1.8 |
| 290- | 14 | 20 | 0.7 |
| 310-340 | 9 | 30 | 0.3 |

**ثم يتم رسم المدرج التكرارى كما يوضحه الشكل(٢-١١) ولرسم المضلع التكرارى باستخدام المدرج التكرارى الموضح فى هذا الشكل فإنه ليست هناك أية تحفظات على توصيل منتصفات قمم المستطيلات الممثلة للفئات الستة الاولى بخطوط مستقيمة حيث ان هذه الفئات متساوية فى الطول(حيث أن طول كل منها ١٠جنيه) وأما فيما يتعلق بالفئة السابعة(٢٩٠وحتى اقل من ٣١٠جنيه) فان التوصيل لنقطة منتصف قمة المستطيل الممثل لهذه الفئة سوف يترتب عليه ان تكون المساحة الواقعة تحت المضلع اكبر من المساحة الواقعة تحت المدرج التكرارى ، وللتغلب على هذه المشكلة فأننا نرسم الخط المستقيم من النقطة(أ) مارا بمنتصف الارتفاع ب جـ (النقص فى الارتفاع) الى أن تقطع قمة المستطيل الممثل للفئة (٢٩٠حتى أقل من ٣١٠ جنيه) وليكن فى النقطة(س) وذلك حتى يتحقق أن يكون المثلث المستبعد يساوى فى مساحته المثلث المضاف.**

**شكل(٢-١١)**

**رسم المضلع التكرارى لتوزيع دخل ١٠٠ اسرة**




**وبالمثل يتم رسم الخط المستقيم بدءا من النقطة(س)الممثلة أخيرا بمنتصف الارتفـاع او المضـلع (دهـ ) ويمتـد علـى اسـتقامته الـى ان يقطـع قمـة المستطيل الممثل للفئـة(٣١٠حتـى أقل ٣٤٠)فـى نقطـة ولـتكن(ص) وذلـك حتى يتحقق أن يكون المثلث المستبعد يساوى فى مساحته المثلث المضاف واخيرا يتم رسم خط مستقيم بدءًا من النقطـة (ص) ومـارا بمنتصف الارتفـاع (و ز) الى أن يقطع المحور الافقى . وبذلك يتم إغلاق الشكل تحت المضلع التكرارى .**

**وفى حقيقة الامر فإن هذا الاجراء يضمن ان تكون المساحة تحت المضلع التكرارى مساوية تماما للمساحة تحت المدرج التكرارى.**

٢- **المنحنى التكرارى :** Frequency Curve

**إذا تـم تمهيـد نقـاط المضـلع التكرارى باليد فإننـا نحصـل علـى شكل منحنـى يسمى بالمنحنى التكرارى ولذلك فإنـه عند رسم المنحنى التكرارى يراعى نفس**

الشروط المتبعة سابقا عند رسم المدرج أو المضلع التكرارى لتوزيع تكرارى غير منتظم حيث يلزم تعديل التكرارات وتؤخذ مراكز الفئات على المحور الافقى والتكرارات الاصلية فى حالة التوزيعات التكرارية المنتظمة او التكرارات المعدلة فى حالة التوزيعات التكرارية الغير منتظمة على المحور الرأسى .

وبصفة عامة إذا كان تمهيد المنحنى جيدا فإن المساحة تحت المنحنى التكرارى سوف تساوى المساحة تحت المضلع التكرارى أو المساحة تحت المدرج التكرارى والتى تعبر هذه المساحات كما بينا فيما قبل عن مجموع تكرارات التوزيع التكرارى .

هذا ونتيجة لأن المنحنى التكرارى لا يمر بجميع نقاط المضلع التكرارى فإن المساحة تحت المنحنى التكرارى قد لا تتساوى مع المساحة تحت المضلع التكرارى . وفى واقع الامر فإنه كلما كانت أطوال فئات التوزيع قصيره وفى نفس الوقت زاد حجم البيانات فأن المضلع التكرارى يتحول تدريجيا الى شكل انسيابى يقترب اكثى فأكثر من شكل المنحنى التكرارى . ويمكن للطالب التحقق من ذلك برسم المضلع التكرارى وكذا المنحنى التكرارى لتوزيع معين. ثم يقوم بإنقاص طول الفئة مرة بعد أخرى فيلاحظ الزيادة المستمرة فى اقتراب المضلع التكرارى من المنحنى التكرارى.

**مثال(١٦) :**

ارسم المنحنى التكرارى لتوزيع أجور العاملين بإحدى الشركات الوارد بياناته في مثال (٤) بالجدول (٢٥-٢) .

**الحل :**

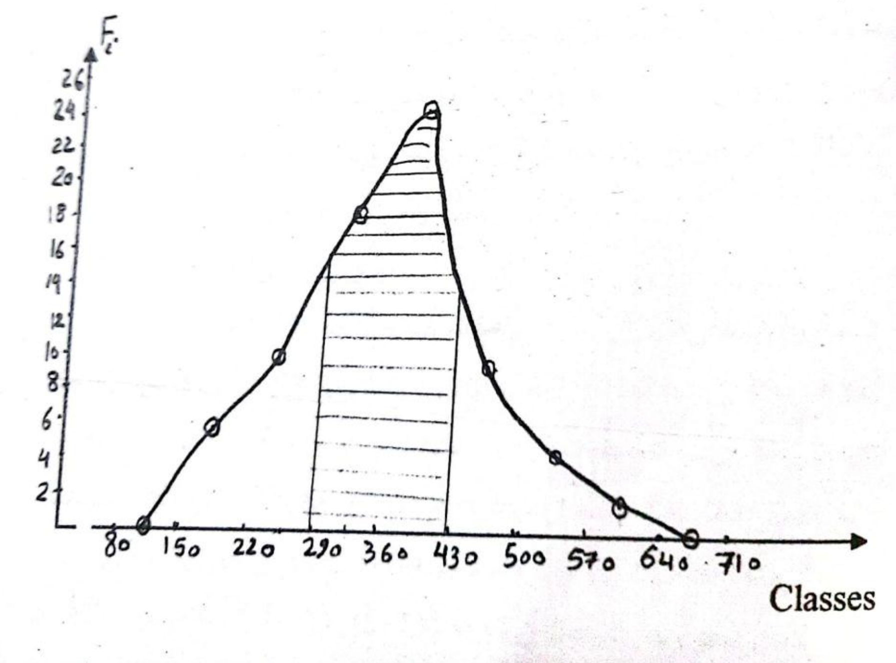
حيث أن توزيع الأجور الشهرية الوارد بالجدول (٢٥-٢) بيانات لتوزيع تكرارى منتظم . لذا يتم رسم المنحنى التكرارى مباشرة من خلال تحديد نقاط تلاقى مراكز فئات التوزيع وما يقابلها من تكرارات أصلية حيث يتم تحديد الفئات او مراكز الفئات على المحور الافقى والتكرارات على المحور الرأسى .

**ويتم تحديد نقاط المنحنى التكرارى كما هو موضح على الرسم المبين فى شكل (٢-١٢).**

**شكل(٢-١٢)**

**المنحنى التكرارى لتوزيع الاجر الشهرى**

**لمجموعة من العاملين بإحدى الشركات**




**وفى المنحنى التكرارى الموضح بالرسم في شكل (٢-١٢) لاحظ ان عدد المشاهدات (التكررات ) التى تقع فيما بين قيمتين من قيم المتغير أو الظاهرة محل الدراسة يتناسب مع المساحة المحصورة تحت المنحنى فيما بين هاتين القيمتين . فعلى سبيل المثال يتناسب عدد العمال الذين تتراوح اجورهم الشهرية مابين ٢٩٠-٤٤٠ مع المساحة المظللة فى الشكل (٢-١٢) .**

**هذا ويمكن رسم المضلع او المنحنى التكرارى لتوزيع تكرارى باستخدام التوزيعات التكرارية النسبية بدلا من التكرارات الاصلية . وتستخدم المنحنيات التكرارية فى مقارنة توزيعين تكراريين أو اكثر . إلا أنه يجب التنويه إلى أنه**

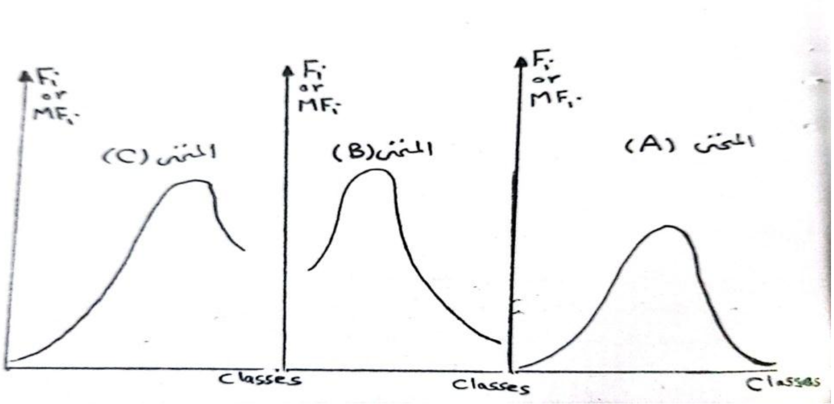
فى حالة تساو ى المجموع الكلى للتكرارات فإن عملية المقارنة تصح مباشرة من خلال رسم المنحنى التكرارى لتلك التوزيعات باستخدامك التكرارات الاصلية فى عملية المقارنة. وفيما عدا ذلك فإن المقارنة يجب أن تجرى على أساس التوزيعات التكرارية النسبية . وتعتبر أشكال المنحنيات التكرارية على درجة عالية من الأهمية وذلك لأنها تستخدم فى الحكم على مدى التواء التوزيع من عدمه(أى مدى قرب التوزيع من التماثل) كما تستخدم فى التعرف على شكل قمة المنحنى (دراسة التفلطح) وهو ما سيرد فى أبواب لاحقة بإذن الله.

**أشكال المنحنيات التكرارية:**

تناولنا فيما سبق كيفية رسم المنحنى التكرارى ورأينا ان شكل المنحنى يتوقف على التوزيع التكرارى الذى يمثله . كما يبرز أهمية دور المنحنيات التكرارية فى استخدامها فى المقارنة فيما بين التوزيعات المختلفة. هذا ومن المتوقع ان نجد عددا كبيرا من الاشكال المختلفة للمنحنيات التكرارية ومع ذلك فإننا سوف نتعرض هنا الى أكثر هذه الاشكال استخداما والموضحة فى شكل(٢-١٣) مرجئين الحديث عن بعض الاشكال الاخرى حتى دراستنا لمقاييس الالتواء او التفلطح فيما بعد فى مقررنا الدراسى بمشيئة الله.

شكل(٢-١٣)

بعض أشكال المنحنيات التكرارية




**هذا ولكى تسهل عملية التمييز فيما بين أشكال المنحنيات الثلاثة (A,B,C) المعطاه فى الجدول (٢-٢٧) والتى تناظر على الترتيب المحنيات التكرارية A,B,C الموضحة فى الشكل(٢-١٣).**

**جدول(٢-٢٧)**

**توزيعات تكرارية مختلفة**

| Classes | 10- | 22- | 34- | 46- | 58- | 70- | 82- | ∑ |
|---|---|---|---|---|---|---|---|---|
| $F_A$ | 4 | 8 | 13 | 15 | 13 | 8 | 4 | 65 |
| $F_B$ | 3 | 13 | 25 | 11 | 7 | 4 | 2 | 65 |
| $F_C$ | 2 | 4 | 7 | 11 | 25 | 13 | 3 | 65 |

**لاحظ أن التوزيعات الثلاثة متساوية من حيث مجموع التكرارات ولكن بتوزيع تكرارات مختلفة بالشكل الذى يخدم الغرض الذى من أجله نود التفرقة فيما بين هذه التوزيعات الثلاث . حيث تمكننا التوزيعات الثلاث من تصنيف الأشكال المختلفة للمنحنيات التكرارية الموضحة فى الشكل(٢-١٣) حيث نجد أن:-**

**<u>التوزيع(ا):-</u>**

**توزيع متماثل حول المحور الرأسى الذى يمر بنقطة النهاية العظمى. اى أن المحور الراسى يقسم المنحنى الى قسمين متطابقين تماما ويسمى المنحنى التكرارى فى هذه الحالة بالمنحنى المتماثل . حيث يكون تزايد او تناقص**

التكرارات متشابها ومنتظما بطريقة متماثلة على جانبى المحور الرأسى والذى يمر بمركز التوزيع . ويمكن ملاحظة ذلك فى الشكل (٢–١٣) ومن الجدول(٢٧–٢) . وبعبارة اخرى فأن المنحنى التكرارى (أ) والموضح فى شكل (٢–١٣) والذي يمثله التوزيع التكرارى (أ) فى الجدول (٢٧–٢) يمثل صورة للتوزيعات المتماثلة Symmetric Distribution

**التوزيع(ب) :-**

وفى هذا التوزيع تتزايد التكرارات أو تتناقص بشكل غير منتظم على جانبى المحور أو العمود المنشأ عند وسط التوزيع . ولذلك فهو توزيع غير متماثل تتزايد فيه التكرارات بسرعة حتى تصل الى القمة وبعدها تتناقص ببطء . ويطلق على هذه التوزيعات والتى يميل فيها التكرارات الكبيرة إلى التركيز عند الفئات الدنيا بالتوزيعات الموجبة الالتواء

Positive Skewness Distributions

والمنحنى التكرارى الممثل لهذه التوزيعات يسمى بالمنحنى غير المتماثل حيث يكون طرفه(ذيله) الايمن أطول من طرفه الايسر (انظر الشكل(٢–١٣) المنحنى(ب)) والتوزيع التكرارى(ب) الموضح فى الجدول(٢٧–٢) يمثل هذا النوع من المنحنيات.

**التوزيع (ج):-**

وفى هذا التوزيع ايضا تتزايد التكرارات او تتناقص بشكل غير منتظم على جانبى المحور المنشأ عند وسط التوزيع . وبالتالى فهو توزيع غير متماثل. ووفقا لهذا النوع من التوزيعات فإن التكرارات تتزايد ببطء حتى تصل الى القمة ثم بعدها تأخذ فى التناقص بسرعة. ومثل هذا التوزيع والذى تميل فيه التكرارات الكبيرة الى التركيز عند الفئات العليا يسمى بالتوزع السالب الالتواء (Negative Skewness Distribution ) والمنحنى التكرارى الممثل لهذا التوزيع هو منحنى غير متماثل حيث يكون الطرف(الذيل) الايسر فيه أطول

من الطرف الايمن (أنظر الشكل (٢-١٣) والمنحنى(جـ) والذي يمثل هذا النوع من المنحنيات التوزيع التكرارى(جـ) الوارد فى جدول(٢٧-٢) .

**مثال(١٧):**

الجدول التالى يوضح التوزيع التكرارى لدرجات مجموعة من الطلاب (ذكورا وإناثا على حده) فى احدى المواد.

**جدول(٢٨-٢)**

**التوزيع التكرارى لدرجات مجموعة من الطلاب**

**(ذكورا وإناثا على حده) فى احدى المواد.**

| Classes | 25- | 35- | 45- | 55- | 65- | 75- | 85- | ∑ |
|---|---|---|---|---|---|---|---|---|
| No.of Males | 20 | 26 | 35 | 43 | 30 | 27 | 19 | 200 |
| No.of Femals | 2 | 4 | 7 | 13 | 12 | 9 | 3 | 50 |

والمطلوب:- المقارنة مابين التوزيعين التكراريين للذكور والاناث وذلك بإستخدام الشكل البيانى الملائم.

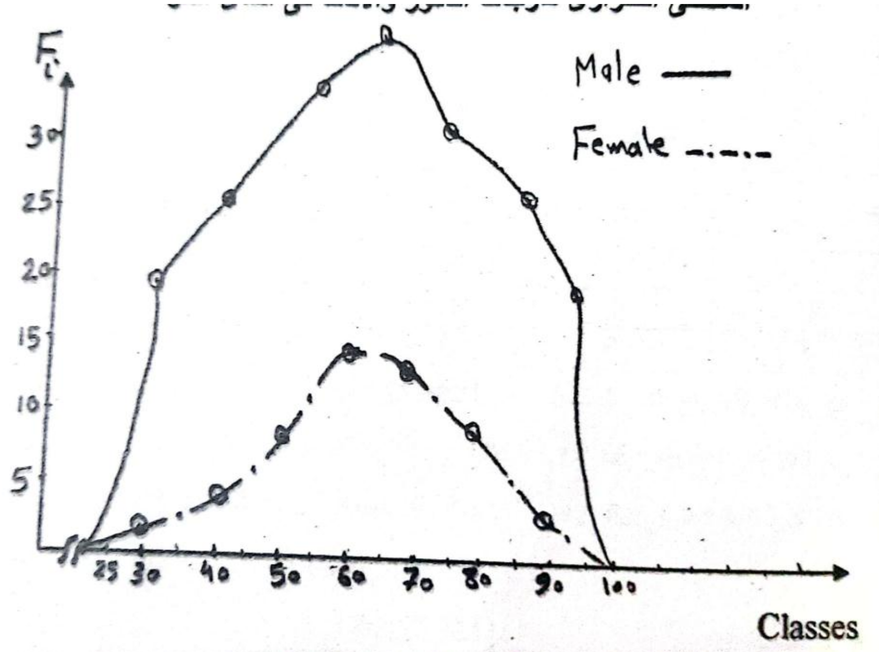
**الحل :-**

حيث أن فئات الدرجات تعبر عن ظاهرة كمية متصلة لذا فالشكل البيانى الملائم لعملية المقارنة فيما بين التوزيع التكرارى للذكور والاناث هو اما المدرج أو المضلع او المنحنى التكرارى . وحيث ان فئات التوزيعين واحدة ومنتظمة لذا يمكن رسم المدرج أو المنحنى أو المضلع التكرارى مباشرة (لكن فى حالة الجداول الغير منتظمة فلابد من الحصول على التكرارات المعدلة أولا) والان دعنا نقوم بالمقارنة من خلال المنحنى التكرارى .

والان عزيزى الطالب لاحظ أن مجموع التكرارات فى حالة الذكور مختلفا عن مجموع التكرارات فى حالة الاناث فإذا تم إجراء المقارنة فيما بين التوزيعين مباشرة دون الحصول على التكرارات النسبية، سيكون النتيجة مقارنة غير واضحة وغير موضوعية نظرا لاختلاف مجموع التكرارات . والشكل(٢-١٤) يوضح رسم المنحنى التكرارى للذكور والاناث مباشرة.

**شكل(٢-١٤)**

**المنحنى التكرارى لدرجات الذكور والاناث فى احدى المواد**




فيتضح من شكل (٢-١٤) ان المقارنة بين التوزيعين غير واضحة إذ أن المساحة تحت كلا من المنحنيين تختلف عن الاخرى نظرا لاختلاف مجموع التكرارات فى كلا من التوزيعين عن الأخر لذا لكى تكون المقارنة موضوعية على أساس سليم فأن المساحة تحت المنحنى الممثل لدرجات الذكور يجب أن تتساوى مع المساحة تحت المنحنى الممثل لدرجات الاناث ولكى يتحقق ذلك لابد من ايجاد التوزيعات التكرارية النسبية لكل من الذكور والاناث حيث تتساوى المساحتين فى هذه الحالة.

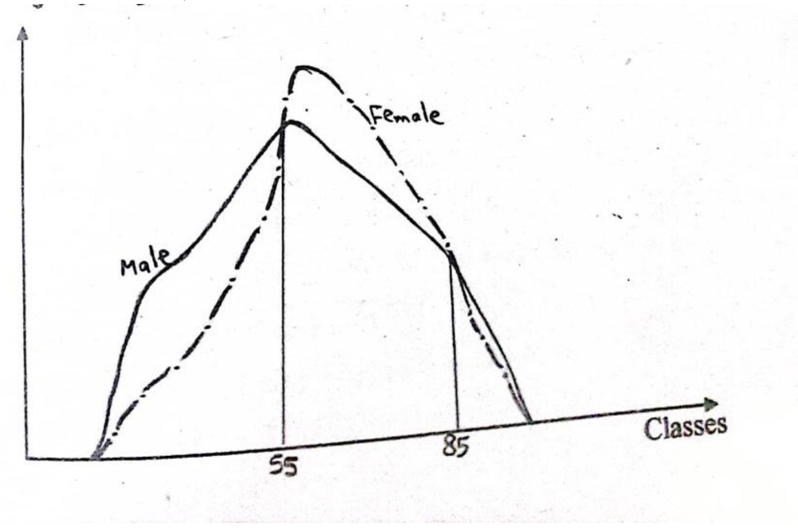
**والجدول (٢٩-٢) يوضح التوزيع التكرارى النسبى لدرجات الطلاب (ذكورا وإناثا) فى إحدى المواد.**

| Classes | Frequencies | | Relative Frequencies % | |
|---|---|---|---|---|
| | Male | Female | $RF_M$ | $RF_F$ |
| 25- | 20 | 2 | 10 | 4 |
| 35- | 26 | 4 | 13 | 8 |
| 45- | 35 | 7 | 17.5 | 14 |
| 55- | 43 | 13 | 21.5 | 26 |
| 65- | 30 | 12 | 15 | 24 |
| 75- | 27 | 9 | 13.5 | 18 |
| 85- | 19 | 3 | 9.5 | 6 |
| $\sum$ | 200 | 50 | 100 % | 100 % |

**ومن بيانات الجدول(٢٩-٢) يمكننا رسم المنحنى الذى يمثل التوزيع التكرارى النسبى لكلا من درجات الذكور والإناث كلا على حده وذلك كما هو موضح بالشكل (٢-١٥) حيث يبين هذا الشكل منحنى التكرارات النسبية لكلا من الذكور والاناث الواردة بالجدول(٢٩-٢).**

**شكل(٢-١٥)**

**منحنى التكرارات النسبية لدرجات الطلاب (ذكورا وإناثا) فى إحدى المواد**



**ومن خلال شكل(٢-١٥) يتضح لنا الاتى:-**

- **ترتفـع نسـبة الـذكور عنـد الحـدود الـدنيا للتوزيـع (وبالتحديـد عنـد الـدرجات التـى تقـل عـن٥٥درجـة) عـن مثيلتهـا للإنـاث حيـث يقـع منحنـى التكـرار النسـبى للـذكور أعلـى مـن منحنـى التكـرار النسـبى للإناث.**
- **ترتفع نسبة الاناث عند الحدود العليا لفئات التوزيـع (وبالتحديد عند الـدرجات التـى تزيـد عن٥٥درجـة)حيث يقـع منحنـى التكـرار النسـبى للإناث اعلى من منحنى التكرار النسبى للذكور .**

**وبصفة عامـة عنـد المقارنـات بإستخدام المـدرج أو المضـلع أو المنحنـى وإن كان يفضل المضلع او المنحنـى التكرارى لكـى تتضـح عمليـة المقارنـة سـواء باستخدام التكرارات الاصلية (فى حالة تساوى مجموع التكرارات) أو التكرارات المعدلة(فى حالة عدم تساوى مجموع التكرارت) فإنـه يتم تحديد نقـاط تقـاطع المضـلعات أو المنحنيـات التكراريـة ويـتم إسـقاط مـن تلـك الـنقط أعمـدة علـى المحور الافقى ثم تتم عملية المقارنة فيما بعد ذلك كما سبق فى الشكل (٢-١٥).**

**المنحنيات التكرارية المتجمعة:-** Cumulative Frequency Curves

**بصفة عامة تستخدم المنحنيات التكرارية المتجمعة لتمثيل الجداول التكراريـة المتجمعـة بيانيـا هـذا بالاضـافة لاسـتخدامها فـى تحديـد عـدد التكـرارات التـى تقابل اقل من قيمة مافى الفئات أو فيما بينها او العكس صحيح.**

**ولمـا كـان هنـاك نوعـان مـن الجـداول التكراريـة المتجمعـة (الجـداول التكراريـة المتجمعـة الصـاعدة والجـداول التكراريـة المتجمعـة الهابطـة) لـذا فإنـه مـن الطبيعى ان يكون هناك نوعين للمنحنيات التكرارية المتجمعة وهى:-**

- **المنحنــى التكـرارى المتجمــع الصــاعد** Ascending Cumulative Curve

– **المنحنى التكرارى المتجمع الهابط** Descending Cumulative Curve

**حيث يستخدم المنحنى التكرارى المتجمع الصاعد لرسم الجدول التكرارى المتجمع الصاعد بيانيا كما يفيد فى حساب عدد التكرارات (أو نسبة التكرارت) التى تأخذ قيمة اقل من قيمة ما(سواء فئة معينة او فيما بين فئتين) من قيم فئات التوزيعات التكراري. أو العكس تفيد فى حساب ما هى القيمة التى يقل عنها عدد معين من التكرارات(او نسبة من التكرارات).**

**ولرسم المنحنى التكرارى المتجمع الصاعد فأننا نختار المحور الرأسى لتمثيل التكرارت المتجمعة الصاعدة بينما يتم تمثيل الفئات على المحور الافقى . حيث يتم تمثيل كل فئة وما يقابلها من تكرار متجمع بنقطة إحداثياها هما الحد الاعلى للفئة والتكرار المتجمع الصاعد لهذه الفئة ، ثم نقوم بتوصيل هذه النقاط بخط منحنى ممهد باليد فنحصل بذلك على المنحنى التكرارى المتجمع الصاعد. اما المنحنى التكرارى المتجمع الهابط فيستخدم لرسم الجدول التكرارى المتجمع الهابط بيانيا كما يفيد فى حساب عدد التكرارات(او نسبة التكرارات) التى تبلغ قيمة ما فى الفئات (أو من بين فئتين)فأكثر او يفيد فى تحديد ما هى القيمة التى يبلغها عددا ما من التكرارات (أونسبة التكرارات) فأكثر . وبأسلوب مماثل يمكن رسم المنحنى التكرارى المتجمع الهابط حيث يتم تمثيل التكرارات المتجمعة الهابطة على المحور الرأسى بينما يمثل المحور الافقى فئات التوزيع التكرارى المتجمع الهابط. ويتم تمثيل كل فئة وما يقابلها من تكرار متجمع هابط بنقطة إحداثياها هما الحد الادنى للفئة والتكرار المتجمع الهابط المقابل لهذه الفئة . ثم يتم توصيل كافة النقاط التى تم تحديدها على الرسم بخط منحنى ممهد باليد فنحصل بذلك على المنحنى التكرارى المتجمع الهابط والمثال التالى يوضح أهمية المنحنيات التكرارية المتجمعة.**

**<u>مثــال(١٨) :-</u> فيمـا يلـى لـديك التوزيـع التكـرارى التـالى بـين توزيـع الاجـر الاسبوعى لخمسين عاملا بإحدى الشركات (بالجنيه) .**

**جدول(٣٠-٢)**

| Classes | 30- | 40- | 50- | 60- | 70- | 80- | 90- | ∑ |
|---|---|---|---|---|---|---|---|---|
| Frequencies | 1 | 4 | 7 | 14 | 11 | 8 | 5 | 50 |

**والمطلوب :-**

**١- رسم المنحنى التكرارى المتجمع الصاعد**

**٢- رسم المنحنى التكرارى المتجمع الهابط**

**٣- بإستخدام المنحنى التكرارى التجمع الملائم حدد مايلى:-**

**أ- ما هو قيمة الاجر الاسبوعى الذى يتقاضى اقل من ٣٥ عاملا.**

**ب- اوجد قيمة الاجر الاسبوعى الذى يحصل عليـه - كحد أدنـى – ثلاثـون عاملاً**

**جـ- مـا هـو عـدد العمال(وكـذا مـاهى نسبة العمـال) الـذين يتقاضـون أجـورا اسبوعية تقل عن٦٤ جنيها.**

**د- ما هو قيمة الاجر الاسبوعى الذى يتقاضى أقل منه :**

**- ٨٠% من إجمالى عدد العمال .**

**- ٥٢% من إجمالى عدد العمال.**

**هـ- حـدد مـا هـو عـدد العمال(وكـذا مـا هـى نسـبة العمـال) الـذين يتقاضـون أجورا أسبوعية تبلغ ٤٥ جنيها على الاقل .**

**و- حـدد قيمـة الاجـر الاسـبوعى الـذى يحصـل عليـه كحـد أدنـى ٥٠% مـن إجمالى عدد العمال**

**<u>الحل :-</u>**

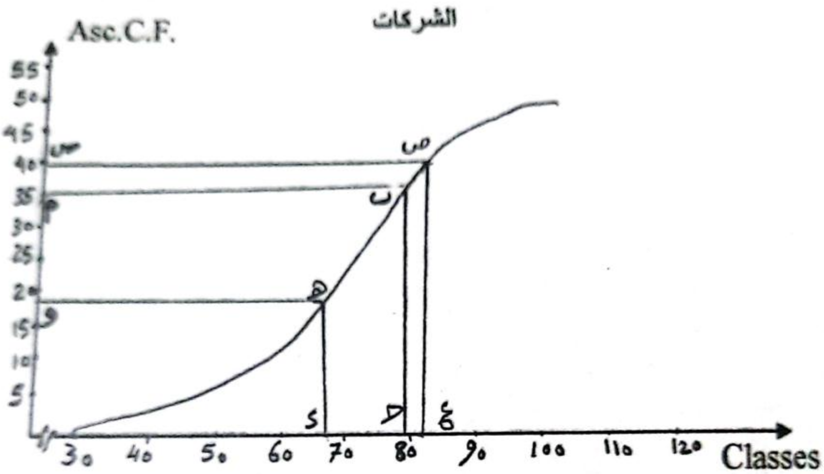
**١- لرسم منحنى التكرارى المتجمـع الصـاعد يـتم أولا ايجـاد الجـدول التكـرارى المتجمع الصاعد وذلك على النحو المبين فى الجدول التالى :-**

**جدول (٣١-٢)**
**التوزيع التكرارى والتوزيع التكرارى المتجمع الصاعد لأجور العمال الاسبوعية فى إحدى الشركات**

| Classes | $F_i$ | Less than the C.U.L | Asc.Cum.Freq |
|---|---|---|---|
| 30- | 1 | Less than 30 | 0 |
| 40- | 4 | Less than 40 | 1 |
| 50- | 7 | Less than 50 | 5 |
| 60- | 14 | Less than 60 | 12 |
| 70- | 11 | Less than 70 | 26 |
| 80- | 8 | Less than 80 | 37 |
| 90- | 5 | Less than 90 | 45 |
| ∑ | 50 | Less than 100 | 50 |

ومن هذا الجدول يتم رسم العمودين الاخيرين لتمثيل المنحنى التكرارى المتجمع الصاعد. والشكل (٢-١٦) يوضح المنحنى التكرارى المتجمع الصاعد للاجور الاسبوعية لمجموعة العاملين المعطاه .

**شكل(٢-١٦)**
**المنحنى التكرارى المتجمع الصاعد لأجور العمال الاسبوعية فى إحدى الشركات**



**٢- لرسم المنحنى التكرارى الهابط يجب أولا ايجاد الجدول التكرارى المتجمع الهابط وذلك على النحو المبين فى الجدول التالى :**

**جدول(٢-٣٢-٢)**

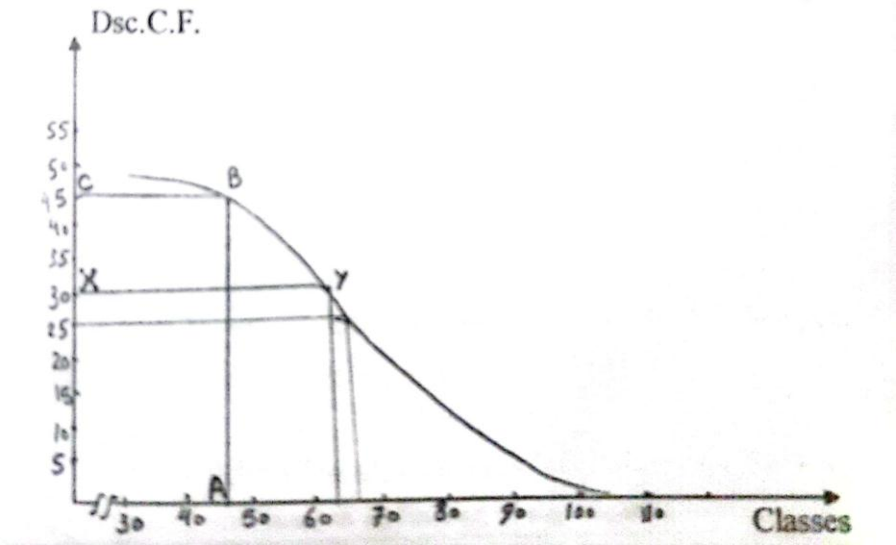
**التوزيع التكرارى والتوزيع التكرارى المتجمع الهابط للأجور الاسبوعية لمجموعة من العاملين بإحدى الشركات**

| Classes | Fi | More than the C.L.L | Desc.Cum.freq |
|---|---|---|---|
| 30- | 1 | More than30 | 50 |
| 40- | 4 | More than40 | 49 |
| 50- | 7 | More than50 | 45 |
| 60- | 14 | More than60 | 38 |
| 70- | 11 | More than70 | 24 |
| 80- | 8 | More than80 | 13 |
| 90- | 5 | More than90 | 5 |
| ∑ | 50 | More than100 | 0 |

**ومن هذا الجدول يتم رسم العمودين الأخيرين لتمثيل المنحنى المتجمع الهابط. والشكل(٢-١٧) يوضح المنحنى المتجمع الهابط للأجور الاسبوعية لمجموعة العاملين المعطاه .**

**شكل(٢-١٧)**

**المنحنى التكرارى المتجمع الهابط لاجورالعمال الاسبوعية فى إحدى الشركات**



**٣- بعـد رسـم المنحنـى التكـرارى المتجمـع الصاعد(شـكل (٢-١٦) والمنحنـى التكـرارى المتجمـع الهـابط (شـكل (٢-١٧) ومـن خـلال المرادفـات السـابقة والتـي تـم الإشـارة إليهـا عنـد دراسـتنا للجـداول التكرارية الصاعدة والهابطة فإن:**

**أ- لتحديد قيمة الاجر الاسبوعى الذى يتقاضى أقل منـه ٣٥ عاملا فيكون ذلـك مـن خـلال المنحنـى التكرارى المتجمـع الصـاعد. حيث يـتم تعيـين موضـع تلـك النقطـة علـى المحـور الرأسـى فـى الشـكل(٢-١٦) ولـتكن النقطة(أ) والتى تمثل عدد العمال فـى صورة تكرارات متجمعة صـاعدة . ومن هذه النقطـة يتم رسـم خطـا عموديـا علـى المحـور الرأسـى وموازيـا للمحـور الافقـى الـى أن يقطـع هـذا الخـط المنحنـى التكـرارى المتجمـع الصاعد فـى نقطـة ولـتكن النقطـة (ب) فنسـقط عمـودا مـن نقطـة إلتقـاء هذا الخط الموازى للمحور الافقى بالمنحنى المتجمع الصاعد ليقطع هذا العمود المحور الافقى(الاجور) فى نقطة ولتكن النقطة (جـ) فتكون هـى النقطةالتى توضـح قيمـة الاجر الـذى يتقاضـى أقل من ٣٥ عـاملا.ومن النقطة(جـ) على الرسم يتضح لنا أن قيمة الاجر الذى يتقاضـى أقل مـن ٣٥ عاملا هو الاجر ٨٧ جنيها تقريبا.**

**ب- لإيجاد قيمة الأجر الـذى يتقاضـاه - كحد أدنـى- ثلاثـون عـاملا فيكـون ذلـك مـن خـلال المنحنـى التكـرارى المتجمـع الهـابط لانـه كمـا سبق وأن أوضحنا أن لفظ كحد أدنـى يكـافئ فـى المعنـى كلمـة فـأكثر . وهنـا يـتم تعيين موضع النقطة ٣٠ عاملا على المحور الرأسـى للمنحنـى التكرارى المتجمع الهابط كما هو موضح فى شكل (٢-١٧) ولـتكن النقطـة (X) والتى تمثل عدد العمال فـى صورة تكرارات متجمعة هابطـة . ومـن هـذه النقطـة يـتم رسـم خطـا عموديـا علـى المحـور الرأسـى وموازيـا للمحـور الافقـى الـى أن يقطـع هذا الخط المنحنى التكرارى المتجمـع الهـابط فـى نقطـة ولـتكن النقطـة(Y) فنسـقط عمـودا مـن النقطـة (Y) علـى المحـور**

الافقى والذى يمثل الحدود الدنيا لفئات التوزيع التكرارى المتجمع الهابط الى ان يقطع هذا المحور الافقى فى نقطة ولتكن (Z) فتكون النقطة (Z) موضحة لقيمة الاجر الذى يتقاضاه كحد أدنى ثلاثون عاملا . ومن النقطة (Z) على الرسم الموضح فى شكل (٢-١٧) يتضح لنا أن قيمة الاجر الذى يتقاضاه كحد أدنى ثلاثون عاملا هو الاجر ٦٦ جنيه تقريبا .

ج- أما لتحديد عدد العمال (وكذا نسبة العمال) الذين يتقاضون أجورًا أسبوعية تقل عن ٦٤ جنيه . فيتم استنتاج المطلوب باستخدام المنحنى التكرارى المتجمع الصاعد ، حيث يتم تحديد موضع الاجر ٦٤ جنيه بنقطة على المحور الافقى لفئات التوزيع التكرارى المتجمع الصاعد الموضحة فى الشكل (٢-١٦) ولتكن النقطة (د) ومن النقطة (د) يتم إنشاء عمودا على المحور الافقى وموازيا للمحور الرأسى الى أن يقع هذا العمود المنحنى التكرارى المتجمع الصاعد فى نقطة ولتكن النقطة (هـ) . ومن النقطة (هـ) يتم رسم خطًا موازيا للمحور الافقى ويمتد على استقامته الى أن يقطع المحور الرأسى الذى يعبر عن التكرارات المتجمعة الصاعدة فى نقطة ولتكن النقطة (و) وهى تعتبر بمثابة الاجر المطلوب. هذا ومن الشكل (٢-١٦) وبالتحديد عند النقطة (و) يتضح لنا أن عدد العمال الذين يتقاضون أجرا اسبوعيا يقل عن ٦٤ جنيه هو ١٨ عاملا . ومن ثم فإن :-

نسبة العمال الذين تقل أجورهم الاسبوعية عن ٦٤ جنيه

= (١٨ ÷ ٥٠) × ١٠٠ = ٣٦%

د- لتحديد قيمة الاجر الذى يتقاضى أقل منه :-

- ٨٠% من إجمالى عدد العمال فحيث أن النسبة ٨٠% من العمال تشكل عدد مطلقا.

$$50 \times \frac{80}{100} = 40$$ عاملا . لذا فإنه يمكن استنتاج قيمة الاجر

الاسبوعى الذى يتقاضى أقل منه ٤٠ عاملا باستخدام المنحنى التكرارى المتجمع الصاعد بنفس آلية المطلوب الاول رقم (أ) من هذا المطلوب من خلال الشكل رقم (٢-١٦). حيث يتم تحديد موضع عدد العمال ٤٠ عاملا على المحور الرأسى وليكن عند النقطة (س) ورسم خط عمودي على المحور الرأسى وموازيا للأفقى الى أن يقطع المنحنى التكرارى المتجمع الصاعد فى نقطة ولتكن النقطة (ص) والتى يتم منها إسقاط عمودا على المحور الافقى وموازيا للمحور الرأسى الى أن يقطع المحور الافقى فى نقطة ولتكن النقطة (ع) . ومنها فإن قيمة الاجر الذى يتقاضاها اسبوعيا أقل من ٤٠عاملا (٨٠% من إجمالى العمال) هو الاجر ٨٤ جنيه تقريبا.

- لتحديد قيم الأجر الذي يتقاضى أقل منه ٥٢% من اجمالى عدد العمال . فحيث ان النسبة ٥٢% من إجمالى عدد العمال تشكل العدد المطلق وهو :

$$50 \times (52 / 100) = 26$$ عاملا

لذا فإنه يمكن استنتاج قيمة الاجر الذى يتقاضى أقل منه ٢٦ عاملا بنفس الاسلوب السابق باستخدام المنحنى التكرارى المتجمع الصاعد كما يوضحه (٢-١٦) حيث ينتج ان الاجر الذى يتقاضى أقل منه ٢٦ عاملا (اى ٥٢% من إجمالى عدد العمال) هو ٧٠جنيه تقريبا .

<u>ملحوظة:-</u>

لاحظ فى الجدول التكرارى المتجمع الصاعد ان عدد العمال٢٦ يشكل أحد التكرارات المتجمعة الصاعدة ويقابله الفئة ٧٠جنيها كحد أقصى للأجر الذى يتقاضاه عدد٢٦ عاملا.

هـ – لتحديد عدد العمال (وكذا النسبة ) الذين يتقاضون أجورا اسبوعية تبلغ ٤٥ جنيها على الاقل (وهذه العبارة تعادل كلمة فأكثر) فأن ذلك يكون من خلال المنحنى التكرارى المتجمع الهابط . حيث يتم تحديد هذا المطلوب بنفس الاسلوب اذى تم به إجابة المطلوب (ب) من نفس المثال . حيث يتم تحديد موضع الاجر ٤٥ جنيها بنقطة على المحور الافقى للمنحنى التكرارى المتجمع الهابط الموضح فى الشكل (٢–١٧) ولتكن النقطة (A) ثم يتم إنشاء عمودا من النقطة (A) على المحور الافقى وموازيا للمحور الرأسى الى أن يقطع هذا العمود المنحنى التكرارى المتجمع الهابط فى نقطة ولتكن (B) ومن النقطة (B) يتم رسم خطا موازيا للمحور الافقى فى اتجاه المحور الرأسى الى ان يقطع المحور الرأسى فى نقطة ولتكن النقطة (C) . فتكون النقطة (C) هى بمثابة عدد العمال الذين يتقاضون على الاقل أجرا اسبوعيا يعادل ٤٥ جنيها . ومن الرسم الموضح فى الشكل (٢–١٧) يتضح لنا أن عدد العمال المطلوب هو ٤٧ عاملا تقريبا ومن ثم فإن :–

النسبة المطلوبة = (٤٧ ÷ ٥٠ ) × ١٠٠ = ٩٤%

و– لتحديد قيمة الاجر الاسبوعى الذى يتقاضاه كحد أدنى ٥٠% من اجمالى عدد العمال . يتم استنتاج المطلوب من المنحنى التكرارى المتجمع الهابط . وذلك بعد تحويل النسبة ٥٠% الى عدد مطلق فحيث أن:

٥٠% من اجمالى عدد العمال = ٥٠ × (٥٠ ÷ ١٠٠ ) = ٢٥ عاملا .

ومن ثم فأن المطلوب الان هو تحديد قيمة الاجر الذى يتقاضاه كحد أدنى (فأكثر) ٢٥عاملا . وبنفس الاسلوب الذى تم به اجابة المطلوب (ب) من نفس المثال فإنه من خلال الشكل(٢–١٧) يتضح لنا أن قيمة الاجرى الذى يتقاضاه كحد أدنى ٥٠% من عدد العمال (٢٥ عاملا) هو الاجر ٦٩ جنيه تقريبا .

٤- **<u>مخطط الجذع والورقة</u> :- Stem-and-Leaf**

**يعتبر مخطط الجذع والورقة من المخططات البيانية المستحدثة فى مجال العرض البيانى للظواهر الكمية المتصلة . وفى هذا الشكل أو المخطط البيانى يتم تقسيم كل مفردة من مفردات الظاهرة أو المتغير محل الدراسة الى جزئين أساسين الاول يسمى الجذع أو الساق (Stem) والثانى يسمى بالورقه( Leaf ) حيث يمثل الجذع الجزء الايسر من من البيانات أما الورقة فتمثل بالجزء الايمن . وهذا المخطط– الجذع والورقه – يشبه بدرجة كبيرة المدرج والمنحنى والمضلع التكرارى . والفارق الوحيد بين هذا المخطط البيانى والرسوم السابقة (مدرج– مضلع– منحنى تكرارى) فى أن التكرارات فى حالة المدرج والمضلع والمنحنى التكرارى تمثل بأعمدة بيانية أو مجموعة من النقاط أما فى حالة مخطط الجذع والورقة فيتم تمثيل القيم الحقيقية للمفردات . وبذلك فأن مخطط الجذع والورقة يعكس معلومات عن طبيعة القيم الموجودة أو الفعلية للظاهرة محل الدراسة . والامثلة التالية توضح كيفية إنشاء المخطط البيانى للساق أو الجذع والورق .**

**<u>مثال(١٩):-</u>إذا كانت لديك مجموعة المفردات التالية والتى تعبر عن الدخل اليومى لمجموعة 15 مفردة :–**

**5 , 7, 12 , 15 , 19 , 20 , 21 , 23 , 25 , 27 , 28 , 29 , 30 , 31,35**

**والمطلوب تمثيل مخطط الجذع أو (الساق) والورقة .**

**<u>الحل :-</u> لتمثيل مخطط الجذع أو الساق والورقة فإننا نقسم البيانات التى تعبر عن الدخل أو الاجر اليومى لمجموعة المفردات المعطاة الى جزئين الاول يعبر عن الجذع أو(الساق) Stem والذى يمثل خانة العشرات والثانى يعبر عن الورق Leaf والذى يمثل خانة الآحاد وكأنه تم تقسيم المتغير الى أقسام أو فئات طول كل منها ١٠جنيه فيعطى الشكل البيانى التالى (شكل ٢–١٨)**

**شكل (٢–١٨)**

**مخطط الجذع(الساق ) والاوراق للدخل اليومى لمجموعة من العاملين**

Stem- and- Leaf Plot

INCOME Stem-and-Leaf Plot

| Frequency | Stem & | Leaf |
|---|---|---|
| 2.00 | 0 . | 57 |
| 3.00 | 1 . | 259 |
| 7.00 | 2 . | 0135789 |
| 3.00 | 3 . | 015 |

Stem width: 10.00
Each leaf: 1 case(s)

**فمن مخطط الجذع والورقة لتلك البيانات كما يوضحها الشكل (٢-١٨) يتضح لنا انه تم رصد تكرارات كل فئة فى الجانب الايسر . حيث ان هناك مفردتين تبلغ دخولهم اليومية أقل من عشرة جنيهات . وهناك سبعة مفردات تبلغ دخولهم مابين العشرة جنيهات وأقل من عشرون جنيها . وهناك سبعة مفردات دخولهم مابين العشرون جنيها وأقل من ثلاثون جنيها وأخيرا هناك ثلاثة مفردات تبلغ دخولهم ثلاثون جنيها فأكثر. أما الجانب الأيمن فيعطى صورة تفصيلية عن قيم المفردات . حيث يتضح الاتى :-**

**- بالنسبة للصف الاول الذى يقع أسفل الجذع أو الساق والاوراق ان هناك مفردتين خانة العشرات (الجذع) تحوى صفرا أما الورق فيحوى (٥)،(٧) وهما نتاج العملية الرياضية التالية:-**

$7 + 10 \times 0 = 7$ ، $5 + 10 \times 0 = 5$

**- بالنسبة للصف الثانى الذى يقع أسفل الجذع والورقة فهناك ثلاثة مفردات كل منهم يحتوى على عشرة جنيهات بالاضافة لمفردات الآحاد التى يمثلها الورق . وهذه المفردات الثلاثة نتاج العملية الرياضية التالية :**

$9 + 10 \times 1 = 19$ , $5 + 10 \times 1 = 15$ , $2 + 10 \times 1 = 12$

وهى المفردات الثلاثة التى تنحصر مابين عشرة جنيهات وأقل من عشرون جنيها .

– بالنسبة لمفردات الصف الثالث الذى يقع أسفل الجذع والورقة فهناك سبعة مفردات كل منهم يحتوى على عشرون جنيها أى اثنين فى خانة العشرات (وهى تعادل ٢٠=١٠×٢) هذا بالاضافة لأرقام الآحاد التى يوضحها الورق . وهذه الارقام السبعة تفصيلهم على النحو التالى :–

$3 + 10 \times 2 = 23$ , $1 + 10 \times 2 = 21$ , $0 + 10 \times 2 = 20$

$9 + 10 \times 2 = 29$ , $8 + 10 \times 2 = 28$ , $7 + 10 \times 2 = 27$

$5 + 10 \times 2 = 25$

وهى بمثابة المفردات السبعة التى تنحصر مابين العشرون جنيها وأقل من ثلاثون جنيها . وهكذا بالنسبة للصف الاخير الذى يقع أسفل الجذع والورق

– فى السطر قبل الاخير من المخطط يعطى صورة للقارئ عن طول الفئة (أى طول فئة الساق أو الجذع) ثم السطر الاخير من هذا المخطط يبين أن كل ورقة تمثل مفردة من مفردات الظاهرة محل الدراسة .

– يمكن من خلال النظر لمخطط الساق والورقة معرفة اقل قيمة وأكبر قيمة فى البيانات .

**<u>مثال(٢٠):-</u>**

**البيانات التالية تمثل أطوال مجموعة من نباتات القطن مقدرة بالسنتيمترات (٥٦ مفردة).**

**80 , 84 , 71 , 72 , 35 , 93 , 91 , 74 , 60 , 63 , 79 , 80 , 70 , 68**
**, 90 , 92 , 80 , 70 , 63 , 76 , 48 , 90 , 92 , 85 , 83 , 76 , 61 ,**
**99 , 83 , 88 , 74 , 70 , 65 , 51 , 73 , 71 , 72 , 95 , 82 , 70 , 33**
**, 37 , 32 , 41 , 44 , 49 , 47 , 50 , 59 , 55 , 53 , 56 , 52 , 64 ,**
**60 , 66**

**<u>والمطلوب :-</u>**

**مستخدما مخطط الجذع والورقة تمثيل أطوال هذه المجموعة من نباتات القطن بيانيا .**

**الحل :-**

**من خلال قراءة بيانات أطوال نباتات القطن يتضح لنا أن أقل قيمة فى الاطوال هى ٣٢ سم وأكبر قيمة فى المفردات هى٩٩ سم . فإذا اعتبرنا طول الفئة ١٠سم فيكون المخطط الخاص بالجذع (أو الساق) والورقة يبدأ أول رقم فى الجذع أو الساق بالرقم(٣) ليعنى أن خانة العشرات (الحد الادنى) هو ثلاثون . وحيث ان المفردات تحتوى على العشرات المتتالية أى ٤٠,٥٠,٦٠,٧٠,٨٠,٩٠ لذا فإن عمود الساق سينتهى بخانة العشرات بالرقم(٩) اما عن الورق فيوضحه الشكل البيانى التالى:-**

**شكل(٢-١٩)**

**مخطط الجذع (أو الساق) والورقة لأطوال عينة من نباتات القطن(بالسنتيمتر)**

TALL Stem-and-Leaf Plot

| Frequency | Stem & | Leaf |
|---|---|---|
| 4.00 | 3 . | 2357 |
| 5.00 | 4 . | 14789 |
| 7.00 | 5 . | 0123569 |
| 9.00 | 6 . | 001334568 |
| 14.00 | 7 . | 00001122344669 |
| 9.00 | 8 . | 000233458 |
| 8.00 | 9 . | 00122359 |

Stem width: 10.00
Each leaf: 1 case(s)

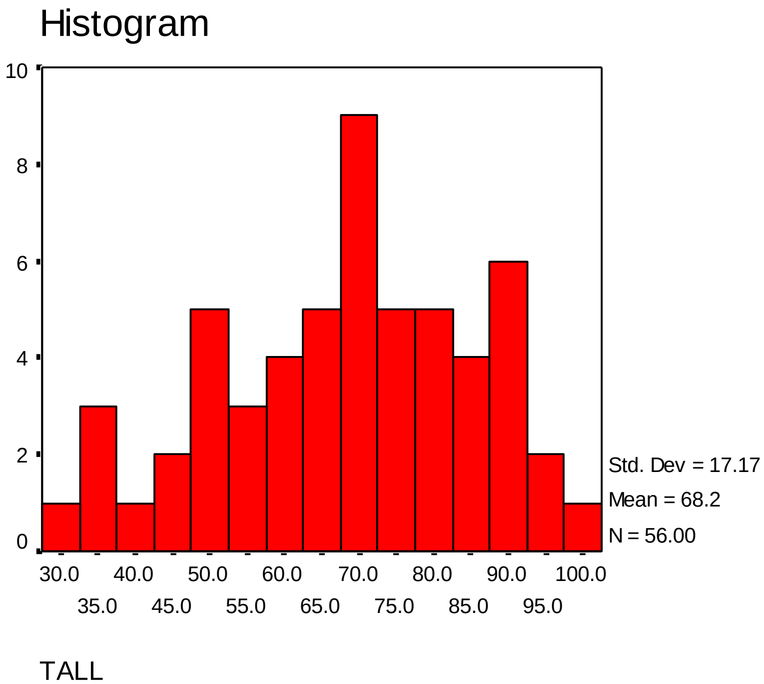
**من خلال شكل (٢-١٩) يتضح لنا أن :-**

- **أصغر قيمة فى البيانات هى ٣٢ وأكبر قيمة فى البيانات هى٩٩ .**

- طول الفئة فى هذا المخطط هو ١٠
- عدد المفردات والتى يمثلها مجموع التكرارات هو ٥٦ مفردة وهو بمثابة حجم العينة المعطى فى المثال .
- من خلال مخطط الجذع والورقة يمكن الحكم على مدى تماثل التوزيع من عدمه .هذا ويمكن ملاحظة الفرق بين مخطط الجذع والورقة والمدرج التكرارى . فالشكل(٢-٢٠) يوضح رسم المدرج التكرارى لأطوال عينة النباتات لهذا المثال .

**شكل (٢-٢٠)**

**المدرج التكرارى لأطوال عينة من نباتات القطن**




**<u>ثالثا :ـ الاشكال أو المخططات البيانية الخاصة بالسلاسل الزمنية :</u>**

Time Series Charts

السلسلة الزمنية هى عبارة عن دراسة لقيم ظاهرة أو متغيرا ما خلال فترات زمنية متساوية (يوم أو شهر أو سنة أو......) ، وهناك طريقتين أساسيتين لعرض بيانات السلسلة الزمنية وهما:-

**١- خريطة الخط البيانى :- Line Chart**

حيث تبين خريطة الخط البيانى مدى تطور ظاهرة ما خلال فترات زمنية معينة . حيث يتم تمثيل عنصر الزمن على المحور الافقى أما قيم الظاهرة محل الدراسة فيتم تمثيلها على المحور الرأسى بمقياس رسم يتناسب مع قيم الظاهرة . وبناءا على خريطة الخط البيانى يتم تمثيل قيمة الظاهرة محل الدراسة عند كل نقطة زمنية بنقطة إحداثياها هما النقطة الزمنية وقيمة الظاهرة عند تلك النقطة . وبعد الانتهاء من توقيع كافة النقاط التى تمثل السلسلة الزمنية للظاهرة يتم توصيل كل نقطتين متتاليتين بخط مستقيم أو أشبه بالمضلع . فيكون الناتج هو عبارة عن الخط البيانى للظاهرة محل الدراسة . والمثال التالى يوضح ذلك .

**مثال (٢١):-**

فيما يلى بيان بتطور عدد العاملين(الذكور) بالوظائف العليا بالدولة وذلك خلال الفترة (١٩٨٠-١٩٨٤)

جدول(٣٣-٢)

| السنة | ١٩٨٠ | ١٩٨١ | ١٩٨٢ | ١٩٨٣ | ١٩٨٤ |
|---|---|---|---|---|---|
| عدد العاملين | ١٤٠٧ | ١٤١٢ | ٢٠٠٣ | ٢١٣٨ | ٢١٨٧ |

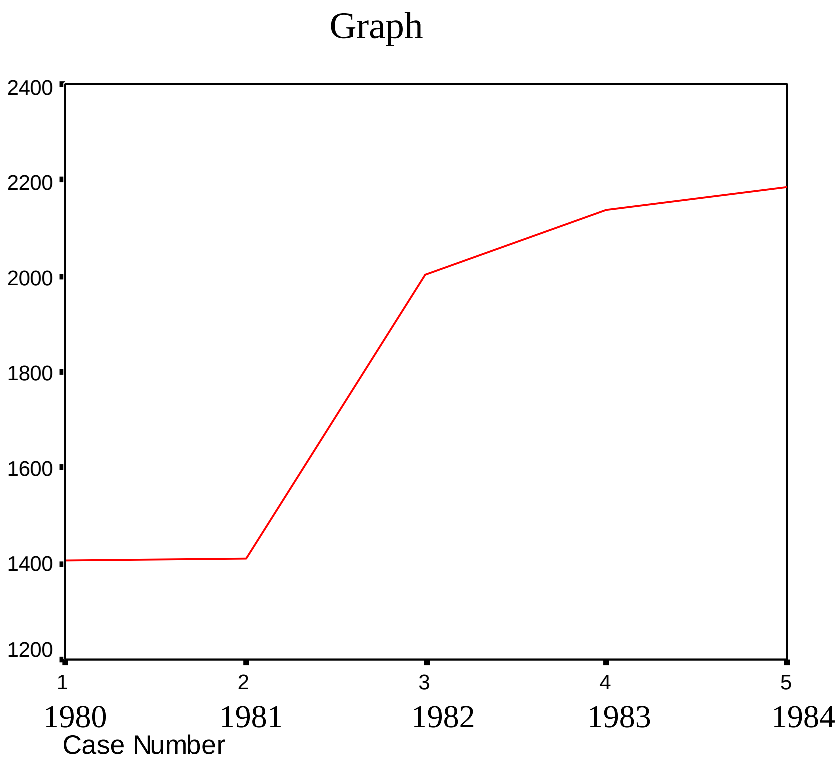
والمطلوب عرض بيانات تلك السلسلة باستخدام خريطة الخط البيانى.

**<u>الحل :-</u>**

**يتم تمثيل السنوات المختلفة للسلسلة على المحور الافقى بينما يمثل عدد العاملين على المحور الرأسى . وحيث أن أصغر قيمة لعدد العاملين هى ١٤٠٧ عاملا فإنه يكون من الملائم كسر المحور الرأسى وذلك حتى يمكن اختيار مقياس رسم مناسب يوضح تطور أعداد العاملين دون ترك مساحة كبيرة من الرسم البيانى غير مستخدمة . لذا فإننا سوف نختار مقياس الرسم للمحور الرأسى هو ١سم/ ١٠٠عاملا مع البدء بالقيمة١٤٠٠ كما يتضح ذلك فى الشكل (٢-٢١) .**

**شكل(٢-٢١)**

**خط بيانى يوضح تطور أعداد العاملين (الذكور) بالوظائف العليا فى الدولة**



**ومن خلال هذا الشكل يمكن استنتاج أن أعداد العاملين بالوظائف العليا فى الدولـة فى زيـادة مستمرة مـع الـزمن وذلـك بمجـرد النظـر الـى الخـط البيـانى الموضح فى الشكل (٢-٢١) .**

**<u>٢- خريطة الشريط :-</u>** Band Chart

**وتعتبر خريطة الشريط بمثابة امتدادا لشكل الخط البيانى السابق الاشارة إليه حيث أنـها تشمل خطين بيانين أو أكثر حسب أقسام الظاهرة . علـى أن يمثـل كـل خـط إحـدى الظـواهر المـراد تمثيلهـا وبحيـث يكـون الفـرق بـين القـيم المتناظرة لهذه الظواهر ذا معنى .**

**ولرسم خريطة الشريط . يتم تمثيل كل ظاهرة من الظاهرتين بخط بيانى مع تلوين المساحة بين الخطين بلون أو تظليل معين إذا كانت الفروق فيما بين القيم المتناظرة تأخذ اتجاها واحدا (موجبا أو سالبا) . أما اذا كانت الفروق موجبـة أحيانـا وسـالبة احيانـا أخـرى فإنـه يلـزم اسـتخدام لـونين مختلفـين أو طـريقتين مختلفتـين فـى التظليـل وذلـك للتفرقـة مـا بـين الفـروق الموجبـة والاخرى السالبة بهدف توضيح مدلول كل منهما .**

**<u>مثال</u> :**

**فيما يلى لديك بيانات سلسلة زمنية للفترة(١٩٦٠-١٩٦٥) لأعداد الطلاب (ذكورا واناثا) فى إحدى الكليات.**

| السنة | ١٩٦٠ | ١٩٦١ | ١٩٦٢ | ١٩٦٣ | ١٩٦٤ | ١٩٦٥ |
|---|---|---|---|---|---|---|
| عدد الذكور | ١٢٠ | ١٥٠ | ١٧٥ | ٢١٠ | ٢٨٠ | ٣٢٠ |
| عدد الاناث | ٩٠ | ١٠٠ | ١٢٠ | ١٥٠ | ١٦٠ | ١٧٥ |

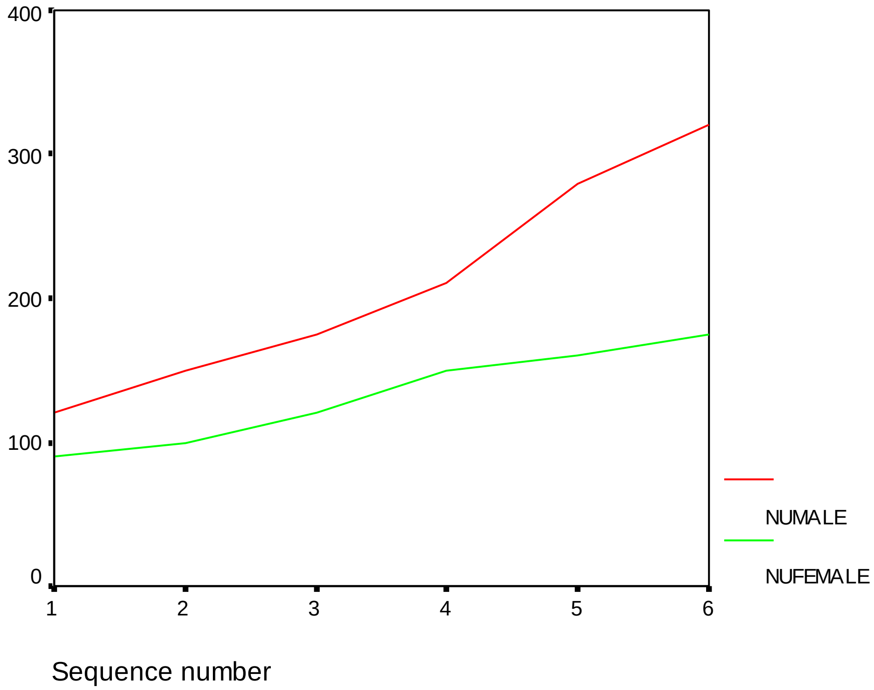
والمطلوب عرض بيانات هذه السلسلة باستخدام خريطة الشريط .

**الحل :**

الشكل البيانى التالى (شكل(٢-٢٢)) يوضح خريطة الشريط لبيانات السلسلة الزمنية المعطاه.

**شكل(٢-٢٢)**

**خريطة الشريط لأعداد الطلاب (ذكور وإناث)فى إحدى الكليات عن الفترة ١٩٦٠-١٩٦٥**




ويلاحظ من خريطة الشريط الموضحة فى الشكل (٢-٢٢) ان عدد الطلاب الذكور يكون دائما أكبر من عدد الطلاب الاناث . كما يتضح من الشكل ايضا ان الفرق بينهما فى زيادة مستمرة من سنة لأخرى .

وأخيرا فهناك أشكال بيانية أخرى لا يسعنا المجال لذكرها حيث أننا نختار منها ما يخدم مجال الدراسة . فهناك ما هو يساهم فى تمثيل التوزيع أو التركيب العمرى للسكان أو ما يسمى بالهرم السكانى ( Population Pyramid ) ومنها ما يستخدم فى الدراسات الاقتصادية والذى يعطى صورة عن مدى التوزيع العادل للثروة أو الناتج القومى على جمهور الافراد أو السكان داخل قطر معين وهو ما يسمى بمنحنى لورنز (Lorenz Curve ) وغير ذلك من الرسوم البيانية الأخرى.

## تمارين

(١) أخذت عينة من ٤٠ طالب وتم تجميع البيانات الخاصة بالالتحاق بالشعب المختلفة حسب التخصص فكانت كما يلى :

| | | | | |
|---|---|---|---|---|
| محاسبة | محاسبة | إدارة | محاسبة | محاسبة |
| إدارة | إدارة | محاسبة | إدارة | محاسبة |
| محاسبة | محاسبة | إقتصاد | محاسبة | إدارة |
| إقتصاد | إدارة | محاسبة | إدارة | إقتصاد |
| محاسبة | محاسبة | إدارة | محاسبة | محاسبة |
| إقتصاد | محاسبة | إقتصاد | إقتصاد | إحصاء |
| إحصاء | إقتصاد | إدارة | إدارة | محاسبة |
| محاسبة | إحصاء | محاسبة | إدارة | إدارة |

والمطلوب وضع هذه البيانات فى جدول تكرارى بسيط وفقا لشعب التخصص.

(٢) أخذت تقديرات ٥٠ طالبا فى مادة الاحصاء فى أحد الفصول الدراسية فكانت كما يلى:

| | | | | |
|---|---|---|---|---|
| ممتاز | مقبول | جيد | ضعيف | جيد |
| ضعيف جدا | ضعيف | مقبول | مقبول | جيد جدا |
| جيد | جيد | جيد جدا | ممتاز | جيد |
| جيد | ممتاز | مقبول | جيد | ضعيف |
| جيد جدا | ضعيف | ضعيف جدا | جيد جدا | مقبول |
| مقبول | جيد جدا | ضعيف | مقبول | جيد |
| مقبول | جيد | جيد | مقبول | ممتاز |
| مقبول | مقبول | جيد جدا | مقبول | ضعيف جدا |
| مقبول | مقبول | مقبول | مقبول | مقبول |
| مقبول | مقبول | مقبول | مقبول | جيد |

**والمطلوب . وضع هذه التقديرات فى جدول تكرارى بسيط حسب التقديرات المختلفة .**

**(٣) فيما يلى تقديرات نجاح ٢٠طالبا فى مادة الرياضيات :**

**جيد جدا، جيد، مقبول، جيد، ممتاز، جيد، جيد جدا، جيد، جيد جدا، مقبول، جيد، جيد جدا، جيد، جيد، جيد، مقبول، جيد جدا، جيد جدا، مقبول، ممتاز.**

**المطلوب :-**

**أ- تبويب هذه البيانات فى جدول تكرارى .**

**ب- تمثيل البيانات بيانيا وذلك بإستخدام شكل الدائرة والاعمدة البيانية.**

**(٤) اعتبر الارقام التالية تمثل الدرجات التى حصل عليها ٥٤ طالبا فى احدى الاختبار(الدرجة العظمى=١٠) .**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ٤ | ٣ | ٥ | ٤ | ٣ | ٥ | ٥ | ٤ | ٥ |
| ٦ | ٥ | ٤ | ٥ | ٤ | ٥ | ٣ | ٤ | ٦ |
| ٥ | ٥ | ٧ | ٤ | ٣ | ٤ | ٣ | ٢ | ٧ |
| ٤ | ٣ | ٦ | ١ | ٣ | ٦ | ٣ | ١ | ٨ |
| ٦ | ٣ | ٥ | ٢ | ٧ | ٥ | ٧ | ٣ | ٤ |
| ٦ | ٥ | ٨ | ٦ | ٤ | ٣ | ٤ | ٤ | ٨ |

**والمطلوب وضع هذه البيانات فى جدول تكرارى بسيط .**

**(٥) فيما يلى تقديرات رسوب ونجاح الطلاب فى مادة الاحصاء فى الكلتين (أ ، ب) فى أحد الاعوام .**

| الكلية/التقديرات | ضعيف جدا | ضعيف | مقبول | جيد | جيد جدا | ممتاز | المجموع |
|---|---|---|---|---|---|---|---|
| الكلية أ | ٧٨ | ٢٠٣ | ٣١١ | ٢٩٦ | ٧٨ | ٢٤ | ٩٩٠ |
| الكلية ب | ٢٧٤ | ٢٣٦ | ٢٢٩ | ٢٩١ | ٤١ | ١٦ | ١٩٨٧ |

**المطلوب: عرض هذه البيانات بالأشكال البيانية المناسبة .**

**(٦) الجدول التكرارى يوضح التوزيع التكرارى لعدد أيام الاجازات التى قام بها ٨٠ موظفا فى إحدى الشركات خلال عام كامل :**

| عدد أيام الاجازة | ١ | ٢ | ٣ | ٤ | ٥ | ٦ | ٧ |
|---|---|---|---|---|---|---|---|
| عدد الموظفين | ٤ | ٧ | ١١ | ١٨ | ٢٢ | ١٠ | ٨ |

**والمطلوب عرض هذه البيانات باستخدام الشكل المناسب .**

**(٧) البيانات التالية تمثل الدرجات التى حصل عليها ١٠٠طالب فى إحدى المواد الدراسية:**

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ٤٥ | ٥٢ | ٢٦ | ٥٠ | ٥٦ | ٦٧ | ٥٧ | ٥٥ | ٣٦ | ٤٥ |
| ٩٣ | ٩٣ | ٤٠ | ٥٤ | ٢٧ | ٥٧ | ٣٤ | ٣١ | ٤٦ | ٧٨ |
| ٢٧ | ٣٥ | ٥٤ | ٨٠ | ٣٠ | ٣٨ | ٥٠ | ٤٤ | ٤٧ | ٢٠ |
| ٧١ | ٢٥ | ٥١ | ٦١ | ١٢ | ٤٢ | ٤٥ | ٥٧ | ٣٣ | ٥٤ |
| ٦٦ | ٦٨ | ١٢ | ٥٢ | ٦٢ | ٦٥ | ٤٣ | ٩٦ | ٢٥ | ٥٦ |
| ٤٢ | ٥١ | ٢٨ | ٥٢ | ٦٧ | ٥٥ | ٥٠ | ٤٠ | ٤٨ | ٧٤ |
| ٩٣ | ٤٩ | ٦٦ | ٦٣ | ٥٩ | ٥٢ | ٣٩ | ٥١ | ٨٣ | ١٥ |
| ٨٤ | ٤٠ | ٤٦ | ٥٠ | ٥٦ | ٥٣ | ٥٠ | ٦١ | ٣١ | ٤٥ |
| ٧٩ | ٥٦ | ٥٥ | ٩١ | ٨٦ | ٥٢ | ٧٥ | ٧٥ | ٥٤ | ٦٧ |
| ٦٣ | ٧٧ | ٥٨ | ٦٢ | ٩٥ | ٨٦ | ٥١ | ٥٥ | ٣٦ | ٥٦ |

**والمطلوب :**

**١- وضع هذه البيانات فى صورة جدول تكرارى بسيط .**
**٢- رسم المدرج التكرارى والمضلع التكرارى لهذا التوزيع .**
**٣- تكوين الجدول التكراري المتجمع الصاعد والنازل .**

**(٨): البيانات التالية تمثل أعمار ٦٠ شخصا بالسنوات :**

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ٢١ | ٤٤ | ٩ | ٣٠ | ١٣ | ٢٥ | ٣٢ | ١٢ | ٣٥ | ١٥ |
| ٣٣ | ٢٤ | ٢٣ | ٢٢ | ١٧ | ٣٦ | ٢٦ | ٢٠ | ١٤ | ٣١ |
| ٢٣ | ١٢ | ٣١ | ١٦ | ٢٧ | ٥ | ١٨ | ٣ | ٢٨ | ٣٨ |
| ١٦ | ٢٥ | ١٤ | ٢٥ | ٢١ | ٢٢ | ١٥ | ٢٧ | ٢١ | ١١ |
| ٤٢ | ٧ | ٢٣ | ٣٧ | ٢٤ | ٣٢ | ٢٦ | ٢٩ | ٢٣ | ١٥ |
| ٣٤ | ٢٩ | ٣٩ | ٢٢ | ١٦ | ١٠ | ٢٨ | ١٧ | ٨ | ٢٩ |

**المطلوب :**

**١- ناقش بإيجاز الخطوات التى تستخدم فى تبويب هذه البيانات ، ثم استخدم هذه الخطوات فى تكوين جدول تكرارى ذى فئات متساوية.**
**٢- باستخدام الجدول التكرارى الناتج فى(أ) كون الجدول التكرارى المتجمع الصاعد ، مرة باستخدام التكرارات الاصلية ومرة أخرى باستخدام التكرارت النسبية.**
**٣- أوجد قيمة العمر الذى يقل عنه ٣٣شخصا.**
**٤- استخدم شكل المدرج التكرارى للتعبير عن التوزيع الناتج فى (أ) ثم علق على شكل التوزيع .**

**(٩): فيما يلى قيمة الارباح (بالآف الجنيهات) لشركتين متنافستين (س،ص) فى مجال الصناعات القطنية وذلك خلال السنوات(١٩٧٠– ١٩٧٥) .**

| السنة | ١٩٧٠ | ١٩٧١ | ١٩٧٢ | ١٩٧٣ | ١٩٧٤ | ١٩٧٥ |
|---|---|---|---|---|---|---|
| الشركة(س) | ١٢ | ١٦ | ٢١ | ٢٣ | ٢٤ | ٣٢ |
| الشركة(ص) | ١٠ | ١١ | ١٣ | ١٨ | ٢٥ | ٢٦ |

**المطلوب :**

**أ- قارن بين أرباح الشركتين بإستخدام الاعمدة المجزأة والاعمدة المتلاصقة**

**ب- إستخدام خريطة الخط البيانى لعرض البيانات الخاصة بأرباح الشركة(س)**

**ج- بين ما هى الشركة الاقدر على المنافسة باستخدام خريطة الشريط .**

**(١٠) فيما يلى درجات ٣٠ طالبا فى كل من مادتي الرياضة والإحصاء :**

| الرياضة | الاحصاء | الرياضة | الاحصاء | الرياضة | الاحصاء |
|---|---|---|---|---|---|
| ٨٣ | ٧٧ | ٨٤ | ٧٦ | ٨٦ | ٨٣ |
| ٨٩ | ٧٢ | ٨١ | ٨٩ | ٧٩ | ٨٤ |
| ٦٦ | ٦٤ | ٩٢ | ٨٧ | ٩٤ | ٩١ |
| ٦٦ | ٧٢ | ٧٢ | ٦٩ | ٨٠ | ٦٧ |
| ٧٠ | ٨٦ | ٧١ | ٧٧ | ٨٥ | ٩٣ |
| ٩٧ | ٩٤ | ٨٢ | ٨١ | ٨٨ | ٨٥ |
| ٦٩ | ٧٨ | ٨٣ | ٩٢ | ٩٦ | ٩٠ |
| ٧٤ | ٧٥ | ٧٥ | ٨٠ | ٧٨ | ٨٨ |
| ٦٥ | ٦٣ | ٧٩ | ٧٤ | ٥٠ | ٥٨ |
| ٦٨ | ٦٧ | ٧٦ | ٨٢ | ٥٢ | ٥٧ |

**والمطلوب وضع هذه البيانات فى صورة جدول تكرارى مزدوج.**

**(١١) البيانات التالية تمثل درجات ٥٠طالبا فى كل من مادتي الرياضيات (س) والإحصاء (ص) :**

| س | ص | س | ص | س | ص | س | ص | س | ص |
|---|---|---|---|---|---|---|---|---|---|
| ٤١ | ٢٠ | ٨٩ | ٧٨ | ٥٦ | ٤١ | ٥٢ | ٤٧ | ٦٢ | ٦٤ |
| ٨٠ | ٧٠ | ٥٠ | ٣٤ | ٨٢ | ٧٣ | ٤٠ | ٤٢ | ٩٣ | ٧٥ |
| ٦٧ | ٥٥ | ٦٥ | ٥٣ | ٦٦ | ٦٣ | ٦٧ | ٦٥ | ٦٠ | ٤٩ |
| ٩٦ | ٨٥ | ٦٩ | ٧٠ | ٩٣ | ٨٢ | ٧٣ | ٦٤ | ٤٧ | ٣٦ |
| ٣٠ | ٢٥ | ٨٢ | ٨٠ | ٨٠ | ٦٠ | ٧٦ | ٦٧ | ٩٠ | ٨٠ |
| ٦١ | ٦٢ | ٤٩ | ٤٦ | ٣٥ | ٣٧ | ٥٨ | ٥٣ | ٥٠ | ٤٣ |
| ٩٨ | ٧٨ | ٧٠ | ٦٩ | ٥٢ | ٥٩ | ٣٨ | ٣٩ | ٦٠ | ٦٠ |
| ٥٠ | ٦٧ | ٧٥ | ٧١ | ٧٦ | ٥٣ | ٧٩ | ٥٩ | ٧٢ | ٦٢ |
| ٨٥ | ٦٤ | ٥٠ | ٥٠ | ٧٢ | ٧٠ | ٥٥ | ٣٦ | ٥١ | ٤٥ |
| ٧٥ | ٦٦ | ٦١ | ٥٨ | ٦٣ | ٦٨ | ٦٣ | ٥٤ | ٦٨ | ٥٧ |

**المطلوب :**

**أ- عرض هذه البيانات فى شكل جدول تكرارى مزدوج وذلك على النحو الاتى :**

- **طول الفئة فى كل من س،ص =١٠**
- **الحد الادنى للفئة الاولى من فئات(س) = ٣٠.**
- **الحد الادنى للفئة الاولى من فئات(ص) = ٢٠.**

**ب- استنتج التوزيع التكراري الهامشى لدرجات الطلاب فى كل من مادتى الرياضيات والاحصاء.**

**ج- قارن بين التوزيعين الناتجين فى (ب) من حيث مستوى درجات الطلاب وذلك بإستخدام المنحنى التكرارى.**

د- باستخدام المدرج التكرارى ، قارن بين شكلى التوزيعين من حيث التماثل والالتواء .

ه- عزز نتائجك فى(جـ) بإيجاد نسبة نجاح الطلاب فى كل من مادتى الرياضيات والاحصاء وذلك باستخدام التوزيع المتجمع المناسب .

(ملحوظة : يعتبر الطالب ناجحا إذا حصل على ٥٠ درجة فأكثر).

و- باستخدام التوزيع المتجمع المناسب ، حدد قيمة الدرجة التى يقل عنها ١٤% من إجمالى عدد الطلاب فى كل من مادتى الرياضيات والاحصاء ثم علق على إجابتك وذلك فى ضوء نتائجك فى (جـ)،(هـ) .

ز- أعد تبويب بيانات درجات الطلاب فى مادتى الرياضيات والاحصاء كل على حده وذلك بغرض ايجاد التوزيع التكرارى لتقديرات نجاح ورسوب الطلاب فى كل مادة على انفراد ، ثم قارن بين التوزيعين الناتجين باستخدام الاعمدة المتلاصقة والاعمدة المجزأة .

ملحوظة : تقديرات النجاح والرسوب تحدد كما يلى :

| التقدير | ضعيف جدا | ضعيف | مقبول | جيد | جيد جدا | ممتاز |
|---|---|---|---|---|---|---|
| فئة الدرجات | صفر- | ٣٥- | ٥٠- | ٦٥- | ٨٠ | ٩٠-١٠٠ |

(١٢) الجدول التالى يمثل عدد ١٠٠شخص موزعين حسب فئات السن المختلفة :

| فئات السن | ٢٠- | ٢٥- | ٣٥- | ٤٥- | ٥٠- | ٥٥-٦٠ | الجملة |
|---|---|---|---|---|---|---|---|
| عدد الاشخاص | ٥ | ٢٢ | ٣٠ | ٢١ | ١٩ | ٣ | ١٠٠ |

والمطلوب :

١- رسم المدرج التكرارى والمضلع التكرارى لهذا التوزيع .
٢- عمل الجدول التكرارى المتجمع الصاعد والنازل ثم أوجد
- عدد الاشخاص الذين يقل عمرهم عن٤٥ سنة بالرسم.
- الحد الادنى لأعمار ٤٠ شخصا بالرسم .

(١٣): فيما يلى التوزيع التكرارى لأعمار ١٠٠ شخص .

| فئات الاعمار | ٥- | ١٠- | ١٥- | ٢٠- | ٢٥- | ٣٥- | ٤٥-٥٥ | المجموع |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| عدد الاشخاص | ٥ | ١٢ | ١٣ | ٢٢ | ١٨ | ١٦ | ١٤ | ١٠٠ |

المطلوب :
أ- إرسم كلا من المدرج التكرارى والمضلع التكرارى والمنحنى التكرارى لهذا التوزيع .
ب- احسب قيمة العمر الذى يقل عنه ٣٠% من إجمالى عدد الموظفين . ثم بين أن هناك نفس النسبة من الموظفين الذين تبلغ أعمارهم ٣٥ سنة على الاقل .
ج- بإستخدام المنحنى التكرارى المتجمع المناسب ، أحسب نسبة الاشخاص الذين تبلغ أعمارهم ١٨ سنة على الاقل.

(١٤) باستخدام التوزيع التكرارى الوارد فى الجدول (١٥-٢) ، والمطلوب :
أ- ارسم المضلع التكرارى والمنحنى التكرارى لبيانات الانفاق .
ب- كون الجدول التكرارى المتجمع الصاعد وذلك بإستخدام التكرارت النسبية ، ثم أوجد قيمة الانفاق الاسبوعى الذى يقل عنه ٨٤% من إجمالى عدد الاسر .

ج- على ضوء نتائجك فى (ب) ، إستنتج قيمة الانفاق الاسبوعى الذى تنفقه – كحد أدنى – ٣٦%من إجمالى عدد الاسر .

(١٥) يوجد فى إحدى الشركات الصناعية ٢٠٠ عامل كان توزيع أجورهم الأسبوعية بالجنيه كالآتى:

| الأجر | ٢٠– | ٢٥– | ٣٠– | ٣٥– | ٤٠– | ٤٥– | ٥٠– | ٥٥– |
|---|---|---|---|---|---|---|---|---|
| العدد | ١٠ | ١٧ | ٢٤ | ٤٣ | ٣٤ | ٣٠ | ٢٣ | ١٩ |

والمطلوب :

١- عرض هذه البيانات بإستخدام المدرج التكرارى والمضلع التكرارى .

٢- رسم المنحنى المتجمع الصاعد ومنه أوجد :

- عدد العمال الذين يتقاضون أجرا أقل من ٤٠ جنيه .
- الحد الاعلى للأجر الذى يبلغه ١٢٠ عاملا .

٣- رسم المنحنى المتجمع النازل ومنه أوجد :

- عدد العمال الذين يتقاضون أجرا قدره ٣٧.٥ جنيه فأكثر .
- الحد الادنى الذى يتقاضاه ١٢٠ عاملا .

(١٦): سحبت عينة حجمها ١٣٠ عامل بشركة السكر وكانت كمية إنتاج العامل مبوبة فى الجدول التكرارى التالى :

| كمية الانتاج | ٢٠– | ٣٠– | ٤٠– | ٥٠– | ٦٠– | ٧٠– ٨٠ | المجموع |
|---|---|---|---|---|---|---|---|
| عدد العمال | ١٤ | ٢٠ | ٢٨ | ٣٦ | ٢٢ | ١٠ | ١٣٠ |

والمطلوب :

أ- رسم المدرج التكرارى والمضلع التكرارى والمنحنى التكرارى للتوزيع السابق.

ب- إيجاد الجدول التكرارى المتجمع الصاعد والهابط .

(١٧): تم اختيار ٤٠ طالب من مدرسة ما لدراسة مستوى الذكاء للطلبة وكانت النتائج كالتالى :

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ٥٥ | ٥٠ | ٤٠ | ٦٠ | ٧٠ | ٥٥ | ٤٠ | ٥٠ | ٦٠ | ٢٠ |
| ٤٥ | ٤٠ | ٣٥ | ٥٠ | ٥٥ | ٥٠ | ٣٥ | ٤٠ | ٦٠ | ٣٠ |
| ٨٠ | ٤٩ | ٦٠ | ٢٨ | ٤٢ | ٥١ | ٦٢ | ٢١ | ٣٦ | ٥٢ |
| ٧٦ | ٥٦ | ٩٣ | ٢٠ | ٣٩ | ٣٢ | ٣٨ | ٤٢ | ٨٢ | ٨٥ |

المطلوب :

أ- ايجاد توزيع تكرارى منتظم لمستوى الذكاء (طول الفئات متساوى).

ب- ايجاد توزيع تكرارى غير منتظم لمستوى الذكاء.

(١٨): فيما يلى كمية الانتاج فى بعض الصناعات بالألف طن :

| المصنع | السكر | الورق | الاسمنت | الحديد | الغزل | زيت البترول |
|---|---|---|---|---|---|---|
| الانتاج | ٢٣٠ | ٨٠٠ | ١٩٠٠ | ٧٠٠ | ٢٠٠ | ٧ |

المطلوب تمثيل بيانات الجدول السابق مستخدمًا في ذلك :

أ- الاعمدة البسيطة .

ب- الدائرة .

(١٩): البيانات التالية تمثل التحويلات السياحية (بالمليون جنيه) وعدد الليالى السياحية (بالمليون) خلال ٢٥ شهرا ، والمطلوب وضع البيانات فى جدول تكرارى مزدوج .

| الليالى السياحية | الليالى السياحية | التحويلات السياحية | الليالى السياحية |
|---|---|---|---|
| ٨٩ | ٣٦ | ٨٢ | ٤٦ |
| ٦٦ | ٣٢ | ٧٤ | ٤٠ |
| ٦٥ | ٣٦ | ٨٥ | ٤٢ |
| ٧٠ | ٤٣ | ٨٨ | ٤٢ |
| ٩٦ | ٤٧ | ٩١ | ٤٣ |
| ٦٨ | ٣٧ | ٨٣ | ٣٠ |
| ٧٤ | ٣٦ | ٧٨ | ٤٥ |
| ٨٤ | ٣٥ | ٥٠ | ٢٧ |
| ٧٢ | ٤٤ | ٩٦ | ٤٦ |
| ٧٣ | ٣٤ | ٨٥ | ٤٦ |
| ٧٢ | ٣٨ | ٨٠ | ٤٦ |
| ٨١ | ٤١ | ٩٤ | ٣٢ |
| | | ٧٩ | ٤٥ |

**(٢٠) أخذت عينة من ٣٠ سائح وتم جمع البيانات الخاصة بالغرض من القدوم فكانت كالآتى:**

| | | | |
|---|---|---|---|
| عقيدة | عقيدة | متعة وترفيه | متعة وترفيه |
| متعة وترفيه | متعة وترفيه | عقيدة | عقيدة |
| عقيدة | عقيدة | رياضة | متعة وترفيه |
| رياضة | متعة وترفيه | عقيدة | عقيدة |
| عقيدة | عقيدة | متعة وترفيه | رياضة |
| رياضة | عقيدة | رياضة | |
| زيارة للأسرة | رياضة | متعة وترفيه | |
| عقيدة | زيارة للأسرة | عقيدة | |
| | | عقيدة عقيدة | |

**والمطلوب :**

**وضع هذه البيانات فى جدول تكرارى بسيط تبعا للغرض من القدوم .**

**(٢١): البيانات التالية تمثل توزيع الانفاق اليومى ل ٢٥٠ سائح فى إحدى محافظات ج.م.ع**

| الإنفـــاق اليومى | ٢٠- | ٢٥- | ٣٠- | ٣٥- | ٤٠- | ٤٥- | ٥٠- | ٥٥- |
|---|---|---|---|---|---|---|---|---|
| عــــــدد السائحين | ١٨ | ٢٧ | ٣٢ | ٥١ | ٤٣ | ٣٥ | ٢٤ | ٢٠ |

**المطلوب :**

**١- عرض هذه البيانات باستخدام المدرج التكرارى والمضلع التكرارى .**

**٢- رسم المنحنى المتجمع الصاعد ومنه أوجد :**

- **عدد السائحين الذين ينفقون أقل من ٤٠ جنيه يوميا .**
- **الحد الادنى للانفاق اليومى لـ ١٢٠ سائحا .**

**٣- رسم المنحنى المتجمع النازل ومنه أوجد .**

- عدد السائحين الذين ينفقون ٣٧.٥ جنيه فأكثر يوميا .
- الحد الادنى للإنفاق اليومى لـ ١٢٠سائحا .

(٢٢): الجدول التالى يوضح التوزيع التكرارى لأجور العمال الاسبوعية (بالجنيهات) فى ثلاثة مصانع مختلفة (س،ص،ع) :

| فئات الاجور | | ٥٠- | ٦٠- | ٧٠- | ٨٠- | ٩٠- | ١٠٠- | المجموع |
|---|---|---|---|---|---|---|---|---|
| عدد العمال | المصنع (س) | ١ | ٣ | ٤ | ١٢ | ١٢ | ٨ | ٤٠ |
| | المصنع (ص) | ٨ | ١٢ | ١٢ | ٤ | ٣ | ١ | ٤٠ |
| | المصنع(ع) | ١٥ | ١٧ | ١٨ | ١٨ | ١٧ | ١٥ | ١٠٠ |

المطلوب :

١- إستخدام الشكل البيانى المناسب للمقارنة بين توزيع الاجور (من حيث المستوى) وذلك فى الحالات الاتية :

- المصنعان : س،ص – المصنعان : س،ع

٢- قارن بين أشكال التوزيعات الثلاثة (من حيث التماثل والالتواء) وذلك بإستخدام المنحنى التكرارى .

٣- بإستخدام التوزيع التكرارى المتجمع المناسب . قارن بين قيمة الاجر الذى يحصل عليه (كحد أدنى) ٥٠% من اجمالى عدد العمال وذلك فى كلا من المصنعين س،ع.

٤- قارن بين قيمة الاجر الذى يحصل على أقل منه ٢٠% من اجمالى عدد العمال فى كل من المصنعين س،ص وذلك بإستخدام التوزيع المناسب للتكرارات المتجمعة .

٥- إحسب النسبة المئوية للعمال الذين يتقاضون أجورأقل من ٨٠ جنيها فى كل من المصانع الثلاثة. علق على نتائجك على ضوء ماحصلت عليه فى (أ).

٦- احسب قيمة الاجر الذى يقل عنه ٤٠ عاملا فى المصنع (ع).

٧- أدمج توزيعى أجور العمال فى المصنعين س،ص فى كل فئة من الفئات المختلفة ، ثم قارن بين شكل التوزيع الجديد وبين شكل التوزيع في كل من هذين المصنعين مستخدمًا المنحنى التكراري مبررًا ما تحصل عليه من نتائج .

(٢٣): فيما يلى لديك التوزيع التالى لأعمار عينة من الاشخاص حجمها ١٠٠ مفردة

| الفئة | اقل من ٣٠ | اقل من ٣٥ | اقل من ٤٥ | اقل من ٦٠ | اقل من ٨٠ |
|---|---|---|---|---|---|
| التكرار | ١٩ | ٢٧ | ٤٨ | ٧٣ | ١٠٠ |

والمطلوب :

أ- رسم المدرج التكرارى .

ب- ماهى نسبة الاشخاص الذين تتراوح أعمارهم مابين ٤٠,٦٧ سنة.

ج- أذا كانت نسبة الاشخاص الذين تتراوح أعمارهم مابين ٣٧ سنة والعمر y تبلغ ١٥% من اجمالى الاشخاص . حدد ماهى قيمة (y) .

د- ماهو الحد الادنى للعمر الذى يبلغه ٧٥% من العمال على الاقل .

# الباب الثالث
## مقاييس النزعة المركزية (المتوسطات)
## Measures of Central Tendency

**أوضحنا فى الباب السابق أن الهدف الاساسى من التحليل الاحصائى هو التعرف على خصائص المجتمع الاصلى محل الدراسة وأستخدمنا الجداول الإحصائية والرسوم البيانية بهدف عرض وتلخيص البيانات، ومما لاشك فيه أن طرق العرض الجدولى للبيانات تقدم وصفا عاما وسريعا للبيانات مما يتفق مع المثل القائل بأن صورة واحدة تساوى ألف كلمة ، إلا أن هناك حدودا – على أية حال – لاستخدام الطرق البيانية والجدولية فى مجال وصف وتحليل البيانات ، فطرق العرض الجدولى والبيانى تعد غير دقيقة وغير كافية لوصف الظواهر محل الدراسة ، فكثيرا مانجد أن الجداول التكرارية لاتزال تحتوى على أعداد كبيرة بحيث يصعب تكوين فكرة واضحة عن الظاهرة ، كما اننا قد نحصل على طرق عرض مختلفة بإختلاف الاشخاص العارضين للظاهرة وبالتالى نصل الى عدة استنتاجات مختلفة لتفس الظاهرة وهذا يتنافى مع الخصائص الجيدة لطرق العرض . ويزداد الامر صعوبة إذا ما أردنا مقارنة مجموعتين أو أكثر من البيانات وثمة امر أخر على جانب كبير من الاهمية يضع حدا لاستخدام الطرق البيانية والجدولية وهو صعوبة الاستفادة منها فى مجال الاستقراء الاحصائى ، وربما اقتصرت فوائدها الاستقرائية على أن يقدم المضلع التكرارى مثلا لعينة من البيانات نقوم بتلخيصها تصورا عن شكل المضلع التكرارى للمجتمع الذى سحبت منه هذه العينة . وفى هذا الباب سنلجأ لتلخيص البيانات فى صورة رقمية وذلك عن طريق حساب بعض المقاييس الاحصائية والتى تعرف باسم المتوسطات وهى التى تصف لنا كيفية توزيع الظاهرة محل الدراسة فى العينة بطريقة دقيقة ومختصرة وتصلح أيضا للمقارنة بين ظاهرتين أو اكثر ، فضلا عن أنها لاتختلف بإختلاف الاشخاص العارضين أو الدارسين للظاهرة. والمتوسط لأية مجموعة من البيانات هو القيمة التى تمثل هذه**

المجموعة وتعبر عنها بصفة عامة بصرف النظر عن الاختلافات الموجودة بينها، وهذه القيمة تقع عادة عند وسط المجموعة او قريبا من مركزها بعد ان ترتب مجموعة البيانات بحسب قيمتها العددية . فمن الملاحظ، انه فى كثير الظواهر السلوكية والاجتماعية تنزع معظم البيانات الى التمركز حول تلك القيمة المتوسطة . فأولئك الذين يتصفون بالنحافة الشديدة مثلاً هم قلة وفى المقابل نجد أن الذين يتصفون بالسمنة المفرطة هم قلة أيضا وذلك قياسا على عامة الناس التى تقع بين هذا وذاك ، وأيضا لو طبقنا أختبار لقياس درجة الذكاء على مجموعة من الطلاب لوجدنا أن الطلاب المتفوقين الموهوبين قلة والطلاب المبتلين بالبلادة قلة وتنزع درجة الذكاء عند معظم الطلاب الى التمركز حول الوسط ، ويسمى ذلك الميل الى التجمع حول هذه القيمة الوسطى بالنزعة المركزية ، لذلك تعرف المتوسطات أحيانا (بمقاييس النزعة المركزية ) والتى يقصد بها ببساطة ميل البيانات للتركز والتراكم حول هذه القيمة المتوسطة. وفى حياتنا اليومية كثيرا مانستخدم كلمة " فى المتوسط" فنتحدث عن الرجل " متوسط الدخل" والشاب" متوسط الثقافة" ، كما نقول أن " إنفاق الاسرة اليومى هو فى المتوسط كذا " ....الخ.

وهذه الاستخدامات الشائعة لكلمة متوسط تعبر عن شعور داخلى معين يحسه ويفهمه كل منا ولايستطيع ترجمته بدقة ، ومقاييس النزعة المركزية هى محاولة لترجمة هذا الشعور بطريقة دقيقة ومحددة تماما .

وعند حساب احد مقاييس النزعة المركزية يجب مراعاة العوامل الاتية :

١- ان تكون طريقة حسابه سهلة وأن يعكس التغير فى الظاهرة ولاتتغير قيمته بتغيير طريقة حسابه .

٢- يراعى أن تكون البيانات متجانسة ، فلا يجوز مثلا حساب دخل الوزير وبعض العمال فى وزارته ، بينما يجوز حساب متوسط السن مثلا للوزير والعمال معا .

**٣- يفضل أن يأخذ المقياس عند حسابه – فى الاعتبار – جميع القيم ولايتأثر بالقيم المطرفة او الشاذة .**

**٤- يجب مراعاة القياس بوحدات متساوية فعند حساب متوسط الدخل عن فترة زمنية طويلة يجب مراعاة تغيير القوة الشرائية للنقود .**

**هذا ويوجد عدة مقاييس للنزعة المركزية سوف نتناولها بالتفصيل لبيان كيفية حسابها فى حالة البيانات غير المبوبة وفى حالة البيانات المبوبة ، هذه المقاييس هى :**

**١- الوسط الحسابى .**

**٢- الوسيط ، وسندرس معه الربيعين الادنى والأعلى .**

**٣- المنوال .**

**٤- الوسط الهندسى .**

**٥- الوسط التوافقى .**

***<u>الوسط الحسابى:-</u>** Arithmetic Mean (or Avarage)

**يعتبر الوسط الحسابى بحق من أهم مقاييس النزعة المركزية وأكثرها إستخداما فى الحياة العملية وذلك لبساطته وسهولة فهمه ووضوح معناه ، وهو الذى يشار له بإسم المتوسط فى لغة التخاطب العادى .**

**ويعرف الوسط الحسابى لمجموعة من القيم على انه خارج قسمة مجموع هذه القيم على عددها وأحيانا يعرف على أنه القيمة التى لو أعطيت لكل مفردة من مفردات المجموعة لكان مجموع هذه القيم الجديدة هو نفس مجموعة القيم الاصلية .**

**<u>(٣-١-١) حساب الوسط الحسابى من المفردات (بيانات غير مبوبة):</u>**

**يمكن حساب الوسط الحسابى لمجموعة من البيانات (المفردات) لظاهرة معينة بعدة طرق تختلف فيما بينها فى الجهد الحسابى فقط لكنها فى النهاية تؤدى للحصول على نفس النتيجة كما سنرى فى التحليل التالى :**

**<u>أولا : الطريقة الحسابية المطولة (المباشرة) :-</u>**

**بصفة عامة إذا كان لدينا مجموعة من المفردات ولتكن** $x_n , \ldots , x_2 , x_1$ **حيث (n) تعبر عن عدد المفردات . فأن الوسط الحسابى لهذه المجموعة من هذه المفردات لنرمز له بالرمز** $\bar{x}$ **هوعبارة عن النسبة مابين مجموع القيم وعددها أى أن :**

$$\bar{x} = (x_1 + x_2 + x_3 + \ldots + x_n) / n = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (١)$$

**حيث يمثل البسط (** $\sum_{i=1}^{n} x_i$ **) مجموع القيم التى يأخذها المتغير محل الدراسة والتى يبلغ عددها (n) مفردة .**

**<u>مثال(١) :</u>**

**أحسب الوسط الحسابى لمجموعة القيم :**

**٥٠٠,٦٠٠,٧٠٠,٧٥٠,٨٠٠**

**<u>الحل :</u>**

**حيث أن :**

$$\bar{x} = \frac{x_1 + x_2 + \ldots\ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

**لذا فإن :**

$$\bar{X} = (800 + 750 + 700 + 600 + 500) / 5 = 670$$

**<u>مثال (٢)</u>**

**احسب الوسط الحسابى لمجموعة القيم :1- , 3- , 0 , 2 , 5 , 9**

**<u>الحل :</u>**

**حيث أن :**

$$\bar{x} = \frac{x_1 + x_2 + \ldots\ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$= (9 + 5 + 2 + 0 + (-3) + (-1)) / 6$$

$$= 12 \div 6 = 2$$

**ثانيا :- طريقة الانحرافات البسيطة:**

**وتهدف هذه الطريقة الى إختصار العمليات الحسابية بعض الشئ وبصفة خاصة حينما يكون عدد المفردات كبير نسبيا مما يؤدى الى تقليل واضح فى الجهد الحسابى المبذول وفى احتمالات الخطأ خاصة حينما لاتتوافر الالات الحاسبة لإستخدامها فى هذا المجال . وتتلخص هذه اطريقة فى إختيار أى قيمة من مفردات الظاهرة (ويفضل القيمة التى تتكرر أكثر من غيرها فى المفردات) واعتبارها بمثابة وسطا فرضيا يتم طرحه من كافة مفردات الظاهرة . ثم إيجاد الوسط الحسابى للإنحرافات البسيطة الناتجة والتى يمكن منها حساب الوسط الحسابى للمفردات الاصلية . وهذه الطريقة تبنى على اساس أن من اهم خصائص الوسط الحسابى أنه يتأثر بالعمليات الجبرية الاربعة . ومن ثم فأن خطوات هذه الطريقة تتلخص فيما يلى :**

- **بأفتراض أن لديك مجموعة القيم** $x_1 , x_2 , \ldots , x_n$
- **يتم أعتبار غحدى القيم بمثابة وسط فرضى ولتكن قيمته (A) .**
  **ويفضل اختيار تلك القيمة التى تتكرر بشكل واضح وتكون قريبة من متوسط مجموعة المفردات محل الدراسة كوسط فرضى . إذ ان هذا سيؤدى الى تبسيط العملية الحسابية بشكل كبير .**
- **يتم طرح قيمة الوسط الفرضى** (A) **من جميع المفردات لنحصل على انحرافات القيم عن هذا الوسط الفرضى ولتكن** $(d_i)$ **حيث** $i=1,2,\ldots,n$
  **أى أن :**

$$d_1 = x_1 - A \quad , d_2 = x_2 - A \; , d_3 = x_3 - A \; ,\ldots\ldots , d_n = x_n - A$$

**او بلغة أخرى فأن :**

$$d_i = x_i - A \qquad \text{for} \qquad i = 1 , 2 , \ldots.. , n$$

**وتسمى قيم ( $d_i$ ) بالانحرافات البسيطة .**

- **يتم إيجاد الوسط الحسابي للانحرافات البسيطة وذلك من خلال قسمة مجموع تلك الانحرافات على عددها أى ان :**

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n}$$

- **فيكون الوسط الحسابى لقيم المتغير (X) هو :**

$$\bar{x} = A + \bar{d} = A + \frac{\sum_{i=1}^{n} d_i}{n} \qquad (٢)$$

**<u>مثال(٣):</u>**

**أحسب الوسط الحسابى للمفردات الواردة بالمثال (١) مستخدما فى ذلك طريقة الانحرافات البسيطة .**

**<u>الحل</u> :**

**بملاحظة المفردات المعطاه 800 , 750 , 700 , 600 , 500 نجد انه لاتوجد مفردة تتكرر أكثر من غيرها لذا دعنا نفترض ان الوسط الفرضى هو القيمة ٧٠٠ اى ان $A = 700$ فيكون لدينا الجدول التالى :**

**جدول (٣-١)**

| $x_i$ | $d_i = x_i - 700$ |
|---|---|
| 500 | 500 – 700 = -200 |
| 600 | 600 – 700 = -100 |
| 700 | 700 – 700 = 0 |
| 750 | 750 – 700 = 500 |
| 800 | 800 - 700 = 100 |
| $\sum$ | -150 |

**وحيث ان :**

$$\bar{X} = A + \frac{\sum d_i}{n}$$

$$= 700 + \frac{(-150)}{5} = 700 - 30 = 670$$

وهو نفس الناتج الذى حصلنا عليه عند إتباع الطريقة الحسابية المطولة .

**ثالثا : طريقة الانحرافات المختصرة (المعدلة) :-**

إمتداد للطريقة السابقة (الانحرافات البسيطة) فإذا كانت قيم الانحرافات البسيطة $(d_i)$ تقبل القسمة على مقدار ثابت وليكن $(B)$ بدون باقى . فإنه يمكن إختصار العمليات الحسابية أكثر عند حساب الوسط الحسابى وذلك على النحو التالى :

- بعد الحصول على الانحرافات البسيطة للمفردات كما سبق فى الطريقة السابقة ولتكن هذه الانحرافات هى $(d_i)$ لكل قيم $i=1,2,\ldots,n$ فإنه يمكن أعتبار عامل مشترك وليكن $(B)$ فيما بين قيم الانحرافات البسيطة وبإعتبار هذا العامل المشترك نحصل على الانحرافات الأكثر إختصارا على النحو المبين التالى :

$$D_1 = \frac{d_1}{B} \ , \quad D_2 = \frac{d_2}{B} \ , \quad D_3 = \frac{d_3}{B} \ , \ \ldots\ldots , \ D_n = \frac{d_n}{B}$$

- لايجاد الوسط الحسابى للقيم الاصلية $x$ اى $\bar{x}$ مستخدمين فى نفس الوقت الانحرافات الاكثر اختصارا (المعدلة) اى $(D_i)$ فإنه يجب رد الحسابات بطريقة عكسية بعد حساب متوسط الانحرافات الاكثر إختصارا أى بعد حساب $\bar{D}$ فإنه يجب ضربه فى مقدار الثابت $(B)$ الذى تم قسمة الانحرافات البسيطة عليه ثم إضافة المقدار الثابت الذى تم طرحه من المفردات أى $(A)$ اى ان :

$$\bar{x} = A + B\bar{D} = A + B\frac{\sum_{i=1}^{n} D_i}{n} \qquad (3)$$

**مثال (٤) :**

**احسب الوسط الحسابى للمفردات الواردة بالمثال(١) بإستخدام الطريقة الاكثر إختصارا (الانحرافات المعدلة) .**

**الحل :**

**لحساب الوسط الحسابى بالطريقة الاكثر إختصارا لاحظ انه فى المثال السابق فأن الانحرافات البسيطة ($d_i$) تقبل القسمة على الثابت (٥٠) . لذا يمكن اتمام عملية حساب الوسط الحسابى من خلال الجدول التالى :**

**جدول (٣-٢)**

| $x_i$ | $d_i = x_i - 700$ | $D_i = d_i \div 50$ |
|---|---|---|
| 500 | -200 | -4 |
| 600 | -100 | -2 |
| 700 | 0 | 0 |
| 750 | 50 | 1 |
| 800 | 100 | 2 |
| $\sum$ | | -3 |

**وحيث أن :**

$$\bar{X} = A + B\left(\frac{\sum D_i}{n}\right)$$

**لذا فإن :**

$$\bar{X} = 700 + 50(-3/5) = 700 - 30 = 670$$

**وهو نفس الناتج الذى حصلنا عليه سواء بالطريقة المطولة أو بإستخدام طريقة الانحرافات البسيطة .**

**وخلاصة ماسبق انه تبين لنا أننا نحصل على نفس القيمة للوسط الحسابى ( $\bar{x}$ ) بغض النظر عن الطريقة المتبعة فى حسابه سواء كانت**

**الطريقة المطولة او طريقة الانحرافات البسيطة او الطريقة الاكثر إختصارا . هذا بالاضافة الى انه اصبح لدينا علما بأن الوسط الحسابى يتأثر حسابه بالعمليات الجبرية الاربعة (الجمع او الطرح أو الضرب أو القسمة) .**

**(٣-١-٢) خصائص الوسط الحسابى :-**

**بصفة عامة هناك مجموعة من الخصائص التى تميز الوسط الحسابى عن غيره من مقاييس المتوسطات نذكر منها مايلى :**

**أولا: مجموع إنحرافات القيم عن وسطها الحسابى يساوى الصفر :**

**اى أن**

$$\sum_{i=1}^{n} ( x_i - \bar{x} ) = 0$$

**ولاثبات صحة هذه الخاصية دعنا نفرض ان المتغير** ($x$) **والذى يأخذ مجموعة القيم** $x_1 , x_2 , .... , x_n$ **متوسطه الحسابى** $\bar{x}$ **هو فتكون إنحرافات القيم** ($x_i$) **عن متوسطها** $\bar{x}$ **هى عبارة عن :**

$$( \mathbf{x_1} - \bar{x} ) , ( \mathbf{x_2} - \bar{x} ) , ( \mathbf{x_3} - \bar{x} ) , ........., (\mathbf{x_n} - \bar{x} )$$

**وبذلك فإن مجموع هذه الانحرافات هو :**

$$\sum_{i=1}^{n} ( x_i - \bar{x} ) = (x_1 - \bar{x} ) + ( x_2 - \bar{x} ) + ........ + (x_n - \bar{x} )$$

$$= (x_1 + x_2 + x_3 + ...... + x_n ) - ( \bar{x} + \bar{x} +...+ \bar{x})$$

**لاحظ ان الحد الثانى عبارة عن مجموع المقدار الثابت** ( $\bar{x}$ ) **عددا** ($n$) **من المرات فيكون مجموعه هو** ( $n\bar{x}$) **. فيكون :**

$$\sum_{i=1}^{n} ( x_i - \bar{x} ) = \sum_{i=1}^{n} x_i - n \bar{x} \qquad (١)$$

**وحيث أن :**

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

**لذا فإن :**

$$n \bar{x} = \sum_{i=1}^{n} x_i$$

**ومن ثم فالبتعويض فى العلاقة (١) فيكون :**

$$\sum_{i=1}^{n} ( x_i - \bar{x}) = n\bar{x} - n\bar{x} = 0$$

**أى أن مجموع إنحرافات القيم متوسطها الحسابى يساوى دائما أبدا صفرا . وهذه الخاصية توضح الدور المركزى الذى يلعبه قيمة الوسط الحسابى للمفردات .**

**<u>ثانيا :- مجموع مربعات إنحرافات القيم عن وسطها الحسابى أصغر مايمكن</u>**

**أى اصغر من مجموع مربعات إنحرافات القيم عن اى وسط فرضى آخر أى أن :**

$$\sum_{i=1}^{n} (x_i-\bar{x})^2 < \sum_{i=1}^{n} (x_i-A)^2$$

**حيث (A) مقدار ثابت أختيارى تختلف قيمته عن الوسط الحسابى $\bar{x}$ ، ولاثبات صحة ذلك فحيث ان :**

$$\sum_{i=1}^{n} (x_i-A) = \sum_{i=1}^{n} \{(x_i-\bar{x}) + (\bar{x}-A)\}^2$$

$$= \sum_{i=1}^{n} (x_i-\bar{x})^2 + 2\sum_{i=1}^{n} (x_i-\bar{x})(\bar{x}-A) + \sum_{i=1}^{n} (\bar{x}-A)^2$$

**وذلك لأن $\bar{x}$ A مقادير ثابتة .**

**وحيث ان مجموع إنحرافات القيم عن وسطها الحسابى صفرا أى ان :**

**$\sum (x_i-x)=0$ لذا فإن الحد الاوسط تتلاشى قيمته ويصبح لدينا :**

$$\sum_{i=1}^{n} ( x_i - A)^2 = \sum_{i=1}^{n} ( \bar{x}_i - x)^2 + n ( \bar{x} - A)^2$$

**وحيث أن (n) تمثل عدد المفردات فهى عبارة عن مقدار موجب . كما أن المقدار $(\bar{x}A)^2$ ) مقدار مربعا فهو مقدار موجب أيضا . وعليه فإن :**

$$\sum_{i=1}^{n} (x_I - A)^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 + \text{Positive Number}$$

أى أن :

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 < \sum_{i=1}^{n} (x_i - A)^2 \qquad (5)$$

**أى أن مجموع مربعات إنحرافات القيم عن وسطها الحسابى أقل مايمكن .**

**<u>مثال (٥):-</u>**

**إذا كان لديك مجموعة المفردات التالية : 3 , 15 , 12 , 8 , 7**

**فالمطلوب :**

**١- أوجد مجموع إنحرافات القيم عن وسطها الحسابى ، وعلق على نتيجتك**

**٢ –اوجد :**

$$\sum_{i=1}^{5} (x_i-10)^2 \quad , \sum_{i=1}^{5} (x_i-7)^2 \quad , \sum_{i=1}^{5} (x_i-\bar{x})^2$$

**وعلق على نتائجك .**

**<u>الحل :</u>**

**١- لايجاد مجموع انحرافات القيم عن وسطها الحسابى دعنا نقوم بحساب الوسط الحسابى للمفردات بأى من الطرق الثلاث التى سبق الاشارة اليها وليكن بالطريقة المطولة :**

**فحيث أن:**

$$\bar{X} = \frac{\sum_{i=1}^{n} xi}{n} = \frac{7+8+12+15+3}{5} = \frac{45}{5} = 9$$

**والجدول التالى يبين الحسابات اللازمة :**

جدول (٣-٣)

| $X_i$ | $x_i-9$ | $(x_i-9)^2$ | $(x_i-7)$ | $(x_i-7)^2$ | $(x_i-10)$ | $(x_i-10)^2$ |
|---|---|---|---|---|---|---|
| 7 | -2 | 4 | 0 | ٠ | -3 | 9 |
| 8 | -1 | 1 | 1 | 1 | 2 | 4 |
| 12 | 3 | 9 | 5 | 25 | 2 | 4 |
| 15 | 6 | 36 | 8 | 64 | 5 | 25 |
| 3 | -6 | 36 | -4 | 16 | -7 | 49 |
| 45 | 0 | 86 | 10 | 106 | -5 | 91 |

$\sum$

**فمن خلال العمود الثانى فى الجدول (٣-٣) يتضح لنا ان مجموع انحرافات القيم عن وسطها الحسابى مساويا للصفر أى ان :**

$$\sum_{i=1}^{n} ( x_i - \bar{x} ) = \sum_{i=1} ( x_i - 9 ) = 0$$

**وهو مايتفق مع الخاصية الاولى من خصائص الوسط الحسابى .**

**٢- من خلال الاعمدة الموضحة فى جدول (٣-٣) يتضح لنا ان :**

$$\sum_{i=1}^{5} ( x_i - \bar{x})^2 = 86 \quad , \quad \sum_{i=1}^{5} ( x_i - 7 )^2 = 106 \, , \, \sum_{i=1}^{5} (x_i-10)^2 = 91$$

**وهو مايفيد أن مجموع مربعات إنحرافات القيم عن وسطها الحسابى أقل مايمكن . وهو مايتفق مع الخاصية الثانية من خصائص اى وسط فرضى اخر .**

**(٣-١-٢) حساب الوسط الحسابى للبيانات المبوبة تكراريا :-**

**سبق وأن ذكرنا فى التمثيل الجدولى للبيانات انه لاختصارا البيانات التى تنحصر فى مدى واسع معين فإنه يتم تقسيم المدى الى مجموعة من الفئات ويتم تسجيل عدد المفردات فى كل فئة فنحصل على مايسمى بالجدول التكرارى .**

**فلاشك انه لحساب الوسط الحسابى من مثل هذه الجداول التكرارية سيكون أسهل وأسرع كثيرا من حسابه من خلال جمع هذا العدد الضخم من**

البيانات. هذا ولحساب الوسط الحسابى للبيانات المبوبة (الجداول التكرارية) فأننا سوف نستخدم الطرق الثلاث السابقة فى حالة البيانات الغير مبوبة (المفردات) وذلك مع إجراء التعديلات اللازمة حتى تلائم هذه الطرق طبيعة عرض البيانات المبوبة فى جداول تكرارية سواء كانت هذه البيانات تمثل متغيرات كمية منفصلة أم متصلة .

**أولا : حساب الوسط الحسابى لجدول تكرارى لمتغير كمى منفصل (متقطع):**

فى حالة وجود جدول تكرارى لظاهرة كمية منفصلة (مثل حجم الاسرة) أو عدد العاملين .....الخ فأن كل قيمة من قيم الظاهرة المعبرة عن هذا المتغير المنفصل تتكرر عددا من المرات . ولذلك فإنه لإيجاد مجموع كافة هذه القيم وهو مايتطلب حساب الوسط الحسابى والذى يساوى مجموع القيم منسوبا الى عددها . فإنه يلزم ضرب كل قيمة فى عدد مرات تكرارها ثم جمع حواصل الضرب ، بعدها يمكن حساب قيمة الوسط الحسابى بقسمة مجموع هذه القيم على عددها . فإذا كانت أقسام الظاهرة التى تظهر فى الجدول التكرارى للمتغير المنفصل هى $x_1 , x_2 , \ldots , x_2$ تكرارات مقابلة لها فى هذا الجدول ولتكن $F_1 , F_2 , \ldots , F_n$ على الترتيب . فأن الوسط الحسابى سيكون عبارة عن :

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i F_i}{\sum_{i=1}^{n} F_i} \qquad (٦)$$

وهنا يمكن ملاحظة أن القانون الوارد بالعلاقة (٦) وهو الخاص بحساب الوسط الحسابى لبيانات متغير منفصل مبوبا فى جدول تكرارى هو نفسه القانون الوارد بالعلاقة (١) الخاص بحساب الوسط الحسابى للبيانات الغير مبوبة (مفردات) ولكن بعد إجراء التعديل اللازم لكى يتلائم مع إختلاف طريقة عرض البيانات . وهذا التعديل يتمثل فى إستبدال (x) فى العلاقة (١)

بـ $(xF)$ فى العلاقة (٦) . وكذلك استبدال $(n)$ فى العلاقـة (١) بـ $(\sum F)$ فـى العلاقـة (٦) أى ان التكرارت تؤخذ فـى الاعتبـار وذلـك مـن خـلال ضـرب كـل قيمة من قيم المتغير المنفصل $(x_i)$ فى عدد مرات تكرارها أى $(F_i)$ .
هذا مايتعلق بطريقـة حساب الوسط الحسابى لبيانـات متغيـر منفصل مبوبـة تكراريـا . وأمـا فـى حالـة كـل مـن طريقـة الانحرافـات البسـيطة و الانحرافـات المعدلة فأنه يجرى أيضا نفس طبيعة التعديل على القوانين الواردة بالعلاقتين (٢)،(٣) لتكـون صـيغة حسـاب الوسـط الحسـابى للمتغيـر المنفصـل فـى تلـك الحالـة على الصورة التالية :

$$\bar{X} = A + \frac{\sum_{i=1}^{r} d_i F_i}{\sum_{i=1}^{r} F_i} \qquad (٧)$$

$$\bar{X} = A + B\,\frac{\sum_{i=1}^{r} D_i F_i}{\sum_{i=1}^{r} F_i} \qquad (٨)$$

حيث ( $r$ ) تمثل عدد اقسام  أو فئات المتغير المنفصل فى التوزيـع التكرارى والمثـال التـالى يوضـح كيفيـة حسـاب الوسـط الحسـابى فـى حالـة الجـدول التكرارى لمتغير منفصل .

**<u>مثال (٦):</u>**

فيما يلـى التوزيـع التكرارى التـالى يبين عدد العمـال الغائبين عن العمل فـى إحدى الشركات خلال شهر نوفمبر ٢٠٠٠ .

**جدول (٣-٤)**

| No.of Absence | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | $\sum$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Frequency | 2 | 1 | 4 | 7 | 3 | 6 | 3 | 2 | 2 | 30 |

**والمطلوب حساب متوسط عدد الغائبين يوميا خلال هذا الشهر فى تلك الشركة .**

**<u>الحل :</u>**

**الجدول التالى يوضح الحسابات اللازمة لحساب متوسط عدد العاملين الغائبين يوميا بإستخدام الطرق الثلاث .**

**جدول(٣-٥)**

| No.of Absence $(x_i)$ | $F_i$ | $X_i\ F_i$ | $d_i=x_i-4$ | $d_i\ F_i$ | $D_i=d_i/2$ | $D_i\ F_i$ |
|---|---|---|---|---|---|---|
| 0 | 2 | 0 | -4 | -8 | -2 | -4 |
| 1 | 1 | 1 | -3 | -3 | -1.5 | -1.5 |
| 2 | 4 | 8 | -2 | -8 | -1 | -4 |
| 3 | 7 | 21 | -1 | -7 | -0.5 | -3.5 |
| 4 | 3 | 12 | 0 | 0 | 0 | 0 |
| 5 | 6 | 30 | 1 | 6 | 0.5 | 3 |
| 6 | 3 | 18 | 2 | 6 | 1 | 3 |
| 7 | 2 | 14 | 3 | 6 | 1.5 | 3 |
| 8 | 2 | 16 | 4 | 8 | 2 | 4 |
| $\sum$ | 30 | 120 | | 0 | | 0 |

*** <u>حساب الوسط الحسابى بالطريقة المباشرة (المطولة):</u>**

**فحيث ان :**

$$\bar{X} = \frac{\sum_{i=1}^{r} x_i\ F_i}{\sum_{i=1}^{r} F_i}$$

**لذا فإن :**

$$\bar{X} = \frac{120}{30} = 4 \qquad \text{(Persons)}$$

- **حساب الوسط الحسابى بإستخدام طريقة الانحرافات البسيطة :-**

**حيث أن :**

$$\bar{X} = A + \frac{\sum_{i=1}^{r} d_i F_i}{\sum_{i=1}^{r} F_i}$$

**وفى الجدول (٣–٥) لاحظ أن الوسط الفرضى الذى تم طرحه من كافة قيم المتغير المنفصل هو A=4 (حيث يفضل دائما أبدا اعتبار الوسط الفرضى هو القيمة التى تتوسط الجدول التكرارى) لذا فإن :**

$$\bar{X} = 4 + \left(\frac{0}{30}\right) = 4 \qquad \text{( Persons )}$$

**لاحظ ان المجموع $d_i f_i$ مساويا للصفر لأن الوسط الفرضى المستخدم فى تلك الطريقة وهو قيمة (A) مساوية للوسط الحسابى .لذا فلابد أن يكون مجموع هذا العمود ($d_i f_i$) مساويا للصفر وذلك لأن مجموع انحرافات القيم عن وسطها الحسابى هو صفرا (خاصية (١)) .**

- **حساب الوسط الحسابى بإستخدام طريقة الانحرافات المعدلة :**

**فحيث ان :**

$$\bar{X} = A + B \frac{\sum_{i=1}^{r} D_i F_i}{\sum_{i=1}^{r} F_i}$$

ومن الجدول السابق فأن : B= 2 , A=4

$$\bar{X} = 4 + 2\left(\frac{0}{30}\right) = 4 \qquad (\text{ Persons })$$

لاحظ ان النتيجة واحدة بأى من الطرق الثلاث .

<u>ثانيا : حساب الوسط الحسابى لبيانات متغير متصل (مبوب تكراريا) :-</u>

عند إيجاد الوسط الحسابى فى حالة التوزيعات التكرارية للمتغيرات المتصلة. فإننا نواجه مشكلة إختفاء بعض المعلومات الاصلية عن الظاهرة محل الدراسة وذلك نتيجة عملية تبويب البيانات تكراريا . فمن الجدول التكرارى محل الدراسة لانعلم القيمة الحقيقية لكل وحدة من وحدات تكرار الجدول . وحيث أننا فى حاجة الى معرفة قيمة المتغير محل الدراسة حيث نتمكن من إيجاد مجموع مفرداته فأننا نلجأ الى افتراض منطقى وهو ماتبنى على أساسه فكرة الجداول التكرارية ألا وهو أن التكرارت تتوزع بإنتظام على مدار طول الفئة (وذلك بشكل متجانس الانتشار) . وهو مايعنى أننا سنعتبر كل وحدة من وحدات التكرار من تكرارت أى فئة من فئات الجدول التكرارى ستأخذ قيمة ثابتة معينة وهى عبارة عن مركز تلك الفئة .

وعلى ذلك فإنه لإيجاد مجموع قيم المتغير المتصل المبوبة بياناته تكراريا فأننا نقوم بضرب مركز كل فئة من فئات التوزيع التكرارى فى التكرار المقابل له ثم يتم تجميع حواصل الضرب ويكون الناتج بمثابة مجموع المفردات (فرضا) أما عن عدد القيم (المقام فى قانون الوسط الحسابى) هنا يتمثل فى مجموع تكرارت التوزيع التكرارى للفئات المختلفة . ومن هنا يتضح لنا أن قوانين حساب الوسط الحسابى فى حالة المتغير المتصل هى نفسها قوانين حساب الوسط الحسابى للمتغير المنفصل ولكن مع فارق واحد وهو أنه فى حالة المتغير المتصل فأننا نلجأ لحساب مراكز الفئات وهى تقابل قيم الظاهرة فى حالة المتغير المنفصل .

وخلاصة ماسبق هو أن القوانين المستخدمة فى ايجاد الوسط الحسابى فى حالة المتغيرات المتصلة هى نفس القوانين المستخدمة فى حالة المتغيرات المنفصلة مع إجراء التعديل اللازم فى مفهوم قيم ($x_i$) فقط والتى تعبر هنا عن مراكز الفئات فى التوزيع التكرارى .

وسنتناول الآن كيفية حساب الوسط الحسابى لتوزيع تكرارى سواء فى حالة الجداول المنتظمة (المتساوية من حيث أطوال الفئات) أو الغير منتظمة .

**مثال (٧):**

فيما يلى الجدول التالى يبين توزيع فئات الاجر اليومى (بالجنيه) فى إحدى مصانع مدينة العاشر من رمضان .

جدول (٣-٦)

| classes | 120- | 140- | 160- | 180- | 200- | 220- | 240-260 | ∑ |
|---|---|---|---|---|---|---|---|---|
| Frequency | 2 | 4 | 11 | 38 | 30 | 10 | 5 | 100 |

والمطلوب: حساب متوسط الاجر اليومى للعامل فى هذا المصنع بإستخدام الطرق الثلاثة .

**الحل:**

١- **بإستخدام الطريقة المباشرة (المطولة) :**

الجدول التالى يوضح الحسابات اللازمة لحساب متوسط الاجر اليومى

جدول (٣-٧)

| Classes | $F_i$ | Midpoint ($x_i$) | $X_i F_i$ |
|---|---|---|---|
| 120- | 2 | 130 | 260 |
| 140- | 4 | 150 | 600 |
| 160- | 11 | 170 | 1870 |
| 180- | 38 | 190 | 7220 |
| 200- | 30 | 210 | 6300 |
| 220- | 10 | 230 | 2300 |
| 240-260 | 5 | 250 | 1250 |
| ∑ | 100 | | 19800 |

وحيث أن:

$$\bar{X} = \frac{\sum_{i=1}^{r} x_i F_i}{\sum_{i=1}^{r} F_i}$$

لذا فإن :

$$\bar{X} = \frac{19800}{100} = 198 \quad \text{(L.E)}$$

**٢- بإستخدام طريقة الانحرافات البسيطة :**

**بإعتبار مركز الفئة التى تتوسط (مركز) التوزيع التكرارى بمثابة وسطا فرضيا يتم طرحه من كافة المراكز . أى بطرح المقدار A=190 من كافة مراكز الفئات ($x_i$) للحصول على الانحرافات ($d_i$) . ثم ضرب هذه الانحرافات البسيطة فى التكرارت المقابلة وايجاد المجموع ، والجدول التالى يلخص الخطوات اللازمة لحساب المتوسط الحسابى :**

**جدول (٣-٨)**

| Classes | $F_i$ | $x_i$ | $d_i = x_i - 190$ | $d_i F_i$ |
|---|---|---|---|---|
| 120- | 2 | 130 | -60 | -120 |
| 140- | 4 | 150 | -40 | -160 |
| 160- | 11 | 170 | -20 | -220 |
| 180- | 38 | 190 | 0 | 0 |
| 200- | 30 | 210 | 20 | 600 |
| 220- | 10 | 230 | 40 | 400 |
| 240-260 | 5 | 250 | 60 | 300 |
| $\sum$ | 100 | | | 800 |

**وحيث أن :**

$$\bar{X} = A + \frac{\sum_{i=1}^{r} d_i F_i}{\sum_{i=1}^{r} F_i}$$

**وعليه فإن :**

$$\bar{X} = 190 + \frac{800}{100} = 198 \qquad (L.E)$$

**هى نفس الناتج بإستخدام الطريقة المباشرة .**

**٣- <u>بإستخدام طريقة الانحرافات المختصرة :-</u>**

**لاحظ فى الجدول السابق ان عمود الانحرافات البسيطة ($d_i$) يقبل القسمة على (٢٠) وهى بمثابة طول الفئة فى هذا الجدول المنتظم .لذا يمكن إعتبار طول الفئة بمثابة وسطا فرضيا آخر وحساب الانحرافات المعدلة ($D_i$) وذلك من خلال قسمة عمود الانحرافات البسيطة على (٢٠) والجدول التالى يوضح الحسابات اللازمة لايجاد متوسط الاجر اليومى .**

**جدول (٣-٩)**

| Classes | $F_i$ | $X_i$ | $d_i=x_i-190$ | $D_i = d_i \div 20$ | $D_i F_i$ |
|---|---|---|---|---|---|
| 120- | 2 | 130 | -60 | -3 | -6 |
| 140- | 4 | 150 | -40 | -2 | -8 |
| 160- | 11 | 170 | -20 | -1 | -11 |
| 180- | 38 | 190 | 0 | 0 | 0 |
| 200- | 30 | 210 | 20 | 1 | 30 |
| 220 | 10 | 230 | 40 | 2 | 20 |
| 240-260 | 5 | 250 | 60 | 3 | 15 |
| $\sum$ | 100 | | | | 40 |

**وحيث أن :**

$$\bar{X} = A + B\left(\frac{\sum_{i=1}^{r} D_i F_i}{\sum_{i=1}^{r} F_i}\right)$$

**حيث أن :** A = 190 , B=20

**لذا فإن :**

$$\bar{X} = 190+20\left(\frac{40}{100}\right) = 190+8 = 198 \quad (L.E)$$

وهى نفس النتائج المستخلصة من الطريقتين السابقتين لكن أهم مايميز هذه الطريقة هى تبسيط أرقام الجدول لأقصى درجة ممكنة لدرجة عدم الحاجة الى الآلآت الحاسبة فى حساب مفردات ومجاميع الجدول اللازم للحساب .

**(٣-١-٣) :- حساب الوسط الحسابى للتوزيع التكرارى النسبى :-**

**فى حالة توافر توزيعات تكرارية نسبية فإنه يمكن إستخدام القوانين السابقة لايجاد الوسط الحسابى لتلك التوزيعات التكرارية النسبية وذلك بعد إجراء التعديل اللازم على صيغ تلك القوانين وذلك على النحو التالى :–**

- **فالبنسبة للقانون المستخدم لحساب الوسط الحسابى بإستخدام الطريقة المباشرة (المطولة) فمن العلاقة (٦) السابق الاشارة اليها فحيث ان :**

$$\bar{X} = \frac{\sum_{i=1}^{r} xi\ Fi}{\sum_{i=1}^{r} Fi} = \sum_{i=1}^{n} \bar{x}_i \left(\frac{Fi}{\sum_{i=1}^{r} Fi}\right)$$

**وحيث أن التكرار النسبى للفئة ( i ) ماهو إلا :**

$$R_i = \frac{F_i}{\sum_{i=1}^{r} F_i}$$

**حيث (r) تمثل عدد فئات التوزيع . لذا فإن:**

$$\bar{x} = \sum_{i=1}^{r} x_i R_i \qquad (٩)$$

اى أن الصيغة الرياضية لايجاد الوسط الحسابى بإستخدام الطريقة المباشرة ماهى إلا عبارة عن مجموع حواصل ضرب مراكز الفئات فى التكرارت النسبية المقابلة لتلك الفئات .

- كذلك وبنفس الاسلوب تصبح صيغة ايجاد الوسط الحسابى بإستخدام طريقة الانحرافات البسيطة وفى حالة توافر تكرارت نسبية على الصورة التالية :

$$\overline{x} = A + \sum_{i=1}^{r} d_i R_i \qquad (10)$$

حيث $(d_i)$ تمثل إنحرافات مراكز فئات التوزيع عن وسط فرضى معين وهو المقدار $(A)$ ويفضل أن يكون هذا المقدار بمثابة مركز الفئة التى تتوسط التوزيع التكرارى أو مركز الفئة التى تقابل أكبر تكرار فى الجدول التكرارى.

- وأخيرا فأن الصيغة الرياضية لحساب الوسط الحسابى بإستخدام طريقة الانحرافات المعدلة (الطريقة الاكثر إختصارا ) فى حالة توافر تكرارت نسبية تصبح على الصورة التالية :

$$\overline{X} = A + B\sum_{i=1}^{r} D_i \quad R_i \qquad (١١)$$

حيث $(D_i)$ تمثل الانحراف المختصر للفئة $(\mathbf{i})$ كما سبق تعريفه .

والمثال التالى يوضح طريقة حساب الوسط الحسابى بإستخدام الطرق الثلاث فى حالة التوزيعات التكرارية النسبية .

<u>مثال(٨):-</u>

بإستخدام الطرق الثلاث لحساب الوسط الحسابى ماهى متوسط قيمة الاجر للتوزيع التكرارى التالى لمجموعة من العاملين فى إحدى الشركات (بالجنيه) :

**جدول (٣-١٠)**

**التوزيع التكرارى النمسبى للأجر النسبى فى إحدى الشركات**

| Classes | 15- | 25- | 35- | 45- | 55-65 |
|---|---|---|---|---|---|
| Relative Fre.($R_i$) | 0.2 | 0.32 | 0.24 | 0.16 | 0.08 |

**الحل :**

**١- ايجاد الوسط الحسابى للدخل اليومى بإستخدام الطريقة المباشرة (المطولة) . والجدول التالى يبين الحسابات اللازمة لايجاد متوسط الاجر اليومى للعامل فى تلك الشركة :**

**جدول (٣-١١)**

| Classes | $R_i$ | $x_i$ | $X_i$ $R_i$ |
|---|---|---|---|
| 15- | 0.2 | 20 | 4.0 |
| 25- | 0.32 | 30 | 9.6 |
| 35- | 0.24 | 40 | 9.6 |
| 45- | 0.16 | 50 | 8.0 |
| 55- | 0.08 | 60 | 4.8 |
| $\sum$ | 1 | | 36 |

**وحيث أن :**

$$\bar{x} = \sum_{i=1}^{r} x_i R_i$$

**لذا فإن :**

$$\bar{X} = 36 \qquad \text{(L.E)}$$

**٢- ايجاد الوسط الحسابى بطريقة الانحرافات البسيطة :-**

**والجدول (٣-١٢) يوضح الحسابات اللازمة لايجاد متوسط الاجر اليومى فى تلك الشركة .**

**جدول(٣-١٢)**

| classes | $R_i$ | $X_i$ | $d_i = x_i - 40$ | $d_i R_i$ |
|---|---|---|---|---|
| 15- | 0.2 | 20 | -20 | -4 |
| 25- | 0.32 | 30 | -10 | -3.2 |
| 35- | 0.24 | 40 | 0 | 0 |
| 45- | 0.16 | 50 | 10 | 1.6 |
| 55-65 | 0.08 | 60 | 20 | 1.6 |
| $\sum$ | 1 | | | -4 |

**وحيث أن :**

$$\bar{x} = A + \sum_{i=1}^{r} d_i R_i$$

**وحيث ان قيمة A=40 (من الجدول) فيكون :**

$$\bar{X} = 40 + (-4) = 36 \qquad \text{(L.E)}$$

**وهى نفس النتيجة بإستخدام الطريقة المباشرة .**

**٣- <u>ايجاد الوسط الحسابى بطريقة الانحرافات المختصرة (المعدلة) :</u>**

**والجدول (٣-١٣) يوضح الحسابات اللازمة لايجاد متوسط الاجر اليومى فى تلك الشركة باستخدام طريقة الانحرافات المعدلة :**

**جدول(٣-١٣)**

| Classes | $R_i$ | $X_i$ | $d_i = x_i - 40$ | $D_i = d_i \div 10$ | $D_i R_i$ |
|---|---|---|---|---|---|
| 15- | 0.2 | 20 | -20 | -2 | -0.4 |
| 25- | 0.32 | 30 | -10 | -1 | -0.32 |
| 35- | 0.24 | 40 | 0 | 0 | 0 |
| 45- | 0.16 | 50 | 10 | 1 | 0.16 |
| 55-65 | 0.08 | 60 | 20 | 2 | 0.16 |
| $\sum$ | 1 | | | | -0.4 |

**وحيث ان :**

$$\bar{x} = A + B \sum_{i=1}^{r} D_i R_i$$

وحيث ان: $A=40$ ، $B=10$ (من الجدول). فيكون :

$$\bar{X} = 40 + 10\,(-0.4) = 40 - 4 = 36 \qquad (L.E)$$

وهى نفس النتيجة التى تم التوصل اليها بإستخدام كلا من الطريقة المباشرة وطريقة الانحرافات البسيطة .

**(٣-١-٤) : الوسط الحسابى المرجح (المتوسط العام أو متوسط المتوسطات)**

Weighted Mean:

عند ايجاد الوسط الحسابى فقد تم التعامل مع جميع المفردات الداخلة فى تكوينه معاملة واحدة . أى أننا أعتبرناها جميعا لها نفس الاهمية .

ويعد ذلك غير صحيحا فى غالبية الاحوال. فكثيرا ماتتفاوت القيم فى أهميتها النسبية . لذا فإنه عند ايجاد الوسط الحسابى فى هذه الحالة يجب ان يتم ترجيح كل قيمة او مفردة من مفردات المجموعة حسب أهميتها بالنسبة لغيرها من باقى المفردات . ويعتبر مقدار الاهمية النسبية لكل مفردة بمثابة وزنا أو ترجيحا ، والمتوسط الناتج فى تلك الحالة يسمى بالوسط الحسابى المرجح . فأذا اهملنا أوزان الترجيح فى تلك الحالة فستكون قيمة الوسط الحسابى مضللة الى حد كبير وبعيدة عن الواقع .

وخلاصة ماسبق هو أنه إذا كانت قيم المتغير (x) ولتكن $x_1,x_2,x_3,\ldots\ldots,x_m$ وكانت أوزانها الترجيحية المقابلة هى $n_1,n_2,n_3,\ldots\ldots,n_m$ فأن الوسط الحسابى المرجح يكون عبارة عن :

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \ldots\ldots\ldots\ldots + n_mx_m}{n_1 + n_2 + \ldots\ldots\ldots\ldots\ldots\ldots + n_m} = \frac{\sum_{i=1}^{m} n_i x_i}{\sum_{i=1}^{m} n_i}$$

**كذلك إذا كان لدينا مجتمعا محل دراسة وليكن حجمه (N) سحبت منه مجموعة من العينات وليكن عددها (m) عينة ذات الحجم $n_1, n_2, \ldots, n_m$ وكانت اوساطها الحسابية هى**

$\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_m$ **على الترتيب فإن . المتوسط العام (او المرجح) ماهو إلا :**

$$\bar{X} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \ldots\ldots\ldots\ldots + n_m\bar{x}_m}{n_1+n_2+\ldots\ldots\ldots\ldots\ldots\ldots\ldots+n_m} = \frac{\sum_{i=1}^{m} n_i\, \bar{x}_i}{\sum_{i=1}^{m} n_i} \qquad (12)$$

**لاحظ أن كل حد من حدود البسط فى العلاقة (١٢) ماهو إلا مجموع مفردات العينات المسحوبة من هذا المجتمع .أما المقام والذى يمثل مجموع مفردات العينات المسحوبة من هذا المجتمع . وفى حالة تساوى أحجام تلك العينات فأنه يمكن اعتبار ($n_i$) مقدار ثابت أى يكون :**

$$\mathbf{n_1=n_2=n_3=\ldots\ldots\ldots\ldots=n_m=n}$$

**فتصبح صيغة العلاقة (١٢) تأخذ الصورة التالية :**

$$\bar{X} = \frac{n(\bar{x}_1+\bar{x}_2+\ldots\ldots\ldots+\bar{x}_m)}{nm} = \frac{\bar{x}_1+\bar{x}_2+\ldots\ldots+\bar{x}_m}{m}$$

$$= \frac{\sum_{i=1}^{m} \bar{x}_i}{m} \qquad (١٣)$$

**أى يصبح المتوسط العام عبارة عن مجموع الاوساط الحسابية لكل العينات مقسوما على عددها . ومن ثم تجدر الاشارة الى ان الوسط الحسابى بمفهومه البسيط ماهو إلا وسطا حسابيا مرجحا بأوزان متساوية .**

**<u>مثال (٩):</u>**

**مصنع به ثلاثة آلات انتاجية :-**

- **تنتج الآله الاولى ٥٠ وحدة يوميا وتباع الوحدة منها بسعر ٢٠جنيه .**
- **وتنتج الآله الثانية ٤٠ وحدة يوميا وتباع الوحدة منها بسعر ٣٠جنيه.**
- **بينما تنتج الآله الثالثة ١٠وحدات يوميا وتباع منها الوحدة بسعر ٤٠جنيها .**

**والمطلوب حساب متوسط سعر بيع الوحدة من إنتاج هذا المصنع عموما .**

**<u>الحل:-</u>**

**لاحظ أنه إذا استخدمنا صيغة الوسط الحسابى العادى للسعر دون أعتبار للكمية التى تنتجها الآله فأن متوسط سعر الوحدة يكون عبارة عن مجموع الاسعار الثلاثة على عددها (٣) أى ان :**

$$\bar{X} = \frac{20 + 30 + 40}{3} = 30 \qquad (L.E)$$

**وهى نتيجة مضللة لاتعكس حقيقة مستوى السعر الحقيقى لمنتجات الآلآت المختلفة الثلاث .**

**اما إذا استخدمنا الكميات المنتجة كأوزان للترجيح فيكون متوسط سعر بيع الوحدة المنتجة للألات الثلاث هى :**

$$\bar{X} = \frac{\sum_{i=1}^{m} n_i x_i}{\sum_{i=1}^{m} n_i} = \frac{20(50) + 30(40) + 40(10)}{50 + 40 + 10}$$

$$= \frac{2600}{100} = 26 \quad (L.E)$$

**ملحوظـة : مـن عيـوب الوسـط الحسـابى المـرجح عمليـة تحديـد الاوزان الترجيحيـة . ففـى غالبيـة الاحيـان تخضـع الاوزان الترجيحيـة لعمليـة التقدير الشخصى مما يؤدى الى اختلاف النتائج من شخص الى أخر .**

**مثال (١٠):**

**يـتم تـدريس مبـادى الاحصـاء الوصـفى لـثلاث شـعب فـى كليـة مـا . وأجـرى أختبار لطلاب تلك الشعب الثلاث فى هذا المقرر فكانـت النتـائج علـى النحـو التالى :**

| Section | Account | Management | Economic |
|---|---|---|---|
| No.of students | 32 | 20 | 40 |
| Mean of test degree | 70 | ٨٥ | 68 |

**والمطلوب حساب متوسط درجة الطالب فى هذا المقرر فى الشعب الثلاث .**

**الحل:-**

**المعطى لدينا المعلومات التالية :**

$n_1 = 32$ , $n_2 = 20$ , $n_3 = 40$

$\bar{x}1=70$, $\bar{x}2=85$ , $\bar{x}3= 68$

**فيكون المتوسط العام لدرجة الطالب لكافة الشعب عبارة عن :**

$$\bar{X} = \frac{\sum_{i=1}^{3} ni\ \bar{x}i}{\sum_{i=1}^{3} ni} = \frac{30(70)+20(85)+40(68)}{32 + 20 + 40} = 72.39 \text{ (degree)}$$

**ملحوظة :**

**لاحظ أن متوسط الدرجات فى حالة اهمال عدد المفردات فى كل شعبة هو**

$$\bar{X} = \frac{70+85+68}{3} = 74.33 \qquad (\text{degree})$$

**وهذه النتيجة غير صحيحة إحصائيا وتعتبر نتيجة مضللة لاتأخذ كافة البيانات المتاحة حول الظاهرة موضع الدراسة فى الاعتبار .**

**<u>ملاحظات حول الوسط الحسابى (المزيا والعيوب):-</u>**

**١- قيمة الوسط الحسابى لاتتغير مهما اختلفت طريقة حسابه وأيا كان الوسط الفرضى الذى يستخدم للتسهيل فى حسابه . هذا ويفضل إستخدام طريقة الانحرافات المختصرة (المعدلة) لحساب الوسط الحسابى فى حالة الجداول المنتظمة او إذا كانت فئات الجدول غير منتظمة لكن بشرط ان تقبل جميع الانحرافات البسيطة القسمة على مقدار ثابت دون تقريب او دون بواقى . أما طريقة الانحرافات البسيطة فيفضل استخدامها فى حالة الجداول غير المنتظمة . أما الطريقة المباشرة (المطولة) فتستخدم غالبا حينما يكون حجم المفردات صغيرا لأنها تحتاج كما رأينا الى عمليات حسابية مطولة واستخدام آلات الحاسبة نظرا لكبر حجم الارقام بإستخدام تلك الطريقة.**

**٢- أهم مايميز الوسط الحسابى كمقياس من مقاييس النزعة المركزية هو أنه يأخذ فى الاعتبار جميع المفردات عند حسابه . لذلك يمكن استخدامه فى عمليات احصائية أخرى ، كما أن قيمته وحيدة بمعنى أنه لأى مجموعة من البيانات لابد ان يوجد لها وسط حسابى وحيد (غير متعدد القيم) وهو ما يجعله أهم مقاييس النزعة المركزية ولذلك فإنه الاكثر استخداما فى التطبيقات العملية لأساليب التحليل الاحصائي .**

**٣- أما مايعاب على الوسط الحسابى أنه لايفضل حسابه فى حالة الجداول التكرارية المفتوحة سواء من أعلى أو من أسفل أو مفتوحة الطرفين . وذلك لأن حسابه فى الجداول المفتوحة سوف يبنى على افتراضات من أجلها يمكن غلق الجدول المفتوح وهو شرط لازم حتى يمكن تحديد**

مراكز الفئات والتى تعتبر اساسا فى حسابه . وهناك من يرى أنه فى ظل الحالات التى يتعذر فيها تحديد حد الفئة المفتوحة (سواء الحد الادنى للفئة الاولى أو الحد الاعلى للفئة الاخيرة) لايكون أمامنا سوى إهمال هذه الفئة وبصفة خاصة إذا كان تكرارها صغير جدا نسبيا بالنسبة لمجموع تكرارت التوزيع التكرارى وربما من الافضل عدم استخدام الوسط الحسابى فى مثل هذه الحالات ويتم استخدام مقياس أخر من مقاييس النزعة المركزية كالوسيط أو المنوال حيث لايتطلب أى منهم معرفة مراكز الفئات المتطرفة (الاولى أو الاخيرة) كما سنرى عند عرضنا لبقية مقاييس النزعة المركزية فى هذا الباب . أما إذا كانت هناك ضرورة ملحة لحساب الوسط الحسابى فإن هناك ثلاثة بدائل يمكن الأخذ بأحدها . وهذه البدائل هى :

أ- إستبعاد الفئات المفتوحة. بتكرارها وحساب وسطها الحسابى لبقية فئات التوزيع. وإن كان هذا البديل يؤدى الى إهمال جزء من المعلومات الخاصة بالظاهرة محل الدراسة خاصة إذا كانت هذه التكرارات كبيرة نسبيا .

ب- إقفال الفئات المفتوحة . اى إفتراض قيمة للحد الادنى للفئة المفتوحة (هذا إذا كان الجدول مفتوحا من أعلى) أو افتراض قيمة للحد الاعلى للفئة المفتوحة (إذا كان الجدول مفتوحا من اسفل) . وفى الواقع فإن إقفال مثل هذه الفئات المفتوحة يتوقف الى حد كبير على التقدير الشخصى . إذ يتطلب خبرة وإلماما بطبيعة الظاهرة موضع الدراسة .

ج- فى حالة التوزيعات التكرارية القريبة من التماثل فإنه يمكن ايجاد قيمة الوسط الحسابى بدلالة الوسيط والمنوال وذلك بموجب العلاقة التالية (لبيرسون) التى تربط بينهما :

$$\text{الوسط الحسابى} = \frac{\text{٣× الوسيط – المنوال}}{\text{٢}}$$

**وسوف نتعرض لتلك النقطة بالتفصيل فى نهاية هذا الباب وذلك عند الحديث عن العلاقة بين مقاييس المتوسطات الثلاثة (الوسط والوسيط والمنوال) وإستخدامها فى الحكم على مدى تماثل التوزيع من عدمه .كما يؤخذ على الوسط الحسابى أنه لايمكن حسابه بيانيا بدقة إلا اذا كان التوزيع متماثلا تماما كما سنرى فيما بعد حيث يكون الوسط الحسابى متساويا مع كل من الوسيط والمنوال وتكون قيمته ماهى إلا نقطة تلاقى العمود النازل من قمة منحنى التوزيع على المحور الافقى وهو ماسنراه فيما بعد . كما أن من أهم عيوب الوسط الحسابى هو انه يتأثر بدرجة كبيرة جدا بالقيم الشاذة أو المتطرفة من البيانات . حيث يتحيز الوسط الحسابى لها ويقترب منها ومبتعدًا بذلك عن وسط المجموعة الحقيقى لدرجة أنه يصبح فى بعض الاحيان غير ممثلا للبيانات محل الدراسة ولذلك لاينصح بإستخدامه فى حالة وجود القيم المتطرفة ، او فى ظل وجود منحنيات أو توزيعات تكرارية شديدة الالتواء . فعلى سبيل المثال إذا افترضنا ان لدينا مرتبات خمسة موظفين هى ١٨٥,٢٠٥,١٩٥,١٨٠,٨٠ جنيها فإن الوسط الحسابى لمرتباتهم هو ١٦٩ جنيها وهو يقل عن مرتب أربعة موظفين من بين الخمسة (أى ٨٠ % منهم) وهذا يعنى أن الوسط الحسابى لايمثل قيم الظاهرة تمثيلا صحيحا من حيث كونه قيمة متوسطه (لاحظ أن القيمة المتطرفة فى هذا المثال هى القيمة ٨٠ جنيه).**

### (٣-٢) <u>الوسيط (Q2)</u> Median

**رأينا حالا عند دراستنا للوسط الحسابى كأحد مقاييس النزعة المركزية انه إذا كانت هناك إحدى القيم الشاذة أو المتطرفة (سواء كبيرة جدا أم صغيرة جدا)**

مقارنة مع باقى المفردات للظاهرة محل الدراسة فأن الوسط الحسابى يتأثر بشدة بهذه القيم الشاذة أو المتطرفة لأن تلك القيمة تؤدى الى ميل وإقصاء الوسط الحسابى تجاهها مبتعدا بذلك عن الوسط الحقيقى للمجموعة مما يؤثر على جودة تمثيله لها كما أن هناك بعض البيانات والتى قد تكون فى صورة وصفية أو ترتيبية ولايوجد معنى لكلمة متوسط كما عرفناها فى مثل هذه البيانات . فمثلا إذا كان لدينا تقديرات مجموعة من الطلاب فليس هناك معنى لكلمة متوسط التقدير فى تلك الحالة . فيأتى الوسيط هنا كأحد مقاييس النزعة المركزية حيث يمكن حسابه لكل أنواع البيانات وصفية أو كمية وحتى منها الترتيبية . كما أنه لايتاثر بوجود القيم الشاذة أو المتطرفة او أن كان هناك أثرا فيكون أثرا طفيفا .

هذا ويعرف الوسيط بأنه القيمة التى تتوسط البيانات وذلك بعد ترتيبها تنازليا أو تصاعديًا . أو بمعنى أخر فالوسيط هو عبارة عن قيمة المفردة التى تتوسط التوزيع بحيث يكون عدد المفردات الاقل منها فى القيمة مساويا لعدد المفردات الاكبر منها فى القيمة . وبمعنى ثالث فالوسيط هو القيمة التى يقل عنها أو يساويها ٥٠% من المفردات ويزيد عنها أو يساويها ٥٠% من المفردات وسوف نستعرض فيما يلى طرق حساب الوسيط فى حالة كل من البيانات الغير مبوبة والمبوبة تكراريا .

**(٣-٢-١) : <u>إيجاد الوسيط من البيانات الغير مبوبة (المفردات) :</u>**

لحساب الوسيط من البيانات الغير مبوبة يتم الآتى :

- ترتيب البيانات ترتيبا تصاعديا أو تنازليا .
- تحديد رتبة وقيمة الوسيط . وهنا نكون أمام أحد الاحتمالين التاليين :

١) إذا كان عدد المفردات (n) فرديا . فإن قيمة الوسيط هى عبارة عن قيمة المفردات ذات الرتبة $((n+1) / 2)$

**٢) أما اذا كان عدد المفردات (n) زوجيا فيوجد مفردتان تتوسطان مفردات الظاهرة محل الدراسة . الاولى رتبتها ( n / 2 ) والثانية التي تليها ورتبتها هى ((n/2)+1) ويكون الوسيط هو عبارة عن الوسط الحسابى لهاتين المفردتين . وفى حالة الظواهر الوصفية كالتقديرات مثلا فإذا كانت أحد المفردتين مثلا جيد والاخرى مقبول فنقول اصطلاحا بأن الوسيط لتك المجموعة هو مابين الجيد والمقبول .**

**<u>مثال (١١):</u>**

**احسب الوسيط فى الحلات الثلاثة التالية :**

**أ) لمجموعة المفردات : ٢٠,٧,١٥,١٠,٩,١٨,٥**

**ب) لمجموعة المفردات : ٢٢,٢٧,٨,١٢,٣٠,٤,١٨,٢٥**

**جـ) لتقديرات الطلاب فى أحد الاختبارات : جيد – ضعيف – جيد جدا – ضعيف جدا – ممتاز – مقبول – جيد – جيد جدا– جيد .**

**<u>الحل :</u>**

**أ) لايجاد الوسيط للمفردات السبعة المعطاه يتم ترتيب المفردات تصاعديا أو تنازليا وليكن الترتيب تصاعدى على الصورة :**

5 , 7 , 8 , 9 , 10 , 15 , 18 , 20

**وحيث أن عدد المفردلات (n=7) وهو عددا فرديا لذا فإن ترتيب الوسيط هو:**

$$\frac{n+1}{2} = \frac{7+1}{2} = 4$$

**ومن ثم فإن قيمة الوسيط هى المفردة الرابعة اى أن الوسيط هو القيمة عشرة . أى ان** $Q_2 = 10$

**ب) بعد ترتيب المفردات تصاعديا تصبح كما يلى :**

4 , 8 , 12 , 18 , 22 , 25 , 27 , 30

وحيث ان عدد المفردات ($n = 8$) عدد زوجيا لذا فإن رتبتى المفردتين اللتين تتوسطان المفردات هما :

$$\frac{n}{2} = \frac{8}{2} = 4 \quad , \quad \frac{n}{2} + 1 = 4 + 1 = 5$$

فيكون الوسيط هو عبارة عن الوسط الحسابى للمفردتين الرابعة والخامسة أى أن :

$$Q_2 = \frac{18+22}{2} = 20$$

جـ) لتحديد وسيط مجموعة التقديرات المعطاه يتم بداية ترتيب هذه المجموعة من التقديرات تصاعديا أو تنازليا . فيكون ترتيب التقديرات تصاعديًا هو :- ضعيف جدا – ضعيف– مقبول – جيد– جيد – جيد – جيد جدا – جيد جدا – ممتاز.

وحيث أن عدد المفردات هو $n=9$ لذا فإن ترتيب الوسيط هو( $(n+1)/2$) اى ($(9+1)/2 = 5$) أى المفردة الخامسة . ومن ثم فيكون الوسيط مساويا للتقدير "جيد".

وكما يتضح مما سبق أن الوسيط لايتأثر بالقيم الشاذة أو المتطرفة إذ لاتدخل قيمتها بطبيعة الحال فى حسابه بينما رأينا أن الوسط الحسابى كان يتحيز أو يميل باتجاهها . ولذلك يعتبر استخدام الوسيط فى مثل هذه الحالات أفضل. وهذا لايعنى أن الوسيط مفضل دائما فأحيانا توجد بعض الصعوبات فى حسابه كما فى حالة المتغيرات المنفصلة والتى لاتأخذ قيما كسرية . فإذا فرضنا ان لدينا ستة أسرعدد افرادها هو **7 , 6 , 7 , 6 , 4 , 3** **10 , 8** , فلما كان عدد المفردات زوجيا فأن الوسيط حسب التعريف السابق وهو الوسط الحسابى للمفردتين الثالثة والرابعة أى ($6+7/2= 6.5$) أى أن :

$Q_2 = 6.5$ وهذه القيمة لاوجود لها ولامعنى لأن تكون قيمة الوسيط لعدد أفراد الاسرة هو ٦.٥ فردا. هذا بالأضافة الى أنه احيانا يفقد الوسيط دلالته إذا كان عدد المفردات صغيرا وبينهما قيم كثيرة متكررة . فإذا كان لدينا درجات خمسة طلاب هى : 80 , 72 , 70 , 70 , 70 فأن الوسيط فى هذه الحالة هو الدرجة ٧٠ ولاتوجد أى قيمة أصغر من قيمة الوسيط ويوجد ٤٠% من المفردات من المفردات اكبر من قيمة الوسيط .

وجدير بالذكر إذا كان عدد المفردات (n) كبيرا فسوف نتجاوز الدقة التامة فى حساب ترتيب الوسيط ويقدر الوسيط فى هذه الحالة على أنه قيمة المفردة ذات الترتيب $n/2$ أيا كانت (n) عددا فرديا أم زوجيا وذلك توخيا للإقتصاد فى الجهد الحسابى . ولعل السبب فى ذلك هو انه فى مثل هذه الحالات يكون من اللازم علينا تبويب البيانات فى صورة توزيع تكرارى ونقوم بحساب الوسيط لهذا التوزيع كما يتضح فى الجزء التالى :

**(٣-٢-٢) : إيجاد قيمة الوسيط فى حالة البيانات المبوبة تكراريا :-**

أوضحنا فيما سبق أن قيمة الوسيط تقع فى منتصف القيم من حيث الترتيب وانه لابد من ترتيب المفردات تصاعديا أو تنازليا لتحديد موقع ثم قيمة الوسيط . وترتيب المفردات فى حالة البيانات الغير مبوبة يقابله تماما الحصول على الجدول التكرارى المتجمع الصاعد او الهابط فى حالة البيانات المبوبة تكراريا . لذا فإنه لأيجاد قيمة الوسيط فى حالة التوزيعات التكرارية يجب الحصول على التوزيع التكرارى المتجمع الصاعد أو الهابط أولا . ثم يتم تحديد رتبة الوسيط والتى تعادل قسمة مجموع التكرارت على ٢ ليتحدد بذلك رتبة الوسيط ومن ثم تحدد الفئة التى سيقع فيها قيمة الوسيط اما بإستخدام قاعدة النسبة والتناسب السابق الاشارة إليها فى الجداول التكرارية المتجمعة أو يمكن ايجاد قيمة الوسيط من خلال قانون يسمى بقانون

**الوسيط والمستنتج أصلا من قاعدة النسبة والتناسب السابق الاشارة اليها .**

**وعلى ذلك تتلخص خطوات تحديد قيمة الوسيط فيما يلى :-**

**١- تكوين الجدول التكرارى المتجمع وليكن التوزيع التكرارى المتجمع الصاعد .**

$$Ran\ Rof(Q_2) = \frac{\sum_{i=1}^{r} Fi}{2}$$

**٢- تحديد رتبة الوسيط وهى عبارة عن**

**حيث** (r) **عبارة عن عدد فئات التوزيع التكرارى المعطى .**

**٣- تتحدد الفئة الوسيطية بالفئة التى تقع فيها رتبة الوسيط التى تم تحديدها فى الخطوة السابقة من الجدول المتجمع ومن ثم يتم تحديد الحد الادنى للفئة الوسيطية**

(Lower Bound or Limit of $Q_2$ Class )

**وكذا التكرار المتجمع الصعد السابق وليكن** F(pre.Asc) **واللاحق وليكن**F(suc,Asc)

**٤- يتم ايجاد قيمة الوسيط إما بقانون النسبة والتناسب كما سبق أو بإستخدام قانون الوسيط حيث أن الوسيط** ($Q_2$) **يمكن تحديده من خلال القانون التالى :-**

$$Q2 = L.B.of\ (Q2)class + \frac{R.of\ (Q2) - F(pre.Asc)}{F(suc.Asc) - F(pre.Asc)} \times L.I.\ of(Q_2)$$

**(١٤)**

**حيث ان :**

- **تمثل الحد الادنى للفئة الوسيطية** * L.B.of (Q2)class
- **تمثل رتبة الوسيط** * R of (Q2)
- **تمثل التكرار المتجمع السابق لرتبة الوسيط** * F (pre.Asc)

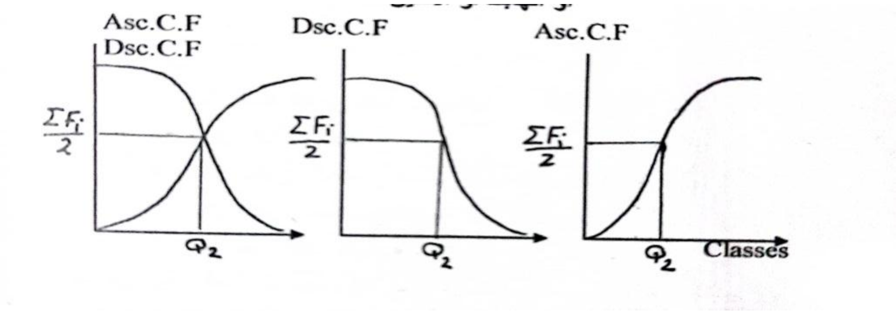
- **تمثل التكرار المتجمع اللاحق لرتبة الوسيط** F (suc.Asc) *
- **تمثل طول الفئة الوسيطية** * L.I of $(Q_2)$class

**اما عن ايجاد قيمة الوسيط بيانيا من خلال رسم المنحنى التكرارى المتجمع الصاعد أو الهابط .ثم يتم تحديد رتبة الوسيط((R.of(Q2)) وهى عبارة عن قسمة مجموع التكرارت على (٢) . ويتم تحديد موقع هذه الرتبة على المحور الرأسى للشكل البيانى الخاص بالمنحنى التكرارى المتجمع الصاعد أو الهابط . ومن هذا الموقع على المحور الرأسى يتم رسم عمود على المحور الرأسى وموازيا للمحور الافقى الى أن يقطع اى من المنحنين التكرارى المتجمع الصاعد او الهابط فى نقطه. يتم اسقاط عمود من تلك النقطة وموازيا للمحور الرأسى ليتقاطع مع المحور الافقى فى نقطة والتى تحدد قيمة الوسيط .أما أذا تم رسم المنحنى التكرارى الصاعد والمنحنى التكرارى المتجمع الهابط فى شكلا بيانيا واحدا. فإن المنحنين يتقاطعا فى نقطة .**

**هذه النقطة إذا تم اسقاط منها عمودا على المحور الافقى تحدد قيمة الوسيط. أما اذا تم اسقاط منها عمودا على المحور الرأسى فهى تحدد رتبة الوسيط. والشكل البيانى التالى يوضح كيفية تحديد قيمة الوسيط بيانيا .**

**شكل (٣-١)**

**ايجاد قيمة الوسيط بيانيا من خلال رسم المنحنى التكرارى المتجمع الصاعد أو الهابط أو الاثنين معا**



**مثال (١٢) :**

**فيما يلى التوزيع التكرارى التالى يوضح الاجر اليومى (بالجنيه) لمجموعة من العاملين بإحدى الشركات :**

**جدول (٣-١٤)**

| Classes | 8- | 18- | 28- | 38- | 48- | 58-68 | $\sum$ |
|---|---|---|---|---|---|---|---|
| Frequency | 8 | 22 | 25 | 20 | 13 | 12 | 100 |

**والمطلوب : أوجد قيمة الوسيط حسابيا وبيانيًا**

**الحل :**

**لايجاد قيمة الوسيط يجب بداية ايجاد الجدول التكرارى المتجمع الصاعد أو الهابط . دعنا نقوم بتكوين التوزيع التكرارى المتجمع الصاعد على النحو التالى :-**

**جدول(٣-١٥)**

**التوزيع التكرارى المتجمع الصاعد للأجر اليومى للعاملين بالشركة**

| Classes | $F_i$ | Less than the U.L of classes | Asc.C.F |
|---|---|---|---|
| 8- | 8 | Less than 8 | 0 |
| 18- | 22 | Less than18 | 8 |
| 28- | 25 | Less than28 | 30 |
| 38- | 20 | Less than38 | 55 |
| 48- | 13 | Less than48 | 75 |
| 58-68 | 12 | Less than58 | 88 |
| $\sum$ | 100 | Less than68 | 100 |

**ثم يتم تحديد رتبة الوسيط حيث ان :**

$$\text{Rank of }(Q2) = \frac{\sum_{i=1}^{r} F_i}{2} = \frac{100}{2} = 50$$

**ومن ثم فإن الحد الادنى للفئة الوسيطية هو L.B=28 والتكرار المتجمع السابق لرتبة الوسيط هو (٣٠) والتكرار المتجمع اللاحق لرتبة الوسيط هو**

**(٥٥) وطول الفئة الوسيطية هو (١٠) جنيه ومن ثم فالبتعويض فى العلاقة (١٤) السابقة فيكون :**

$$(Q_2) = 28 + \left(\frac{50 - 30}{55 - 30}\right) \times 10$$

$$= 28 + ( 20 / 25 ) \times 10 = 36 \qquad (L.E)$$

**أى أن وسيط أجر العامل لهذا التوزيع هو ٣٦ جنيه .**

**<u>ملحوظـة</u> : يمكـن ايجـاد الوسـيط حسـابيا بإسـتخدام قاعـة النسـبة والتناسـب السـابق الاشـارة اليهـا وذلـك بعـد تحديـد رتبـة الوسـيط فيكـون لـدينا الرسـم الكروكى التالى :–**

Less than 28 (L.E) $\longrightarrow$ 30 persons

Less than $(Q_2)$ (L.E) $\longrightarrow$ 50 persons

Less than 38 (L.E) $\longrightarrow$ 55 persons

**فيكون :**

$$\frac{55- 30}{55-50} = \frac{38- 28}{38- Q_2}$$

**وعليه فإن :**

$$5 = \frac{10}{38-Q2} \quad , \text{ i.e. }, \quad 5(38-Q_2) = 10 \quad , \quad 190 - 10 = 5Q_2$$

**فتكون قيمة الوسيط هى :**

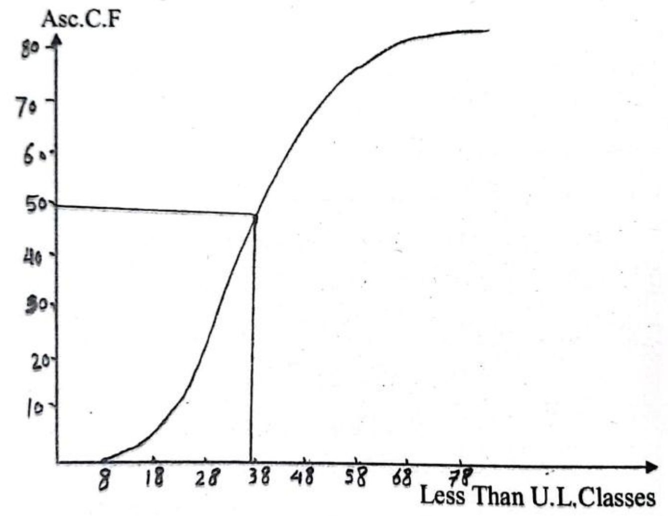
$$Q_2 = \frac{180}{5} = 36 \qquad (L.E)$$

**وهى نفس النتيجة المستخلصة من خلال قانون الوسيط .**

**أما عن ايجاد قيمة الوسيط بيانيا فسوف نحدد قيمة الوسيط من خلال رسم المنحنى التكرارى المتجمع الصاعد تاركين للقارئ تحديده بإستخدام المنحنى التكرارى المتجمع الهابط بالإضافة لتحديد الوسيط بإستخدام المنحنيين المتجمعين الصاعد والهابط معا .**

**شكل (٣-٢)**

**ايجاد قيمة الوسيط من خلال المنحنى التكرارى المتجمع الصاعد**




**<u>ملاحظات حول الوسيط :-</u>**

**من اهم الخصائص او المزايا التى يتمتع بها الوسيط عن غيره من مقاييس النزعة المركزية :**

١- **يتميز الوسيط بأنه لايتأثر بالقيم الشاذة (Outliers) أو المتطرفة (Extreme Points) لأن من خلال تعريفنا للوسيط بأنه يعتمد على المفردة أو القيمة التى تتوسط البيانات أى فى منتصفها . وعلى ذلك فهو يفضل على الوسط الحسابى فى حالة التوزيعات الملتوية أو شديدة الالتواء .**

٢- **يمكن ايجاد الوسيط فى حالة التوزيعات التكرارية المفتوحة مالم يكن التوزيع شديد الالتواء بحيث تقع رتبة الوسيط فى الفئة الاولى او الاخيرة من التوزيع .**

٣- يمكن ايجاد الوسيط بيانيا من خلال رسم المنحنى التكرارى المتجمع الصاعد او المتجمع الهابط أو الاثنين معا .

٤- كما يتميز الوسيط عن باقى مقاييس النزعة المركزية بأن مجموع الانحرافات المطلقة للقيم عن الوسيط أقل من مجموع الانحرافات المطلقة عن اى مقياس اخر خلاف الوسيط. أى أن:

$$\sum_{i=1}^{n} | x_i-Q_2| < \sum_{i=1}^{n} (x_i-A)$$

حيث أن $(A)$ أى مقدار ثابت تختلف قيمته عن الوسيط . وهذه النتيجة غير شائعة الاستخدام فى المجالات التطبيقية .

**أما عن عيوب الوسيط فتتمثل فيما يلى :**

١- من أهم عيوب الوسيط أنه يعتمد على قيمة واحدة تقع فى منتصف المفردات بعد ترتيبها ولايدخل فى حسابه باقى المفردات كما هو الحال فى الوسط الحسابى .

٢- كذلك يعاب على الوسيط أنه لايخضع للعمليات الجبرية خضوع الوسط الحسابى .كذلك ترانا تجاوزنا الدقة التامة عند تحديد رتبة الوسيط فى حالة التوزيعات التكرارية فأعتبرنا أن رتبة الوسيط هى ناتج خارج قسمة مجموع التكرارت على (٢) سواء كان هذا المجموع $(\sum F_i)$ عددا زوجيا أو فرديا .

٣- خلافا لما سبق فإنه يمكن حساب الوسط الحسابى العام لعدة مجموعات جزئية أو عدة عينات بإستخدام الاوساط الحسابية لهذه المجموعات او تلك العينات . فإن هذا الامر لايتحقق فى حالة الوسيط بمعنى إننا لانستطيع حساب الوسيط العام لعدة مجموعات جزئية أو عدة عينات إذا علم الوسيط لكل مجموعة جزئية أو كل عينة.

**(٣-٢-٣) شبيهات الوسيط :**

هناك بعض المقاييس الاحصائية التى تشبه الوسيط فى تعريفها وفى طريق حسابها هى :

- **الربيعين** Quartiles
- **العشيرات** Deciles
- **المئينيات** Percentiles

**وسوف نتناول دراسة تلك المقاييس بشئ من التفصيل على النحو التالى :**

**أولا : الربيعين (الربيع الادنى والربيع الاعلى):-**

**ويطلق على الربيعين الادنى والاعلى أسم الربيع الاول( First Quartile) والربيع الثالث (Third Quartile).**

**هذا ويمكن تعريف الربيع الادنـى (او الاول) أى ($Q_1$) بانـه القيمـة التـى يقل عنهـا ربـع المفـردات أو القيم ويزيـد عنهـا ثلاثـة أربـاع القـيم . أمـا الربيـع الاعلى (أو الثالث) أى ($Q_3$) فيمكن تعريفه بأنه القيمة التى يقل عنها ثلاثة أرباع القيم ويزيد عنها ربع القيم . ومن خلال هذا المنطلق فأنـه يتم تعريف الوسيط بأنه بمثابة الربيع الثانى أى ($Q_2$) حيث انه القيمة التـى يقل عنهـا نصف القيم ويزيد عنها النصف الاخر من تلك القيم .**

**هذا ويتشابه الربيعيين الادنى والاعلى مع الوسيط فى طريقة ايجادهما سواء كـان ذلـك بالطريقـة الحسـابية (سـواء مـن خـلال القـانون او قاعـدة النسـبة والتناسب) أو بالطريقة البيانية . ويمكن توضيح ذلك كما يلى :**

**أ) ايجاد الربيعيين ($Q_3$,$Q_1$) حسابيا :**

**لايجـاد الـربيعيين حسـابيا فأنـه يمكـن اسـتخدام القـانون (١٤) المسـتخدم فـى حسـاب الوسـيط مـع اخـتلاف قسـمى المقيـاس الـذى يـتم حسـابه وذلـك علـى النحو التالى :**

$$Q_1 = L.B\ of\ (Q_1)class + \left(\frac{R.of(Q_1) - F_{(pre.Asc)}}{F_{(suc.Asc)} - F_{(pre.Asc)}}\right) \times L.I.\ of\ (Q_1)Class \qquad (١٥)$$

$$Q_3 = L.B \text{ of } (Q_3)\text{class} + \left(\frac{R.of(Q_3)-F_{(pre.Asc)}}{F_{(suc.Asc)}-F_{(pre.Asc)}}\right) \times L.I. \text{ of } (Q_3)\text{Class} \quad (١٦)$$

**هذا ويمكن استخدام قاعدة النسبة والتناسب السابق الاشارة اليها فى تحديد قيمة كل من الربيع الادنى ($Q_1$) والربيع الاعلى ($Q_3$) .**

**ب) <u>ايجاد قيمة الربيعيين ($Q_3$,$Q_1$) بيانيا :-</u>**

**يمكن ايجاد قيمة الربيعين ($Q_1$) ، ($Q_3$) بأسلوب مشابه تماما للأسلوب المتبع فى حالة الوسيط سواء كان ذلك بإستخدام المنحنى التكرارى المتجمع الصاعد أو المتجمع الهابط أو المنحنيين معا .حيث يتم تحديد موقع رتبة الربيعيين الادنى والاعلى بنقطة على المحور الرأسى للشكل البيانى للمنحنى التكرارى المتجمع الصاعد مثلا . ثم يتم منهما رسم خطين موازيين للمحور الافقى يقطع كلا منهما المنحنى التكرارى المتجمع الصاعد فى نقطة . ومن نقطتى التقاطع مع المنحنى التكرارى المتجمع الصاعد نسقط عمودين على المحور الافقى ليقطعانه فى نقطتين الاولى تمثل قيمة الربيع الادنى ($Q_1$) والثانية تمثل قيمة الربيع الاعلى ($Q_3$) .**

**ملحوظه:- فى حالة استخدام المنحنى التكرارى المتجمع الهابط لاحظ أن من خلال دراستنا لمفهوم الربيعيين الادنى والاعلى فإن :**

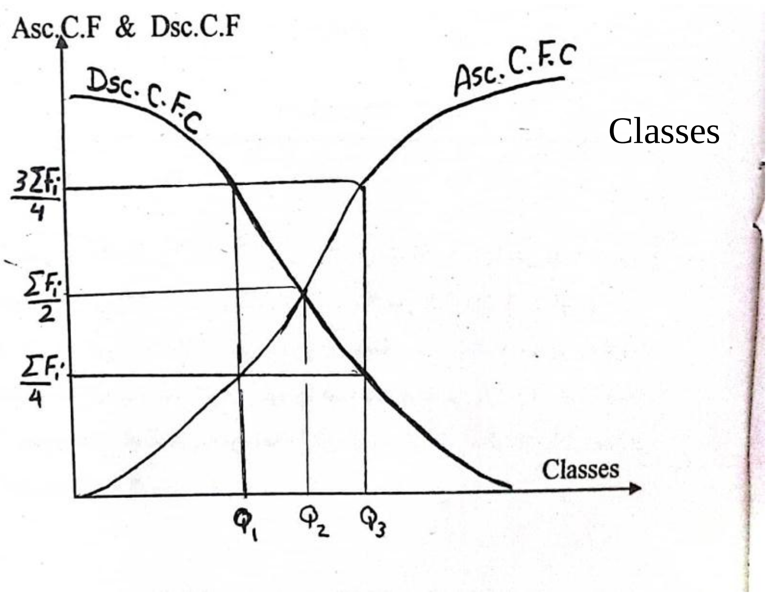
**الربيع الادنى هو القيمة التى يقل عنها ٢٥% من المفردات ويزيد عنها ٧٥% من المفردات . أى انه إذ تم رسم المنحنيين التكراريين المتجمعين الصاعد والهابط معا وبتحديد موقع النقطة التى تعبر عن ترتيب ($Q_1$). فاذا تم منها رسم خطا موازيا للمحورالافقى الى ان يقطع كل من المنحنى التكرارى المتجمع الصاعد والهابط . فيكون العمود النازل من نقطة التقاطع مع المنحنى التكرارى المتجمع الصاعد على المحور الافقى معبرة عن قيمة الربيع الادنى . أما العمود النازل من نقطة التقاطع مع المنحنى التكرارى**

**المتجمع الهابط فتعطى قيمة الربيع الثالث($Q_3$) والشكل البيانى (٣–٣) يوضح ذلك :**

**شكل (٣–٣)**

**ايجاد الربيعين $Q_3,Q_1$**

**من خلال رسم المنحنيين المتجمعيين الصاعد والهابط معا**




**كذلك فإن من نقطة ترتيب الربيع الاعلى ($\frac{3\sum F_i}{4}$) اذا تم رسم خطا موازيا للمحور الافقى فإن**

**هذا الخط يقطع أولا المنحنى التكرارى المتجمع الهابط عند نقطة تقابل تماما قيمة الربيع الاول $Q_1$ . أما نقطة تلاقى هذا الخط الموازى مع المنحنى التكرارى المتجمع الصاعد فهى تقابل قيمة الربيع الثالث $Q_3$ .**

**وفى نهاية تناولنا للربيعيين الادنى $Q_1$ والاعلى $Q_3$ فإنه يمكن القول بأن قيمة الربيعين تعطيان مؤشرا عن كيفية انتشار البيانات ومدى تفاوتها وهو ما سيتضح عند دراستنا لمقاييس التشتت فى الباب التالى من هذا الكتاب . هذا بالإضافة الى ان هناك معلومة يمكن استنتاجها من قيمتى الربيعين**

**يوضحهما الشكل البيانى (٣-٤) فإذا تصورنا ترتيب المفردات ترتيبًا تصاعديا مثلا فإن الشكل (٣-٤) يوضح التعريف الذى تناولناه سابقا لمعنى**

**شكل (٣-٤)**

**نسبة المشاهدات أو المفردات التى تقع مابين الربيعيين**

$Q_3,Q_1$

Ascending

$x_{(1)}$ $Q_1$ $Q_2$ $Q_3$ $x_{(n)}$

**الربيعيين . فحيث أن هناك ٢٥% من المفردات تقل قيمتها عن قيمة الربيع الادنى $Q_1$ وان هناك نفس النسبة اى ٢٥% من المفردات تزيد قيمتها عن الربيع الاعلى $Q_3$ . فإنه يمكن القول أن نصف عدد المشاهدات او المفردات للظاهرة محل الدراسة تتراوح قيمتها فيما بين قيمتى $Q_3,Q_1$ . لذا فأن الحصول على قيمة الربيعيين تمكننا من الحصول على معلومات أكثر مما لو استخدمنا قيمة الوسيط بمفرده .**

**<u>مثال (١٣):</u>**

**أوجد قيمة الربيعيين الادنى والاعلى حسابيا بالطرق المختلفة وبيانيا للتوزيع التكرارى المعطى فى المثال السابق .**

**<u>الحل :</u>**

**١- ايجادة قيم الربيعيين $Q_3,Q_1$ حسابيا .**

**أ- باستخدام قانون الوسيط مع إجراء التعديل اللازم عليه . فبعد الحصول على الجدول التكرارى المتجمع الصاعد كما هو مبين فى الجدول(٣-١٥) . يتم تحديد رتبة كل ربيع على حده على النحو التالى :**

$$\text{Rank of } (Q_1) = \frac{\sum F_i}{4} \times 1 = \frac{100}{4} = 25 \,,$$

$$\text{Rank of } (Q_3) = \frac{\sum F_i}{4} \times 3 = \frac{3x\ 100}{4} = 75 \,,$$

**ومن ثم فيتطبق كل من القانونين(١٥)،(١٦) السابق الاشارة اليهما فيكون:**

$$Q_1 = L.B \text{ of } (Q_1)Class + \left(\frac{R.of(Q_1)-F_{(pre.Asc)}}{F_{(suc.Asc)}-F_{(pre.Asc)}}\right) \times L.I. \text{ of } (Q_1)Class$$

$$= 18 + \left(\frac{25 - 8}{30 - 8}\right) x\ 10$$

$$= 18 + \left(\frac{17}{22}\right) \times 10 = 25.727 \qquad (L.E)$$

**أما عن قيمة $Q_3$ فالبنظر لعمود التكرارت المتجمعة الصاعدة فى الجدول (٣-١٥) يتضح لنا أن رتبة الربيع الاعلى (٧٥)والتى تم تحديدها فى الخطوة السابقة ماهى إلا أحد التكرارات المتجمعة الصاعدة وبالتحديد امام الفئة (٤٨) جنيه فى عمود الفئات المتجمعة الصاعدة .لذا فإننا لسنا فى حاجة لتطبيق أى قانون لحساب قيمة الربيع الاعلى $Q_3$ وتكون القيمة مباشرة ماهى إلا الفئة المتجمعة الصاعدة المقابلة أى أن :**

$Q_3 = 48$ (L.E)

**ب- <u>ايجاد قيمة الربيعيين بإستخدام قاعدة النسبة والتناسب</u> :-**

**وبالتحديد سوف يتم ايجاد قيمة $Q_1$ فقط بإستخدام قاعدة النسبة والتناسب لأن قيمة $Q_3$ يتم تحديدها مباشرة فى هذا المثال كما سبق وأن أوضحنا .**

**فلإيجاد قيمة $Q_1$ لاحظ ان الرتبة (٢٥) الخاصة بترتيب $Q_1$ تقع مابين التكرارين المتجمعيين الصاعديين 30 , 8 ومن ثم فإن فئة الربيع الادنى المقابلة لتلك التكرارت المتجمعة هى من (١٨) جنيها الى أقل من (٢٨) جنيها. أى أن :**

Less than 18 (L.E) $\longrightarrow$ 8 persons

Less than (Q1) (L.E) $\longrightarrow$ 25 persons
Less than 28 (L.E) $\longrightarrow$ 30 persons

**ومن ثم فإن :**

$$\frac{28 - 18}{28 - Q_1} = \frac{30 - 8}{30 - 25}$$

**وعليه فأن :**

$$\frac{10}{28 - Q_1} = \frac{22}{5}$$

**أى أن :**

$$50 = 22\ (28 - Q_1)\ ,\ 22Q_1 = 616 - 50$$

**فيكون :**

$$Q_1 = \frac{566}{22} = 25.272 \qquad \text{(L.E)}$$

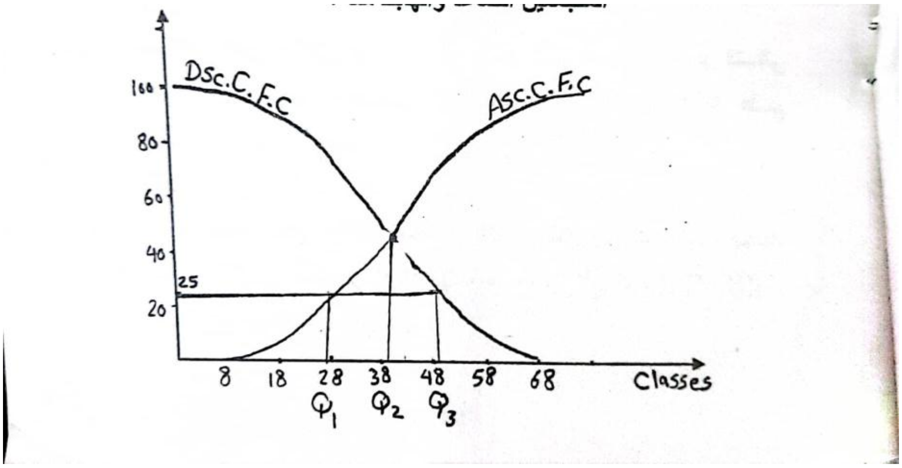
**وهي نفس النتيجة السابقة .**

**٢- <u>ايجاد الربيعيين بيانيا :</u>**

**يمكن تحديد قيمة الربيعيين الادنى والاعلى بيانيا من خلال إما رسم المنحنى التكرارى المتجمع الصاعد أو رسم المنحنى التكرارى المتجمع الهابط او رسم المنحنيين التكرارين المتجمعيين الصاعد والهابط معا . وسنترك للطالب الطريقتين الاولى والثانية ونقوم بإيجاد قيمة الربيعيين من خلال رسم المنحنيين معا كما هو موضح فى الشكل (٣-٥)**

**شكل (٣-٥)**

**ايجاد قيمة الربيعيين $Q_1$ , $Q_3$ من خلال رسم المنحنيين المتجمعين الصاعد والهابط معا .**




**<u>ثانيا : العشيرات :-</u>** Deciles

**يمكننا تعريف العشير الاول First Decile ($D_{(1)}$) بأنه القيمة التى يقل عنها عشر القيم أو المفردات ويزيد عنها تسعة أعشار تلك القيم . وهكذا يمكن تعريف العشيرات الاخرى . هذا ويتم ايجاد العشيرات المختلفة بنفس الاسلوب المتبع فى حالة الوسيط والربيعيين الادنى والاعلى مع مراعاة انه عند تحديد رتبة العشير فإنه يتم قسمة مجموع التكررات على ١٠ ثم ضرب ناتج القسمة فى رتبة العشير المطلوب ايجاد قيمته . وعلى ذلك فإنه لإيجاد رتبة العشير الثالث مثلا يتم قسمة مجموع التكرارت ($\sum F_i$) على( ١٠) ثم ضرب ناتج القسمة فى (٣) والتى تمثل العشير الثالث. ويتم استخدام القانون (١٤) او (١٥)او (١٦) مع اجراء التعديلات اللازمة ، والتى تتلائم مع مفهوم العشير المطلوب حسابه . هذا ويمكن ايضا ايجاد قيمة العشيرات باستخدام قانون النسبة والتناسب أو الرسم من خلال المنحنى التكرارى المتجمع الصاعد أو الهابط أو المنحنيين معا كما هو الحال فى حالة الوسيط والربيعيين الادنى والاعلى .**

**<u>مثال(١٤):</u>**

**بإستخدام التوزيع التكرارى الوارد فى المثال (١٢) احسب العشير الثانى للأجر اليومى لمجموعة العاملين . ثم احسب العشير الخامس وعلق على نتيجتك التى حصلت عليها .**

**<u>الحل:</u>**

**لايجاد العشير الثانى والعشير الخامس للتوزيع التكرارى يتم بداية ايجاد الجدول التكرارى المتجمع الصاعد (كما هو موضح فى الجدول (٣-١٥)) ثم يتم تحديد رتبة العشير . حيث أن :**

**رتبة العشير الثانى :**

$$\text{Rank of } (D_1) = \frac{\sum Fi}{10} \times 2 = \frac{100}{10} \times 2 = 20 ,$$

**كذلك فأن رتبة العشير الخامس هى :**

$$\text{Rank of } (D_5) = \frac{\sum Fi}{10} \times 5 = \frac{100}{10} \times 5 = 50 ,$$

**لاحظ أن رتبة العشير الخامس هى نفس رتبة الوسيط . ومن ثم فلنتوقع أن تكون قيمة العشير الخامس مساوية لقيمة الوسيط تماما .**

**ولحساب قيمة كل من العشير الثانى والعشير الخامس فأمامنا أما الطرق الحسابية (القانون أو قاعدة النسبة والتناسب ) بالاضافة للطريقة البيانية من خلال رسم اى منحنى تكرارى متجمع . وللدقة دعنا نقوم بحساب العشيرين المطلوبين بإستخدام قاعد النسبة والتناسب وذلك على النحو المبين التالى :**

- **فبالنسبة للعشير الثانى فإن الرتبة (٢٠)تقع مابين التكرارين المتجمعين الصاعدين ٣٠,٨ فى جدول (٣-١٥) وعليه فإن فئة**

**العشير الثانى هى من ١٨ الى ٢٨ جنيه وعليه يكون لدينا الشكل التالى :**

Less than 18 (L.E) ---------------------- 8 persons

Less than ($D_2$) (L.E)--------------------- 20 persons
Less than 28 (L.E) ----------------------- 30 persons

**وعليه فإن :**

$$\frac{30-8}{30-20} = \frac{28-18}{28-D_2}$$

**فيكون :**

$$22\ (28 - D_2) = 100$$

**أى ان :**

$$D_2 = \frac{516}{22} = 23.45 \quad (L.E)$$

**ولاحظ أن قيمة العشير الثانى $D_2$ تقع مابين حدى الفئات المتجمعة الصاعدة أى مابين ٢٨,١٨ جنيه أى داخل فئة العشير السابق تحديدها . وهذا يعنى أن ٢٠% من العمال تقل أجورهم عن ٢٣.٤٥ جنيه وأن ٨٠% تزيد اجورهم اليومية عن ٢٣.٤٥ جنيه.اما العشير الخامس فحيث ان رتبته هى ٥٠ وهى مساوية لرتبة الوسيط فستكون النتيجة مساوية لقيمة الوسيط كما سبق أن أوجدناها وهى قيمة ٣٦ جنيه . أى أن**

$D_5 = Q_2 = 36$ (L.E)

**وهو مايفيد أن ٥٠% من العمال تقل أجورهم عن ٣٦ جنيه وأن ٥٠% من العمال تزيد أجورهم عن ٣٦ جنيه.**

**<u>ثالثا : المئينيات</u> :-** Percentiles

**بنفس الدلالة السابقة مع اختلاف المفهوم فيمكن تعريف المئين الاول $P_1$ بأنه القيمة التى يقل عنها ١% من عدد القيم ويزيد عنها ٩٩% من باقى**

**المفردات وعليه فإن المئين السابع والعشرون مثلا $P_{27}$ وهو عبارة عن القيمة التى يقل عنها ٢٧% من المفردات أو القيمة التي يزيد عنها ٧٣% من باقى الفردات . وهكذا يمكن تعريف العديد من المئينيات بنفس الاسلوب . ولايجاد قيم المئينيات يمكن ايضا استخدام قانون الوسيط (١٤) وذلك بعد اجراء التعديلات اللازمة عليه لكى يتفق ومفهوم المئين المطلوب . كما يمكن حساب المئينيات بإستخدام قاعدة النسبة والتناسب وذلك بعد تحديد رتبة المئين فى تلك الحالة . كما يمكن حساب المئينيات بيانيا من خلال المنحنى التكرارى المتجمع الصاعد او الهابط من خلال رسم المنحنيين معا .**

**<u>مثال (١٥)</u> :باستخدام التوزيع التكراري التالي لأجور ٥٠ عاملاً اسبوعياً (بالجنيه) بإحدى الشركات المطلوب حساب المئتين الثامن والستون .**

**جدول ( ٣ – ١٦ )**

| Classes | 30- | 40- | 50- | 60- | 70- | 80- | 90- | $\sum$ |
|---|---|---|---|---|---|---|---|---|
| Frequency | 1 | 4 | 7 | 14 | 11 | 8 | 5 | 50 |

**<u>الحل :-</u>لحساب المئين الثامن والستون يجب أولاً إيجاد التوزيع التكراري المتجمع الصاعد أو الهابط ثم تحديد رتبة المئين ثم تحديد القيمة من خلال القانون أو قاعدة النسبة والتناسب أو بيانياً . وسوف نقوم بحسابه من خلال القانون تاركين للطالب التأكد من صحة النتائج باستخدام قاعدة النسبة والتناسب أو الطريقة البيانية .**

**جدول ( ٣ – ١٧ )**

| Classes | $F_i$ | Less than the U.L. of Classes | Ass.C.F |
|---|---|---|---|
| 30- | 1 | Less than 40 | 1 |
| 40- | 4 | Less than 50 | 5 |
| 50- | 7 | Less than 60 | 12 |
| 60- | 14 | Less than 70 | 26 |
| 70- | 11 | Less than 80 | 37 |
| 80- | 8 | Less than 90 | 45 |
| 90- | 5 | Less than 100 | 50 |
| $\sum$ | 50 | | |

**وحيث أن رتبة المئتين الثامن والستون هي :**

$$\text{Rank of } P_{(68)} = (\sum F_i \ / 100) \times 68$$
$$= 50 \times 68 = 34$$

**والرتبة (٣٤) تنحصر ما بين التكرارين المتجمعين الصاعدين ٢٦ ، ٣٧ . لذا فإن فئة المئتين الثامن والستون هي من (٧٠) حتى أقل من ٨٠ جنيه وعليه فإن :-**

$$P_{(68)} = \text{L.B.of } P_{68} \text{ Class} + \frac{(\text{R.of } P_{68} - F_{(pre.Asc.)})}{F_{(suc.Asc.)} - F_{(pre.Asc)})} \times \text{ L.I. of } P_{68} \text{ Class}$$

$$= 70 + \frac{(34 - 26)}{(37 - 26)} \times 10$$

$$= 70 + (8 \div 11) \times 10 = 77.27 \qquad (\text{L.E})$$

**وهو ما يعنى أن ٦٨% من إجمالي عدد العمال تقل أجورهم الاسبوعية عن ٧٧.٢٧ جنيه .**

**استخدام المئنيات :-**

**لاشك أن إيجاد قيم المئنيات يعتبر مفيد جداً في بعض النواحي التطبيقية ولنضرب على ذلك أمثلة من واقع حياتنا العملية :-**

**١- في حالة الطلاب المتقدمين لشغل أماكن المدن الجامعية ( الإسكان الجامعي ) في إحدى الجامعات . لو افترضنا أنه وفقاً لعدد الأماكن المتاحة فإن سيتم قبول ٧٣% فقط من إجمالي عدد الطلاب المتقدمين وذلك باستخدام النسبة المئوية للنجاح كمعيار لقبول الطالب أو رفضه.**

**في هذه الحالة فإن قيمة المئين السابع والعشرون أي $P_{(27)}$ هي النسبة المئوية للنجاح التي لو حصل عليها الطالب أو أكبر منها فإنه يكون من المقبولين في المدينة الجامعية وفيما عدا ذلك يتم يرفض الطالب .**

**٢- في حالة الطلاب المتقدمين إلى مكتب التنسيق وكانت رغباتهم الأولى هي الالتحاق بإحدى كليات التجارة مثلاً فإذا كانت أعداد هؤلاء الطلاب**

المتقدمون هو ٢٠٠٠ طالباً على سبيل المثال . وكان إجمالي الأماكن المتاحة في هذه الكلية هو ١٦٤٠ طالباً ( أي ٨٢ % من إجمالي المتقدمين ) . فإن في هذه الحالة يمكن تحديد نسبة النجاح التي لو حصل عليها الطالب أو أكبر منها فإنه يقبل وفيما عدا ذلك فإنه يرفض طلبه . وتحدد هذه النسبة بقيمة المئين الثامن عشر أي $P_{(18)}$ .

٣- إذا كان الهدف من إجراء أحد أبحاث ميزانية الأسرة هو تحديد قيمة الدخل الشهري الذي يقل عنه ٦٧% (على سبيل المثال) من إجمالي عدد الأسر التي شملها البحث . فإن قيمة هذا الدخل الشهري تتحدد بإيجاد المئين السابع والستون أي $P_{(67)}$ .

٤- المئين الثالث والخمسون $P_{(53)}$ هو القيمة المستهدفة إذا كان الهدف من إجراء إحدى الدراسات السكانية هو تحديد قيمة العمر الذي يقل عنه ٥٣% (على سبيل المثال) من إجمالي الأشخاص الذين شملهم البحث .

**<u>ملاحظات على الوسيط وشبيهاته :</u>**

بعد أن استعرضنا طرق إيجاد الوسيط وبعض المقاييس الشبيهة به مثل الربيعيين الأدنى والأعلى والعشيرات والمئينيات فإنه تجدر الإشارة لبعض النقاط التالية :

١- إنه وفقًا لتعريف الوسيط والمقاييس الشبيهة له فإنه الوسيط يقسم المساحة تحت المنحنى التكراري للظاهرة إلى قسمين متساويين . بينما تنقسم هذه المساحة إلى أربعة أقسام متساوية باستخدام قيم الربيعين الأدنى والأعلى والوسيط (الربيع الثاني) .

وفي حالة التوزيعات المتماثلة تكون قيمة الوسيط مساوية تمامًا لمتوسط قيمتي الربيع الأدنى والأعلى للتوزيع أي أن الوسيط يقع في منتصف المسافة فيما بين الربيعين الأدنى والأعلى أي أنه في حالة التوزيع المتماثل يكون قيمة الوسيط عبارة عن :

$$\text{الوسيط } (Q_2) = \frac{\text{الربيع الأدنى } (Q_1) + \text{الربيع الأعلى } (Q_3)}{2}$$

ومدى الإنحراف (أي القرب أو البعد) عن صحة هذه العلاقة يعطي مؤشرًا عن مدى قرب أو بعد التوزيع عن التماثل . أي أنه نستطيع القول بأن مقارنة قيمتي الربيعين الأدنى والأعلى بقيمة الوسيط يمكن أن تعطي مؤشرًا لدرجة تماثل التوزيع التكراري . وهو ما سنتعرض له بالدراسة عند دراستنا لمقاييس التشتت والالتواء فيما بعد .

٢- كذلك فإنه في حالة التوزيعات المتماثلة تكون قيمة الوسيط دائمًا مساوية لمتوسط أي زوج من أزواج العشيرات أو المئينيات وأن كل زوج من العشيرات أو المئينيات من تلك الأزواج يقع على بعد مساوٍ من الوسيط . فعلى سبيل المثال تكون قيمة الوسيط لمتوسط قيمتي الأزواج التالية من المقاييس :

- العشير الثاني والعشير الثامن – وكذلك العشير الرابع والعشير السادس .
- المئين الرابع والخمسون والمئين السادس والأربعون .
- المئين الثامن والستون والمئين الثاني والثلاثون .

وهكذا .

فباعتبار الزوج الأول من هذه المقاييس فإنه يمكن التحقق من كلاً من العشير الثاني والعشير الثامن يقع على بعد متساوي من الوسيط .

أي أن :–

الوسيط $(Q_2)$ – العشير الثاني $(D_{(2)})$ = العشير الثامن $(D_{(8)})$ – الوسيط $(Q_2)$

ويمكنك عزيزي الطالب التحقق من صحة ذلك من خلال حساب تلك المقاييس للتوزيع التكراري التالي :

**جدول (٣–١٨ )**

**التوزيع التكراري لدرجات عينة من الطلاب في**

**إحدى المواد بإحدى الكليات**

| Classes | 10- | 22- | 34- | 46- | 58- | 70- | 82- | ∑ |
|---|---|---|---|---|---|---|---|---|
| Frequency | 4 | 8 | 13 | 15 | 13 | 8 | 4 | 65 |

**٣- في حين أن الوسيط لا يساوي الوسط الحسابي للربيعين الأدنى والأعلى والوسط الحسابي لأي زوج من أزواج العشيرات أو المئينيات إلا في حالة التوزيعات المماثلة فقط فإن العلاقات التالية صحيحة دائماً لكافة التوزيعات ( المتماثلة أو الملتوية ) فيما يتعلق بترتيب تلك المقاييس . فمثلاً : في أي توزيع تكراري نجد أن الوسط الحسابي لرتبتي أي زوج من تلك المقاييس يكون عبارة ترتيب الوسيط . فمثلاً :**

$$\text{Rank of }(Q_2) = \frac{\text{Rank of }(Q_1) + \text{Rank of }(Q_3)}{2}$$

**وذلك سواء كان التوزيع متماثلاً أو غير متماثل.ويمكن اثبات ذلك كما يلي : فحيث أن :**

$$\frac{\text{Rank of }(Q_1) + \text{Rank of }(Q_3)}{2} = \frac{(\sum F_i)/4 + (3\sum F_i)/4}{2}$$

$$= \frac{4 \times \sum F_i}{4 \times 2}$$

$$= \frac{\sum F_i}{2} = \text{Rank of }(Q_2)$$

٤- في حالة التوزيعات التكرارية النسبية فإن رتبة الوسيط هي (0.5) أو 50% سواء باستخدام التوزيع التكراري المتجمع النسبي الصاعد أو المتجمع النسبي الهابط . وكذلك فإن ترتيب الربيعين هما 0.25 ، 0.75 أو 25% ، 75% في حالة التوزيع التكراري المتجمع الصاعد . وهكذا لباقي المقاييس الشبيهة بالوسيط فمثلاً العشير الثالث ترتيبه هو 0.3 والمئين الستون ترتيبه هو 0.6 أو 60% وهكذا

## (٣-٣) المنوال Mode

بصفة عامة يمكن تعريف المنوال بأنه القيمة الأكثر شيوعاً أو تكراراً أي أنه هو القيمة التي تتكرر أكثر من غيرها من المفردات .

فمثلاً إذا كانت لدينا مجموعة من المفردات ( الغير مبوبة ) ولتكن درجات مجموعة من الطلاب هي ١٥ ، ١٢ ، ١٥ ، ٢٠ ، ١٤ ، ١٥ ، ٦ ، ٨ ، ١٥ فإن القيمة المنوالية لدرجة هؤلاء الطلاب هي الدرجة (١٥) وذلك لأنها تكررت أكثر من غيرها . في حين أن مجموعة الدرجات التالية : ١٥ ، ١٨ ، ١٠ ، ٧ ، ٥ ، ٤ ، ٣ لا توجد لها قيمة منوالية حيث لا توجد درجة متكررة أكثر من غيرها . أما مجموعة الدرجات ١٥ ، ٨ ، ٥ ، ٥ ، ٤ ، ٣، ٨ ، ٥ فلها قيمتين منواليتين هما ٥ ، ٨ درجة حيث تكررت كل منهما عدداً متساوياً من مرات التكرار وهو ثلاث مرات لكل واحد منهما .

وكمثال آخر إذا كان لدينا ٥٠٠ شخصاً موزعين حسب حالتهم التعليمية وممارستهم للأنشطة الرياضية كما يلي :-

جدول ( ٣ – ١٩ )

| الحالة التعليمية / ممارسة الرياضة | متعلم | أمي |
|---|---|---|
| يمارس الأنشطة الرياضية | ١١٠ | ٩٥ |
| لا يمارس الأنشطة الرياضية | ١٩٥ | ١٠٠ |

فإن منوال هذا التوزيع التكراري المزدوج هو " متعلم ولا يمارس الأنشطة الرياضية فهي الصفة السائدة وذلك لأن تكرارها ١٩٥ هو أعلى تكرار في

خلايا الجدول التكراري المزدوج ( جدول الاقتران ) أي أعلى من تكرار كل من الصفات الثلاث الأخرى التي بالجدول .

وهكذا يتضح لنا أنه لمجموعة من البيانات قد لا يوجد لها قيمة منوالية . وإن وجد فقد يكون المنوال وحيد القيمة أو ربما تتعدد القيم المنوالية لمجموعة البيانات وذلك بعكس الحال في حالة الوسط الحسابي والوسيط وشبيهاته حيث نجد أن لأي مجموعة من البيانات لابد وأن يكون لها وسط حسابي واحد أو وسيط واحد .

لذا فإنه جدير بالذكر والملاحظة فإن دور المنوال في علم الاحصاء لا يرقى إلى دور الوسط الحسابي أو الوسيط كمقياس للنزعة المركزية . إذ يهتم بالمنوال عادة أصحاب الأعمال الصناعية والتجارية والدعاية والتسويق . فالقيمة الأكثر تكرار لها مغزى اقتصادي هام بالنسبة لهم . فالمنتج الأكثر رواجاً في صناعة معينة يجذب اهتمام أصحاب هذه الصناعة لزيادة انتاجه وعمل المزيد من الدعاية له . كما يهتم الباحثون في العلوم السلوكية بالمنوال باعتباره قابلاً للحساب في جميع أنواع البيانات .

هذا ويمكن تقدير قيمة المنوال من التوزيعات التكرارية ( البيانات المبوبة ) بشكل تقريبي بطريقة حسابية أو بالرسم شأنه في ذلك شأن الوسيط .

<u>( ٣ – ٣ – ١ ) إيجاد قيمة المنوال في حالة البيانات المبوبة تكرارياً :</u>

إن عملية إيجاد المنوال لمجموعة من البيانات غير المبوبة ( المفردات ) تتمثل كما ذكرنا حالاً في مجرد البحث عن تلك المفردة التي تتكرر أكثر من غيرها . ومن ثم فهي تنطوي على قدر كبير من البساطة والوضوح . ولكن الأمر ليس كذلك في حالة البيانات المبوبة (في جداول تكرارية) . إذ أن الجدول لبيانات ظاهرة ما لا يعطي حقيقة كل مفردة على حدة وبالتالي لا يمكن تحديد تلك القيمة التي تتكرر أكثر من غيرها . وأفضل ما يمكن عمله في هذا الشأن كتحديد قيمة المنوال هو اختيار تلك الفئة التي تحتوي على

أكبر تكرار أصلي ( في حالة الجداول المنتظمة أي المتساوية من حيث أطول الفئات ) أو أكبر تكرار معدل ( في حالة الجداول الغير منتظمة وهي ما تسمى بالفئة المنوالية Model Class . وبعد ذلك يمكن تحديد قيمة المنوال داخل الفئة المنوالية باستخدام عدة أساليب فتزيد قيمة المنوال عن بداية الفئة المنوالية وتقل عن نهايتها . هذا ويميل أو ينحاز المنوال إلى بداية أو نهاية الفئة المنوالية تبعاً لتكرار الفئتين السابقة للفئة المنوالية واللاحقة لها . ويتم تحديد قيمة المنوال داخل الفئة المنوالية بإحدى الطرق الثلاث التالية :

**أولاً : طريقة مركز الفئة المنوالية :-**

فكما سبق وأن أوضحنا يمكن تعريف الفئة المنوالية على أنها الفئة المقابلة لأكبر تكرار أصلي ( في حالة التوزيعات التكرارية المنتظمة ) أو أكبر تكرار معدل ( في حالة التوزيعات التكرارية الغير منتظمة ) . ووفقاً لطريقة مركز الفئة المنوالية يتم تحديد الفئة المنوالية ومن ثم يتم تحديد الحد الأدنى والأعلى للفئة المنوالية وبالتالي يمكن تحديد مركز هذه الفئة فتكون هي بمثابة قيمة المنوال . أي أن النوال ( Mode ) أو (M) هو عبارة عن

$$\text{Mode(M)} = \frac{(\text{L.B} + \text{u. B}) \text{ for the Model Class}}{2}$$

حيث : L.B هو عبارة عن الحد الأدنى للفئة المنوالية ،

U.B هو عبارة عن الحد الأعلى للفئة المنوالية .

**مثال (١٦) :-**

الجدول التالي يبين توزيع أعمار ٦٠ بطارية حسب مدة صلاحية كل منها بالشهور :

**جدول (٣-٢٠ )**

| Classes | 6- | 9- | 12- | 15- | 18-21 | ∑ |
|---|---|---|---|---|---|---|
| Frequency | 4 | 12 | 20 | 18 | 6 | 60 |

**والمطلوب حدد قيمة المنوال مستخدماً طريقة الفئة المنوالية .**

**الحل**

**بداية وبملاحظة طول الفئة في هذا التوزيع يتبين لنا أن الجدول التكراري متساوي من حيث أطوال الفئات بمعنى أن التوزيع التكراري توزيعاً منتظماً لذا يتم حساب المنوال مباشرة . وبملاحظة قيم التكرارات نجد أن أكبر تكرار هو ٢٠ لذا فإن الفئة المنوالية التي تقابل أكبر تكرار هي الفئة من (١٢) حتى أقل من (١٥) شهر ومن ثم فإن المنوال هو :**

$$\text{Mode}(M) = \frac{(\text{L.B} + \text{U. B})\ \text{for the model class}}{2}$$

$$= \frac{12 + 15}{2} = \frac{27}{2} = 13.5 \qquad (\text{ Month })$$

**مستخدمين في ذلك طريقة مركز الفئة المنوالية كأحد طرق حساب المنوال .**

**ملحوظة :-**

**إذا كان التوزيع التكراري غير منتظم فلابد من حساب التكرارات المعدلة أولاً . وتكون الفئة المنوالية هي الفئة المقابلة لأكبر تكرار معدل في هذه الحالة .**

**مثال (١٧) :**

**الجدول التالي يبين التوزيع التكراري لدرجات مائة طالب في مادة الإحصاء :**

**جدول ( ٣–٢١ )**

| Classes | 10- | 25- | 30- | 40- | 50- | 60- | 80- | 90-100 | ∑ |
|---|---|---|---|---|---|---|---|---|---|
| Frequency | 5 | 7 | 13 | 21 | 19 | 16 | 10 | 9 | 60 |

**والمطلوب حساب المنوال مستخدماً طريقة مركز الفئة المنوالية .**

**<u>الحل :</u>**

**حيث أن الجدول ( ٣–٢١ ) توزيع تكراري غير متساوي من حيث طول الفئة ( توزيع غير منتظم . لذا فإنه يجب حساب التكرارات المعدلة أولاً . وفيما يلي الجدول التالي يوضح التكرارات المعدلة لهذا التوزيع .**

**جدول ( ٣–٢٢ )**

| Classes | $F_i$ | Length($L_i$ ) | Modified Fre. ( $MF_i$ ) = $F_i$ / $L_i$ |
|---|---|---|---|
| 10-<br>25-<br>30-<br>40-<br>50-<br>60-<br>80-<br>90-100 | 5<br>7<br>13<br>21<br>19<br>16<br>10<br>9 | 15<br>5<br>10<br>10<br>10<br>20<br>10<br>10 | 0.33<br>1.4<br>1.3<br>2.1<br>1.9<br>0.8<br>1.00<br>0.9 |
| $\Sigma$ | 100 | | |

**ومن خلال عمود التكرارات المعدلة يتضح لنا أن أكبر تكرار معدل هو (٢.١) لذا فالفئة المقابلة له ( الفئة المنوالية ) هي من (٤٠) حتى أقل من (٥٠) . لذا فإن قيمة المنوال باستخدام طريقة مركز الفئة المنوالية هي :**

$$(M) = \frac{(L.B + u.\ B)}{2} = \frac{40 + 50}{2} = 45 \ \ (\text{degee})$$

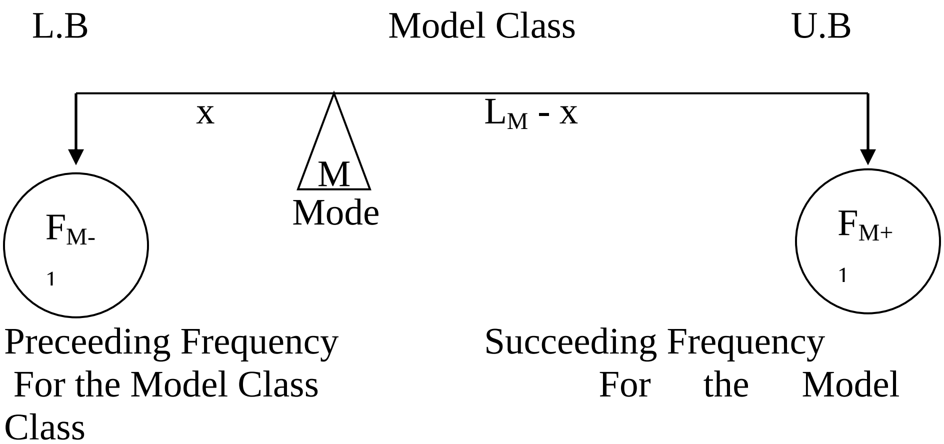
**<u>ثانياً الرافعة :</u>**

**وهذه الطريقة تعتبر الفئة المنوالية بمثابة ذراع أفقي تتجاذبه قوتان (أي رافعة من الدرجة الأولى ) أحدهما التكرار السابق للفئة المنوالية و الأخرى التكرار اللاحق للفئة المنوالية . فإذا كان التكرار السابق للفئة المنوالية أكبر من التكرار اللاحق للفئة المنوالية فإن قيمة المنوال تكون قريبة من الحد**

الأدنى للفئة المنوالية . والعكس صحيح أي أنه إذا كان تكرار الفئة اللاحقة للفئة المنوالية أكبر من تكرار الفئة السابقة للفئة المنوالية فإن قيمة المنوال تكون قريبة من الحد الأعلى للفئة المنوالية . أما إذا تساوت القوتان أي تساوي التكرار السابق واللاحق للفئة المنوالية فإن قيمة المنوال تكون عند مركز الفئة المنوالية .

فإذا رمزنا لتكرار الفئة المنوالية بالرمز $F_M$ ولتكرار الفئة السابقة للفئة المنوالية بالرمز $F_{M-1}$ ولتكرار الفئة اللاحقة للفئة المنوالية بالرمز $F_{M+1}$ وللحد الأدنى للفئة المنوالية بالرمز $(l.B)_M$ ولطول الفئة المنوالية بالرمز $L_M$ . فإن يمكن تصور شكل الرافعة من الدرجة الأولى كما يوضحها الشكل ( ٣-٦ ) . حيث يمثل هذا الشكل خطاً أفقياً يمثل رافعة طولها يساوي الفئة المنوالية أي ($L_M$) . فإذا اعتبرنا أن نقطة التوازن أو الارتكاز هي بمثابة نقطة المنوال . وبافتراض أن نقطة التوازن ( أي قيمة المنوال ) تبعد مسافة قدرها ($x$) عن بداية الفئة المنوالية فيكون بعد المنوال عن نهاية الفئة المنوالية ما هو إلا طول الفئة المنوالية أي $L_M$ مطروحاً منها البعد ($x$) .

شكل (٣-٦)




هذا ويتطبيق قانون الرافعة والذي يحوي في مؤداه كما نعلم : -

**القوة × ذراعها = المقاومة × ذراعها**

**نستطيع الحصول على البعد ( x ) كالآتي :-**

$$X \times (F_{M-1}) = (L_M - X) \times (F_{M+1}) \qquad (I)$$

**أي أن :**

$$x\,(F_{M-1} + F_{M+1}) = L_M(F_{M+1})$$

**ومنها فإن :**

$$X = \frac{F_{M+1}}{F_{M-1} + F_{M+1}} \times (L_M) \qquad (II)$$

**فيكون قيمة المنوال ما هي إلا بداية الفئة المنوالية مضافاً إليها قيمة x الناتجة سواء من العلاقة (II) أو من خلال حل المعادلة (I) . أي أن المنوال يكون مساوياً :**

M = L.B. for the Model Class + X

**أو باستخدام الرموز الواردة في العلاقة II فإن :**

$$M = L.B.\ (\text{Model Class}) + \frac{F_{M+1}}{F_{M-1} + F_{M+1}} \times L_M \qquad (III)$$

**هذا ويجب أن ننوه بالتأكيد على أن قبل حساب المنوال بأي من طرق حسابه يجب التحقق من مدى انتظام الجدول من عدمه . بمعنى إذا كان الجدول منتظماً يتم حساب المنوال مباشرة وإلا يجب حساب التكرارات المعدلة ( في حالة عدم الانتظام ) ثم حساب المنوال .**

**وهذه الطريقة وإن كانت تتميز بأنها تأخذ في الاعتبار التكرار السابق والتكرار اللاحق للفئة المنوالية إلا أنها تتجاهل تكرار الفئة المنوالية ذاتها**

**<u>مثال (١٨) :</u>**

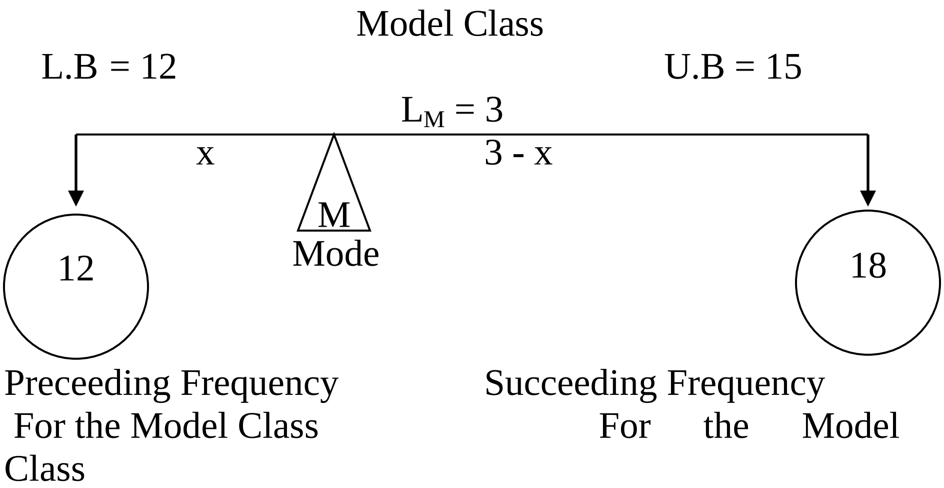
**احسب قيمة المنوال باستخدام طريقة الرافعة للتوزيع التكراري المعطى في مثال (١٦) .**

**<u>الحـل</u>:- بدايـة حيـث أن التوزيـع التكـراري الـوارد فـي مثـال (١٦) توزيعـاً منتظماً لذا يتم حساب المنـوال مباشرة . وباستخدام طريقـة الرافعة فإنـه من خلال الجدول (٣-٢٠) يتضح لنا الآتي :**

- **الفئـة المنوالية** Model Class = ( 12-15 )
- **الحد الأدنى للفئـة المنوالية** L . B for the Model Class = 12
- **طول الفئـة المنوالية** $L_M$ = 3
- **التكرار السابق للفئـة المنوالية** $F_{M-1}$ = 12
- **التكرار اللاحق للفئـة المنوالية** $F_{M+1}$ = 18

**وبافتراض أن المنوال (M) هو نقطة التوازن لرافعة من الدرجة الأولى ويبعد مسافة قدرها (x) عن بداية الفئـة المنوالية فنكون بصدد الشكل التالي:**

**شكل ( ٣ – ٧ )**




**فيكون :**

$$12X = 18(3 - X)$$
$$30x = 45$$
i.e., $X = 54 \div 30 = 1.8$ ( Month )

**فتكون قيمة المنوال :**

$$M = L.B \text{ for ( Model Class )} + X$$
$$= 12 + 1.8 = 13.8 \quad \text{( Month)}$$

**<u>مثال (١٩) :-</u>**

**احسب قيمة المنوال باستخدام طريقة الرافعة للتوزيع التكراري المعطى في مثال (١٧) .**

**<u>الحل :-</u>**

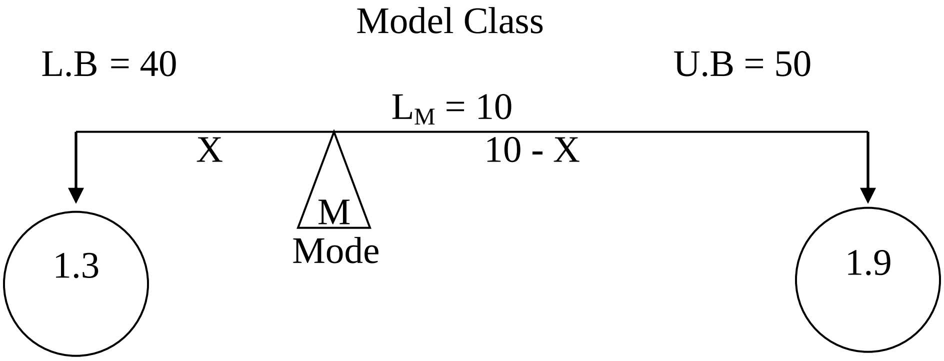
**بداية حيث أن التوزيع التكراري المعطى في مثال (١٧) توزيعاً تكرارياً غير منتظم لذا يجب حساب التكرارات المعدلة أولاً قبل حساب قيمة المنوال .**

**لذا يجب استخدام بيانات الجدول ( ٣ – ٢٢ ) في تحديد البيانات اللازمة لتحديد قيمة المنوال ومن هذا الجدول فإن :**

Model Class = ( 40 – 50 )
$L_M$ = 10
$F_{M-1}$ = 1.3
$F_{M+1}$ = 1.9
L . B for the Model Class = 40

**وعليه فإنه يمكن تصور صور الرافعة في الشكل ( ٣–٨) . فبافتراض أن المنوال (M) هو نقطة التوازن لرافعة من الدرجة الأولى ويبعد مسافة قدرها (x) عن بداية الفئة المنوالية فنكون بصدد الشكل التالي :**

**شكل ( ٣ – ٨ )**




**ولاستنتاج قيمة X فحيث أن :**

**القوة × ذراعها = المقاومة × ذراعها . لذا فإن :**

$$1.3\,X = 1.9\,(10 - X)$$
$$3.2\,x = 19$$
$$X = 19 \div 3.2 = 5.9375$$

**وعليه تكون قيمة المنوال هي :**

$$M = L . B ( \text{Model Class} ) + X$$
$$= 40 + 5.9375$$
$$= 45.9375 = 45.94 \quad ( \text{Degree} )$$

**لاحظ أنه في مثالنا هذا وبمجرد النظر إلى بيانات الجدول بعد تعديل التكرارات كما هو في جدول ( ٣-٢٢ ) فيمكن استنتاج أن قيمة المنوال لابد وأن تزيد عن ٤٥ درجة ( وهو بمثابة مركز الفئة المنوالية ) وذلك لأن التكرار المعدل اللاحق للفئة المنوالية أكبر من التكرار المعدل السابق للفئة المنوالية .**

**ثالثاً : طريقة الفروق لبيرسون :-**

**قلنا سابقاً أن من أهم ما يؤخذ على طريقة الرافعة بأنها تأخذ في الاعتبار كلاً من التكرار السابق واللاحق للفئة المنوالية عند تحديدها لقيمة المنوال مع إهمال تكرار الفئة المنوالية ذاتها بالرغم من أنه أكثر تأثيراً في تحديد قيمة المنوال من تكراري الفئتين السابقة واللاحقة للفئة المنوالية . لذلك جاء كارل بيرسون واعتبر أن الفئة المنوالية كما لو كانت رافعة تتجاذبها قوة تتمثل في الفرق ما بين تكرار الفئة المنوالية ( $F_M$) والتكرار السابق للفئة المنوالية ( $F_{M-1}$) ومقاومة عبارة عن الفرق ما بين تكرار الفئة المنوالية وبين تكرار الفئة اللاحقة للفئة المنوالية ( $F_{M+1}$ ) ويحدث التوازن عند النقطة التي تمثل قيمة المنوال ( $M$ ) .**

**فإذا اعتبرنا أن الفرقة الأولى $D_1$ هو عبارة عن القوة أي أن :**

$$D_1 = F_M - F_{M-1}$$

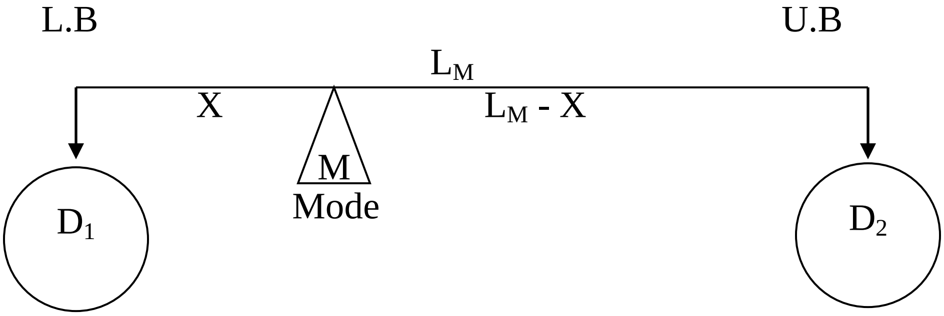
**وأن الفرق الثاني $D_2$ هو عبارة عن المقاومة أي أن : -**

$$D_2 = F_M - F_{M+1}$$

**فتأخذ الرافعة الشكل البياني التالي :-**

**شكل ( ٣-٩ )**

Model Class




**فيكون :**

$$\frac{X}{D_1} = \frac{L_M - X}{D_2}$$

**ومنها تكون :**

$$X = \frac{D_1}{D_1 + D_2} \times (L_M)$$

**وعليه تكون قيمة المنوال عبارة عن بداية الفئة المنوالية مضافاً إليها بعد المنوال عن بداية الفئة المنوالية أي :**

$$M = L.B \text{ for ( Model Class )} + X$$

$$= L.B(M) + \frac{D_1}{D_1 + D_2} \times L_M \qquad \text{(IV)}$$

**والأمثلة التالية توضح كيفية حساب المنوال باستخدام طريقة الفروق لبيرسون .**

**<u>مثال (٢٠) :</u>**

**احسب قيمة المنوال باستخدام طريقة الفروق للتوزيع التكراري المعطى في مثال (١٦) .**

**الحل :-**

**بدايـة حيـث أن التوزيـع التكـراري منتظمـاً لـذا يمكـن حسـاب المنـوال مباشـرة وحيث أن من خلال الجدول ( ٣-٢٠ ) يتضح لنا البيانات التالية :**

Model Class = ( 12 – 15 )

$L_M$ = 3

$D_1$ = 20 – 12 = 8

$D_2$ = 20 – 18 = 2

**هذا ومن خلال العلاقة** IV **فإن قيمة المنوال تكون عبارة عن :**

$$M = 12 + 3 \left( \frac{8}{8 + 2} \right)$$

$$= 12 + 2.4$$

$$= 14.4 \qquad (\text{Month})$$

**مثال (٢١) :**

**احسب قيمة المنوال باستخدام طريقـة الفروق للتوزيـع التكراري المعطى فـي مثال ( ١٧ ) .**

**الحل :-**

**بملاحظة التوزيع التكراري الوارد في الجدول (٣-٢١ ) نجد أنـه توزيعاً تكرارياً غير منتظم لذا يجب الحصول على التكرارات المعدلة قبل حساب قيمة المنوال . لذا فمن بيانات التكرارات المعدلة الموضحة في جدول ( ٣-٢٢ ) يتضح لنا أن أكبر تكرار معدل هو (٢.١) وعليه فإن :**

**التالية :**

Model Class = ( 40 - 50 )

$L_M$ = 10

$D_1 = F_M - F_{M-1} = 2.1 – 1.3 = 0.8$

$D_2 = F_M - F_{M+1} = 2.1 – 1.9 = 0.2$

**وعليـه يـتم اسـتنتاج المنـوال مـن الصـيغة التاليـة باسـتخدام طريقـة الفـروق لبيرسون والتي تأخذ الصورة :**

$$M = L.B \text{ for ( Model Class )} + L_M \times \left( \frac{D_1}{D_1 + D_2} \right)$$

$$= 40 + 10 \left( \frac{0.8}{0.8 + 0.2} \right)$$

$$= 40 + 8 = 48 \qquad \text{( Degee )}$$

**ملاحظات :-**

١- بمراجعة نتائجنا في الأمثلة السابقة الخاصة بالمنوال لاحظ أن كل طريقة من الطرق الثلاث ( مركز الفئة المنوالية – الرافعة – الفروق ) أعطت قيمة مختلفة للمنوال وهذا أمراً طبيعياً ويرجع ذلك إلى أن هذه التوزيعات التي تم إيجاد المنوال لها توزيعات غير متماثلة فالاختلاف ما بين النتائج يرجع لاختلاف طرق حساب المنوال في كل واحدة عن الأخرى ونتيجة لعدم تساوي تكرار الفئة السابقة للفئة المنوالية مع تكرار الفئة اللاحقة لها أدى ذلك إلى اختلاف النتائج فيما بين الطرق الثلاث . وتتساوى نتائج الطرق الثلاث في حالتي التوزيعات المتماثلة أو التوزيعات التي تتساوى فيها التكرار الأصلية ( في حالة الجداول المنتظمة ) السابقة واللاحقة للفئة المنوالية أو التكرارات المعدلة السابقة واللاحقة للفئة المنوالية ( في حالة الجداول الغير منتظمة ) .

٢- يمكن القول أن طريقة مركز الفئة المنوالية هي أقل الطرق الثلاثة التي تحدد قيمة المنوال من حيث الدقة . وأن طريقة الفروق لبيرسون تُفضل على طريقة الرافعة من حيث تقديرها لقيمة المنوال وذلك لأن طريقة الفروق تستخدم معلومات أكثر من تلك التي تستخدمها طريقة الرافعة في تحديد قيمة المنوال .

٣- إذا تساوى التكرار السابق للفئة المنوالية أي ($F_{M-1}$) مع التكرار اللاحق للفئة المنوالية أي ($F_{M+1}$) فإن بعد المنوال عن بداية الفئة المنوالية

( أي قيمــة (X) ) ســوف يأخــذ نفــس القيمــة فــي الطــرق الــثلاث . وبالتالي فإن كل من هذه الطرق سيعطي نفس القيمة للمنوال وتكون قيمة ( X ) في هذه الحالة عبارة عن خارج قسمة طول الفئة المنواليـة أي ( $L_M$ ) على (2) . وبمعنى أكثر بساطة إذا تساوى التكرار السابق واللاحق للفئة المنوالية فإن قيمة المنوال للطرق الثلاث تكون متساوية وتقدر قيمة المنوال في هذه الحالـة بمركز الفئـة المنواليـة . أي أنـه إذا كان :

$$F_{M-1} = F_{M+1}$$

فإن قيمةالمنوال تكون عبارة عن مركز الفئة المنوالية. أى أن :

$$\text{Mode} = \left( \frac{(L . B + U.B) \text{ for the Model Class}}{2} \right)$$

**(٣-٣-٢) إيجاد قيمة المنوال بيانياً من خلال رسم المدرج التكراري :**

رأينا فيما سبق أن حساب المنوال يعتمد على محددات رئيسية وهي الفئة المنوالية وتكرار الفئـة المنواليـة وتكرار الفئـة السابقة للفئـة المنواليـة وتكرار الفئـة اللاحقـة للفئـة المنواليـة . لذلك فإنـه لإيجـاد المنـوال يراعى أولاً تحديد طبيعة المتغير أو الظاهرة محل الدراسة . فإذا كان المتغير الإحصائي متغيراً منفصلاً وموزعاً توزيعاً تكرارياً فإن المنـوال يكـون عنـد تلاقـي العمـود الـذي يأخذ أعلى تكرار في فئات أو أقسام الظـاهرة مـع المحـور الأفقـي ليبين قيمـة الظاهرة التي تقابل أكبر تكر ( أي المنوال ) .

أما في حالة المتغيرات الكمية المتصلة فإنـه لتحديد قيمـة المنـوال بيانيـاً يـتم رسم المدرج التكراري للتوزيع المعطى أو بالتحديد يكتفـي رسـم الفئـات الـثلاث التـي تشـكل أهميـة عنـد تحديـد المنـوال وهـي رسـم الفئـة المنواليـة والفئـة السـابقة للفئـة المنواليـة والفئـة اللاحقـة لهـا . وهنـا يكـون لـدينا ثـلاث مستطيلات متلاصقة أوسطها يعبر عن الفئـة المنواليـة والسـابق لهـا يمثـل

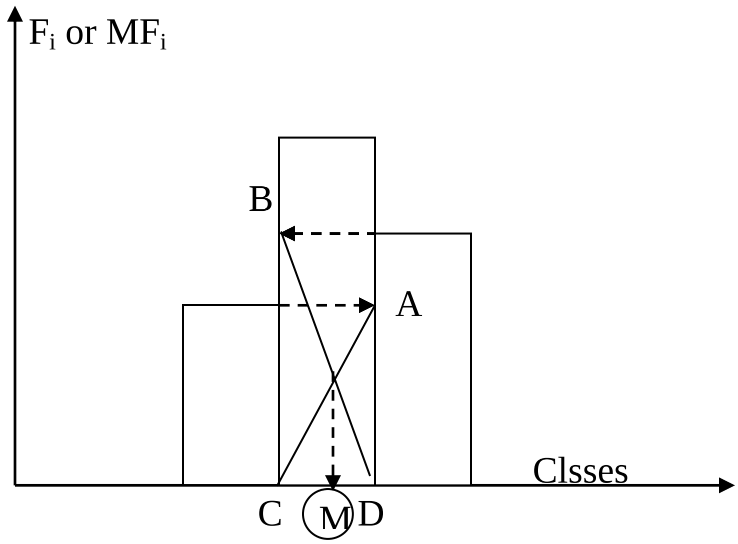
**الفئة السابقة للفئة المنوالية واللاحق لها يمثل الفئة اللاحقة للفئة المنوالية فيكون :**

**أ- <u>وفقاً لطريقة الرافعة :-</u>**

**بعد رسم الفئات الثلاث محل الاهتمام في المدرج التكراري ( الفئةالمنوالية والفئة السابقة والفئة اللاحقة للفئة المنوالية ) حسب مقياس الرسم المناسب يتم استكمال قيم المستطيلات الممثلة للفئتين السابقة واللاحقة للفئة المنوالية بخطوط مستقيمة في اتجاه الفئة المنوالية إلى أن يقطعا هذه الفئة في نقطتين ولتكن A . B على الترتيب . ثم يتم توصيل النقطة (A) ببداية قاعدة مستطيل الفئة المنوالية على المحور الأفقي ولتكن النقطة (C) وكذلك يتم توصيل النقطة (B) بنهاية قاعدة مستطيل الفئة المنوالية على نفس المحور ولتكن النقطة (D) . يتقاطع الخطان (AC) و (BD ) في نقطة ولتكن (E) نسقط منها عموداً على المحور الأفقي إلى أن يقطعه في نقطة لتحدد هذه النقطة قيمة المنوال M بطريقة الرافعة كما هو موضح في الشكل (٣-١٠) .**

**( شكل ٣-١٠)**

**تحديد قيمة المنوال بطريقة الرافعة بيانياً**

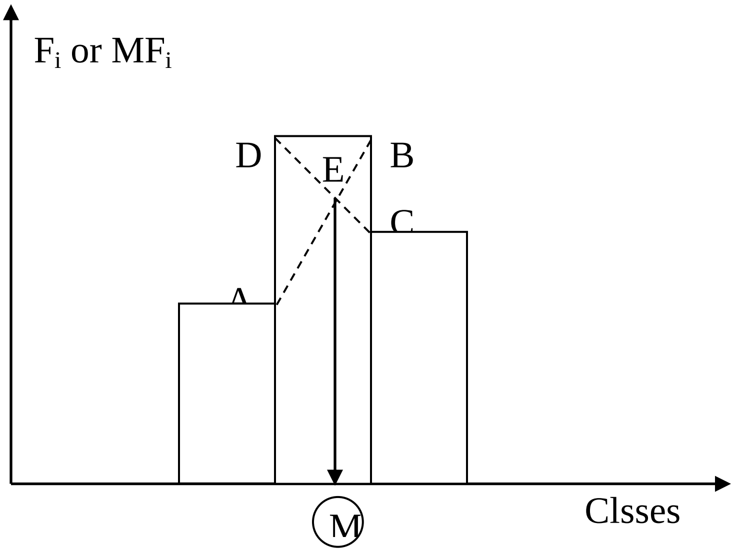


**<u>ب- وفقاً لطريقة الفروق :-</u>**

**يتم توصيل القمم العليا للمستطيلات الثلاثة المتلاصقة بخطين مستقيمين بطريقة عكسية . حيث يتم توصيل نهاية قمة المستطيل الممثل للفئة السابقة للفئة المنوالية بنهاية قمة المستطيل الممثل للفئة المنوالية بخط مستقيم . ثم يتم توصيل بداية قمة الفئة المنوالية ببداية قمة الفئة اللاحقة للفئة المنوالية بخط مستقيم آخر . يتقاطع الخطين المستقيمين المرسومين في نقطة يتم إسقاط منها عموداً على المحور الأفقي ليقطع هذا العمود المحور الأفقي عند نقطة تمثل قيمة المنوال كما يوضحه شكل (٣-١١) .**

**شكل (٣-١١)**

**إيجاد قيمة المنوال بطريقة الفروق بيانياً**




**وعلى الطالب إيجاد قيمة المنوال بيانياً للمثالين (١٦) ، (١٧) باستخدام كل من طريقتي الرافعة والفروق .**

**<u>مثال (٢٢) :</u>**

**فيما يلي لديك الجدول التالي يعبر عن توزيع الدخل اليومي (بالجنيه) لمجموعة من العاملين بإحدى الشركات**

جدول ( ٣- ٢٤ )

| Classes | 5- | 10- | 20- | 40- | 50-60 |
|---|---|---|---|---|---|
| Frequency | 30 | 150 | 240 | 50 | 20 |

والمطلوب إيجاد قيمة المنوال حسابياً بكافة الطرق المستخدمة لحسابه .

**الحل :-**

بملاحظة الجدول التكراري المعطى نجد أنه توزيعاً تكرارياً غير منتظم لذا يجب الحصول على التكرارات المعدلة قبل حساب المنوال . والجدول (٣-٢٥ ) يعطي الحسابات اللازمة لحساب التكرارات المعدلة .

| Classes | $F_i$ | $L_i$ | Modified Frequency ($MF_i$ ) | |
|---|---|---|---|---|
| -5 | 30 | 5 | 30÷5 = 6 | $F_{M-1}$ |
| -10 | 150 | 10 | 15 | $F_M$ |
| -20 | 240 | 20 | 12 | $F_{M+1}$ |
| -40 | 50 | 10 | 5 | |
| 50-60 | 20 | 10 | 2 | |

ومن خلال الجدول (٣-٢٥ ) فإن حساب المنوال باستخدام :-

**أولاً : طريقة مركز الفئة المنوالية فإن :**

$$\text{Mode(M)} = \left( \frac{\text{( L . B + U.B) for the Model Class}}{2} \right)$$

وحيث أن أكبر تكرار معدل هو (١٥) لذا فإن الفئة المنوالية هي :-

Model Class = ( 10 - 20 )

وعليه فإن المنوال :

$$M = \frac{(10 + 20)}{2} = 15 \qquad \text{(L.E)}$$

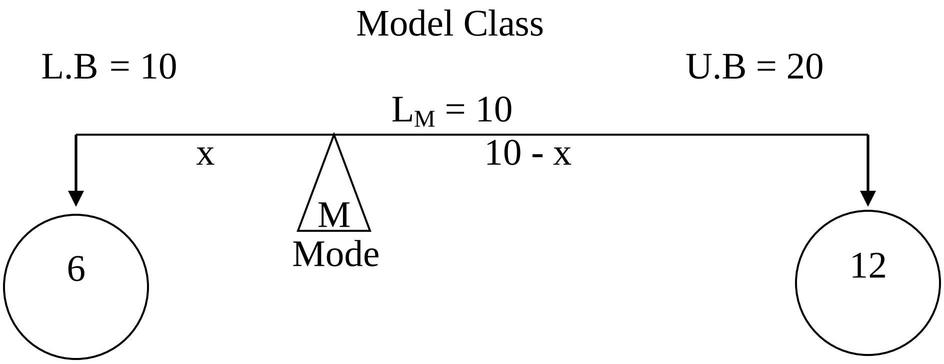
**ثانياً : طريقة الرافعة فإن :-**

من خلال الجدول السابق يتضح لنا أن :

Model Class = ( 10- 20 ) ,
L . B for the Model Class = 10 ,
$F_{M-1}$ = 6 ,
$F_M$ = 15 ,

$F_{M+1} = 12$

وعليه فإنه يمكن تصور الرسم الكروكي التالي :




وطبقاً لقاعدة الرافعة فإن :

$6x = 12(10 - x)$

i.e,. $18x = 120$

$$X = \frac{120}{18} = 6.667 \qquad \text{(L.E)}$$

M = L . B for the Model Class + x

= 10 + 6.667

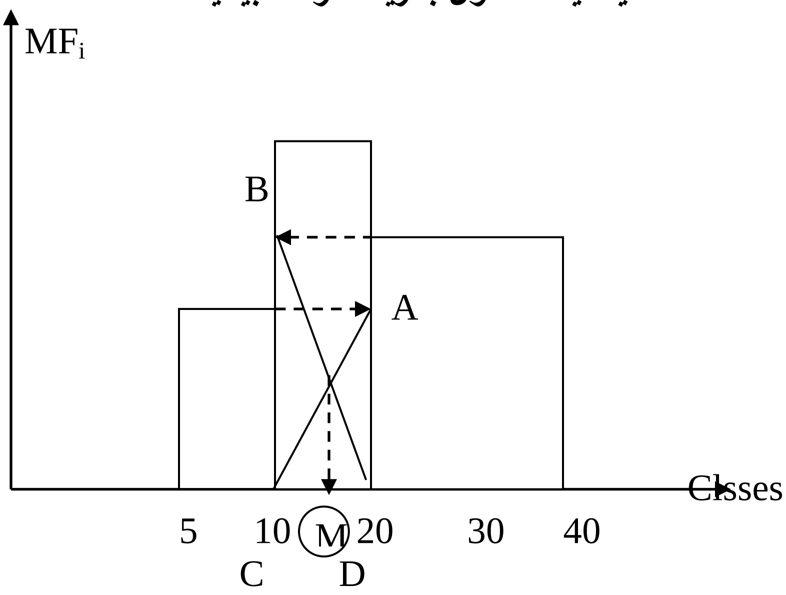
= 16.667 (L.E)

وبيانياً باستخدام طريقة الرافعة يتم رسم المدرج أو بالتحديد الفئات محل الاهتمام ( المنوالية والسابقة واللاحقة للفئة المنوالية ) . والرسم التالي يوضح كيفية حساب المنوال .

**شكل (٣-١٢)**

**تحديد قيمة المنوال باستخدام طريقة الرافعة بيانياً**

**تحديد قيمة المنوال بطريقة الرافعة بيانياً**




**ومن الرسم البياني فإن قيمة المنوال هي** $M = 16.667$

**ثالثاً :- طريقة الفروق لبيرسون فإن :-**

L . B for the Model Class = 10 ,

$F_{M-1} = 6$ ,

$F_M = 15$ ,

$F_{M+1} = 12$ ,

$D_1 = 15 - 6 = 9$ ,

$D_2 = 15 - 12 = 3$

**ومن ثم فتكون قيمة المنوال هي :-**

$$M = L.B \text{ for the Model Class} + L_M \left( \frac{D_1}{D_1 + D_2} \right)$$

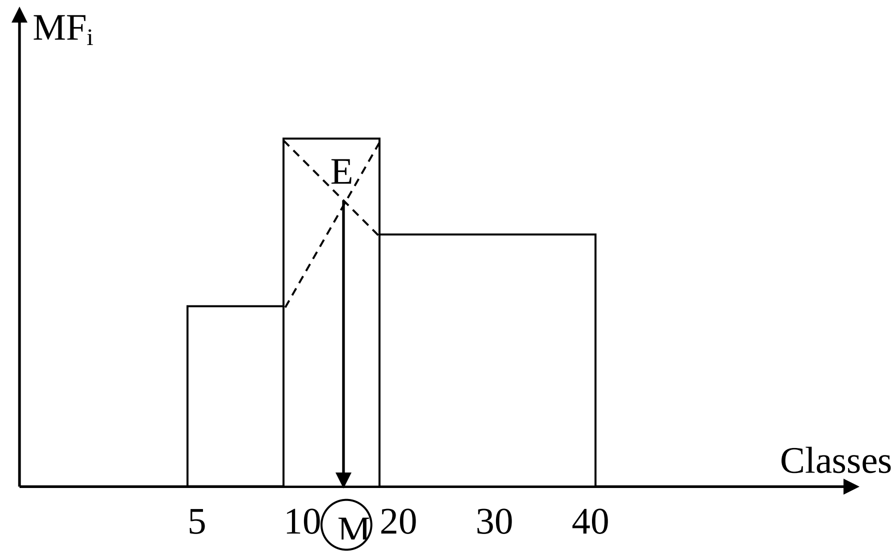
$$= 10 + 10 \left( \frac{9}{9 + 3} \right)$$

$$= 10 + \frac{90}{12} = 17.5 \qquad (L . E)$$

**أما عن تقدير المنوال بيانياً باستخدام طريقة الفروق فالشكل ( ٣-١٣ ) يوضح إيجاد قيمة المنوال :**

**شكل ( ٣-١٣ )**

**تحديد قيمة المنوال باستخدام الفروق بيانياً**




**ومن الرسم يتضح أن :**

$$M = 17.5 \qquad (L.E)$$

**<u>ملاحظات حول المنوال :</u>**

**١- المنوال هو الصفة أو الظاهرة الغالبة في بيانات وصفية أو رقمية ، والصفة الغالبة تعني أنها الصفة التي تتحقق في عدد من المفردات التي نصنفها يفوق عدد المفردات المحققة لأية صفة أخرى . ولا تعني بالضرورة أنها الصفة الاتي تتحقق في أغلبية العناصر أي في أكثر من ٥٠ % منها .**

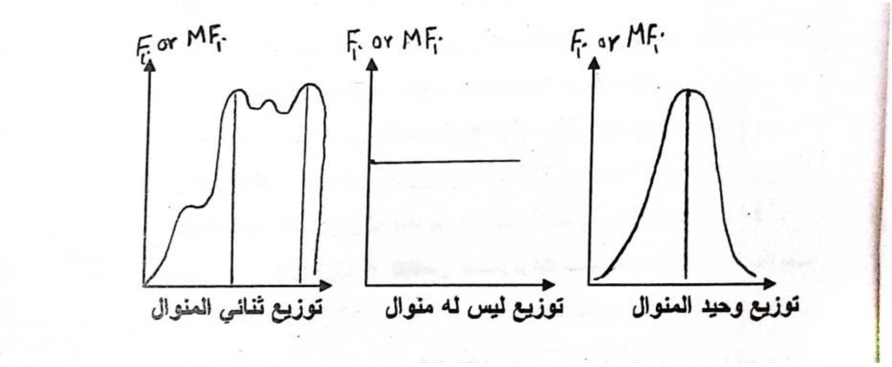
**٢- المنوال هو الصفة أو الظاهرة الأكثر تكراراً وليس تكرار تلك الظاهرة .**

٣- يتميز المنوال بسهولة حسابه خاصة من توزيعات تكرارية إذ أننا نحتاج إلى الفئة المنوالية والفئتين السابقة للفئة المنوالية واللاحقة لها ، ولا يتأثر حسابه بالقيم المتطرفة أو الشاذة .

٤- حينما تكون القياسات نوعية كما هو الحال مثل النوع أو الجنسية أو الديانة فيفضل استخدام المنوال على أساس أنه يمثل الظاهرة الأكثر ظهوراً أو تكراراً .

٥- يعاب على المنوال أنه لا يدخل في حسابه كل مفردات التوزيع ، وتختلف قيمة المنوال باختلاف طريقة حسابه وإن كانت طريقة الفروق أكثر دقة من طريقة الرافعة لأنها تأخذ في الاعتبار تكرار الفئة المنوالية وتكرار الفئة قبل المنوالية وتكرار الفئة بعد المنوالية عند حساب المنوال.

٦- يمكن إيجاد المنوال من المنحنى التكراري ( أو المضلع التكراري ) حيث يمثل المنوال نقطة النهاية العظمى للمنحنى ( أو المضلع التكراري ) ، فإذا كان للمنحنى ( أو المضلع ) التكراري الذي يمثل توزيع ظاهرة ما نقطة نهاية عظمى فيوجد للظاهرة منوال واحد كما في شكل (٣-١٤ – أ ) . أما إذا كان المنحنى ( أو المضلع ) التكراري ليس له نقطة نهاية عظمى حيث يوجد لكل صفة أو نوع التكرار نفسه كما يتضح في شكل ( ٣-١٤ – ب ) فنقول عندئذ بعدم وجود منوال ولا نقول أن كل صفة أو نوع هي في حد ذاتها منوال . وإذا كان للمنحنى ( أو المضلع ) التكراري قيمتين أو أكثر كما في شكل ( ٣ – ٥ – جـ ) فهذا يعني وجود منوالين أو أكثر لهذا التوزيع . وفي الحالتين الأخرتين ( لايوجد منوال أو يوجد أكثر من منوال ) لا يصلح المنوال لأن يكون مقياساً للنزعة المركزية .



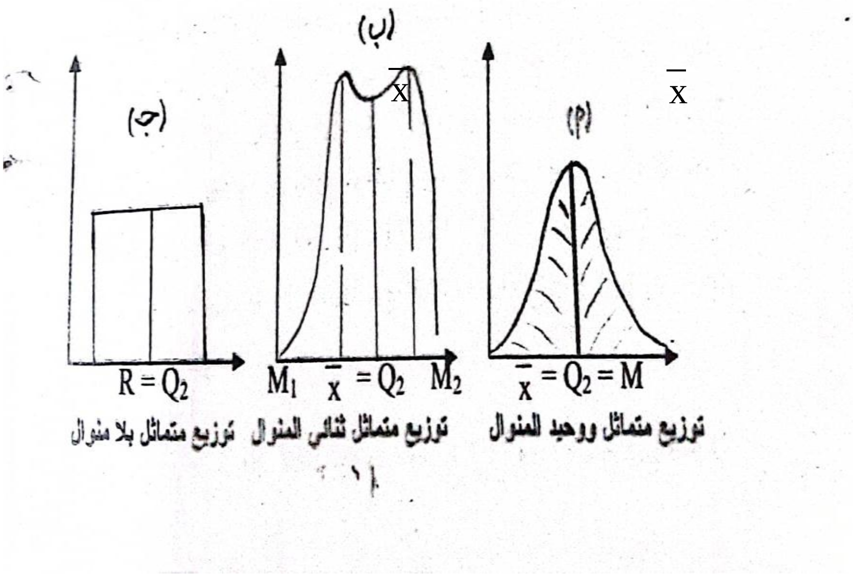

**(٣-٣-٣ ) العلاقة بين الوسط الحسابي والوسيط والمنوال**

**رأينا فيما سبق أن الوسط الحسابي يدخل في حسابه كل القيم لذلك فإنه يتأثر تأثراً بالغاً بالقيم الشاذة أو المتطرفة ، أما الوسيط فيتحدد من خلال المواقع النسبية للقيم بعضها من بعض أي أنه يتحدد من خلال رتب هذه القيم ، ولنأخذ على سبيل المثال القيم** 10 , 11 , 12 ,13 , 14 **فوسطها الحسابي ووسيطها يساوي** 12 **فإذا أضفنا إليها قيمة واحدة ولتكم** 120 **نجد أن الوسط الحسابي**

**سيصبح = ( 180 / 6 ) =** 30 **بينما الوسيط سيصبح** 12.5 **فالوسط الحسابي قد زاد بمقدار** 18 **في حين زاد الوسيط بمقدار** 0.5 **فإضافة المفردة السادسة إلى مجموعة القيم لن تزيد قيمة الوسيط إلا بمقدار نصف مهما كان قيمة تلك المفردة ولكن الزيادة في الوسط الحسابي ستصبح أكثر كلما زادت قيمه المفردة السادسة المضافة . أما المنوال فإنه يتأثر بالقيم الشائعة ويتجه نحو قمة التوزيع . ويقع الوسيط بين الوسط الحسابي والمنوال ، إذ أنه أقل من الوسط الحسابي تأثراً بالقيم الشاذة وأقل من المنوال تأثراً بالقيم الشائعة .**

**وتوجد علاقة بين الوسط الحسابي والوسيط والمنوال لمتغير ما تعتمد على شكل توزيع هذا المتغير ، مؤدى هذه العلاقة هو أنه تتساوى قيم المقاييس الثلاثة الوسط الحسابي والوسيط والمنوال عندما يكون التوزيع متماثلاً**

ووحيد المنـوال . ويكـون التوزيـع متمـاثلاً Symetric Distribution إذا كـان المنحنــى التكـراري لــه ينقسـم عنـد القيمــة المتوسـطة لــه إلــى قسـمين متطابقين ويكون تزايد أو تناقص التكرارات متشابهاً ومنتظماً بطريقة متماثلة على جانبي المحور الرأسي المقام عند القيمة المتوسطة كما يتضح لنـا مـن الشكل (٣–١٥– أ ) . وقد يكون التوزيع متماثلاً




شكل (٣ – ١٥ )

وثنائي المنوال وفي هذه الحالـة فإن الوسط الحسابي يسـاوي الوسيط بينمـا يوجد منـوالين يختلفـان عن قيمـة الوسط الحسابي والوسيط كمـا يتضـح مـن الشـكل ( ٣ – ١٥– ب ) . وقـد يكـون التوزيــع متمـاثلاً بـلا منـوال فيظـل الوسط الحسابي مساوياً للوسيط كما في الشكل ( ٣–١٥–جـ ) .

وفـي حالـة التوزيعات الملتويـة حيث تميل التكرارات الكبيـرة إلـى التركـز فـي جانــب مـن المتوسـط وتنتشـر التكـرارات الصـغيرة بعيـداً علـى شـكل ذيـل فـي الجانب الآخر منـه ويكون اتجاه الذيل هو اتجـاه الالتـواء ، فـإذا كـان الـذيل على اليسار قلنا أن التوزيع ملتو إلى اليسار أي سالب الالتواء Negative Skewness أما إذا كان الذيل على اليمين قلنا أن التوزيع ملتو إلـى اليمين

أي موجب الالتواء Positive Skewness وفي التوزيعات الملتوية يقع الوسيط دائماً بين الوسط الحسابي والمنوال . والوسط الحسابي – كما رأينا – شديد الحساسية للقيم الشاذة لذلك نراه مائلاً إلى اتجاه الذيل بينما يرشح المنوال نفسه عند قمة التوزيع .

فإذا كان التوزيع ملتوياً جهة اليسار ( أي سالب الالتواء ) حيث تميل التكرارات الكبيرة إلى التركز عند فئات التوزيع العليا ويكون اتجاه الذيل على اليسار كما يتضح من الشكل (٣–١٦–أ ) فإن وضع مقاييس النزعة المركزية الثلاثة كالآتي :

الوسط الحسابي < الوسيط < المنوال

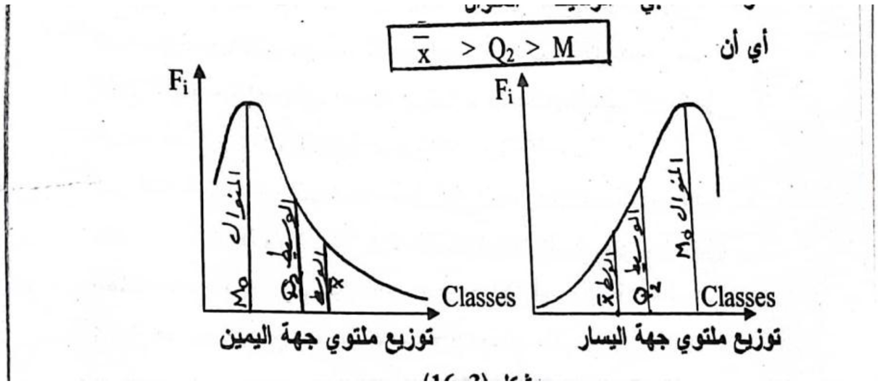
أي أن $\overline{x} < Q_2 < M$

أما إذا كان التوزيع ملتوياً جهة اليمين ( أي موجب الالتواء ( حيث تمثل التكرارات الكبيرة إلى التركز عند فئات التوزيع الدنيا ويكون اتجاه الذيل على اليمين كما يتضح من الشكل ( ٣–١٦ – ب ) فتكون العلاقة بين مقاييس النزعة المركزية الثلاثة هي :

الوسط الحسابي > الوسيط > المنوال

أي أن : $\overline{x} > Q_2 > M$

شكل (٣ – ١٦)



وعلى ذلك فإن هناك علاقة بين قيمة مقاييس النزعة المركزية الثلاثة وشكل التوزيع كالآتي :

فإذا كان

الوسط الحسابي – الوسيط < صفر يكون التوزيع سالب الالتواء

= صفر يكون التوزيع متماثلاً

أو الوسط الحسابي – المنوال > صفر يكون التوزيع موجب الالتواء

وفي حالة التوزيعات القريبة جداً من التماثل تتحقق العلاقة الآتية بين المقاييس الثلاثة السابقة وهي :

الوسط الحسابي – المنوال = ٣( الوسط الحسابي – الوسيط )

وإذا استخدمنا الرموز فإن :

$$\bar{x} - M = 3\,(\bar{x} - Q_2)$$

وتفيد هذه العلاقة في تقدير قيمة المنوال بطريقة ثالثة بالإضافة إلى طريقتي الرافعة والفروق كما تمكن من تقدير قيمة مقربة للوسط الحسابي في حالة التوزيعات التكرارية المفتوحة بعد معرفة قيمة الوسيط والمنوال لهذه التوزيعات بشرط أن تكون هذه التوزيعات قريبة من التماثل .

**مثال ( ٢-٢٠ )**

إذا علمت أن الجدول الآتي يمثل توزيع تكراري قريب من التماثل :

| Classes | Less than 6 | 6- | 12- | 18- | 24 and more | $\sum$ |
|---|---|---|---|---|---|---|
| $F_i$ | 8 | 20 | 40 | 25 | 7 | 100 |

والمطلوب تقدير قيمة الوسط الحسابي .

**الحل :**

الجدول التكراري مفتوح من أعلى ومن أسفل لذلك لا يمكن إيجاد الوسط الحسابي مباشرة ، وحيث أن التوزيع قريب من التماثل كما يتضح من خلال توزيع التكرارات (أو كما هو معطى) فيمكن إيجاد الوسيط والمنوال ثم نستخدم العلاقة السابقة لايجاد الوسط الحسابي .

| Classes | | Less than th U.L | Asc.C.F |
|---|---|---|---|
| Less than 6<br>6-<br>12-<br>18-<br>-24and more | 8<br>20<br>40<br>25<br>7 | Less than 6<br>Less than 12<br>Less than 18<br>Less than 24<br>Less than the U.L<br>For the lost Class | 8<br>28<br>78<br>93<br>100 |
| $\sum$ | 100 | | |

**ولايجاد الوسيط فإن :**

**ترتيب الوسيط**

$$\text{Ran of } (Q_2) = \frac{\sum F}{2} = \frac{100}{2} = 50$$

**ومن ثم تكون قيمة الوسيط هي :**

$$Q_2 = 12 + 6 \left(\frac{50 - 28}{68 - 28}\right) = 15.3$$

**أما لإيجاد المنوال وليكن بطريقة الفروق فنعلم أن :**

$F_M = 40$ ,

Model class = 12 – 18 ,

$F_{M-1} = 20$ ,

$F_{M+1} = 25$ ,

$D_1 = F_M - F_{M-1} = 40 - 20 = 20$ ,

$D_2 = F_M - F_{M+1} = 40 - 25 = 15$

**ومن ثم فإن :-**

$$M = \text{L . B for the Model Class} + L_M \left(\frac{D_1}{D_1 + D_2}\right)$$

$$= 12 + 6\left(\frac{20}{20+15}\right) = 15.43$$

وحيث أن التوزيع قريب من التماثل لذا يصح تطبيق قاعدة بيرسون التالية:

$$\bar{x} - \underline{M} = 3(\bar{x} - Q_2)$$

i.e., $\bar{x} - 15.43 = (3(x-153)$

وبحل المعادلة ينتج أن : $\bar{x} = 15.235$

وتجدر الإشارة إلى أن الاختيار بين مقاييس النزعة المركزية الثلاثة ( الوسط الحسابي والوسيط والمنوال ) للتعبير عن القيمة المتوسطة للتوزيع قد يتوقف أحياناً على الغرض الذي نبتغيه من المقياس : ولبيان ذلك نفرض أن شركة تدفع لعينة من موظفيها رواتب شهرية ( بالجنيه ) كما يلي:

400 , 5600 , 400 , 400 , 7200 , 5000 , 400 , 1400

فلهذه القيم نجد أن :

$\bar{x}$ = 2600 جنيهًا ، $Q_2$ = 900 جنيهًا ، M = 400 جنيهًا

وكما هو واضح فإن هذه القيم الثلاث بتعبيرها عن متوسط الرواتب الشهرية بهذه الشركة تقدم انطباعات مختلفة تمامًا ، وأن كلا من الوسط الحسابي والمنوال لا يعبران بموضوعية عن حقيقة الرواتب في الشركة ، ولو أن مراقبًا للحسابات أراد أن يظهر الشركة بمظهر الذي يدفع رواتب متدنية جدًا في المتوسط لاختار المنوال وهو 400 جنبهًا مقياسًا للنزعة المركزية ، في حين أن مدير الشركة سيختار الوسط الحسابي وهو 2600 جنيهًا مقياسًا للنزعة المركزية ليثبت أن رواتب الشركة مرتفعة ، ولو أن باحثًا محايدًا أراد أن يعبر بموضوعية عن حقيقة الرواتب بالشركة فسيختار الوسيط وهو 900 جنيهًا مقياسًا للنزعة المركزية .ويظل الوسط الحسابي في أغلب الأحوال مقياسًا هامًا للنزعة المركزية يتمتع بخصائص جيدة تؤهله لأن يستخدم على نطاق واسع في علم الإحصاء كما سيتضح لنا

**فيمـا بعـد ، ولـو أخـذنا عينـات مختلفـة ذات أحجـام متسـاوية مـن مجتمـع الدراسـة وحسـبنا لكـل عينــة الوسـط الحسـابي والوسـيط والمنـوال لوجـدنا أن التغيير من عينـة إلى أخرى هو أقل في قيم الوسط الحسابي منـه في قيم كل من الوسيط والمنوال ، ممـا يعنـي أن الوسط الحسـابي أكثـر اسـتقرارًا مـن كـل من الوسيط والمنوال عبر عينات نسحبها من مجتمع الدراسة .**

## (٣-٤) الوسط الهندسي Geometric Mean (G.M)

**يمثل الوسط الهندسي أحد مقاييس النزعة المركزية ويفضل استخدامه عند حساب النسب أو المعدلات وكذلك عند تقدير عدد السكان بين سنوات التعداد .**

**ويعرف الوسط الهندسي لمجموعة من القيم بأنه الجذر النوني لحاصل ضرب هذه القيم .**

### (٣-٤-١) الوسط الهندسي في حالة البيانات غير المبوبة :

**إذا كان لدينا المتغير س يأخذ القيم $X_1$ ، $X_2$ ، ........ ، $X_n$ الموجبة والغير صفرية فإن الوسط الهندسي لهذه القيم وسنرمز له بالرمز (G.M) يأخذ الصورة :**

$$(G.M) = \sqrt[n]{x_1 \,.\, x_2 \,.\, x_3 \,\ldots\ldots\, x_n}$$

$$= \sqrt[n]{\pi_{i=1}^{n} X_i}$$

**وبأخذ لوغاريتم الطرفين فإن** $\text{Log}\,(G.M) = {}^{1}/_{n} \sum^{n} \text{Log}\, X_i$

**وعلى ذلك فإن لوغاريتم الوسط الهندسي لمجموعة من القيم هو الوسط الحسابي للوغاريتمات تلك المجموعة من القيم ، ويكون الوسط الهندسي (G.M) هو العدد المقابل Log (G.M) الذي نحصل عليه من جدول الأعداد المقابل للوغاريتمات أو من خلال العملية العكسية للوغاريتم في الآلة الحاسبة ، وبصفة عامة فإن الوسط الهندسي لمجموعة من القيم دائمًا أقل من الوسط الحسابي ويمكن إثبات ذلك رياضيًا على النحو التالي :**

**الوسط الحسابي =** $\bar{x} = \dfrac{x_1 + x_2}{2}$

**أما الوسط الهندسي** $(G,M) = \quad G.M = \sqrt{X_1 X_2}$

**وحيث أن :**

$$(X_1 - X_2)^2 > 0$$

**أي أن :**

$$X_1 - 2\sqrt{X_1X_2} + X_2 > 0$$

$$2\sqrt{X_1X_2} < X_1 + X_2$$

i.e,

$$\sqrt{X_1X_2} < \frac{X_1 + X_2}{2}$$

**أي أن** $G.M < \bar{x}$

**وبذلك يكون الوسط الهندسي أصغر من الوسط الحسابي لنفس المجموعة من القيم .**

**<u>مثال (٣-٢١)</u>**

**احسب الوسط الهندسي للنسب**

43% , 12% , 7% , 6%

**<u>الحل :</u>**

$$G.M = \sqrt[4]{6 \times 7 \times 12 \times 43}$$

$$\begin{aligned} \text{Log}(G.M) &= \tfrac{1}{4}(\text{Log } 6 + \text{Log } 7 + \text{Log } 12 + \text{Log } 43) \\ &= \tfrac{1}{4}(0.7782 + 0.8451 + 1.0792 + 1.6335) \\ &= \tfrac{1}{4}(4.336) = 1.084 \end{aligned}$$

**بالكشف في جدول الأعداد المقابلة للوغاريتمات نجد أن** $G.M = 12.13$

**إي أن :** $(G.M) = 12.13\%$

**في حين أن الوسط الحسابي لهذه النسب = 17 % وهذا يتفق مع ما هو معروف من أن الوسط الهندسي لمجموعة من القيم الموجبة أصغر من الوسط الحسابي لها .**

**مثال (٣-٢٢ )**

**إذا كان تعداد السكان في إحدى المدن عام 1985 هو 1500000 نسمة وكان تعداد السكان بها في عام 1995 هو 2000000 نسمة ، احسب تقدير عدد السكان بالمدينة عام 1990 .**

**الحل : إذا استخدمنا الوسط الحسابي لتقدير عدد السكان نجد أن :**

$$\bar{X} = \frac{1500000 + 2000000}{2} = 1750000$$

**إلا أن هذه النتيجة تعد مضللة لأنها تعني أن الزيادة في عدد السكان متساوية في كل عام وهذا غير صحيح لأنه بازدياد عدد السكان فإن معدل النمو السكاني يتغير بمعدل متزايد لذلك فإنه يفضل استخدام الوسط الهندسي لتقدير عدد السكان حيث نجد أن الوسط الهندسي هو :**

$$G.M = \sqrt{1500000 \times 2000000} = 1732050 \text{ Person}$$

**والمثال التالي يوضح أن الوسط الهندسي يكون أكثر دقة من الوسط الحسابي عندما تكون البيانات في صورة معدلات أو نسب .**

**(٣-٤-٢) الوسط الهندسي في حالة البيانات المبوبة**

**إذا كان لدينا توزيع تكراري وكانت $x_i$ هي مراكز الفئات للتوزيع ، $F_i$ هي التكرارات المناظرة حيث ( $i = 1,2,\ldots, r$ ) ، r هي عدد فئات التوزيع ، فإن الوسط الهندسي لهذا التوزيع هو :**

$$G.M = \sqrt[\sum]{(x_1)^{Fi} \cdot (x_2)^{F2} \ldots\ldots (x_r)^{Fr}}$$

$$\text{Log (G.M)} = {}^{1}/_{\sum Fi} (F_1 \text{ Log } + F_2 \text{ Log } x_2 + \ldots\ldots + F_r \text{ Log } x_r)$$

$$= {}^{1}/_{\sum Fi} (F_1 \text{ Log } x_i )$$

**مثال (٣-٢٤)**

**احسب الوسط الهندسي للتوزيع التكراري التالي :**

| classes | -55 | -65 | -75 | 85-95 | ∑ |
|---|---|---|---|---|---|
| $F_i$ | 10 | 30 | 40 | 20 | 100 |

**الحل :**

**لحساب الوسط الهندسي نكون الجدول التالي :**

| classes | $F_i$ | $x_i$ | $\text{Log } x_i$ | $F_i \log x_i$ |
|---|---|---|---|---|
| 55- | 10 | 60 | 1.7782 | 17.782 |
| 65- | 30 | 70 | 1.8451 | 55.353 |
| 75- | 40 | 80 | 1.9031 | 76.124 |
| 85-95 | 20 | 90 | 1.9542 | 39.084 |
| $\sum$ | 100 | | | 188.343 |

$$\text{Log(G.M)} = \frac{\sum F_i \log x_i}{\sum_i^r F_i} = \frac{188.343}{100} = 1.88343$$

**ومن جدول اللوغاريتمات نجد أن :**

G.M = 76.38

**الوسط التوافقي :** Harmonic Mean (H . M)

**الوسط التوافقي يمثل أحد مقاييس النزعة المركزية ويفضل استخدامه في حساب معدل التغيير أو معدل السرعة بالنسبة للزمن أو بلغة أخرى حينما تكون البيانات في صورة معدلات مرتبطة بالزمن ، ويعرف الوسط التوافقي لمجموعة من القيم بأنه مقلوب الوسط الحسابي لمقلوبات القيم .**

**(٣-٥-١) الوسط التوافقي في حالة البيانات غير المبوبة**

**إذا كان لدينا مجموعة القيم** $x_1, x_2, \ldots, x_n$ **الغير صفرية فإن الوسط التوافقي – ونرمز له بالرمز** (H . M) **يحسب كالآتي :**

$$(H\,.\,M) = \frac{n}{{}^1/_{x_1} + {}^1/_{x_2} + \ldots + {}^1/_{x_n}} = \frac{n}{\sum {}^1/_{x_i}}$$

**مثال (٣-٢٥)**

**احسب الوسط التوافقي والوسط الهندسي للقيم التالية :**

**1 , 5 , 25 , 5 وماذا تستنتج ؟**

**الحل :**

**الوسط التوافقي هو :**

$$(H.M) = \frac{4}{\frac{1}{1} + \frac{1}{5} + \frac{1}{25} + \frac{1}{5}} = \frac{4}{36/25} = 2.78$$

$$(G.M) = \sqrt[4]{1 \times 5 \times 25 \times 5} = \sqrt[4]{625} = 5$$

**ونستنتج أن الوسط التوافقي أصغر من الوسط الهندسي .**

**مثال(٣-٢٦)**

**إذا كانت المسافة من المدينة أ إلى المدينة ب هي 40 كم قطعها القطار بسرعة 80 كم/ساعة والمسافة من المدينة ب إلى المدينة جـ هي 60 كم قطعها القطار بسرعة 90 كم / ساعة والمسافة من المدينة جـ إلى المدينة د هي 30 كم قطعها القطار بسرعة 120 كم/ساعة . احسب متوسط سرعة القطار خلال الرحلة .**

**الحل :**

**إذا أردنا تقدير متوسط سرعة القطار في الساعة باستخدام الوسط الحسابي والوسط التوافقي لمعرفة أيهما أكثر ملائمة فإن :**

**أ – باستخدام الوسط الحسابي**

**متوسط سرعة القطار في الساعة هو:**

$$\overline{X} = \frac{80+90+120}{3} = 96.66 \quad \text{K.M/hour}$$

**<u>ب- باستخدام الوسط التوافقي :</u>**

**فإن متوسط سرعة القطار هو :**

$$X = \frac{3}{{}^{1}/_{80} + {}^{1}/_{90} + {}^{1}/_{120}} = 96.66 \qquad K.M/hour$$

**لمعرفة أي المقياسين أكثر ملائمة نجد أن :**

القطار قطع مسافة من المدينة أ إلى المدينة ب في زمن قدره ( 40 ÷ 80) = 0.5 ساعة.

وقطع المسافة من المدينة ب إلى المدينة جـ في زمن قدره ( ٦٠ ÷ ٩٠= 0.67 ساعة )

وقطع المسافة من المدينة جـ إلى المدينة د في زمن قدره $\frac{30}{120}$ =0.67 ساعة

**وبالتالي فإن :**

**الزمن الإجمالي للرحلة =** 0.5 + 0.67 + 0.25 = 1.42 **ساعة**

**المسافة الإجمالية للرحلة =** 40 + 60 + 30 = 130 **كم /ساعة**

**متوسط سرعة القطار في الساعة =** $\frac{130}{1.42}$ = 91.55 **كم/ ساعة .**

**ويتضح من ذلك أن الوسط التوافقي أكثر ملائمة لقياس متوسط سرعة القطار من الوسط الحسابي ، إذ أنه يعطي تقدير لمتوسط السرعة أكثر قربًا من متوسط السرعة الحقيقية للقطار ، والوسط التوافقي لأي مجموعة من القيم يكون دائمًا أقل من الوسط الهندسي وبالتالي أقل من الوسط الحسابي.**

**أي أن :**

$H . M < G.M < x$

**(٣-٥-٢) الوسط التوافقي في حالة البيانات المبوبة :**

**إذا كان لدينا توزيع تكراري وكانت** $xi$ **هي مراكز الفئات للتوزيع ،** $F_i$ **هي التكرارات المناظرة لها فيتم حساب الوسط التوافقي للتوزيع التكراري كما يلي:**

- **نوجد مقلوب مراكز الفئات ، أي** $^1/_{xi}$
- **نضرب مقلوب مركز كل فئة في التكرار المقابل له ، أي نوجد** $(^1/_{xi})Fi$ .
- **نقسم** $\sum^r_1 F_i$ **على** $\sum^r_1 (^{Fi}/_{xi})$ **فنحصل على الوسط التوافقي**
- **أي أن :**

$$H.M = \frac{\sum^r_1 Fi}{\sum^r_1 (^{Fi}/_{xi})}$$

**مثال ( ٣-٢٧)**

**احسب الوسط التوافقي للتوزيع التكراري التالي :**

| classes | -5 | -15 | -25 | 35-45 | $\sum$ |
|---|---|---|---|---|---|
| Frequency | 20 | 50 | 90 | 40 | 200 |

**الحل :**

**لحساب الوسط التوافقي يتم تكوين الجدول التالي :**

| classes | $F_i$ | $x_i$ | $1/x_i$ | $F_i / x_i$ |
|---|---|---|---|---|
| -5 | 20 | 10 | 1.0 | 2 |
| -15 | 50 | 20 | 05.0 | 2.5 |
| -25 | 90 | 30 | 0.033 | 3 |
| 35-45 | 40 | 40 | 0.025 | 1 |
| $\sum$ | 200 | | | 8.5 |

**فيكون الوسط التوافقي عبارة عن :**

$$H.M = \frac{\sum^r_1 F_i}{\sum^r_1 (^{Fi}/_{xi})} = \frac{200}{8.5} = 23.53$$

**ملاحظات حول الوسط الهندسي والوسط التوافقي**

١- لأي مجموعة من البيانات يلاحظ أن الوسط الحسابي أكبر من أو يساوي الوسط الهندسي والوسط الهندسي أكبر من أو يساوي الوسط التوافقي

أي أن :

الوسط الحسابي ≥ الوسط الهندسي ≥ الوسط التوافقي

وإذا استخدمنا الرموز فإن :

$$\bar{x} \geq G.M \geq H.M$$

٢- كل من الوسط الهندسي والوسط التوافقي يأخذ في الاعتبار كل المفردات عند الحساب وكلاهما أقل تأثراً من الوسط الحسابي بالقيم المتطرفة أو الشاذة ، ويعتبر الوسط الهندسي أنسب المتوسطات في حالة النسب والمعدلات .

٣- يعاب على كل من الوسط الهندسي والوسط التوافقي أن كلاهما صعب في حسابه ولا يمكن إيجاده من الجداول التكرارية المفتوحة حيث لا يمكن فيها تحديد مراكز الفئات ، كما لا يمكن حساب كلاً منهما بالرسم .

**مثال ( ٣-٢٨ )**

فيما يلي توزيع تكراري يبين درجة تلوث الهواء ( مقاسه بالميكروجرام في المتر المكعب ) في ٨٠ مدينة كبيرة في العالم :

| فئات درجة التلوث | ١٠- | ٢٠- | ٣٠- | ٤٠- | ٥٠- | ٦٠- | ٧٠-٨٠ |
|---|---|---|---|---|---|---|---|
| عدد المدن | ٩ | ١١ | ١٦ | ٢٠ | ١٢ | ٨ | ٤ |

**والمطلوب :**

١- حساب الوسط الحسابي والوسط الهندسي والوسط التوافقي والتحقق من أن الوسط الحسابي > الوسط الهندسي > الوسط التوافقي .

٢- حساب الوسيط والمنوال والربيعين الأدنى والأعلى بالحساب وبالرسم .

**٣- ما هي نسبة المدن ( في عينة الدراسة ) التي تزيد فيها درجة التلوث عن ٤٥ ميكروجرام / م٣**

**٤- ما هي درجة التلوث التي تكون عندها ٣٠% من المدن في عينة الدراسة أكثر نظافة ؟ .**

**<u>الحل :</u>**

**١- لحساب الوسط الحسابي والوسط الهندسي والتوافقي يتم تكوين الجدول التالي :-**

| classes | $F_i$ | $x_i$ | $d_i = x_i - 45$ | $D_i = \frac{d_i}{10}$ | $D_i F_i$ | $\text{Log } x_i$ | $F_i \log x_i$ | $\frac{F_i}{x_i}$ |
|---|---|---|---|---|---|---|---|---|
| -10- | 9 | 15 | -30 | -3 | -27 | 1.1761 | 10.5849 | 0.6 |
| 20- | 11 | 25 | -20 | -2 | -22 | 1.3979 | 15.3769 | 0.44 |
| 30- | 16 | 35 | -10 | -1 | -16 | 1.5441 | 24.7056 | 0.4571 |
| 40- | 20 | 45 | 0 | 0 | 0 | 1.6532 | 33.064 | 0.4444 |
| 50- | 12 | 55 | 10 | 1 | 12 | 1.7404 | 20.8848 | 0.2182 |
| 60- | 8 | 65 | 20 | 2 | 16 | 1.8129 | 14.5032 | 0.1231 |
| 70-80 | 4 | 75 | 30 | 3 | 12 | 1.8751 | 7.5004 | 0.0533 |
| | 80 | | | | -25 | | 126.6198 | 2.3361 |

$$X = A + B \left(\frac{\sum Di\ Fi}{\sum Fi}\right) = 45 + 10 \left(\frac{-25}{80}\right) = 41.875$$

**وبفرض أن الوسط الهندسي هو G.M فإن :**

$$\text{Log(G.M)} = \frac{\sum Fi \log xi}{\sum Fi} = \frac{126.6198}{80}$$

**من جدول الأعداد المقابلة للوغاريتمات نجد أن :**

G.M = 38.25 **ميكروجرام/م٣**

**أما عن الوسط التوافقي فإن :**

$$H.M = \frac{\sum F_i}{\sum (F_i / x_i)} = \frac{80}{2.3361} = 34.2451$$ **ميكروجرام/م٣**

**ويتضح لنا من النتائج الثلاث السابقة أن الوسط الحسابي أكبر من الوسط الهندسي والوسط الهندسي أكبر من الوسط التوافقي .**

**٢- لللحساب الوسيط والربيعين الأدنى والأعلى يلزم تكوين الجدول التكراري المتجمع الصاعد :**

| classes | $F_i$ | Less than the class U.L | Asc.c.F |
|---|---|---|---|
| 10- | 9 | Less than 20 | 9 |
| 20- | 11 | Less than 30 | 20 |
| 30- | 16 | Less than 40 | 36 |
| 40- | 20 | Less than 50 | 56 |
| 50- | 12 | Less than 60 | 68 |
| 60- | 8 | Less than 70 | 76 |
| 80-70 | 4 | Less than 80 | 80 |
| $\sum$ | 80 | | |

**لحساب الوسيط أي** $Q_2$ **فإن :**

$$\text{Rank of } (Q_2) = \frac{\sum Fi}{2} = \frac{80}{2} = 40$$

$$Q_2 = 40 + 10\left(\frac{40 - 36}{56 - 36}\right) = 42$$ **ميكروجرام/م٣**

**لحساب الربيع الأدنى** ($Q_2$) **فإن :**

$$\text{Rank of } (Q_1) = \frac{\sum Fi}{4} = \frac{80}{4} = 20$$

**وحيث أن الترتيب ٢٠ يوجد مباشرة في عمود التكرار المتجمع الصاعد فيكون الربيع الأدنى هو القيمة التي تناظر هذا الترتيب من الجدول ، أي أن:**

$Q_1$ = ٣٠ ميكروجرام/م³ .

**لحساب الربيع الأعلى أي $Q_3$ فإن :**

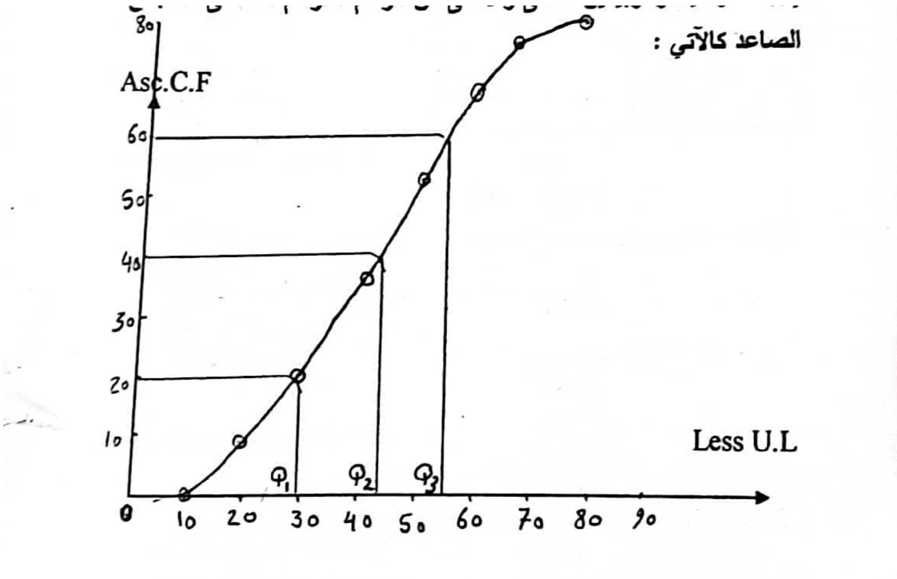
$$\text{Rank of } (Q_3) = \frac{3\sum F}{4} = 3 \times \frac{80}{4} = 60$$

**وعليه تكون قيمة** $Q_3$ **هي : −**

$$Q_3 = 50 + 10\left(\frac{60 - 56}{68 - 56}\right) = 53.33$$ **ميكروجرام/م³**

**لإيجاد الوسيط والربيعين الأدنى والأعلى من الرسم ، نرسم المنحنى المتجمع الصاعد كالآتي :**

Asc.C.F




Less U.L

**ومن الرسم نجد أن :**

$Q_1 = 30$ , $Q_2 = 42$ , $Q_3 = 53.33$

**<u>المنوال :</u> حيث أن الجدول التكراري منتظم فيحسب المنوال من الفئات والتكرارات الأصلية مباشرة دون تعديل للتكرارات ، ويمكن حساب المنوال بطريقة الرافعة أو بطريقة الفروق ، فإذا استخدمنا نجد أن :**

$F_M = 20$

$F_{M-1} = 16$

$F_{M+1} = 12$

**فئة المنوال هي من 40 إلى أقل من 50 أي أن :**

Model class = (40-50)

**ومن ثم فإن الفرق الأول هو**

$D_1 = F_M - F_{M-1} = 20 - 16 = 4$

**والفرق الثاني هو**

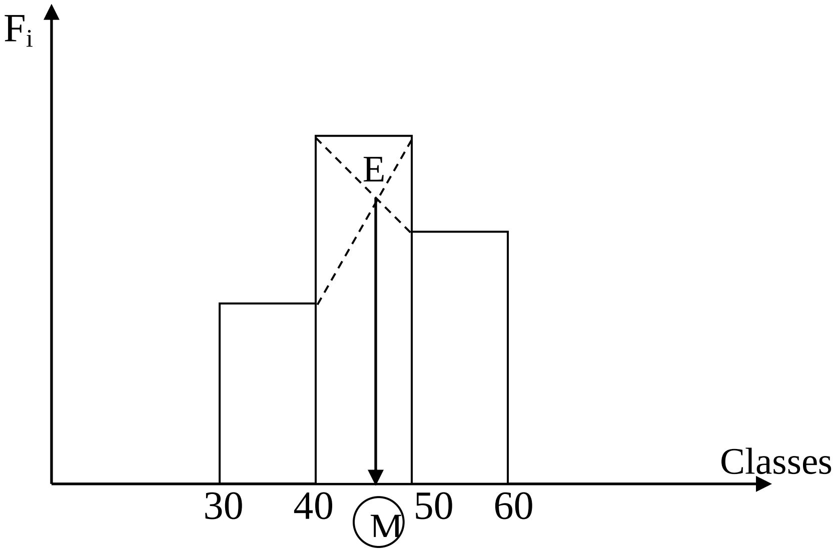
$D_2 = F_M - F_{M+1} = 20 - 12 = 8$

**فتكون قيمة المنوال هي :**

$$\text{Mode} = \text{L . B for the Model Class} + L_M \left( \frac{D_1}{D_1 + D_2} \right)$$

$$= 40 + 10 \left( \frac{4}{4+8} \right) = 43.33$$ **ميكروجرام**

**أما لإيجاد المنوال بيانياً فإن يتم رسم ثلاثة مستطيلات متلاصقة تمثل الفئة المنوالية والفئى السابقة عليها والفئة اللاحقة لها كالآتي :**




**ومن الرسم نستنتج أن :**

Mode = 43.33 **ميكروجرام /م٣**

**٣- لإيجاد مسبة المدن التي تزيد فيها درجة التلوث عن 45 ميكروجرام/م٣ من المنحنى المتجمع الصاعد السابق رسمه على المحور الأفقي وعند النقطة 45 نرسم خطاً رأسياً يقابل المنحنى المتجمع الصاعد في نقطة نرسم منها خطاً أفقياً يقابل المحور الرأسي عند النقطة 47 ، وبالتالي فإن :**

**عدد المدن التي تقل فيها درجة التلوث عن 45 ميكروجرام /م٣= 47 مدينة**

**إذن :**

**عدد المدن التي تزيد فيها درجة التلوث عن 45 ميكروجرام /م٣**

= 80 − 47 = 33 **مدينة** .

**نسبة المدن التي تزيد فيها درجة التلوث عن 45 ميكروجرام /م٣** =

$$\frac{33}{80} \times 100 = 41.25\%$$

**٤- لإيجاد درجة التلوث التي تكون عندها 30% من المدن التي في عينة الدراسة أكثر نظافة فحيث أن نسبة 30% من المدن هي :**

**نسبة 30% من المدن في عينة الدراسة =** $\frac{30}{100} \times 80 = 24$ **مدينة**

**لذا فإنه من المنحنى المتجمع الصاعد نرسم على المحور الرأسي وعند النقطة ٢٤ خطا أفقياً يقابل المنحنى الصاعد في نقطة نرسم منها خطا رأسياً يقابل المنحنى الأفقي عند** النقطة 32 **وبالتالي فإن :**

**درجة التلوث التي عندها 30% من المدن في عينة الدراسة تعد أكثر نظافة هي 32 ميكروجرام /م$^{٣}$ .**

## تمارين الباب الثالث

١- الجدول التالي يبين توزيع الطلاب حسب أيام الغياب خلال العام الماضي

| عددأيام الغياب | ٠ - | ٥- | ١٠- | ١٥- | ٢٠- | ٢٥- | ٣٠- | ٣٥- | ٤٠- | مجـ |
|---|---|---|---|---|---|---|---|---|---|---|
| عدد الطلبة | ٢٩ | ١٩٥ | ٢٤١ | ١١٧ | ٥٢ | ١٠ | ٦ | ٣ | ٢ | ٦٥٥ |

والمطلوب هو حساب الوسط الحسابي والوسيط لهذا التوزيع .

٢- الجدول التالي يبين الأجور التي يتحصل عليها من مصنعين أ ، ب حيث أن كل مصنع يستخدم ٥٠ عاملاً

| Wages Classes | No. of Employees | |
|---|---|---|
| | A | B |
| 45- | 1 | 1 |
| 50- | 0 | 0 |
| 55- | 1 | 0 |
| 60- | 5 | 5 |
| 65- | 6 | 15 |
| 70- | 12 | 10 |
| 75- | 14 | 9 |
| 80- | 7 | 6 |
| 85- | 2 | 2 |
| 90- | 1 | 1 |
| 95- | 0 | 0 |
| 100- | 1 | 0 |
| $\sum$ | 50 | 50 |

والمطلوب هو مقارنة الوسيط للأجور في المصنعين .

**٤- الجدول التالي يوضح توزيع الدرجات التي حصل عليها عدد معين من الطلاب في امتحان إحدى المواد الدراسية :**

| Degree | 30- | 35- | 40- | 45- | 50- | 55- | 60- | 65- |
|---|---|---|---|---|---|---|---|---|
| No. of Student | 12 | 18 | 22 | 27 | 17 | 23 | 19 | 8 |

**والمطلوب حساب**

**أ – الوسط الحسابي .**

**ب – الوسيط .**

**جـ – المنوال .**

**٥– الجدول التالي يوضح توزيع عينـة من الأسر في إحدى البلاد حسب فئات الانفاق الشهري بالجنية :**

| Monthly Exp. Classes | 5- | 25- | 45- | 65- | 85- | 105- | 125- | 145- |
|---|---|---|---|---|---|---|---|---|
| No. of Families | 4 | 6 | 15 | 22 | 13 | 7 | 8 | 3 |

**والمطلوب :**

**١- حساب قيمة الوسط الحسابي لإنفاق هذه الأسر .**

**٢- حساب قيمة الوسيط .**

- **من الجدول التكراري المتجمع الصاعد والنازل .**
- **بالرسم من المنحنى المتجمع الصاعد والنـازل كلٍ على حدة ثم من المنحنيين معاً .**

**٣ – حساب قيمة المنوال .**

**– باستخدام طريقة الحساب .**

**– بالرسم من المدرج التكراري .**

**٤– حساب الوسط الهندسي .**

**٥ – حساب الوسط التوافقي .**

**٦- الجدول التالي يوضح توزيع عدد من المحلات التجارية حسب جملة المبيعات السنوية بآلاف الدولارات .**

| جملة المبيعات | ١١٠- | ١١٥- | ١٢٠- | ١٢٥- | ١٣٠- | ١٣٥- |
|---|---|---|---|---|---|---|
| عدد المحلات | ٤ | ٢٠ | ٣٨ | ٢٤ | ١٠ | ٤ |

**والمطلوب هو :**

**١- حساب الوسط الحسابي لجملة المبيعات .**

**٢- حساب الوسيط.**

**٧- احسب متوسط الطول بالمليمتر ) للبيانات التالية :**

| الطول بالمليمتر | ١٩٨ | ١٩٩ | ٢٠٠ | ٢٠١ | ٢٠٢ |
|---|---|---|---|---|---|
| التكرار | ١ | ٤ | ١٧ | ٢ | ١ |

**٨- فيما يلى توزيع درجات الحرارة خلال مجموعة من الأيام :**

| درجات الحرارة | ١٦ | ١٧ | ١٨ | ١٩ | ٢٠ | ٢١ | ٢٢ |
|---|---|---|---|---|---|---|---|
| التكرار | ١ | ٤ | ٦ | ٨ | ٩ | ١ | ١ |

**والمطلوب حساب الوسط الحسابي لدرجة الحرارة .**

**٩- الجدول التالي يمثل توزيع مستويات أسعار ١٠٠ سلعة .**

| الغلة | ١١٥- | ١٢٥- | ١٣٥- | ١٤٥- | ١٥٥- | إجمالي |
|---|---|---|---|---|---|---|
| التكرار | ١٠ | ٢٠ | ٤٠ | ٢٠ | ١٠ | ١٠٠ |

**والمطلوب أحسب مايلى :**

- **الوسط الحسابي .**
- **الوسيط .**
- **المنوال .**

- **الوسط الهندسي والوسط التوافقي .**

**١٠- الجدول التالي يبين لأطوال لعدد معين من السيدات يعملن في إحدى المصانع .**

| الطول بالسنتيمتر | ١٥٦- | ١٥٩- | ١٦٢- | ١٦٥- | ١٦٨-١٧٠ |
|---|---|---|---|---|---|
| العدد | ٢ | ٩ | ١٨ | ٧ | ٣ |

**والمطلوب حساب الوسط الحسابي والوسط الهندسي والوسط التوافقي للطول.**

**١١- فيما يلى الجدول التكراري لمبيعات المنتج ( أ ) حسب السعر:**

| سعر المنتج أ | ٧٥- | ٧٧- | ٧٩- | ٨١- | ٨٣- | ٨٥- | المجموع |
|---|---|---|---|---|---|---|---|
| مبيعات المنتج أ | ٣ | ٢٣ | ٥٢ | ١٥ | ٧ | ٢ | ١٠٢ |

**والمطلوب حساب :**

**أ - الوسط الحسابي للمبيعات**

**ب - الوسط الحسابي بالطريقة المختصرة .**

**جـ- الوسيط .**

**د - القيمة الأكثر شيوعاً للمبيعات حسابياً وبيانياً .**

**١٢- الجدول التالي يبين توزيع العمال حسب الدخل**

| فئات الدخل | أقل من ٢٠ | ٢٠- | ٣٠- | ٤٠- | ٥٠- | ٦٠ فأكثر | المجموع |
|---|---|---|---|---|---|---|---|
| عدد العمال | ٥ | ١٤ | ٢٠ | ١٥ | ١٠ | ٦ | ٧٠ |

**المطلوب إيجاد :**

**أ – مقياس مناسب للنزعة المركزية موضحاً سبب ذلك .**

**ب – مقياس مناسب للنزعة المركزية عن طريق الرسم وحسابياً .**

**١٣– من الجدول السابق في التمرين رقم (٧) إذا علمت أن أقل قيمة للدخل هو ١٠ جنيهاً وأكثر قيمة ٧٠ جنيهاً فأوجد مقياس آخر مناسب للنزعة المركزية .**

**١٤ – المطلوب حساب الوسط الحسابي والوسيط والمنوال للقيم التالية :**

**٦ ، ٣٧ ، ٢١ ، ٥٨ ، ٣٦ ، ٣٣ ، ٤٥**

**١٥– كانت متوسطات درجات الطلبة في الإحصاء في ثلاث اختيارات أ ، ب ، جـ هي ٧٥ ، ٨٢ ، ٨٤ درجة ، فإذا كان عدد الطلبة في كل من هذه الفصول هو ٣٢ ، ٢٥ ، ١٧ طالباً على التوالي ،**

**المطلوب حساب الوسط الحسابي لجميع الفصول .**

**١٦ – الآتي بيان بتوزيع أفراد عينتين من نزلاء أحد فنادق الدرجة الأولى بجمهورية مصر العربية بحسب فئات تكاليف الإقامة بأحد الليالي السياحية.**

- **فئات تكلفة الإقامة بالجنيه : ١٠٠–١٥٠–٢٥٠–٣٠٠–٣٥٠–٤٠٠**
- **عدد أفراد عينة الفندق الأولى : ٨ ٢٠ ١٢ ١٠ ٥**
- **عدد أفراد عينة الفندق الثاني : ٦ ١٥ ١٢ ٧ ٢**

**المطلوب :**

**١- رسم المدرج التكراري لبيانات عينة الفندق الأول .**

**٢- رسم المنحنى المتجمع الصاعد لبيانات عينة الفندق الثاني وأيضاً الوسيط فيه .**

**٣- إيجاد الوسط الحسابي والوسيط والمنوال لعينة الفندق الأول .**

**٤- إيجاد المنوال بالرسم فقط لعينة الفندق الثاني .**

**١٧- فيما يلي التوزيع التكراري لـ ٣٠٠ أسرة حسب دخولها الأسبوعية (بالجنيه) :**

| Weekly Inc. Classes | 25- | 30- | 35- | 40- | 45- | 50- | 55- | ∑ |
|---|---|---|---|---|---|---|---|---|
| No. of Families | 14 | 30 | 45 | 68 | 62 | 49 | 32 | 300 |

**<u>والمطلوب :-</u>**

**أ- حساب الوسط الحسابي والوسيط والمنوال .**

**ب- حساب الربيعين الأدنى والأعلى .**

**ج- ادرس شكل التوزيع مرة باستخدام الوسط الحسابي والوسيط والمنوال ومرة أخرى باستخدام الربيعين الأدنى والأعلى والوسيط ، قارن بين النتيجة في الحالتين .**

**د- احسب قيمة الدخل التي يقل عنها ٦٠% من إجمالي عدد الأسر .**

**ه- احسب النسبة المئوية لعدد الأسر التي يقل دخلها عن ٤٣ جنيهاً .**

**و- حساب العشير السابع والمئين الخامس والثمانين . اذكر بعض النتائج الأكثر تفصيلاً التي يمكن الحصول عليها باستخدام قيمة الربيعين الأدنى والأعلى والوسيط ، العشير السابع ، المئين الخامس والثمانين .**

**ز- احسب عدد الأسر الذي يتراوح دخلها الاسبوعي ما بين ٣٧ ، ٤٦ جنيهاً .**

**١٧ – البيانات التالية هي قيمة الإيجارات الشهرية (بالجنيه) لعينة من المساكن حجمها ٥٠ شقة .**

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 112 | 87 | 76 | 95 | 90 | 105 | 86 | 99 | 84 | 103 |
| 101 | 94 | 104 | 87 | 100 | 67 | 109 | 75 | 118 | 88 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 122 | 72 | 112 | 92 | 65 | 115 | 80 | 111 | 74 | 112 |
| 90 | 101 | 92 | 87 | 108 | 90 | 97 | 93 | 104 | 98 |
| 120 | 86 | 103 | 77 | 98 | 84 | 117 | 69 | 123 | 94 |

**والمطلوب :**

**أ- حساب الوسط الحسابي والوسيط لهذه البيانات .**

**ب- عرض هذه البيانات في صورة جدول تكراري متساوي الفئات طول كل فئة ١٠ مبتدئا بالفئة ٦٥ – . ثم احسب الوسط الحسابي والوسيط من الجدول التكراري الناتج . وفسر ما قد يوجد من فروق بين المتوسطين في أ ، ب .**

**ج- احسب المنوال من الجدول التكراري الناتج في ( ب ) وذلك باستخدام طريقتي الفروق والرافعة .**

**د- احسب نسبة الشقق التي تبلغ قيمتها الإيجارية ٩٥ جنيهاً على الأقل مستخدماً القيم الفعلية للبيانات ، وكذلك المنحنى التكراري المتجمع الملائم مع المقارنة بين النتيجة في الحالتين .**

**ه- احسب نسبة الشقق التي يتراوح إيجارها الشهري ما بين ٧٠، ٩٠ جنيهاً .**

**١٨- الجدول التالي يوضح التوزيع التكراري لدرجات الطلاب بإحدى كليات التجارة حسب نظام الدراسة ( انتظام – انتساب ) وذلك في دور مايو ١٩٨٧.**

| فئات الدرجات<br>نظام الدراسة | صفر- | ٢١- | ٣٠- | ٣٩- | ٤٨- | ٥٤-٦٠ | المجموع |
|---|---|---|---|---|---|---|---|
| انتظام | ١٦٦ | ١٤٨ | ١٧٠ | ١٦٧ | ٣٦ | ١٥ | ٧٠٢ |
| انتساب | ١٠٨ | ٨٨ | ٥٩ | ٢٤ | ٥ | ١ | ٢٨٥ |

**والمطلوب :**

أ- حساب نسبة النجاح لكل من الطلاب المنتظمين والطلاب المنتسبين .

ب- قارن بين القيمة المتوسطة في كل من نظامي الدراسة مستخدماً الوسط الحسابي والوسيط والمنوال . استخدم الأشكال البيانية المناسبة للتعبير عن هذه البيانات .

ج- احسب نسبة الطلاب اللذين تنحصر درجاتهم ما بين ٣٥، ٤٥ درجة لكل من الطلاب المنتظمين والطلاب المنتسبين ، ثم قارن بين النسبة في الحالتين .

١٩ – فيما يلي التوزيع التكراري لأجور العمال الاسبوعية بالجنيه في مصنعين أ ، ب لصناعة المنسوجات القطنية :

| Wages Classes | 40- | 50- | 60- | 70- | 80- | 90- | 100- | 110- | ∑ |
|---|---|---|---|---|---|---|---|---|---|
| A | 3 | 4 | 12 | 25 | 20 | 9 | 5 | 2 | 80 |
| B | 4 | 10 | 23 | 33 | 28 | 12 | 7 | 3 | 120 |

والمطلوب :

أ– قارن بين الوسط الحسابي والوسيط والمنوال لأجور العمال في كل من المصنعين ثم استخدم نتائجك في دراسة شكل التوزيع في كل من المصنعين

ب- احسب نسبة العمال الذين يحصلون على ٧٤ جنيهاً على الأقل في كل من المصنعين ثم قارن بينهما .

ج- احسب الوسط الحسابي لأجور العمال في المصنعين معاً وذلك باستخدام ما يلي خاصية الوسط الحسابي :

- تكوين الجدول التكراري الذي يمثل توزيع العمال في المصنعين معاً ( وذلك بالحصول على مجموع تكراري المصنع أ والمصنع ب بالنسبة لكل فئة ) .
- قارن بين النتيجة في الحالتين .

**د – احسب نسبة العمال الذين تتراوح أجورهم ما بين ٦٠ ، ٩٥ جنيهاً في كل من المصنعين . قارن بين النسبة في الحالتين .**

**٢٠ – الجدول التالي يوضح التوزيع التكراري لـ أسرة حسب الحجم :**

| Family Size | 1 | 2 | 3 | 4 | 5 | 6 | ∑ |
|---|---|---|---|---|---|---|---|
| No. of Families | 4 | 5 | 8 | 12 | 8 | 3 | 40 |

**<u>والمطلوب</u> : هو إيجاد قيمة متوسطة لحجم الأسرة باستخدام المتوسط الملائم**

**٢١ – فيما يلي التوزيع التكراري لدرجات ١٠٠ طالب في مادة الإحصاء في كل من كليتي التجارة والتربية في إحدى الجامعات .**

| Degree Classes | 10- | 20- | 30- | 40- | 50- | 60- | 70- | 80- | 90- |
|---|---|---|---|---|---|---|---|---|---|
| No. of Stu. Commerce | 3 | 5 | 9 | 15 | 23 | 26 | 14 | 9 | 6 |
| No. of Stu. Education | 2 | 9 | 12 | 18 | 25 | 14 | 11 | 6 | 3 |

**<u>والمطلوب :</u>**

**أ – قارن بين الوسط الحسابي والوسيط والمنوال في كل من الكليتين .**

**ب– احسب النسبة المئوية لعدد الطلاب الحاصلين على تقدير جيد فأكثر في كل من الكليتين وقارن بينهما ( الحد الأدنى لتقدير جيد هو ٦٥ درجة ) .**

**جـ – قارن بين عدد الطلاب الذين تنحصر درجاتهم بين ٤٢ ، ٧٦ درجة في كل من الكليتين .**

**د – قارن بين شكل التوزيعين باستخدام الربيعين الأدنى والأعلى والوسيط حقق نتائجك برسم المنحنى التكراري لكل من التوزيعين .**

**هـ – اوجد العشير الثالث والمئين الرابع والثلاثين في كل من كلية التجارة وكلية التربية مع المقارنة .**

**و– احسب نسبة الرسوب في مادة الاحصاء في كل من الكليتين مع المقارنة بينهما .**

**٢٢ – احسب الوسط الحسابي العام من البيانات الآتية**

$n_1 = 10$ , $\bar{x}_1 = 25$ , $n_2 = 12$ , $\bar{x}_2 = 15$

**٢٣ – إذا كان :**

$n_{1 = 10}$ , $\bar{x}_1 = 10$ , $n_2 = ?$ , $\bar{x}_2 = 20$

**فإذا كان الوسط الحسابي العام للعينتين هو 16 ، احسب حجم العينة الثانية.**

**٢٤– الجدول التالي يوضح توزيع ٨٠ أسرة حسب انفاقها اليومي بالجنيه :**

| Expenditure Classes | 12- | 14- | 16- | 18- | 22- | 26- | $\sum$ |
|---|---|---|---|---|---|---|---|
| No. of Families | 7 | 11 | 17 | 22 | 25 | 8 | 80 |

**والمطلوب :**

**أ – احسب الوسط الحسابي والوسيط والمنوال .**

**ب – احسب الوسيط باستخدام المنحنى التكراري المتجمع الهابط .**

**جـ – باستخدام المنحنى التكراري المتجمع الهابط في ب ، احسب نسبة الأسر التي يتراوح إنفاقها اليومي ما بين ١٥ ، ١٩ جنيها .**

**د– من النتائج أ – ب ، ادرس تماثل التوزيع .**

٢٥- التوزيع التكراري التالي يوضح الدخل الاسبوعي (بالجنيهات ) لعينة حجمها ٥٠ أسرة .

| Income Classes | 40- | 50- | 60- | 70- | 80- | 90-100 | ∑ |
|---|---|---|---|---|---|---|---|
| No. of Families | 6 | 8 | 11 | 14 | 7 | 4 | 50 |

والمطلوب :

أ- حساب كل من الوسط الحسابي والوسيط .

ب- احسب قيمة الدخل الشهري الذي يقل عنه ٤٠% من إجمالي عدد الأسر .

ج- استخدم شكل المدرج التكراري للتعبير عن هذا التوزيع .

٢٦ – فيما يلي التوزيع التكراري للأجور الأسبوعية لمائة عامل في إحدى الشركات :

| Weekly Wages Classes | 35- | 45- | 55- | 65- | 75- | 85-95 | ∑ |
|---|---|---|---|---|---|---|---|
| No. of Employees | 9 | 13 | 8 | 26 | 19 | 15 | 100 |

والمطلوب :

أ- حساب كل من الوسط الحسابي .

ب- تكوين الجدول التكراري المتجمع الصاعد .

ج- باستخدام طريقة الفروق ، احسب المنوال .

د- ٢٨ – فيما يلي التوزيع التكراري لدرجات ٤٠ طالباً في مادة الإحصاء :

| Degree Classes | 25- | 35- | 45- | 55- | 65- | 75- | 85-95 |
|---|---|---|---|---|---|---|---|
| No. of Student | 3 | 6 | 8 | 9 | 7 | 5 | 2 |

**والمطلوب :**

**أ – حساب الوسط الحسابي .**

**ب- باستخدام طريقة الرافعة ، احسب المنوال .**

**ج- باستخدام التوزيع التكراري المتجمع الملائم ، احسب ما يلي :**

- **نسبة رسوب الطلاب ( علماً بأن الطالب يعتبر راسباً إذا حصل على أقل من ٥٠ درجة )**
- **إذا كان عدد الطلاب الذين تتراوح درجاتهم ما بين ( الدرجة ٤٥ . الدرجة x ) يمثل ٦٥ % من إجمالي عدد الطلاب ، احسب قيمة الدرجة x .**

**د – من نتائجك في أ ، ب ناقش بإيجاز شكل التوزيع من حيث التماثل والالتواء .**

# الباب الرابع

## *مقاييس التشتت*

## *Measures of Dispersion*

**فى الباب السابق لاحظنا أن مقاييس النزعنة المركزية أو المتوسطات (الوسط الحسابى- الوسيط – المنوال – الوسط الهندسى – الوسط التوافقى ) تمثل قيمة متوسطة تتجمع حولها معظم البيانات الخاصة بظاهرة ما وتعبر عنها بصورة عامة . والآن فالسؤال الذى يطرح نفسه فى هذا المجال هو : هل يكفى معرفة مقياس من مقاييس النزعة المركزية ليعطى مؤشراً حقيقياً للبيانات محل الدراسة ؟ وهل يمكن من خلال مقاييس النزعة المركزية مقارنة هذه البيانات بمجموعة أخرى من البيانات؟**

**فى واقع الأمر أن هذا النوع وحدة من تلك المقاييس – مقاييس النزعة المركزية لا يكفى لإعطاء صورة متكاملة عن توزيع المتغير محل الدراسة. إذ يتطلب ذلك نوعاً آخر من أنواع المقاييس الإحصائية ليعطى فكرة واضحة ودقيقة عن درجة أو مدى تجانس توزيع الظاهرة أو بمعنى آخر مدى تركز المفردات حول قيمة معينة عادة ما تكون هى قيمة الوسط الحسابى أو وسيط الظاهرة محل الدراسة أو مدى إنتشارها أو بتاعدها عن هذه القيمة المتوسطة . كما أن مقاييس النزعة المركزية وحدها لا تكفى لإتمام عملية المقارنة فيما بين مجموعتين أو أكثر من البيانات. فقد تكون المقارنة غير كاملة إن لم تكن مضللة. فقد تشترك مجموعتين أو أكثر فى المتوسط الحسابى ولكنها قد تختلف فى درجة تجانسها. ولبيان ذلك دعنا نفترض أن لدينا مجتمعين أ ، ب وكان الدخل الأسبوعى (بالجنية) لخمسة أفراد من كل من المجتمعين على النحو المبين التالى:**

الدخل الأسبوعى فى المجتمع (أ) هو ٥ ، ٥٠ ، ١٠٠ ، ١٥٠ ، ١٩٥ جنيه وأن الدخل الاسبوعى فى المجتمع (ب)هو٩٨ ، ٩٩ ، ١٠٠ ، ١٠١ ، ١٠٢ جنيه . فنجد أن الوسط الحسابى للدخل الاسبوعى لكل من المجتمعين أ ، ب هو مساوياً ١٠٠ جنيه . كما أن وسيط الدخل الأسبوعى لكل من المجتمعين أ ، ب أيضاً مساوياً ١٠٠ جنيه . أى أنه وفقاً لمقاييس النزعة المركزية فإننا نجد أن البيانات فى المجتمعين أ ، ب لا يختلفان من حيث متوسط الدخل الاسبوعى . ولكن هذا مخالفاً لواقع البيانات الفعلية فى المجتمعين أ ، ب حيث يوجد هناك تفاوتاً واضحاً فى توزيع مفردات الدخل داخل كل مجتمع عن الآخر. فالبيانات فى المجتمع الأول (أ) يتراوح فيها الدخل الأسبوعى فيما بين ١٩٥ جنيه و ٥ جنيهات أى أن المدى هو :-

**Range = Maximum value – Minimum value.**

**= 195 – 5 = 190 (L.E)**

بينما نجد أن فى المجتمع الثانى (ب) يترواح فيه الدخل الأسبوعى فيما بين ١٠٢ ، ٩٨جنيه. اى أن المدى هو :

**Range = 102 – 94 = 4 (L.E)**

فيتضح لنا من خلال المقارنة المبدئية أن بيانات المجتمع (أ) غير متجانسة بمعنى أنها مبعثرة أو أكثر تشتتاً من مفردات المجتمع الثانى (ب) .

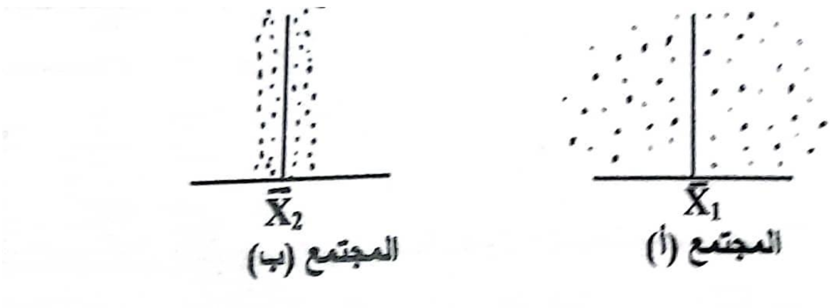
فكما هو واضح قد تتساوى مقاييس النزعة المركزية لمجموعتين (أو أكثر من البيانات بما يوحى أنهما متطابقين أو متقاربين . بينما هما فى حقيقة الأمر يختلفان عن بعضهما إختلافاً جذرياً واضحاً. ومن ثم يتضح لنا من هذا أن مقاييس النزعة المركزية وحدها لا تصلح لوصف مجموعة من المفردات وصفاً

كاملاً ودقيقاً بالإضافة الى أنها وحدها لا تصلح للمقارنة بين مجموعتين أو أكثر وخصوصاً إذا اختلفت تلك المجموعات من حيث وحدات قياسها.

وبصورة عامة فأن التشتت لأى مجموعة من المفردات يقصد به مدى التباعد أو الأختلاف فيما بين قيم تلك المفردات فهذا التشتت يكون صغيراً إذا كان الاختلاف بين قيم مفردات الظاهرة محل الدراسة قليلاً اى إذا كانت مفردات الظاهرة تحتشد وتتجمع حول متوسطاتها أو قريباً منه (فإذا تساوت جميع القيم للظاهرة محل الدراسة فإن التشتت ينعدم أى يكون مساوياً للصفر) فى حين يكون التشتت كبيراً إذا ما كان الاختلاف فيما بين قيم المفردات كبيراً أى إذا كانت المفردات تنتشر بعيداً على جانبى الوسط الحسابى (أو مقياس النزعة المركزية) ولذلك فكما يعنينا أن نقدر القيمة المتوسطة للظاهرة محل الدراسة من خلال أى من مقاييس النزعة المركزية فإنه لا يقل عنه أهمية أن نتمكن من قياس درجة تجانس أو عدم تجانس (أى تشتت) توزيع هذه الظاهرة محل الدراسة . ولذلك سوف نستخدم معايير كمية لقياس المدى الذى تنتشر عليها البيانات أو لقياس شدة تبعثر القيم أو التفافها حول متوسطها . وتعرف هذه المقاييس بمقاييس التشتت. فهى تعنى دراسة مدى تباين أو تشتت المفردات الذى يحدث حول إحدى القيم المتوسطة (مقاييس النزعة المركزية) . ففى الشكل (٤-١) نجد أن مفردات المجتمع (أ) أكثر تشتتاً (أو أقل تجانساً ) من مفردات المجتمع (ب) .

شكل (٤-١)

يان مدى تجانس وتشتت المفردات




ويلاحظ أنه حينما يكون التشتت صغيراً فإن القيمة المتوسطة تعتبر تقديراً جيداً أو مأموناً. إما إذا كان التشتت كبيراً فإن القيمة المتوسطة لا تعتبر تقديراً جيداً أو بمعنى تعتبر تقديراً غير مأموناً لتمثيل الظاهرة محل الدراسة .

هذا ويوجد عدة مقاييس للتشتت وهى تختلف فيما بينها من حيث الدقة وطريقة الحساب. فكما سيتضح لنا فيما بعد إنه لأختيار المقياس الأفضل يجب ان نبحث عن توافر بعض الخصائص التى سبق أن ذكرناها عند الكلام عن المتوسطات من حيث سهولة فهم المقياس وأن يأخذ فى الاعتبار جميع المفردات عند حسابه وإمكانية وضعه فى صورة جبرية مما يسهل معه أن يخضع للعمليات الحسابية وقلة تأثرة بإختيار العينة والقيم الشاذه أو المتطرفة فى قيم المفردات وسوف ندرس لهذا الغرض نوعين أساسين من مقاييس التشتت هما:-

النوع الأول : وهى ما تسمى بمقاييس التشتت المطلقة وتشمل المدى ونصف المدى الربيعى والأنحراف المتوسط والانحراف المعيارى.

النوع الثانى : وهو ما يسمى بمقاييس التشتت النسبية وهى تشمل ما يسمى بمعاملات الاختلاف سواء النسبى أو الربيعى والتى سيرد دراستنا لها فى هذا الباب.

**هذا وسوف نتعرض لمختلف هذه المقاييس بنفس الطريقة التى أتبعناها فى الباب السابق حيث يتم التعرف على كيفية حساب تلك المقاييس فى حالة توافر بيانات مفردة (غير مبوبة) ثم فى حالة البيانات المبوبة تكرارياً .**

***أولاً :- مقاييس التشتت المطلقة :-***

**Measures of Absolute Dispersion**

**وتستخدم هذه المقاييس فى دراسة مقارنة تشتت مجموعتين (أو أكثر) متحدتين فى وحدة القياس : بمعنى تستخدم هذه المقاييس المطلقة للتشتت مثلاً فى دراسة مقارنة تشتت درجات مجموعتين مختلفين من الطلاب أو الطالبات فى مادة معينة.**

**وبالإضافة إلى ان مقاييس التشتت المطلقة تقيس تشتت البيانات أو بعثرتها فأهم ما يميزها أنها تأخذ فى النهاية وحدة قياس هى نفس وحدة قياس القيم الأصلية كما هو الحال فى حالة مقاييس النزعة المركزية فإذا كانت القيم الأصلية مثلاً بالكيلومتر فستكون النتيجة النهائية لقيمة هذه المقاييس بالكيلومتر أيضاً.**

**وتتمثل مقاييس التشتت المطلقة فيما يلى:-**

١- المدى : Range

**ويعرف المدى لمجموعة من المفردات كما سبق بأنه الفرق ما بين أكبر القيم وأصغرها أو يعرف بالقيمتين الصغرى والكبرى معاً ويستخدم فى بيان مدى الإتساع الذى تنتشر عليه المفردات. أى أن المدى :-**

- **فى حالة المفردات (البيانات الغير مبوبه) : هو :**

Range = Maximum value – Minimum value.

**فإذا كان لدينا متغيراً يأخذ مجموعة القيم $X_1$ ، $X_2$ ، ..... ، $X_n$ وتم ترتيبها تصاعدياً أو تنازلياً وليكن $X_{(1)}$ ، $X_{(2)}$ ، ..... ، $X_{(n)}$ حيث يتم التعبير عن أصغر القيم بالقيمة $X_{(1)}$ وأكبر قيمة بالقيمة $X_{(n)}$ فأن المدى يكون عبارة عن:**

Range = Maximum value – Minimum value.

$$= X_{(n)} - X_{(1)}$$

**وبلغه أخرى نقول أن المدى للمتغير $(X)$ هو من $X_{(1)}$ حتى $X_{(n)}$ :**

***فمثلاً*** **: إذا كان لدينا مجموعة القيم : ٢٠ ، ١٥ ، ٣٨ ، ١٢ ، ٣٠ فإن المدى لتلك المجموعة هو :**

**Range = 38 - 12 = 26**

**أو نقول بأن المدى من ١٢ حتى ٣٨ وإن كان هذا التعبير الأخير هو الأفضل حيث أنه يعطى معلومات أكثر من التعبير الأول (القانون) .**

*** المدى فى حالة البيانات المبوبة تكرارياً :-**

**إذا كان لدينا توزيعاً تكرارياً فإن هناك مجموعة من الصيغ التى يمكن من خلالها حساب المدى .**

**الأولى: والتى تفيد بأن المدى هو عبارة عن الفرق ما بين الحد الأعلى للفئة الأخيرة والحد الأدنى للفئة للتوزيع التكرارى المعطى أى أن :-**

Rang = U.L for the latter class – L.L for the first class

**حيث :**

**تمثل الحد الاعلى للفئة الاخيرة** U.L for the latter class

**تمثل الحد الادنى للفئة الاولى** L.L for the first class

**والثانى: والتى يمكن من خلالها حساب المدى ايضا وذلك من خلال ايجاد الفرق بين مركزى الفئة الاخيرة والفئة الاولى للتوزيع التكرارى. فوفقنا لهذا التعريف فان المدى يمكن كتابته على الصورة :**

**Range = Midpoint for the Latter Class – Midpoint for the First Class**

**ويعتبر المدى بذلك من أسهل مقاييس التشتت المطلق وأسرعها فى الحساب وهو مفيد أكثر فى العمليات الصناعية وخصوصاً فى مجال إحصاءات مراقبة جودة الإنتاج حيث تؤخذ عينات صغيرة ومتساوية من حيث الحجم (ومن ثم فأن حجم العينة لن يؤثر على المدى وهو ما يعيبه كمقياس لدرجة التشتت) وذلك على فترات متقاربة ونحتاج الى حساب سريع لتشتت كل عينة من تلك العينات المسحوبة من قطاع الإنتاج. كذلك نستخدم المدى فى دراسة التغيرات فى درجات الحرارة اليومية والتغيرات التى تنشأ فى الأجور أو الأسعار خلال فترة زمنية معينة. والمدى بهذا التعريف كمقياس من مقاييس التشتت يعاب عليه ما يلى :-**

- **أنه يستخدم قيمتين فقط من البيانات فى حسابه ويهمل باقى قيم الظاهرة محل الدراسة. لذلك فهو يعتبر أقل مقاييس التشتت من حيث الدقة فى التعبير عن تشتت مجموعة أو المقارنة فيما بين تشتت مجموعتين أو أكثر.**
- **نظراً لإعتماد المدى فى حسابه على المفردتين المتطرفتين فى القيمة** Extreme points **فأنه يتأثر جداً بالقيم المتطرفة أو الشاذه إذ أن وجود قيمة شاذه فى مجموعة قيم الظاهرة محل الدراسة قد يسبب زيادة كبيرة فى المدى فيستدل منه على أن مفردات الظاهرة متشتتة جداً بينما قد تكون مفرداتها كلها تقريباً – أى فيماً عدا القيمة الشاذه أو المتطرفة – متقاربة . فمثلاً إذا كان لدينا مجموعة القيم ١٥، ١٨ ، ٢٠ ، ١٣ ، ١٩ فإن المدى لهذه القيم مساوياً :**

Range = 20 – 13 = 7

**فـإذا مـا أضـفنا قيمــة واحـده ولـتكن المفـردة ١٠٠ مـثلاً لمجموعـة المفـردات السابقة فإن المدى سيصبح :-**

Range = 100 – 13 = 87

**أى أن المدى تغيرت قيمتـه لتصبح ٨٧ بـدلاً من ٧ نتيجـة إضـافة المفـردة التـى تأخذ القيمة الشاذه أو المتطرفة عن باقى المفردات (100).**

- **يعتبر المدى حساس جداً بالنسبة لحجم العينـة. فغالباً ما يزداد المـدى نتيجـة زيادة حجم العينـة . ولذلك فلا يعتمد الإحصـائيون علـى المـدى لتقدير التغيـر فـى المجتمع إذ أن قيمته تتوقف على حجم العينـة.**
- **المدى يتغير كثيراً من عينة لأخرى – وهذا الأمر يكون دائمـاً غيـر مرغـوب فيه لدى الإحصائي – وهذا الإختلاف أكثر من الأختلاف الـذى نحصـل عليـه مـن بعض مقاييس التشتت الأخرى.**
- **يصعب حساب المدى فـى حالـة الجداول التكراريـة المفتوحـة حيث لا يكـون معلومـاً الحد الأدنى للفئة الأولى أو الحد الأعلى للفئـة الأخيرة هذا بالإضافة لعدم إمكانيـة حسابه بيانيـاً.**

**هذا وبالرغم من هذه العيوب المتعددة إلا أن المدى كما سبق وأن ذكرنـا يستخدم علـى نطـاق واسـع فـى مجـال واسـع فـى مجـال دراسـات ضـبط جـودة الإنتـاج** Quality control**. والسبب فى ذلك يرجع إلى أنه كما ذكرنـا أنـه فـى مثـل هـذه الدراسـات (دراسـات ضـبط الجـودة) فإننـا نحتـاج إلـى حسـاب مقيـاس للتشـتت بسرعة . والمـدى بطبيعـة الحـال هو أسرع مقيـاس يمكـن إيجـاده ويعطـى دلالـه علـى التشتت. هذا وتجدر الإشارة إلى أنه فـى مثـل تلـك الحـالات فأنـه المـدى يـتم حسابه أكثر من مره وذلك بإستخدام عدد كبير من العينـات وذلك للتقليل ما أمكن**

**من فرص الحصول على قيم شاذه أو متطرفة. وبصفة عامه فأنه المدى عادة ما يستخدم كمقياس للتشتت فى حالة الحاجه إلى حساب مقياس سريع لتشتت البيانات دون الإهتمام بالدقة فى القياس وكذلك عندما يكون لقيم المفردات المتطرفة أهمية خاصة فى البحث يراد إبرازها من خلال الدارسة أو البحث.**

- **شبيهات المدى :-** Quasi - Ranges

**للتغلب على مشكلة القيم المتطرفة أو الشاذه والتى يتأثر بها المدى فى حسابه بشكل كبير فقد أتجه الإحصائيون إلى مقاييس بديلة شبيهه بالمدى ولكنها تهمل القيم المتطرفة. ووفقاً لهذه البدائل فأن قيم المفردات يتم ترتيبها تصاعدياً ثم يتم حذف عدداً متساوى من القيم عند طرفى مجموعة المفردات ثم يتم حساب المدى لبقية القيم. وتسمى هذه المقاييس بشبيهات المدى ويمكن تعريفها على النحو التالى:**

First Range = $X_{(n-1)} - X_{(2)}$

Second Range = $X_{(n-2)} - X_{(3)}$

**وهكذا بالنسبة لبقية مقاييس شبيهات المدى حيث يتم التعريف المدى الرائى أو ذو الرتبة (r) بأنه على الصورة:**

$R\ \underset{\sim}{th}$ Range = $X_{(n-r)} - X_{(r+1)}$

***فمثلاً***

**إذا كان لدينا مجموعة البيانات التالية :**

**١٢ ، ١١ ، ١٠ ، ١٣ ، ١٤ ، ١٧ ، ١٥ ، ١٨ ، ٨٤ ، ١٦.**

**فأنه يمكن إيجاد المدى وشبيهاته وذلك على النحو التالى :-**

**فبعد ترتيب المفردات تصاعدياً مثلاً فإن البيانات تأخذ الصورة التالية :-**

**٨٤ ، ١٨ ، ١٧ ، ١٦ ، ١٥ ، ١٤ ، ١٣ ، ١٢ ، ١١ ، ١٠**

**فيكون المدى هو :**

$$\text{Range} = X_{(n)} - X_{(1)} = X_{(10)} - X_{(1)} = 84 - 10 = 74$$

**أما شبيهات المدى فهى عبارة عن :**

**فالمدى الأول :** $\textbf{First Range} = X_{(n-1)} - X_{(2)}$

$\textbf{or 1st Range} = X_{(9)} - X_{(2)} = 18 - 11 = 7$

**والمدى الثانى هو :**

$\textbf{Second range} = X_{(n-2)} - X_{(3)}$

$\textbf{or 2}^{\textbf{nd}}\textbf{ Range} = X_{(8)} - X_{(3)}$

$= 17 - 12 = 5$

**والمدى الثالث هو :**

$\textbf{Third Range} = X_{(n-3)} - X_4$

$\textbf{3}^{\textbf{rd}}\textbf{ Range} = X_{(7)} - X_4$

$= 16 - 13 = 3$

**لاحظ مدى الأثر على شبيهات المدى من حيث عدم تأثر مثل هذه المقاييس بالقيم الشاذه أو المتطرفة أى أنه بإستخدام شبيهات المدى فإننا نكون قد تغلبنا على مشكلة القيم الشاذه أو المتطرفة.**

**هذا وتجدر الإشارة إلى أن مقاييس شبيهات المدى يمكن إستخدامها سواء كانت عدد المفردات أو القيم صغيراً أو كبيراً. ففى حالة كبر عدداً القيم أو المفردات فإنه يمكن إيجاد المدى ما بين العشير التاسع والعشير الأول أو بين العشير الثامن والعشير الثانى وذلك على سبيل المثال . وكذلك فإنه إذا كان عدد المفردات كبيراً كبراً كافياً فإنه يمكن إيجاد المدى بين المئين الخامس والتسعين**

والمئين الخامس أو بين المئين السادس والتسعون والمئين الرابع على سبيل المثال. ومثل هذه المقاييس كثيراً ما تستخدم فى مجالات بحوث التربية وعلم النفس على وجه الخصوص. ولعل أشهر مقاييس شبيهات المدى وأكثرها إستخداماً هو المدى بين الربيع الثالث ($Q_3$) والربيع الأول ($Q_1$) وهو ما يسمى بالمدى الربيعى . وسوف نتعرض لهذا المقياس بشئ من التفصيل فيما يلى .

٢- نصف المدى الربيعى : **Semi – Interquartile Range**

لتلافى عيوب المدى المطلق يوجد مقياس آخر من مقاييس التشتت يحاول معالجة حساسية المدى للقيم الشاذه أو المتطرفة وذلك من خلال إهمال ربع البيانات (أى 25%) من البداية وكذلك إهمال ربع المفردات من النهاية وذلك بعد ترتيب المفردات تصاعدياً أو تنازلياً أولاً. ثم يتم الحصول على المدى فيما بين الى 50% الباقية من البيانات حتى نتخلص من أثر القيم الشاذه أو المتطرفة سواء فى بداية أو نهاية البيانات. هذا المقياس يسمى بالمدى الربيعى أو الأنحراف الربيعى والذى إذا ما تم قسمته على (2) نحصل على مايسمى بنصف المدى الربيعى. أى أن :

$$\text{Semi. Int. qu. Range} = \frac{Q_3 - Q_1}{2}$$

ويتميز نصف المدى الربيعى بأنه سهل الحساب والفهم والتطبيق كما أنه لا يتأثر نوعاً ما بالقيم الشاذه أو المتطرفة. هذا بالإضافة إلى أنه المقياس الوحيد من مقاييس التشتت المطلقة الذى يمكن حسابه من الجداول التكرارية المفتوحة. وبعبارة أخرى فإن نصف المدى الربيعى يفضل حسابه كمقياس من مقاييس التشتت المطلقة فى حالة الجداول التكرارية المفتوحة سواء من أعلى

**أو أسفل أو مفتوحة الطرفين . إلا أنه يعاب عليه أنه يعتمد كذلك فى حسابه على قيمتين فقط مهملاً باقى القيم شأنه فى ذلك شأن المدى فضلاً عن أنه يتأثر بحجم العينة شأنه فى ذلك أيضاً شأن المدى كما أنه لا يخضع خضوعاً تاماً للعمليات الجبرية إذ أنه لا يمكن حساب نصف المدى الربيعى العام لعدة مجموعات جزئية إذا علم نصف المدى الربيعى لكل مجموعة جزئية.**

**وفى النهاية يجب أن نؤكد على أن جودة نصف المدى الربيعى كمقياس من مقاييس التشتت المطلق يعتمد على درجة تركيز المفردات أو القيم عند قيمة الربيعين الأدنى والأعلى ( $Q_1$ ، $Q_3$). فإذا كان هناك فراغ أو تبعثر أو ثغرات فى البيانات حول الربيعين فإنه يصبح نصف المدى الربيعى غير ملائماً لقياس تشتت الظاهرة محل الدراسة.**

***<u>مثال :-</u>***

**فيما يلى لديك التوزيع التكرارى لدرجات مجموعة من الطلاب فى اختبار مادة الإحصاء.**

**جدول (٤-١)**

| Classes | 0- | 10- | 20- | 30- | 40-50 | Σ |
|---|---|---|---|---|---|---|
| Frequency | 5 | 10 | 24 | 12 | 9 | 60 |

**والمطلوب :-**

**١- أحسب المدى . والمدى الأول.**

**٢- أحسب نصف المدى الربيعى لهذا التوزيع.**

***<u>الحل :</u>***

**حيث أن البيانات المعطاه لتوزيعاً تكرارياً لذا فأن :**

**١- المدى هو :**

Range = Midpoint for the latter class- Midpoint for the first class

= 45 – 5 = 40 (degree) , or

Range = U.L. for the latter class – L.L for the first class

= 50 – 0 = 50 (degree )

**لاحظ إختلاف القيم ليس إلاً نتيجة أختلاف مفهوم الصيغة الرياضية التى يتم حساب المدى منها .**

**أما عن شبيهات المدى وبالتحديد المدى الأول فيتم حسابه من خلال إيجاد الفرق ما بين مركز الفئة قبل الأخيرة ومركز الفئة الثانية أو الحد الأعلى للفئة قبل الأخيرة والحد الأدنى للفئة الثانية . أى أن المدى الأول هو أما أن يكون :**

$1^{st}$ Range = 35 – 15 = 20 (degree)

or = 40 – 10 = 30 (degree)

**اما لحساب نصف المدى الربيعى فيجب فى البداية ايجاد الجدول التكرارى المجتمع الصاعد او الهابط وذلك على النحو التالى :**

**جدول ( ٤ – ٢ )**

| Classes | Frequency | Less than the U.Lfor classes | Asc.c.f |
|---|---|---|---|
| 5 – | 5 | Less than 0 | 0 |
| 10 – | 10 | Less than 10 | 5 |
| 20 – | 24 | Less than 20 | 15 ← $Q_1$ |
| 30 – | 12 | Less than 30 | 39 |
| 40 – 50 | 9 | Less than 40 | 51 ← $Q_3$ |
| | | Less than 50 | 60 |
| $\sum$ | 60 | | |

**ثم تحديد رتبة كل ربيع ومن ثم تطبيق القانون فى تحديد كل قيمة من قيم الربيعين اى ان :**

$$\text{Rank of } Q_1 = \frac{\sum F}{4} = \frac{60}{4} = 15$$

**وبملاحظة عمود التكرار المجتمع الصاعد نجد ان رتبة الربيع الادنى جاءت بصورة صريحة معبرة عن احد تلك التكرارات المتجمعة الصاعدة لذا فان قيمة الربيع الادنى تمثل الرقم المقابل للفئات المجتمعة اى ان**

$Q_1 = 20$ degree

**اما عن الربيع الثالث فان :**

**رتبة الربيع الثالث هى :**

$$\text{Rank of } (Q_3) = \frac{\sum F}{4} \times 3$$

$$= \frac{60}{4} \times 3 = 45$$

**وبملاحظة رتبة الربيع الثالث (45 ) تقع ما بين التكرارين المجتمعين 51,39 ومن ثم فان بداية فئة الربيع الثالث هى 30 درجة وبتطبيق الصيغية الحسابية للربيع الثالث فان :**

$$Q_3 = \text{L.L for } (Q_3) \text{ class} + \frac{\frac{3\sum F}{4} - \text{Pre. Asc. C.F. for the rank of } (Q_3)}{\text{Suc. Asc. C.F. for} - \text{Pres. Asc. C.F. for the rank of } (Q_3)} \times LQ_3$$

**اى ان :**

$$Q_3 = 30 + 10\left(\frac{45\text{-}39}{51\text{-}39}\right) = 30 + 10\left(\frac{6}{12}\right)$$

i.e.,

$Q_3 = 35$ (degree)

**وعليه فان نصف المدى الربيعى يكون :**

$$\text{Semi – Int. Qu. Range} = \frac{Q_3 - Q_1}{2}$$

$$= \frac{35\text{-}20}{2} = 7.5 \quad \text{(degree)}$$

**واخيرا فبالاضافة الى ان مقياس نصف المدى الربيعى كمقياسا للتشتت المطلق يتغلب على مشكلة القيم المتطرفة فى البيانات كما انه يمكن ايجاده من الجداول التكرارية المفتوحة فان هذا المقياس – نصف المدى الربيعى يكون ملائما فى حالة التوزيعات التكرارية الملتوية ( اللهم الا اذا كان التوزيع ملتو التواءا شديدا بحيث تتركز ربع القيم فى الفئة الاولى او الفئة الاخيرة مما يصعب معه حساب الربيعين الادنى والاعلى ) حينئذ يصعب حساب نصف المدى الربيعى ومع ذلك فان نصف المدى الربيعى يؤخذ عليه انه غير حساس للقيم الاقل من الربيع الادنى ( $Q_1$ ) والقيم الاكبر من الربيع الاعلى وكذلك القيم التى تنحصر بينها وهو ما يعنى ان هذا المقياس – مثل المدى – يقتصر فى حسابة على قيمتين فقط هما $Q_1$ , $Q_3$ فى حالة نصف المدى الربيعى وعلى قيمتى $X_{(1)}$ ، $X_{(n)}$ فى حالة المدى .**

**وفيما يلى نتعرض بالدراسة الى مقياسين آخرين من مقاييس التشتت المطلقة والذين لا يعتمدان فى حسابهما على قيمتين فقط وانما يعتمدان على جميع القيم ومن ثم لا يهملان جزءا من المعلومات المتاحة عن الظاهرة محل الدراسة وهذان المقياسان هما متوسط الانحرافات المطلقة ( او ما يسمى بالانحراف المتوسط ) والانحراف المعيارى .**

**٣- الانحراف المتوسط : Mean Deviation (MD)**

**او متوسط الانحرافات المطلقة : Mean of Absolute Deviations**

**رأينا فيما سبق ان كل من المدى ونصف المدى الربيعى يستخدم فى حساب قيمتين فقط يختلف موضعهما باختلاف المقياس المسحوب ويهمل باقى القيم اما الانحراف المتوسط فهو ياخذ فى اعتباره جميع مفردات التوزيع عند الحساب ويعرف الانحراف المتوسط لمجموعة من القيم بانه عبارة عن متوسط الانحرافات المطلقة لهذه القيم عن وسطها الحسابى ($\bar{X}$ ) هو عبارة عن الفرق بين تلك القيمة وبين الوسط الحسابى مع اهمال اشارة الفرق اى ان :**

$d_i = |\ X_i - \bar{X}\ |$ where $i= 1,2,\ldots,n$

**وقد تم اعتبار الفرق المطلق اى أهملت الاشارة هنا وذلك لانه وفقا لخاصية الوسط الحسابى فان مجموع انحرافات القيم عن وسطها الحسابى يساوى صفرا ( وهو ما سبق اثباته عند دراستنا لخصائص الوسط الحسابى فى الباب السابق ). ولذلك فان الانحراف المتوسط يتم حسابه على اساس اعتبار الانحرافات المطلقة للقيم عن وسطها الحسابى او الوسيط اى اهمال الاشارة التى تنتج عن انحراف اى قيمة من القيم عن الوسط الحسابى او الوسيط لتلك القيم والا فان هذا المقياس سوف تساوى قيمته الصفر دائما ايا كانت مجموعة البيانات ان لم يتم اخذ الفرق المطلق لتلك الانحرافات. اى ان الانحراف المتوسط هو عبارة عن :**

$$\text{Mean Deviation (MD)} = \frac{\sum_{i=1}^{n} |\ X_i - \bar{X}\ |}{n}$$

**ويتضح من تعريفنا للانحراف المتوسط ان قيمته تعتمد على مدى تقارب او تباعد المفردات عن الوسط الحسابى ( او الوسيط ) .**

***<u>الانحراف المتوسط فى حالة البيانات الغير مبوبة ( المفردات ) :</u>***

**سبق وان ذكرنا حالا انه اذا كان لدينا متغيرا يأخذ مجموعة القيم $X_1 , X_2 , X_n,$ فان الانحراف المتوسط عن الوسط الحسابى يعرف كالاتى :**

$$\text{Mean Deviation about Mean (MD)} = \frac{\sum_{i=1}^{n} \left| X_i - \bar{X} \right|}{n}$$

**كما ان الانحراف المتوسط عن الوسيط يعرف على الصورة التالية :**

$$\text{Mean Deviation about median} = \frac{\sum_{i=1}^{n} \left| X_i - \bar{X} \right|}{n}$$

**حيث تغنى $\sum|X_i - \bar{X}|$ ، $\sum|X_i - Q_2|$ مجموع الانحراف المطلق ( اى بعد إهمال اشارة كل انحراف ) عن الوسط او عن الوسيط على الترتيب .**

**هذا ويمكن حساب الانحراف المتوسط عن اى قيمة متوسطة اخرى كالمنوال او الوسط الهندسى التوافقى الا ان هذا نادر الحدوث ويعتبر المقياس الاكثر استخداما هو الانحراف المتوسط عن الوسط الحسابى او الوسيط . واذا ذكر الانحراف المتوسط بدون تخصيص بانه حول اى من قيم المتوسطات المختلفة فيكون المقصود به هو الانحراف المتوسط عن الوسط الحسابى .**

***<u>مثال :</u>***

**احسب الانحراف المتوسط عن الوسط الحسابى مرة وعن الوسيط مرة اخرى لمجموعة البيانات التالية : 150,131,139,135,134,137,133**

***الحل :-***

**لايجاد الانحراف المتوسط عن الوسط الحسابى يتم حساب الوسط الحسابى للقيم وذلك على النحو التالى :**

$$\bar{X} = \frac{\sum X}{n} = \frac{133 + 137 + \dots\dots\dots + 131 + 150}{7}$$

$$= \frac{959}{7} = 137$$

**والجدول التالى يعطى الفرق فيما بين محموعة المفردات والوسط الحسابى ومن ثم الانحرافات المطلقة لتلك الفروق :**

**جدول (٤-٣ )**

| $X_i$ | $d_i = X_i - 137$ | $d_i = \lvert X_i - \bar{X} \rvert$ |
|---|---|---|
| 133 | 133-137= - 4 | 4 |
| 137 | 0 | 0 |
| 134 | - 3 | 3 |
| 135 | - 2 | 2 |
| 139 | 2 | 2 |
| 131 | - 6 | 6 |
| 150 | 13 | 13 |
| 259 | + 15-15=0 | 30 |

**ومن خلال نتائجنا فى الجدول السابق فتكون قيمة الانحراف المتوسط عن الوسيط هى عبارة عن :**

$$Mean\ Deviation = \frac{\sum|X_i - \bar{X}|}{n} = \frac{\sum|d_i|}{n} \quad i.e.,$$

$$MD_{\bar{X}} = \frac{30}{7} = 4.28$$

أما لايجاد الانحراف المتوسط عن الوسيط فيجب ترتيب المفردات وليكن تصاعديا لتحديد قيمة الوسيط اولا فتكون مجموعة القيم على الصورة :

**131, 133, 134, 135, 137, 139, 150**

وحيث ان عدد المفردات (n) = 7 وهو عددا فرديا لذا فان رتبة الوسيط

$$Rank\ of\ Q_2 = \frac{n+1}{2} = \frac{7+1}{2} = 4$$

اى ان الوسيط هو عبارة عن المفردات الرابعة للبيانات بعد ترتيبها اى ان قيمة الوسيط هى 135 اى ان $Q_2=135$ وعليه فانه لحساب قيمة الانحراف المتوسط عن الوسيط يتم تكوين الجدول التالى لحساب الفروق المطلقة للقيم عن الوسيط .

**جدول (٤-٤)**

| $X_i$ | $d_i = X_i - Q_2 = X_i - 135$ | $\lvert d_i \rvert = \lvert X_i - 135 \rvert$ |
|---|---|---|
| 131 | -4 | 4 |
| 133 | -2 | 2 |
| 134 | -1 | 1 |
| 135 | 0 | 0 |
| 137 | 2 | 2 |
| 139 | 4 | 4 |
| 150 | 15 | 15 |
| 959 | | 28 |

ومن خلال نتائج جدول ( ٤-٤ ) فان :

الانحراف المتوسط من الوسيط هو :

$$\textit{Mean Deviation} = \frac{\sum|X_i - Q_2|}{n} = \frac{\sum|d_i|}{n}$$

*about Median*

$$i.e., MDQ2 = \frac{28}{7} = 4$$

- **الانحراف المتوسط فى حالة البيانات المبوبة :**

**فـى حالـة التوزيعـات التكراريـة فانـه لايجـاد الانحـراف المتوسـط يـتم حسـاب الانحرافات المطلقة وذلك من خلال ايجاد الفرق فيما بين كل قيمة من قيم مراكز فئـات التوزيـع وبـين الوسـط الحسـابى او الوسـيط حسـب المطلـوب آخـذين فـى الاعتبار تكرارات فئات التوزيع اى ان :**

**الانحراف المتوسط عن الوسط :**

$$\textit{Mean Deviation} = \frac{\sum|X_i - \bar{X}|\, F_i}{\sum F_i}$$

$$MD_{\bar{X}} = \frac{\sum|d_i|F_i}{\sum F_i}$$

**حيث $X_i$ , $F_i$ تعبـر عن مركـز وتكـرار الفئـة (i) علـى الترتيـب حيـث $i = 1,2,\ldots..n$، n تعبر عن عدد فئات التوزيع . حيث ان :**

$$\bar{X} = \frac{\sum X_i\, F_i}{\sum F_i}$$

**اما عن الانحراف المتوسط عن الوسيط فهو عبارة عن :**

$$\textit{Mean Deviation} = \frac{\sum|X_i - Q_2|F_i}{\sum F_i}$$

**about Median**

**مثال :**

**احسب الانحراف المتوسط عن الوسط الحسابى للتوزيع التكرارى التالى والذى يمثل توزيع اعمار مجموعة معينة من موظفى احد الادارات بالجامعة(بالسنة):**

**جدول (٤-٥)**

| Classes | 20- | 28- | 36- | 44- | 52-60 | $\sum$ |
|---|---|---|---|---|---|---|
| Frequency | 4 | 12 | 17 | 4 | 3 | 40 |

**الحل :**

**لحساب الانحراف المتوسط عن الوسط الحسابى يتم حساب الوسط الحسابى اولا :(وفى حالة الجداول المنتظمة يفضل حسابه بطريقة الانحرافات المعدلة ) ومن ثم نكون بصدد الجدول التالى (٤-٦ ) والذى يوضح الحسابات اللازمة لحساب الوسط الحسابى ومن ثم الانحراف المتوسط عن قيمة الوسط الحسابى**

**جدول (٤-٦)**

| **Classes** | $F_i$ | $X_i$ | $d_i=X_i-40$ | $D_i=\frac{D_i}{8}$ | $D_i F_i$ | $\lvert X_i-\bar{X}\rvert$ | $\lvert X_i-\bar{X}\rvert F_i$ |
|---|---|---|---|---|---|---|---|
| 20- | 4 | 24 | -16 | -2 | -8 | 14 | 56 |
| 28- | 12 | 32 | -8 | -1 | -12 | 6 | 72 |
| 36- | 17 | 40 | 0 | 0 | 0 | 2 | 34 |
| 44- | 4 | 48 | 8 | 1 | 4 | 10 | 40 |
| 52-60 | 3 | 56 | 16 | 2 | 6 | 18 | 54 |
| $\sum$ | 40 | | | | -10 | | 256 |

**فيكون :**

$$\overline{X} = A + B\,\frac{\sum D_i F_i}{\sum F_i}$$

$$= 40 + 8\left(\frac{-10}{40}\right) = 40 - 2 = 38 \qquad (year)$$

وعليه فان الانحراف المتوسط على الوسط :

$$MD_{\overline{X}} = \frac{\sum |X_i - \overline{X}| F_i}{\sum F_i} = \frac{256}{40} = 64 \qquad (year)$$

**مثال:** احسب الانحراف المتوسط عن الوسيط للتوزيع التكرارى التالى:

جدول (٤-٧)

| Classes | Less than 40 | 40- | 48- | 54- | 60-80 | $\sum$ |
|---|---|---|---|---|---|---|
| Frequency | 10 | 15 | 50 | 22 | 3 | 100 |

**الحل :** لحساب الانحراف المتوسط عن الوسيط فيجب حساب الوسيط اولا ثم يتم حساب الانحراف المتوسط عن الوسيط والجدول ( ٤-٨) يوضح الحسابات اللازمة لذلك :

جدول (٤-٨)

| Classes | Fi | Xi | Less than the classes upper limit | ASC. C.F | $\lvert X_i - Q_2\rvert$ = $\lvert X_i - 51\rvert$ | $\lvert X_i - Q_2\rvert^{F_i}$ |
|---|---|---|---|---|---|---|
| 30- | 10 | 35 | Less than 40 | 10 | 16 | 160 |
| 40- | 15 | 44 | Less than48 | 25 | 7 | 105 |
| 48- | 50 | 51 | Less than54 | 75 | 0 | 0 |
| 54- | 22 | 57 | Less than60 | 97 | 6 | 132 |
| 60-80 | 3 | 70 | Less than80 | 100 | 19 | 57 |
| $\sum$ | 100 | | | | 48 | 454 |

$$\boldsymbol{Rank\ of\ Q_2 = \frac{\sum F}{2} = \frac{100}{2} = 50}$$

**وحيث ان ترتيب الوسيط (50) يقع فى نقطة المنتصف ما بين التكرارين المتجمعين الصاعدين 25، 75 لذا فان قيمة الوسيط تنصف البعد فيما بين الحدين الأدنى والاعلى للفئة الوسيطية اى تنصف البعد ما بين 48 , 54 اى ان:**

$$\boldsymbol{Q_2 = \frac{48-54}{2} = 51 \qquad (year)}$$

**( يمكنك عزيزى القارئ التأكد من صحة قيمة الوسيط باستخدام صيغة القانون المعروف والمستخدم لحساب الوسيط أو باستخدام قاعدة النسبة والتناسب )**

**وعليه فان الانحراف المتوسط عن الوسيط هو :**

$$\text{Mean Deviation About median} = \frac{\sum |\ X_i - Q_2\ |\ F_i}{\sum F_i}$$

**ومن خلال الجدول السابق فان :**

$$MD_{Q2} = \frac{454}{100} = 4.45 \qquad (\text{year})$$

**وأخيرا فان الانحراف المتوسط كمقياس للتشتت المطلق يعتبر غير شائع الاستخدام وذلك لعدة اسباب هى :**

**١- عدم سهولة العمليات الحسابية اللازمة لايجاده**

**٢- صعوبة التعامل معه جبرا او حسابيا فعلى سبيل المثال لو كان متوفرا لدينا عددا من العينات مسحوبة من مجتمع من المجتمعات عبر فترات زمنية متفاوته واردنا تقدير الانحراف المتوسط للمجتمع بدلاله الانحرافات المتوسطة المحسوبة لهذه العينات لما امكننا ذلك وفى تلك الحالة فاننا نكون مضطرين الى**

**ادماج هذه العينات جميعا فى عينة كلية ثم حساب الانحراف المتوسط لهذه العينة الكلية .**

**هذا وتجدر الاشارة الى انه ليس المقصود بكلامنا هذا ان الانحراف المتوسط قليل الاهمية كمقياس للتشتت فهو يعتبر مقياس جيد اذا ما كنا نود الحصول على مقياس للتشتت بهدف المقارنة بين مجموعتين او اكثر تتحد فى وحده القياس .اما اذا كان الهدف هو ايجاد مقياس للتشتت لاستخدامه فيما بعد فى عمليات التحليل الاحصائى فاننا يجب ان نستخدم مقياسا غيره يكون اكثر ملائمة لهذا الغرض.**

**ولقد كانت الفكرة الاساسية وراء هذا المقياس هى اننا اخذنا الانحرافات المطلقة عن الوسط لتكون كلها موجبة بغض النظر عن اشارتها وذلك تجنبا لمشكلة ان مجموع انحرافات القيم عن وسطها الحسابى يساوى صفر دائما ابدا. مما ينعكس بدوره على ان الانحراف المتوسط سوف ياخذ فى كل الحالات القيمة صفرا وذلك اذا تم التعامل مع الانحرافات باشارتها الفعلية كما هى دون اعتبارا القيمة المطلقة لتلك الانحرافات .**

**والسؤال الذى يطرح نفسه الان اليس هناك بديل آخر للتغلب على اوجه القصور التى يعانى منها الانحراف المتوسط ؟**

**فى واقع الامر يمكن التغلب على تلك المشكلة بايجاد مربعات تلك الانحرافات حيث يكون مجموع المربعات دائما ذا قيمة موجبة ومن ثم فبأيجاد متوسط هذه المربعات فاننا نحصل على ما يسمى بالتباين وهو ما سوف نتعرض له الان بالتفصيل وكذا الانحراف المعيارى .**

## ٤ - التباين والانحراف المعيارى :

Variance and Standard Deviation

**سبق ان راينا ان انحرافات القيم عن وسطها الحسابى تعطى صورة جيدة لدرجة تباين مفردات الظاهرة موضع الدراسة. كما بينا ايضا ان مجموع انحرافات القيم عن وسطها الحسابى يساوى الصفر دائما وللتغلب على ذلك كانت وسيلتنا فى الانحراف المتوسط هى اهمال اشارات الانحراف واعتبارها جميعا موجبة اى الاعتماد على الانحرافات المطلقة وهذا الاهمال للاشارات لا يستند كما سبق ان اوضحنا الى اى منطق رياض وللتغلب على ذلك وتلافيا للتعقيدات التى يسببها وجود القيم المطلقة للانحرافات عند حساب الانحراف المتوسط يمكن اللجوء الى حل اخر للمشكلة وذلك من خلال ايجاد مربع تلك الانحرافات لتصبح كل الانحرافا ت موجبة ثم نوجد حاصل جمع هذه المربعات لتلك الانحرافات ونوجد متوسطها لنحصل على ما يسمى بالتباين وهو ما يرمز له بالرمز ($\sigma^2$).**

**فالتباين يعرف بانه عبارة عن متوسط مربعات انحرافات القيم عن وسطها الحسابى اما الانحرافات المعيارى والذى يرمز بالرمز ($\sigma$) والذى يعتبر من اهم مقاييس التشتت المطلقة فما هو الا الجزر التربيعى ( الموجب ) للتباين اى ان الانحراف المعيارى ما هو الا الجزر التربيعى ( الموجب ) للوسط الحسابى لمربعات انحرافات القيم عن وسطها الحسابى وعادة ما يستخدم الانحراف المعيارى ($\sigma$) كمقياس للتشتت عندما لا يكون قياس تشتت الظاهرة هو نهاية المطاف بل بداية لعملية التحليل الاحصائى الاكثر اهمية والمتعلقة بالاستنتاج الاحصائى لمقاييس المجتمع محل الدراسة وترجع تلك الاهمية الى خضوعه للطرق الرياضية وسهولة التعامل معه جبريا فهو يستخدم على نطاق واسع**

للغاية فى نظرية التقدير الاحصائى واختبارات الفروض الاحصائية كاحد المهام الرئيسية لفرع الاحصاء الاستدلالى – الاحصاء التطبيقى - وفيما يلى نوضح كيفية حساب الانحراف المعيارى ومن ثم التباين فى حالة المفردات وكذا فى حالة البيانات المبوبة تكراريا .

**الانحراف المعيارى للبيانات غير المبوبة ( المفردات ) :**

اذا كان لدينا متغيرا وليكن $X$ ياخذ مجموعة القيم $X_1, X_2, \ldots\ldots X_n$ لظاهرة معينة وكان متوسطها الحسابى هو $(\bar{X})$ فان انحرافات هذه القيم عن وسطها الحسابى هى :

$$(X_1-\bar{X}),\ (X_2-\bar{X}),\ (X_3-\bar{X}),\ \ldots\ldots\ldots,\ (X_n-\bar{X})$$

ومن ثم يكون التباين عبارة عن متوسط مربعات انحرافات القيم عن وسطها الحسابى اى ان :

$$\sigma^2 = \frac{1}{n}\left[(X_1-\bar{X})^2 + (X_2-\bar{X})^2 + \ldots\ldots + (X_n-\bar{X})^2\right]$$

اما عن الانحراف المعيارى $\sigma$ فما هو الا الجذر التربيعى للتباين والعكس صحيح اى ان التباين ما هو الا مربع الانحراف المعيارى اى ان :

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n}}$$

والسبب فى اخذ الجزر التربيعى للتباين $\sigma^2$ للحصول على الانحراف المعيارى هو اننا كما يتضح من الصيغة السابقة بانه تم تربيع انحرافات القيم عن الوسط الحسابى ولكى يتم الرجوع الى الوحدات الاصلية بعد عملية التربيع السابقة فانه يجب علينا اخذ الجذر التربيعى لتلك الانحرفات فيكون التشتت بذلك مقيسا بنفس وحدات القيم الاصلية ولهذا السبب فهناك من جمهور الاحصائيين من

**يعتبر ان التباين لا يعتبر مقياسا للتشتت المطلق اما الانحراف المعيارى فهو مقياس التشتت المطلق واهمها نظرا لاهميته فى مجال التحليل الاحصائى .**

**هذا وعلي غرار ما تم في حالة الوسط الحسابي من حيث امكانية حسابه بالطرق الثلاث ( المباشرة – والانحرافات البسيطة والانحرافات المعدله ) فانه يمكننا حساب الانحراف المعياري لظاهرة معينه بنفس الطرق الثلاث مع الاختلاف البسيط في الجهد الا ان هذه الطرق الثلاث تعطي في النهاية نتيجة واحده لمقدار تشتت تلك الظاهره .**

**وقبل تناولنا للطرق الثلاث المستخدمه في حساب الانحراف المعياري او التباين دعنا نعرض المثال التالي لبيان كيفية حساب تلك المقاييس باستخدام الصيغه الاساسيه السابقة .**

***مثال :*احسب الانحراف المعياري لمجموعه القيم التالية :**

٦ ، ١٥ ، ٨، ٩، ١٢ ، ٣

***الحل : -***

**لاستخدام الصيغه الرياضية السابقه في حساب الانحراف المعياري يتم حساب الوسط الحسابي لتلك المفردات بداية : فحيث ان**

$$\bar{X} = \frac{\sum X}{N} = \frac{6+10+8+9+12+3}{6} = \frac{48}{6} = 8$$

**ثم دعنا نقوم بأجراء الحسابات اللازمة لصيغة الانحراف المعياري ( σ) من خلال الجدول التالي :**

جدول ( ٤ – ١٠)

| $X_i$ | $(X_i - \bar{X})$ | $(X_i - \bar{X})^2$ |
|---|---|---|
| 6 | 8 – 6 = -2 | 4 |
| 10 | 10 – 8 = 2 | 4 |
| 8 | 0 | 0 |
| 9 | 1 | 1 |
| 12 | 4 | 16 |
| 3 | -5 | 25 |
| 48 | -7+7= 0 | 50 |

لابد من تحقيق ذلك

وعليه فأن

$$\sigma = \sqrt{\frac{\sum(X_i - X)^2}{n}} = \sqrt{\frac{50}{6}}$$

$$= \sqrt{8.333} \simeq 2.887$$

اي ان الانحراف المعياري لتلك القيم هو تقريبا 2.887 اما التباين فهو عبارة عن الناتج تحت الجزر التربعي اي ان $\sigma^2 = 8.333$

وجدير بالذكرانه بفضل استخدام هذه الصيغه الرياضيه لحساب الانحراف المعياري او التباين اذا كان الوسط الحسابي للقيم لا يحتوي علي قيما كسرية وأن كانت هذه الصيغه تستخدم جهدا حسابيا لا مبرر له حيث تسهل الصيغه التاليه عنها .

**الطريقة المباشرة لحساب الانحراف المعياري :**

**في كثير من الاحيان وعند حساب الانحراف المعياري بالصيغه السابقة فأنه غالبا ما نجد ان قيمة الوسط الحسابي تأخذ قيمة كسرية مما يصعب معه تعقيد الحسابات اللازمه لحساب الانحراف المعياري او التباين بتلك الصيغه هذا بالاضافه الي التقريب المستخدم غالبا والذي قد يؤدي الي عدم دقة في النتائج لذا ينصح بالتغلب علي ذلك من خلال ما يلي :**

**فحيث ان :**

$$\boldsymbol{\sigma} = \sqrt{\frac{\sum(X-\bar{X})^2}{n}}$$

$$\sum(X_i-\bar{X})^2 = \sum_{i=1}^{n}(X_i^2 - 2\,\bar{X}\,X_i + \bar{X}^2)$$

$$= \sum_{X_i{}^2} - 2\,\bar{X}\sum_{Xi} + n\,\bar{X}^2$$

$$= \sum_{X_i{}^2} - 2\,n\,\overline{X^2} + n\,\bar{X}^2$$

$$= \sum_{X_i{}^2} - n\,\bar{X}^2$$

$$= \sum_{X_i{}^2} - \frac{(\sum X_i)^2}{n}$$

**وعليه فإنه :-**

$$\boldsymbol{\sigma} = \sqrt{\frac{1}{n}\left[\sum_{X_i{}^2} - \frac{(\sum xi)^2}{n}\right]}$$

$$= \sqrt{\frac{\sum \overset{2}{X}i}{n} - \left(\frac{\sum Xi}{n}\right)^2} \qquad \mathbf{(2)}$$

**ومما لا شك فيه ان هذه الصيغه الاخيرة للانحراف المعياري اسهل بكثير من الصورة السابقة . حيث لا يتطلب حساب الانحراف المعياري وفقا للصورة**

الاخيرة طرح قيمـة الوسط الحسـابي مـن كل قيمـة مـن قيم المتغير ومـا يترتـب علي ذلـك مـن صـعوبة فـي اجراء العمليـات الحسـابية . وأنمـا يتطلـب وفقـا لتلـك الصيغه مجرد حساب مجموع القيم اي $\sum x_i$ ومجموع مربعات هذه القيم اي . $(\sum x_i)$

ففي مثالنا فأن :-

$\sum_x = 48 \quad , \quad \sum x^2 = 434$

فيكون الانحراف المعياري عبارة عن :

$$\sigma^2 = \sqrt{\frac{\sum X^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

$$= \frac{434}{6} - \left(\frac{48}{6}\right)^2 = \sqrt{8.333} = 2.887$$

اما عن التباين فهو مربـع الانحـراف المعيـاري او النـاتج تحت الجـزر التربيعـي اي ان :

$$\sigma^2 = 8.333$$

وهي نفس النتائج السابقه .

وفـي هذا المثـال قد لا تبـرز اهميـة صيغة القـانون الاخير بشكل ملمـوس حيث سـهولة اسـتخدامه وذلـك لان الوسـط الحسـابي لمجموعـة القـيم يأخـذ صـورة كسرية مما يجعل معه طريقة حساب الانحراف المعياري بأتبـاع الصيغه الاولـي امرا اكثر تعقيدا . ولكن وكما سبق وأن ذكرنـا فـأن الصيغه الثانيـة تبسط كثيـرا من العمليات الحسابية في الحـالات التي يكون قيمـة الوسط الحسـابي لمجموعـه المفردات تحتوي علي كسور مما يجعل معه ايجاد الانحرافات عنه وكذا مربعات

**هذه تلانحرافات عملية معقدة تنطوي علي احتمالات كبيرة للوقوع في اخطاء حسابية .**

***مثال : -***

**احسب الانحراف المعياري لمجموعة المفردات التالية :**

**3, 1, 9, 8, 2, 4, 6**

**وذلك من خلال استخدامك :**

**أ - لطريقة الانحرفات عن الوسط الحسابي**

**ب – للقيم الاصلية**

**وماذا نستنتج بين النتيجتين . علق علي نتائجك .**

***الحل:***

**لحساب الانحراف المعياري بالطريقتين سواء بالانحرافات سواء بالنحرافات عن الوسط الحسابي وباستخدام القيم الصلية دعنا اولا نقوم بحساب الويط الحسابي للمفردات . فحيث ان**

$$\overline{X} = \frac{\sum x}{n} = \frac{3 + 1 + 9 + \ldots\ldots\ldots + 6}{7} = \frac{33}{7} = 4.71$$

**ثم دعنا نجري الحسابات اللازمة للطريقتين من خلال الجدول التالي :**

| $X_i$ | $X_i^2$ | $(xi - \bar{X})$ | $(x_i - \bar{x})^2$ |
|---|---|---|---|
| 3 | 9 | 3 - 4.71 = -1.71 | 2.9241 |
| 1 | 1 | 1 - 4.71= 13.71 | 13.7641 |
| 9 | 81 | 4.29 | 18.4041 |
| 8 | 64 | 3.29 | 10.8241 |
| 2 | 4 | - 2.71 | 7.3441 |
| 4 | 16 | - 0.71 | 0.5041 |
| 6 | 36 | 1029 | 106641 |
| 33 | 211 | 0.03 $\simeq$0 | 55.4287 |

**وعلي ذلك فان :**

**قيمة الانحراف المعياري باستخدام :-**

**ا – طريقة الانحراف عن الوسط هي :**

$$\sigma = \sqrt{\frac{\sum(X-\bar{X})^2}{n}} = \sqrt{\frac{55.4287}{7}} = \sqrt{7.918385714} = 2.813962636$$

**إما بإستخدام القيم الأصلية فإن**

$$\sigma = \sqrt{\frac{\sum X^2}{n} - \left(\frac{\sum \quad x}{n}\right)^2} = \sqrt{\frac{211}{7} - \left(\frac{33}{7}\right)^2}$$

$$= \sqrt{7.918367347} \simeq 2.813959372$$

**لاحظ ان الاختلاف فيما بين قيمة الانحراف المعياري ياتي في الرقم الخامس من بعد العلامه العشرية والسبب في ذلك راجع الي اخطاء التقريب التي صاحبت طريقة الانحرافات عن الوسط الحسابي حيث احتوي الوسط الحسابي**

علي قيمة كسرية مما يترتب عليه ان ناتج الجمع الخاص بمجموع انحرافات القيم عن الوسط الحسابي ماويا 0.03 اي تقريبا مساويا للصفر لكنه لم يساوي الصفر بصورة صريحة وهو راجع ايضا لخطأ التقريب في قيمة الوسط كذلك . والخلاصه ان الوسط الحسابي باستخدام القيم الاصلية يعتبر اكثر دقة من طريقة الانحرافات حيث لا تتضمن هذه الطريقة اى تقريب نهائيا اللهم الا في الناتج النهائي اذا اردنا ذلك وكما يتضح لنا من المثالين السابقين فان اذا كان الوسط الحسابي قيمته عددا صحيحا

( اي لا يحتوي علي فيمة كسرية ) فسوف يتساوي قيمة الانحراف المعياري باستخدام اي من الاسلوبين سواء باستخدام طريقة الانحرافات حول الوسط او باستخدام القيم الاصلية لانه لن يكون هناك اي عمليات تقريب وأن كان من المفضل استخدام الطريقة الخاصه بالقيم الاصلية لانها تحتوي علي عمليات حسابية اقل كما سبق ان اوضحنا . اما اذا كان الوسط الحسابي يحتوي علي قيمة كسرية فانه يفضل حساب الانحراف المعياري باستخدام القيم الاصلية . لانها تتجنب اخطاء التقريب في الحسابات وبالتالي نتيجه اكثر دقة .

***<u>خصائص الانحراف المعياري :</u>***

١ – لا يتاثر الانحراف المعياري بعمليتي الجمع او الطرح : بمعني اذا كان لدينا قيم المتغير $x$ هي مجموعه القيم $x_1, x_2, \ldots\ldots, x_n$ فان :-

الوسط الحسابي لتلك القيم هو :

$$\bar{X} = \frac{\Sigma x}{n}$$

كما ان الانحراف المعياري لها هو :

$$\sigma_X = \sqrt{\frac{\sum(X-\bar{X})^2}{n}}$$

فاذا تم طرح مقدار ثابت وليكن ( a ) من كل قيمة من القيم الاصلية فان قيم الانحرافات الناتجه حول هذا المقدار الثابت هي :

$d_1 = X_1 - a, d_2 = X_2 - a, \ldots\ldots , X_n = d_n - a$

ومن ثم فان :

الوسط الحسابي للقيم $d_i$ حيث $i = 1 , 2 ,\ldots\ldots,n$ هو :

$$\bar{d} = \frac{\sum di}{n}$$

$$= \frac{\sum(X_i - a)}{n} = \frac{\sum X_i - na}{n}$$

$$= \frac{\sum X}{n} - a = \bar{X} - a$$

كما ان الانحراف المعياري لمجموعة القيم $d_i$ هو :

$$\sigma_{di} = \sqrt{\frac{\sum(di-\bar{d})^2}{n}}$$

وبالتعويض في المعادله الاخيره عن( $d_i = x_i - a$ ) ، ( $\bar{d} = \bar{x} - a$ ) ينتج لنا ان الانحراف المعياري لانحرافات القيم $x$ عن الوسط الفرض ( $a$ ) . هو :

$$\sigma_{di} = \sqrt{\frac{\sum(X_i - a - \bar{X} + a)^2}{n}}$$

$$= \sqrt{\frac{\sum(X_i - \bar{X})^2}{n}}$$

**اي ان :**

$$\sigma_{di} = \sigma_{Xi}$$

**اي ان الانحراف المعياري لانحرافات القيم xi عن الوسط الفرضي ( a ) هي نفسها الانحراف المعياري لقيم الاصلية . اي ان الانحراف المعياري لم يتأثر بعملية الطرح . ولذا يمكن إثبات عدم تأثر الانحراف المعياري بالعملية العسكية للطرح ( اي الجمع ) . ومن ثم فإن الانحراف المعياري لا يتأثر بعملية الجمع او الطرح . ولعل السبب في ذلك يرجع الي ان قيمة الانحراف المعياري يقيس مدي تباعد او تقارب القيم عن بعضها ( اي درجة التشتت ) فمما لا شك فيه ان البعد ما بين المفردات سيظل ثابتا اذا تم طرح مقدار ثابت وكذلك يظل هذا التباعد ثابتا اذا تم اضافة مقدار ثابت لتلك القيم الاصلية .**

**٢- الانحراف المعياري يتأثر بعمليتي الضرب والقسمه**

**ويمكن اثبات ذلك رياضيا بنفس الالية السابقة المستخدمه في الخاصية الاولي وذلك علي النحو التالي :**

**فنفرض ان لدينا محموعة القيم $x_1$ , $x_2$ , ..... , $x_N$ حول او للمتغير ( x ) فان :**

**الوسط الحسابي لمجموعة تلك المفردات هو :**

$$\bar{X} = \frac{\sum x}{n}$$

**وكذلك فإن الانحراف المعيارى لها هو :**

$$\sigma_X = \sqrt{\frac{\sum(X-\bar{X})^2}{n}}$$

**فإذا تم ضرب مقدار ثابت وليكن (a) فى كل قيمة من قيم المتغير (X) حيث أنـه قيمة ( a ) لا تساوي الصفر اي ( $a \neq 0$ ) فان مجموع الصيغ الجديده ما هي الا :** $D_1 = ax_1, D_2 = ax_2, \ldots\ldots, D_n = ax_n$

**فيكون الوسط الحسابى لها هو :**

$$\bar{D} = \frac{\sum D_i}{n} = \frac{\sum(ax_i)}{n} = a\,\frac{\sum x}{n}$$

**أى أن :**

$$\bar{D} = a\bar{x}$$

**أما الإنحراف المعيارى لتلك القيم فهو عبارة عن :-**

$$\sigma_D = \sqrt{\frac{\sum(Di-\bar{D})^2}{n}} = \sqrt{\frac{\sum(ax_i-a\bar{x})^2}{n}}$$

**أى أن :**

$$\sigma_D = \sqrt{\frac{a^2\ \sum(X_i-\bar{X})^2}{n}}$$

$$= a\sqrt{\frac{\sum(X_i-\bar{X})^2}{n}} = a\,\sigma_x$$

**وهو ما يفيد ان ان الانحراف المعياري للقيم الجديدة ( اي بعد ضرب القيم الاصلية في مقدار ثابت وهو المقدار a) مـا هو الا حاصل ضرب الانحـراف المعيـاري للقيم الاصلية ( اي قبل عملية الضرب ) مضروبا في المقدار الثابت (a ) . ومن ثم فان**

$$\sigma_x = \frac{\sigma_D}{a}$$

**اي ان الانحراف المعياري للقيم الاصلية ($x_i$) = الانحـراف المعيـاري للقيم النـاتجـة بعد اجراء عملية ضرب المفردات الاصلية في المقدار القابت (a)**

**كذلك يتم التعامل مع العملية العكسية لعملية الضرب ( اي عملية القسمة ) نفس معاملة عملية الضرب . بمعني انه اذا تم قسمه قيم المتغير ( $x_i$ ) وليكن ( a ) حيث $(a \neq 0)$**

**فانه الانحراف المعياري للقيم الاصلية اي $\sigma_x$ سوف يكون مساويا للانحراف المعياري للقيم الجديدة ( اي d x) مضروبا في هذا المقدار الثابت ( a) هذا ويمكن بيان مدي صحة هاتين الخاصيتين السابقتين من خلال المثال البسيط التالي :**

**افرض ان لديك المفردات الخمس التاليه 400,400,300,200,100 فان :**

- **الوسط الحسابي لتلك المفردات الاصلية هو** $\bar{X} = \frac{\sum x}{n}$

$$= \frac{100+200+300+400+500}{5} = \frac{1500}{5} = 300$$

**كما ان الانحراف المعياري لتلك المفردات لتلك المفردات الاصلية هو :**

$$\sigma_x = \sqrt{\frac{\sum(x-\bar{x})^2}{2}}$$

$$= \sqrt{\frac{(100-300)^2 + (200-300)^2 + ........ + (500-300)^2}{5}} = 141.2413562$$

**والآن دعنا نفرض أنه تم طرح مقدار ثابت وليكن 100 من المفردات الخمسة أى أن (a=100). فإن مجموعة الانحرافات للقيم الأصلية عن الوسط الفرض 100 هى**

$$d_i = 0, 100, 200, 300, 400$$

**فيكون الوسط الحسابى لتلك الانحرافات هو**

$$\bar{d} = \frac{0+100+200+300+400}{5} = 200$$

كما أن الإنحراف المعيارى لتلك الإنحرافات هو :

$$\sigma_{di} = \sqrt{\frac{(0-200)^2+(100-200)^2+.........+(400-200)^2}{5}} = 141.4213562$$

اي ان اثر التغير فقط وقع علي قيمة الوسط الحسابي اما الانحراف المعياري فلا يتاثر بالطرح او الجمع لكن يتاثر فقط بالضرب او القسمة وللتاكد من تاثره بعملية الضرب والقسمة : دعنا نقوم بقسمة المفردات محل هذا المثال علي مقدار الثابت ( 100 ) فتكون نواتج قسمة المفردات الاصلية علي 100هي :

$D_i = 1,2,3,4,5,$

فان الوسط الحسابي لتلك القيم الجديده هو :

$$\bar{D} = \frac{1+2+3+4+5}{5} = \frac{15}{5} = 3$$

اما الانحراف المعياري للقيم الجديدة ( اي لمجموعة القيم $D_i$ )

فهو عباره عن :

$$\sigma_D = \sqrt{\frac{\sum(Di-\bar{D})^2}{n}}$$

$$= \sqrt{\frac{(1-3)^2+2-3^2+.........+(5-3)^2}{5}} = 1.414213562$$

أى أن :-

$$\bar{X} = 300 \quad = 100\ (3) \quad = a.\bar{D}$$

وكذلك فأن :-

$$\sigma_x = 141.4213562$$

$$= 100\,(1.414213562) \quad = a.\sigma_D$$

اي ان الوسط الحسابي وكذا الانحراف المعياري للقيم الاصلية هو عبارة عن حاصل ضرب المقدار ( a ) في الوسط الحسابي والانحراف المعياري للقيم الجديدة علي الترتيب .

ومن خلال دراستنا لهاتين الخاصيتين وللمثال المبسط السابق يمكننا الان دراسة طريقتي الانحراف البسيطة والانحرافات المعدلة واللتين تعتمدان بصفة اساسية في تطبيقاتها علي هاتين الخاصيتين للانحراف المعياري .

***طريقة الانحرافات البسيطة :-***

وتهدف هذه الطريقة كما سبق وان لااينا عند دراستنا للوسط الحسابي الي اختصار العمليات الحسابية وخصوصا عندما يكون عدد المفردات كبيرا مما يؤدي الي تحقيق وفره في الجهد الحسابي وتقليل احتمالات الخطأ حيث يتم طرح وسطا فرضيا مناسبا من مجموعة القيم الاصلية بحيث نحصل علي اصغر اعداد ممكنه تسمي بالانحرافات البسيطة . ويتم حساب الانحراف المعياري لهذه الانحرافات فيكون هو نفسه الانحراف المعياري للقيم الاصلية . اذا ان الانحراف المعياري لا يتأثر بعمليتي الطرح او الجمع ( وفقا للخاصية الاولي من خصائص الانحراف المعياري ) كما سبق ان اوضحنا ذلك .

فاذا اعتبرنا ان قيم الانحرافات هي $d\ i = xi - a$ حيث $i = 1,2,\ldots,n$ تمثل الانحرافات البسيطة حول الوسط الفرض (a) فان :

$$\sigma_X = \sigma_d = \sqrt{\frac{\sum di^2}{n} - \left(\frac{\sum di}{n}\right)^2}$$

والوسـط الفـرض ( a ) يـتم اختيـاره بالشـكل الـذي يـؤدي الـي تبسـيط العمليـات الحسابية .

## طريقة الانحرافات المعدلة ( الاكثر اختصارا ) :-

بعد الحصول علي الانحرافات البسيطة $d_i = x_i - a$ فاذا كانت جميع الانحرافات البسيطة $d_i$ تقبل القسمة علي وسط فرض اخر ليكن بدون بواقي . فأنـه يتم قسمة قيم الانحرافات البسيطة $(d_i)$ على الوسط الفرضى $(d)$ فنحصل علي الانحرافات المعدلة $D_i$ حيث ان

$$D_i = \frac{X_i - a}{b} = \frac{di}{b} \qquad where \quad i = 1, 2, \ldots\ldots, n$$

والانحرافات المعدلة لا شك انها تحقق وفرا اكثر في الجهد الحسابي حيث تؤدي الي تبسيط قيم المفردات ويكـون الانحـراف المعيـاري للقيم الاصلية $(Xi)$ مـا هـو الا عبـارة عن الاتـزان المعيـارى للانحـراف المعدلـة$(Di)$تعد ضربها فـي قيمـة الوسط الفرضي $(b)$ حيث ان الانحراف المعياري يتأثر بعملية الضرب والقسمة كما سبق وان اسلفنا (الخاصية الثانية من خصائص الانحرافات المعيارية ) اى ان :

$$\sigma_X = b\,\sigma_d = b\sqrt{\frac{\sum di^2}{n} - \left(\frac{\sum di}{n}\right)^2}$$

***<u>مثال :</u>***

احسب الانحراف المعيارى لمجموعة القيم :-

**625,700,650,725**

وذلك باستخدام

أ- القيم الاصلية

ب- طريقة الانحرافات البسيطة

**ج- طريقة الانحرافات المعدلة**

**الحل :-**

**أ- حساب النحراف المعيارى باستخدام القيم الاصلية : الجدول التالى يوضح الحسابات اللازمة لحساب النحراف المعيارى من واقع استخدامنا للقيم الاصلية:**

| $X_i$ | $X_i^2$ |
|---|---|
| 625 | 390625 |
| 700 | 490000 |
| 650 | 422500 |
| 725 | 525625 |
| 2700 | 1828750 |

**وحيث أن :**

**فإن:**

$$\sigma = \sqrt{\frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2}$$

$$\sigma = \sqrt{\frac{1828750}{4} - \left(\frac{2700}{4}\right)^2} = 39.53$$

**ب- حساب الانحراف المعيارى باستخدام طريقة الانحرافات البسيطة : يتم اختيار اى من قيم المفردات بمثابة وسط فرض ولتكن a=650 والجدول التالى**

يوضح الحسابات اللازمة لحساب الانحراف المعيارى باستخدام طريقة الانحراف البسيطة :

| $X_i$ | di = xi - 365 | $d_i^2$ |
|---|---|---|
| 625 | 625-650= -25 | 625 |
| 700 | 50 | 2500 |
| 650 | 0 | 0 |
| 725 | 75 | 5625 |
| | 100 | 8750 |

وحيث أن :

$$\sigma = \sqrt{\frac{\sum di^2}{n} - \left(\frac{\sum di}{n}\right)^2}$$

فإنه

$$\sigma = \sqrt{\frac{8750}{4} - \left(\frac{100}{4}\right)^2} = 39.53$$

وهى نفس النتيجة التى تم الحصول عليها فى (أ) لاحظ اثر تبسيط الحسابات فى هذه الطريقة عن السابقة .

<u>ج – حساب الانحراف المعيارى باستخدام طريقة الانحرافات المعدلة :</u>

بملاحظة عمود قيم (di ) فى الطريقة السابقة يلاحظ انه يمكن قسمة قيم الانحراف البسيطة على الوسط الفرضى b=25 دون بواقى لذا يتم تكوين الجدول التالى لايجاد المجاميع اللازمة لحساب الانحراف المعيارى باستخدام طريقة الانحرافات المعدل على النحو التالى :

| Xi | $d_i = X_i - 650$ | $d_i = \frac{di}{25}$ | $d_i^2$ |
|---|---|---|---|
| 625 | -25 | -1 | 1 |
| 700 | 50 | 2 | 4 |
| 650 | 0 | 0 | 0 |
| 725 | 75 | 3 | 9 |
| | | 4 | 14 |

وعليه فإن

$$\sigma_X = b\,\sigma_D = b\sqrt{\frac{\sum Di^2}{n} - \left(\frac{\sum Di}{n}\right)^2}$$

$$= \sqrt{\frac{14}{4} - \left(\frac{4}{4}\right)^2} = 39.53$$

وهى نفس النتيجة التى حصلنا عليها فى (أ) ، (ب) مع تبسيط اكثر فى العمليات الحسابية وجهد التجميع .

- **حساب الانحراف المعيارى فى حالة البيانات المبوبة :**

على غرار ما تم فى حساب الانحراف المعيارى فى حالة البيانات الغير مبوبة فانه يمكن حساب الانحراف المعيارى فى حالة البيانات المبوبة تكراريا بأآ من الطرق الثلاث السابقة ( الطريقة المباشرة ( او القيم الصلية ) – طريقة الانحرافات البسيطة – طريقة الانحرافات المعدلة ) وذلك مع اعتبار ان مراكز فئات التوزيع هى قيم (Xi ) ويراعى ترجيحها بالتكرارات المناظرة لتلك المراكز او تلك الفئات اى (Fi ) .

**١ – الطريقة المباشرة لحساب الانحراف المعيارى فى حالة البيانات المبوبة :**

اذا كانت لدينا مراكز فئات توزيع تكرارى هى $X_1,X_2,.....,X_n$ وما بناظرها من تكرارات $F_1,F_2....,F_n$ فمن المعروف ان كل مركز فئة يعتبر مكررا عدد من المرات مساويا لتكرار تلك الفئة المناظرة وفى هذه الحالة فانه لحساب الانحراف المعيارى باستخدام الطريقة المباشرة يتم ضرب مركز الفئات ( $X_i$ ) فى التكرارات المناظرة لها ( $F_i$) ثم ايجاد مجموع النواتج لحاصل الضرب وقسمة هذا المجموع على مجموع التكرارات لتعطى بذلك الوسط الحسابى للتوزيع التكرارى اى ان

$$\overline{X} = \frac{\sum X_i F_i}{\sum F_i}$$

**وبعد حساب قيمة الوسط الحسابى للتوزيع يتم ايجاد انحرافات قيم مراكز فئات التوزيع ( $X_i$) عن الوسط الحسابى للتوزيع ( $\bar{X}$ ) اى نوجد ( $(X_i - \bar{X})$ ) لكل فئة من فئات التوزيع ثم يتم ايجاد مربعات انحرافات المراكز عن الوسط الحسابى. ثم ضرب نواتج مربعات انحرافات المراكز عن وسطها الحسابى فى التكرارات المقابلة لها ثم ايجاد المجموع والذى يأخذ الصورة التالية:**

$$\sum(X_i - \bar{X})^2 \, F_i$$

**فيكون الانحراف المعيارى للتوزيع التكرارى عبارة عن :**

$$\sigma = \sqrt{\frac{\sum(X_i - \bar{X})^2 F_i}{\sum F_i}}$$

**ويمكن تبسيط هذه الصورة الاخيرة كما سبق لتأخذ الصورة التالية :**

$$\sigma = \sqrt{\frac{\sum X_i^{\,2} F_i}{\sum F_i} - \frac{(\sum X_i F_i)^2}{\sum F_i}}$$

**ويفضل استخدام هذه الصيغة عن السابقة لها لحساب الانحراف المعيارى خصوصا اذا كانت قيمة الوسط الحسابى تحتوى على قيما كسرية كما سبق وان بينا ذلك .**

**٢ – <u>طريقة الانحراف البسيطة</u> :**

**وهنا يتم اخذ احد مراكز الفئات بمثابة وسط فرضى وليكن (a ) ويفضل اختيار قيمة الوسط الغرض (a ) على ان يكون مركز الفئة التى تتوسط فئات التوزيع التكرارى او مركز الفئة الذى يقابل اكبر تكرار وذلك لتبسيط العمليات الحسابية**

بشكل ملحوظ فاذا افترضنا ان انحرافات مراكز فئات التوزيع التكرارى عن الوسط الفرض بمثابة القيم $(d_i)$ حيث :

$d_i = x_i - a \qquad i=1,2,\ldots\ldots n$

وتعبر $(n)$ عن عدد فئات التوزيع التكرارى ثم يتم ايجاد حاصل ضرب الانحرافات ( di ) فى التكرارات المقابلة ومن ثم المجموع اى $\sum di\ fi$ ومرة اخرى يتم ضرب $di\ fi$ فى عمود الانحرافات (di) ثم الجمع لنحصل على المجموع $\sum d_i^2 F_i$ فيكون الانحراف المعيارى عبارة عن

$$\sigma_X = \sigma_d = \sqrt{\frac{\sum di^2 F_i}{\sum F_i} - \left(\frac{\sum_{di} F_i}{\sum F_i}\right)^2}$$

### ٣ – طريقة الانحرافات المعدله ( الاكثر اختصارا ) :

اذا كانت قيم الانحراف البسيطة $(d_i)$ المحسوبة فى الطريقة السابقة تقبل القسمة على مقدار اوسط فرضى آخر وليكن $(b)$ بدون بواقى ( وغالبا ما تكون قيمة b بمثابة طول الفئة فى حالة الجدول التكرارية المنتظمة ) فانه يفضل قسمة الانحرافات البسيطة $(d_i)$ على الوسط الفرضى $(b)$ لينتج لنا الانحرافات المختصرة ولتكن di حيث :

$$D_i = \frac{X_i - a}{b} = \frac{d_i}{b} \qquad where: \quad i = 1, 2, \ldots\ldots, n$$

فانه يتم ضرب قيم الانحرافات المختصرة $(D^i)$ فى التكرار المقابل ثم ايجاد المجموع اى $(\sum D_i F_i)$ ومرة اخرى يتم ضرب عمود $(D_i F_i)$ فى عمود الانحرافات المختصرة $(d_i)$ ثم الجمع لينتج $\sum D_i^2\ F_i$ ثم يتم حساب الانحراف المعيارى من الصيغة التالية :

$$\sigma = b\,\sigma_d = b\sqrt{\frac{\sum Di^2\,F_i}{\sum F_i} - \left(\frac{\sum_{Di}\,F_i}{\sum F_i}\right)^2}$$

وكما راينا عند دراستنا للوسط الحسابى فانه يفضل استخدام طريقة الانحرافات المعدلة اذا كان التوزيع التكرارى منتظما .

***مثال :***

فيما يلى التوزيع التكرارى التالى يبين عدد الوحدات المعيبة التى وجدت فى عينة مكونة من مائة صندوق من الوحدات المصنعة :

**جدول ( ٤-١١)**

| No. of defective units | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Frequency | 13 | 15 | 18 | 25 | 20 | 19 |

المطلوب اوجد الانحراف المعيارى لعدد الوحدات المعيبة .

***الحل :***

لاحظ ان الجدول المعطى عبارة عن توزيعا تكراريا لظاهرة كمية منفصلة ( متقطعة ) لذا يتم اعتبار عدد الوحدات المعيبة بمثابة مراكز الفئات لهذا التوزيع ( Xi ) ومن ثم فالجدول التالى يبين الحسابات اللازمة لايجاد الانحراف المعيارى لعدد الوحدات المعيبة ( باستخدام الطريقة المباشرة ) :

جدول ( ٤ – ١٢ )

| $X_i$ | $F_i$ | $X_i F_i$ | $X_i^2 F_i$ |
|---|---|---|---|
| 0 | 13 | 0 | 0 |
| 1 | 15 | 15 | 15 |
| 2 | 18 | 36 | 72 |
| 3 | 25 | 75 | 225 |
| 4 | 20 | 80 | 320 |
| 5 | 19 | 95 | 475 |
| $\sum$ | 100 | 301 | 1107 |

فيكون الانحراف المعيارى لعدد الوحدات المعيبة فى الصندوق هو :

$$\sigma = \sqrt{\frac{\sum X_i^2 F_i}{\sum F_i} - \left(\frac{\sum Xi\, Fi}{\sum F_i}\right)^2}$$

اى ان:

$$\sigma \sqrt{\frac{1107}{100} - \left(\frac{301}{100}\right)^2} = 1.42 \text{ unite}$$

***مثال :***

احسب الوسط الحسابى والانحراف المعيارى للتوزيع التكرارى التالى :

| Classes | 12- | 16- | 30- | 40- | 50-70 | $\sum$ |
|---|---|---|---|---|---|---|
| Frequency | 8 | 13 | 15 | 10 | 4 | 50 |

***الحل :*** لحساب الوسط الحسابى والانحراف المعيارى لهذا التوزيع يتم تكوين الجدول التالى :

**جدول ( ٤-١٥ )**

| Classes | $F_i$ | Mid point(xi) | $d_i=x_i-35$ | $d_i\ f_i$ | $d_i^2\ F_i$ |
|---|---|---|---|---|---|
| 12- | 8 | 14 | -21 | -168 | 3528 |
| 16- | 13 | 23 | -12 | -156 | 1872 |
| 30- | 15 | 35 | 0 | 0 | 0 |
| 40- | 10 | 45 | 10 | 100 | 1000 |
| 50-70 | 4 | 60 | 25 | 100 | 2500 |
| $\sum$ | 50 | | | -124 | 8900 |

**وعليه فان الوسط الحسابى باستخدام طريقة الانحرافات البسيطة هو :**

$$\bar{X} = a + \frac{\sum d_i\, Fi}{\sum Fi}$$

$$= \ 35 + \left(\frac{-124}{50}\right) = 35 - 2.48 = 32.52$$

**اما الانحراف المعيارى باستخدام طريقة الانحرافات البسيطة فهو :**

$$\sigma_x = \sigma_d = \sqrt{\frac{\sum di^2\, F_i}{\sum F_i} - \left(\frac{\sum di\, F_i}{\sum F_i}\right)^2}$$

$$= \sqrt{\frac{8900}{50} - \left(\frac{-124}{50}\right)^2}$$

## مثال :

**فيما يلى التوزيع التكرارى التالى لاطوال عينة من طلاب كلية التجارة بالسنتيمتر .**

جدول ( ٤ -١٦ )

| Classes | 140- | 150- | 160- | 170- | 180-190 | $\sum$ |
|---|---|---|---|---|---|---|
| Frequency | 6 | 10 | 22 | 7 | 5 | 50 |

والمطلوب احسب الوسط الحسابى والانحراف المعيارى لطول الطالب بالعينة وكذا تباين هذا التوزيع :

**الحل :**

حيث ان التوزيع التكرارى المعطى توزيعا تكراريا منتظما ( اى متساوى من حيث اطوال فئاته ) لذا يفضل استخدام طريقة الانحرافات المعدلة والجدول ( ٤-١٧ ) يبين الحسابات اللازمة لايجاد كلا من الوسط الحسابى والانحراف المعيارى لطول الطالب فى العينة باستدام طريقة الانحرافات المعدلة : .

جدول ( ٤ – ١٧ )

| Classes | $F_i$ | $X_i$ | di=Xi-165 | $D_i = \frac{d_i}{10}$ | Di fi | $D_i^2 F_i$ |
|---|---|---|---|---|---|---|
| 140- | 6 | 145 | -20 | -2 | -12 | 24 |
| 150- | 10 | 155 | -10 | -1 | -10 | 10 |
| 160- | 22 | 165 | 0 | 0 | 0 | 0 |
| 170- | 7 | 175 | 10 | 1 | 7 | 7 |
| 180-190 | 5 | 185 | 20 | 2 | 10 | 20 |
| $\sum$ | 50 | | | | -5 | 61 |

وحيث ان صيغة القانون المستخدم لحساب الوسط الحسابى باستخدام طريقة الانحرافات المعدلة هى :

$$\bar{X} = a + b\,\frac{\sum Di\,Fi}{\sum Fi}$$

لذا فان :

$$\bar{X} = 165 + 10\left(\frac{61}{50}\right) = 177.2 \qquad (C.M)$$

اما عن صيغة الانحراف المعيارى باستخدام طريقة الانحرافات المعدلة فهى :

$$\sigma_X = b\,\sigma_D = b\sqrt{\frac{\sum di^2 F_i}{\sum F_i} - \left(\frac{\sum_{di} F_i}{\sum F_i}\right)^2}$$

$$= 10\sqrt{\frac{61}{50} - \left(\frac{-5}{50}\right)^2} = 11\ (C.M)$$

اما عن التباين ($\sigma^2$) فهو عبارة عن مربع الانحراف المعيارى

اى ان $\sigma^2 = (11)^2 = 121 \qquad (C.M)$

- **الانحراف المعيارى المصحح لشبرد ( تصحيح الانحراف المعيارى للبيانات المبوبة )**

سبق وان راينا انه قد تؤدى عملية تبويب البيانات فى الجداول التكرارية الى فقدان بعض المعلومات عن الظاهرة محل الدراسة حيث لا يكون معروفا كل مفرده من المفردات بالتحديد وذلك نتيجة لتجميع قيم المفردات داخل فئات يكون معلوما الحد الادنى والحد الاعلى فقط لكل منها ويفترض ان مفردات كل فئة تتركز فى مركزها او بمعنى اخر ان الجداول التكرارية منشأة تحت افتراض ان قيم المفردات تتوزع بانتظام ( اى متجانسة الانتشار) على مدار كل فئة من فئات الجدول وهو ما يدعونا الى استخدام مركز كل فئة كقيمة تأخذها كل مفردة من المفردات داخل هذه الفئة وهذا الافتراض غير صحيح الا انه يعتبر افتراضا تقريبيا الغرض منه تسهيل العمليات الحسابية وهو ما يؤدى بطبيعة الحال الى اخطاء فى قيم المقاييس الاحصائية وبالتحديد تلك المقاييس الاحصائية

( كالوسط والانحراف المعيارى ) المحسوبة من البيانات المبوبة سوف لا تنطبق أو لا تتساوى تماما مع قيمة المقاييس المحسوبة من القيم الاصلية قبل تبويبها . وهو ما يستلزم تصحيح المقاييس المحسوبة من التوزيعات التكرارية ولكن هذا التصحيح يتوقف على عوامل متعددة منها شكل التوزيع وطبيعة لمقياس المستخدم فى الحساب .

وبصفة عامة يمكن القول انه فى حالة الوسط الحسابى فان اخطاء التبويب قليلة القيمة وذلك لانه وان كانت المفردات لا تقع جميعها عند مراكز الفئات الا انها تقع على جانبى تلك المراكز بحيث تكون بعض القيم اقل من مركز الفئة بينما يزيد البعض الاخر فى قيمته على قيمة هذا المركز وذلك بالصورة التى يمكن معها افتراض ان الفروق الموجبة تميل الى ان تتساوى فى مجموعها مع الفروق السالبة وعليه فان القيمة المحسوبة للوسط الحسابى من البيانات الاصلية مباشرة قد لا تختلف كثيرا عن قيمة الوسط الحسابى المحسوب بعد تبويب هذه البيانات فى جدول تكرارى .

اما بالنسبة للانحراف المعيارى فان الامر يختلف بعض الشئ عنه فى حالة الوسط الحسابى حيث ان تجميع مربعات الفروق بين قيم المفردات ووسطها الحسابى يميل الى التراكم وليس الى التلاشى ( حيث ان مربعات الفروق السالبة تكون موجبة ) وبالتالى فان قيمة الانحراف المعيارى المحسوبة من بيانات مبوبة فى جدول تكرارى تميل الى ان تختلف عن قيمة الانحراف المعيارى فى حالة حسابه من البيانات الاصلية مباشرة هذا ويزداد هذا الاختلاف كلما زاد طول الفئات للتوزيع التكرارى .

وتحاشيا لمسئلة الاختلاف فيما بين قيمة الانحراف المعيارى للمفردات الاصلية عن الانحراف المعيارى لتلك المفردات بعد تبويبها فقد قام شبرد **Shepperd**

بدراسة ذلك على التوزيعات المتصلة التى تلامس طرفا كل منها المحور الافقى واوجد تصحيحا Correction للانحراف المعيارى وذلك من خلال طرح خارج قسمة مربع طول الفئة فى هذا التوزيع على ( 12 ) اى طرح المقدار $\left(\frac{L^2}{12}\right)$ من صيغة قانون الانحراف المعيارى المعروفة اى ان صيغة الانحراف المعيارى المصحح لشبرد تأخذ الصورة التالية :

$$\sigma_{cor.} = \sqrt{\frac{\sum X^2 F}{\sum F} - \left(\frac{\sum X F}{\sum F}\right)^2 - \frac{L^2}{12}}$$

أى أن:

$$\sigma_{cor.} = \sqrt{\sigma^2 - \frac{L^2}{12}}$$

حيث ان $\sigma_{cor}$ نقصد بها الانحراف المعيارى المصحح لشبرد .(L) تعبر عن طول الفئة الغالب للتوزيع .

هذا وتجدر الاشارة الى ان يمكن التجاوز عن هذا التصحيح متى كان طول الفئات فى التوزيع التكرارى قصيرا وخاصة اذا لم يكن هدفنا هو الحصول على تقدير الانحراف المعيارى على مستوى عال من الدقة . ويمكنك عزيزى القارئ التحقق من انه كلما قصر طول الفئة فى التوزيع التكرارى كلما كان الفارق ما بين الانحراف المعيارى المصحح والانحراف المعيارى الغير مصحح قليلا ومن جهة اخرى فان قيمة التصحيح عادة ما تكون طفيفة الاهمية اذا كان محموع التكرارات قليل نسبيا ( اقل من ١٠٠٠ مفرده مثلا ) او اذا كان عدد الفئات قليلا ( اقل من ٢٠ فئة مثلا ) كما انه يجب ملاحظة ان هذا التصحيح لا يسرى على التوزيعات التى لا تلامس منحنياتها التكرارية المحور الافقى عند طرفيها .

وعلى سبيل المثال فانه لا يجوز استخدام هذا التصحيح اذا كان منحنى التوزيع التكرارى على شكل (U ) او اذا كان التوزيع شديد الالتواء ( كالتوزيع المعروف بالاسى السالب او الاسى الموجب ) او فى حالة التوزيعات المستطيلة.

**مثال :**

احسب الانحراف المعيارى المصحح للتوزيع الوارد فى المثال السابق جدول ( ٤-١٦ ).

**الحل :**

من خلال نتائجنا فى المثال السابق اتضح لنا ان التباين لهذا التوزيع قيمته (121 ) اى ان $\sigma^2 = 121$ وحيث ان طول الفئة فى هذا التوزيع هو L=10

لذا فان الانحراف المعيارىالمصحح هو :

$$\sigma_{cor.} = \sqrt{\sigma^2 - \frac{L^2}{12}} = \sqrt{121 - \frac{(10)^2}{12}}$$

$$= \sqrt{112.6667} = \sqrt{10.614} \quad (C.M)$$

- <u>**الانحراف المعيارى فى حالة التوزيعات التكرارية النسبية :-**</u>

يمكن استخدام القوانين السابقة المستخدمة فى ايجاد الانحراف المعيارى للتوزيعات التكرارية النسبية وذلك بعد اجراء بعض التعديلات الطفيفة عليها بما يتلائم مع مفهوم التكرارات النسبية كما هو الحال فى حالة الوسط الحسابى.

فوفقا للطريقة المباشرة فى حساب الانحراف المعيارى فأن :-

$$\sigma = \sqrt{\frac{\sum X_i^2\ F_i}{\sum F_i} - \left(\frac{\sum X_i^2\ F_i}{\sum F_i}\right)^2}$$

(١)

$$= \sqrt{\sum X_i^2\ \frac{F_i}{\sum F_i} - \left(\sum X_i \frac{F_i}{\sum F_i}\right)^2}$$

وحيث ان التكرار النسبى ما هو الا تكرار الفئة منسوبا الى مجموع التكرارات اى ان :

$$R_i = \frac{F_i}{\sum F_i} \qquad where \quad i = 1, 2, 3, \ldots., n$$

حيث (n) تمثل عدد فئات التوزيع .

لذا فأنه بالتعويض عن التكرارات النسبية فى القانون (١) ينتج ان :

$$\sigma = \sqrt{\sum X_i^2 R_i - \left(\sum X_i R_i\right)^2}$$

وبالمثل فأنه وفقا لطريقة الانحرافات البسيطة تصبح صيغة قانون الانحراف المعيارى فى حالة الجداول التكرارية النسبية هى :

$$\sigma = \sqrt{\sum d_i^2 R_i - \left(\sum d_i R_i\right)^2}$$

اما فى حالة طريقة الانحرافات المعدلة فأن الانحراف المعيارى للتوزيع التكرارى النسبى هو :

$$\sigma = b\sqrt{\sum D_i^2 R_i - \left(\sum D_i R_i\right)^2}$$

***مثال :*** **التوزيع التالى يمثل التوزيع التكرارى النسبى لدرجات الحرارة المتوقعة لاحد الايام حسب المناطق المختلفة بالمملكة العربية السعودية:**

**جدول ( ٤ – ١٨ )**

| Classes | 17.5 - | 22.5 - | 27.5 - | 32.5 - | 37.5 - 42.5 |
|---|---|---|---|---|---|
| Relative freq. | 0.12 | 0.20 | 0.28 | 0.25 | 0.15 |

**والمطلوب :**

**١- حساب متوسط درجة الحرارة المتوقعة بالمملكة فى هذا اليوم .**

**٢- حساب الانحراف المعيارى لدرجة الحرارة المتوقعة فى هذا اليوم .**

***الحل :***

**لاحظ ان التوزيع المعطى فى جدول ( ٤-١٨ ) توزيعا منتظما لذا يفضل استخدام طريقة الانحرافات المعدلة فى حساب كلا من الوسط الحسابى والانحراف المعيارى والجدول ( ٤-١٩) يبين الحسابات اللازمة لذلك :**

**جدول ( ٤-١٩)**

| Classes | $R_i$ | $X_i$ | $D_i = \frac{X_i - 30}{5}$ | $D_i R_i$ | $D_i R_i^2$ |
|---|---|---|---|---|---|
| 17.5- | 0.12 | 20 | -2 | -0.24 | 0.48 |
| 22.5- | 0.2 | 25 | -1 | -0.2 | 0.2 |
| 27.5- | 0.28 | 30 | 0 | 0 | 0 |
| 32.5- | 0.25 | 35 | 1 | 0.25 | 0.25 |
| 37.5-42.5 | 0.12 | 40 | 2 | 0.3 | 0.6 |
| Σ | 1 | | | 0.11 | 1.53 |

وعليه فأن :

الوسط الحسابى لدرجة الحرارة المتوقعة فى هذا اليوم هو :

$$\bar{x} = a + b \sum_{i=1}^{n} D_i R_i = 30 + 5\,(0.11)$$

$$= 30.55 \qquad (degree)$$

أما الانحراف المعيارى فهو :

$$\sigma = b\sqrt{\sum D_i^2 R_i - \left(\sum D_i R_i\right)^2}$$

$$= 5\sqrt{1.53 - (0.11)^2} = 6.16 \qquad (degree)$$

**ملاحظات حول مقاييس التشتت المطلقة :-**

١- يعتبر الانحراف المعيارى من اهم مقاييس التشتت المطلقة واكثرها استخداما فى الاحصاء وذلك نظرا لانه ياخذ فى الاعتبار جميع المفردات فى الحساب فضلا على انه يخضع للمعالجة الجبرية ويستخدم فى معالجات اخرى لذلك يفضل استخدام الانحراف المعيارى حينما لا يكون قياس تشتت الظاهرة هو نهاية المطاف او نهاية التحليل الاحصائى بل انه بداية لعمليات احصائية اخرى اكثر اهمية كعملية الاستدلال الاحصائى او مراقبة جودة الانتاج Quality Control

٢- يعتبر الانحراف المعيارى للعينة احسن تقدير للانحراف المعيارى الحقيقى للمجتمع الذى سحبت منه العينة وذلك لان اخطاء المعاينة بالنسبة له اصغر من

بقية مقاييس التشتت اذا ان قيمتة لا تختلف كثيرا من عينة لاخرى مسحوبة من نفس المجتمع محل الدراسة .

٣- رغم ان الانحراف المعيارى ياخذ فى الاعتبارجميع قيم المفردات عند قياس التشتت الا انه اكثر تأثيرا بالقيم الشاذة او المتطرفة كما هو الحال بالنسبة للوسط الحسابى كاحد مقاييس المتوسطات الذى يتأثر بدرجة كبيرة بالقيم الشاذة او المتطرفة واما نصف المدى الربيعى فهو اقل هذه المقاييس تأثيرا بالقيم الشاذة او المتطرفة ولذلك فيفضل حساب نصف المدى الربيعى كاحد مقاييس التشتت المطلق فى حالة التوزيعات شديدة الالتواء .

هذا وبالرغم من ان الانحراف المتوسط اقل تأثيرا بالقيم الشاذة او المتطرفة الاانه من الممكن تقليل تأثير القيم الشاذة على الانحراف المتوسط وذلك من خلال استخدام الانحراف المتوسط عن الوسيط بدلا من الوسط الحسابى فى ايجادة .

٤- يعاب ايضا على الانحراف المعيارى انه لا يمكن استخدامه فى التوزيعات التكرارية المفتوحة كما لا يمكن حساب بيانيا وفى هذه الحالة يفضل استخدام نصف المدى الربيعى لدراسة تشتت التوزيع .

٥- اختلاف قيم مقاييس التشتت المطلق للتوزيع التكرارى الواحد امرا طبيعيا ومتوقعا وذلك نظرا لاختلاف تلك المقاييس من حيث المفهوم او التعريف فكل مقياس له قانونه الخاص به . وهذا فى حد ذاته يعطى مؤشرا الى انه فى حالة مقارنة تشتت ظاهرة واحدة فى عدة مجتمعات او مقارنة عدة ظواهر لها نفس وحدات القياس فأنه يلزم استخدام مقياس واحد فقط من هذه المقاييس وذلك بما يتناسب مع طبيعة البيانات وذلك حتى لا تكون النتائج المتحصل عليها مضللة .

٦- فى حالة التوزيعات المتماثلة او القريبة من التماثل ( بسيطة الالتواء ) هناك علاقات تجريبية تربط ما بين مقاييس التشتت المطلق منها على سبيل المثال :

- الانحراف المتوسط = $\frac{4}{5}$ الانحراف المعيارى
- نصف المدى الربيعى = $\frac{2}{3}$ الانحراف المعيارى.

ومن هاتين العلاقتين يمكن التوصل لعلاقة تقريبية تربط ما بين الانحراف المتوسط ونصف المدى الربيعى وهى :

الانحراف المتوسط = ١.٢ نصف المدى الربيعى

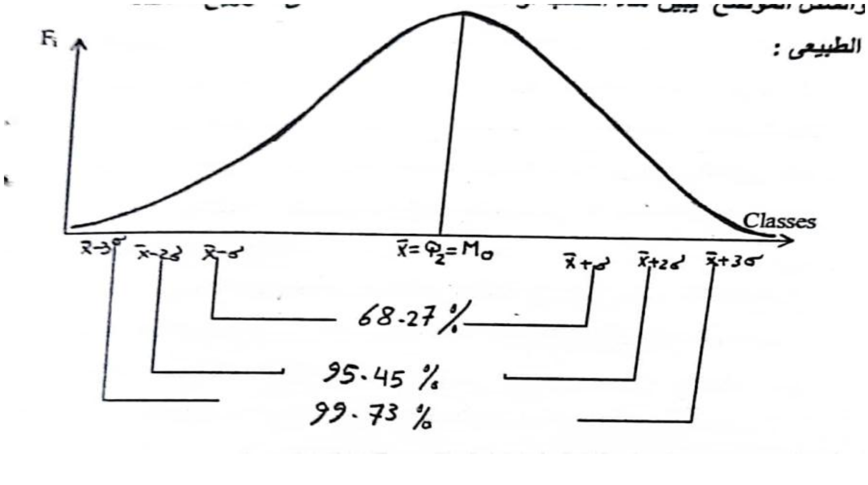
اخيرا فمن اهم خصائص التوزيع المعتاد الطبيعى **Normal pistribution** ما يلى :

* ان ٦٨.٢٧% من جميع المفردات تقع فى المدى ما بين $(\bar{X} \mp \sigma)$

* ان ٩٥.٤٥% من جميع المفردات تقع فى المدى ما بين $(\bar{X} \mp 2\sigma)$

* وان ٩٩.٧٣ من جميع المفردات تقع فى المدى ما بين $(\bar{X} \mp 3\sigma)$

والشكل الموضح يبين هذه النسب او المساحات تحت منحنى التوزيع المعتاد الطبيعى :



وجدير بالذكر فى هذا الخصوص ان كثيرا من الظواهر الطبيعية او المتغيرات تفترب او توزيعها المعتاد الطبيعى وذلك اذا كان حجم او عدد المشاهدات كبيرة نسبيا ( نظرية الحد المركزية **Centeral Limit Theorem** ) .

**ثانيا : مقاييس التشتت النسبية ( معاملات الختلاف ) :**

**Cofficients of Variation.**

راينا فيما سبق ان مقاييس التشتت المطلقة والتى عرضناها سابقا تاخذ نفس تمييز الوحدات الاصلية للظاهرة محل الدراسة ومن ثم فأن مقاييس التشتت المطلقة باختلاف مسمياتها تستخدم فى المقارنة بين مجموعتين ( او اكثر ) متحدتين فى وحدة القياس اى ان هذه المقاييس لا تصلح للمقارنة بين تشتت مجموعتين ( او اكثر ) مختلفين فى وحدات القياس كالاوزان ( بالكيلوجرام مثلا ) والاعمار ( بالسنوات مثلا ) فليس من المعقول اجراء مقارنة بين الكيلوجرامات والسنوات ونقول ان الكيلوجرامات اقل او اكثر تشتتا من السنوات اضف الى ذلك انه قد تتحد ظاهرتين او اكثر من وحدات القياس لكن طبيعة الظاهرتين قد تختلف اختلافا جوهريا فيما بين المجموعتين المطلوب المقارنة بينهما .فمثلا بافتراض ان هناك مجموعة من القيم تعبر عن الدخل الشهرى لمجموعة من المستثمرين او رجال الاعمال ( دخول مرتفعة جدا ) وهناك مجموعة اخرى من القيم تعبر عن الدخل الشهرى لمجموعة من العمال (دخول بسيطة ) فأن هذا يؤدى بدوره الى انه بالطبع يختلف متوسط الدخل الشهرى للمجموعتين اختلافا كبيرا وبالتالى فأن استخدام مقاييس التشتت المطلقة لمقارنة تشتت هاتين المجموعتين فى هذه الحالة سوف يؤدى بطبيعة الحال الى نتائج مضللة واستنتاجات خاطئة بالمرة .

ومن هنا فمقارنة تشتت المجموعتين باستخدام مقاييس التشتت المطلقة تقوم على اساس غير سليم وذلك نظرا لانه من المتوقع ان تكون المجموعة ذات الوسط الحسابى الكبير ذات انحراف معيارى كبير نسبيا والعكس صحيح دون ان تعكس هذه المقاييس درجة التشتت الفعلية فيما بين قيم مفردات كل مجموعة. وفى مثل تلك الحالات فأنه يلزم التخلص من اثر الاختلاف فى وحدات القياس من جانب وكذلك التخلص من اثر الاختلاف فى قيمة الوسطين الحسابيين من جانب اخرذلك من خلال استخدام مقاييس نسبية للتشتت تقيس التشتت بوحدات من الوسط الحسابى وهى ما يسمى بمعاملات الاختلاف . هذه المقاييس النسبية تجرى عملية المقارنة على اساس سليم لا يعتمد على وحدات القياس المستخدمة ولا يتأثر بكون قيم المفردات كبيرة ام صغيرة . فهذه المقاييس النسبية تحقق فائدتين هامتين وهما :

- تتخلص المقاييس النسبية من اثر اختلاف قيمة المتوسطات الذى ربما قد يوجد ما بين المجموعات.
- كما انها تاتى على شكل نسبة مئوية خالية من التمييز وبالتالى يمكن استخدامها لمقارنة تشتت مجموعتين متحدتين او مختلفين فى وحدات قياسها .

وبصفة عامة فأن معامل الاختلاف ياخذ الصورة التالية :

$$\text{معامل الاختلاف} = \frac{\text{مقياس للتشتت المطلق}}{\text{مقياس من مقاييس المتوسطات}} \times ١٠٠$$

على ان يكون مقياس المتوسطات المستخدمة فى المقام له نفس طبيعة (أو طريقة حساب ) مقياس التشتت المستخدمة فى البسط : لذا فأن هناك مقياسين اساسيين لدراسة التشتت النسبى وهما :

١- <u>معامل الاختلاف ( بدلاله الوسط الحسابى والانحراف المعيارى ) :</u>

وهذا المعامل يعتبر اكثر معاملات الاختلاف انتشارا وهو ما يسمى بمعامل الاختلاف النسبى وهو عبارة عن النسبة ما بين الانحراف المعيارى والوسط الحسابى للظاهرة محل الدراسة اى ان :

$$\text{معامل الاختلاف} = \frac{\text{الاختلاف المعيارى}}{\text{الوسط الحسابى}} \times ١٠٠$$

أى أن :

**Cofficient of variation**

$$C_V = \frac{\sigma}{\bar{\bar{X}}} \times 100$$

هذا ويفضل استخدام هذا المعامل فى حالة التوزيعات التكرارية المتماثلة او القريبة من التماثل لكن ما يعيب هذا المقياس هو انه لا يمكن حسابه من الجداول التكرارية المفتوحة اذا انه يعتمد فى حسابه على ايجاد كل من الوسط الحسابى الانحرافى المعيارى وكلاهما لا يمكن حسابه من الجداول التكرارية المفتوحة كما ان هذا المقياس يعطى نتائج مضلله اذا كان التوزيع شديد الالتواء وذلك نظرا لتحيز كل من الوسط الحسابى والانحراف المعيارى للقيم الشاذه او المتطرفة فى مثل هذه الحالة .

٢- <u>معامل الاختلاف الربيعى **Quartile coefficient of variation**</u>

اذا لم يكن فى الامكان حساب الوسط الحسابى والانحراف المعيارى للتوزيع التكرارى كما هو الحال فى حالة التوزيعات التكرارية المفتوحة اوفى حالة التوزيعات التكرارية شديدة الالتواء نظرا لوجود قيم شاذة او متطرفة فأنه

يفضل حساب معامل الاختلاف الربيعى وهو عبارة عن النسبة ما بين نصف المدى الربيعى والوسيط اى ان :-

معامل الربيعى الاختلاف = $\frac{\text{نصف المدى الربيعى}}{\text{الوسيط}}$ × ١٠٠

$$Q.C_V = \frac{Q_1 - Q_1}{2\ Q_2} \times 100$$

هذا وفى حالة ما اذا كان التوزيع محل الدراسة قريبا من التماثل فأنه قيمة الوسيط تساوى تقريبا الوسط الحسابى للربيعين الادنى والاعلى .

ومن ثم يكون :

$$Q_2 = \frac{Q_1 + Q_3}{2}$$

وبالتعويض عن قيمة الوسيط فى صيغة معامل الاختلاف الربيعى السابقة يكون:

$$Q.C_V = \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$$

ويفضل استخدام اى من هاتين الصيغتين فى حالة الجداول التكرارية المفتوحة وهناك صيغ اخرى لمعاملات الاختلاف تقيس التشتت النسبى للظاهرة موضع الدراسة ولكنها غير شائعة الاستخدام وهى :-

* معامل الاختلاف باستخدام الانحراف المتوسط عن الوسط وياخذ الصورة التالية:-

فى حالة المفردات

$$C_V.M_D = \frac{\frac{\sum|X - \bar{X}|}{n}}{\bar{X}} \times 100$$

**or**

**فى حالة البيانات المبوبة :**

$$= \frac{\frac{\sum|X_i - \bar{X}|}{\sum F_i}}{\bar{X}} \times 100$$

- *معامل الاختلاف بدلالة الانحراف المتوسط عن الوسيط وياخذ الصورة التالية :*

$$C_V.M_Q = \frac{\frac{\sum|X_i - Q_2|}{n}}{Q_2} \times 100$$

**ويتم تعديل الصيغة فى حالة الجداول التكرارية كما هو فى الصيغة الخاصة بالمعامل بدلالة الانحراف المتوسط عن الوسط من خلال ترجيح الفروق المطلقة بتكرارات الفئات المقابلة .**

**وفى نهاية هذا الجزء يجب ان ننوه الى ان قيمة معامل الاختلاف سوف تختلف باختلاف المعادلة المستخدمة فى الحساب وهذا امرا طبيعيا نظرا لاختلاف الاساس الرياضى لكل معادلة من المعادلات السابقة ولذلك يجب استخدام صيغة واحدة لاحد تلك المعاملات بما يتناسب مع طبيعة البيانات وذلك عند دراسة مقارنة التشتت النسبى ما بين المجموعات المختلفة .**

**مثال :**

**اجريت دراسة لمجموعة مكونة من ٨٠ عاملا باحد المصانع شملت العمر والاجر الاسبوعى وحصلنا على البيانات التالية :**

**بالنسبة للعمر فأن : الوسط الحسابى = ٣٥ ، التباين = ٤٩**

**اما بالنسبة للاجر الاسبوعى فكان موزعا على النحو التالى :**

| Classes | 65- | 69- | 73- | 77- | 81- 85 | Σ |
|---|---|---|---|---|---|---|
| Frequency | 6 | 14 | 42 | 10 | 8 | 80 |

والمطلوب تحديد اى من الظاهرتين ( العمر ام الاجر الاسبوعى ) الاكثر تشتتا ؟

**<u>الحل</u> :**

حيث ان المطلوب المقارنة ما بين تشتت ظاهرتين مختلفتين فى وحدات القياس لذا يفضل استخدام احد مقاييس التشتت النسبية وحيث ان الجدول التكرارى مغلق الطرفين لذا يفضل استخدام صيغة معامل الاختلاف النسبى لدراسة المقارنة فيما بين تشتت هاتين الظاهرتين :

فبالنسبة لظاهرة العمر فأن معامل الاختلاف النسبى هو :

$$C_V = \frac{\sigma}{\bar{X}} \times 100 = \frac{7}{35} \times 100 = 20\ \%$$

اما بالنسبة لظاهرة الدخل الاسبوعى فيجب حساب الوسط الحسابى والانحراف المعيارى لذا يتم تكوين الجدول التالى :

جدول ( ٤- ٢٠)

| Classes | $F_i$ | $X_i$ | $D_i=X_i-75$ | $D_i = \frac{di}{4}$ | $D_i\ F_i$ | $D_i^2\ F_i$ |
|---|---|---|---|---|---|---|
| 65- | 6 | 67 | -8 | -2 | -12 | 24 |
| 69- | 14 | 71 | -4 | -1 | -14 | 14 |
| 73- | 42 | 75 | 0 | 0 | 0 | 0 |
| 77- | 10 | 79 | 4 | 1 | 10 | 10 |
| 81-85 | 8 | 83 | 8 | 2 | 16 | 32 |
| Σ | 80 | | | | 0 | 80 |

وعليه فأن الوسط الحسابى والانحراف المعيارى للدخل الاسبوعى هما :

$$\overline{X} = a + b\,\frac{\sum DiFi}{\sum Fi}$$

$$= 75 + 4\left(\frac{0}{80}\right) = 75 \quad (L.E)$$

$$\sigma = b\sqrt{\frac{\sum D_i^{\,2}\,F_i}{\sum Fi} - \left(\frac{\sum Di\,F_i}{\sum Fi}\right)^2}$$

$$= 4\sqrt{\frac{80}{80} - \left(\frac{0}{80}\right)^2} \quad = 4(L.E)$$

ومن ثم فأن معامل الاختلاف النسبى للدخل الاسبوعى هو:

$$C_V = \frac{\sigma}{\overline{X}} \times 100 = \frac{4}{75} \times 100 = 5.3\ \%$$

وبمقارنة معاملى الاختلاف للعمر (20%) والدخل الاسبوعى (5.2%) نجد ان توزيع العمر اكثر تشتتا من توزيع الاجر او الدخل الاسبوعى .

***مثال :***

| Classes | Less than 5o | 50- | 60- | 70- | 80 and more |
|---|---|---|---|---|---|
| Frequencies | 5 | 15 | 55 | 20 | 5 |

فيما يلى لديك توزيع درجات مجموعة من الطلاب فى مادة الاحصاء :

والمطلوب حساب معامل الاختلاف المناسب لهذا التوزيع .

***الحل :*** حيث ان الجدول التكرارى الذى يعبر عن توزيع الدرجات جدولا تكراريا مفتوح الطرفين . لذا يتعين حساب معامل الاختلاف الربيعى كذلك فأنه من شكل توزيع التكرارات يتضح من الجدول ان التوزيع قريبا من التماثل . لذا يمكن

حساب الربيعين الادنى والاعلى فقط دون الحاجة لحساب الوسيط اى يتم استخدام صيغة معامل الاختلاف الربيعى المستخدمة للتوزيعات القريبة من التماثل وهى:

$$Q.C_V = \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$$

والجدول التالى يبين التوزيع التكرارى المجتمع الصاعد اللازم لايجاد قيمة كل من $Q_1$ , $Q_3$ :

| Classes | $F_1$ | Less then the classes U.L | Asc.C.F |
|---|---|---|---|
| Less then 50 | 5 | Less then 50 | 5 |
| 50- | 12 | Less then 50 | 20 |
| 60- | 55 | Less then 70 | 75 |
| 70- | 20 | Less then 80 | 95 |
| 80 and more | 5 | Less then the U. l. of the Last class | 100 |
| E | 100 | | |

وحيث ان :

$$Rank\ of\ Q_1 = \frac{\sum F}{4} \times 1 = \frac{100}{4} = 25$$

$$Rank\ of\ Q_3 = \frac{\sum F}{4} \times 3 = \frac{100}{4} \times 3 = 75$$

وعليه فأن :

$$Q_1 = 60 + \left(\frac{25 - 20}{75 - 20}\right) \times 10$$

$$= 60 + \left(\frac{5}{55}\right) \times 10 = 60.9 \qquad (degree)$$

$$Q_3 = 75 \qquad (degree)$$

لاحظ ان ترتيب $Q_3$ كان بمثابة احد التكرارت المستجمعة الصاعدة لذا فأن قيمة $Q_3$ هى القيمة المقابلة لهذا التكرار وهى القيمة 75 درجة وعليه فأن معامل الاختلاف الربيعى هو :

$$Q.C_V = \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$$

$$= \frac{75 - 60.9}{75 + 60.9} \times 100 = 6.59\ \%$$

***مثال :-***

فى دراسة لمستوى الاجور الشهرية للعمال فى قطاع صناعة الغزل والنسيج وقطاع الصناعات الدوائية تم اخذ عينة عشوائية من عمال كل قطاع من القطاعيين وحصلنا على النتائج التالية لكل منهما :

| العينة الاولى ( من قطاع الغزل والنسيج ) | العينة الثالثة (من قطاع الصناعات الدوائية) |
|---|---|
| $\sum_{i=1}^{30} X_{1i} = 12000$ | $\sum_{i=1}^{50} X_{2i} = 22500$ |
| $\sum_{i=1}^{30} X_{1i}^{i} = 4817280$ | $\sum_{i=1}^{50} X_{2i}^{i} = 10141200$ |

والمطلوب :

١- احسب الوسط الحسابى والانحرافى المعيارى لكل عينة .

٢- قارن بين العينات من حيث التشتت موضحا ايها الاكثر تشتتاً .

٣- اذا تم ادماج العينين فى عينة واحدة فاحسب متوسط العينة الجديدة وكذا معامل الاختلاف لها .

**الحل :-**

أ - حساب الوسط الحسابى والانحراف المعيارى لكل عينة على حده فأنه بالنسبة للعينة الاولى :

$$\bar{X}_1 = \frac{\sum_{i=1}^{30} X_{1i}}{30} = \frac{12000}{30} = 400 \ \ (L.E)$$

اما الانحراف المعيارى للعينة الاولى فهو:

$$\sigma_1 = \sqrt{\frac{\sum x_{1i}^2}{n} - \left(\frac{\sum x_{1i}}{n}\right)^2}$$

بالنسبة للعينة الثانية :

$$\bar{X}_1 = \frac{\sum_{i=1}^{50} X_{2i}}{50} = \frac{22500}{50} = 450 \ \ (L.E)$$

$$\sigma_1 = \sqrt{\frac{\sum x_{2i}^2}{n} - \left(\frac{\sum x_{2i}^2}{n}\right)^2} = \sqrt{\frac{10141200}{50} - \left(\frac{22500}{50}\right)^2}$$

$$= 18 \quad (L.E)$$

ب- لمقارنة تشتت المجموعتين لتحديد ايهما الاكثر تشتتا لاحظ ان هناك اختلاف فى معيارين يحولا دون استخدم مقاييس التشتت المطلق فى عملية المقارنة وهما اختلاف حجم العينتين اولا اى اختلاف قيمة (n) هذا بالاضافة الى الاختلاف الجوهرى فيما بين مستوى الاجور فى القطاعين والذى تعكسه

المجاميع المعطاه لذا يفضل ان تتم المقارنـة من خلال احد معاملات الاختلاف لمقاييس تدرس التشتت النسبى وعليه فأن :

$$C_V \, for \, the \, first \, sample = \frac{\sigma_1}{\bar{X}_1} \times 100$$

$$= \frac{24}{400} \times 100 = 6\,\%$$

اما عن معامل الاختلاف النسبى للعينة الثانية فهو :

$$C_V \, for \, the \, second \, sample = \frac{\sigma_2}{\bar{X}_2} \times 100$$

$$= \frac{18}{45} \times 100 = 4\,\%$$

ومن خلال مقارنـة معاملى الاختـلاف النسـبى يتضـح لنـا ان اجـور العينـة الاولـى تعتبر الاكثر تشتتا ( اى الاقل تجانسا ) من اجور العينـة الثانيـة مستخدمين فى عملية المقارنة معامل الاختلاف النسبى كاحد مقاييس التشتت النسبى

جـ - <u>اذا تم دمج العينتين فى عينة واحدة فأن :</u>

حجم العينة الجديدة ( الكلية ) هو :

$$n = n_1 + n_2 = 30 + 50 = 80$$

والوسط الحسابى للعينة الجديدة هو :

$$\bar{X} = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2}{n_1 + n_2}$$

اما عن تباين العينة الجديدة ( الكلية ) فأننا سوف نورد صيغة التباين المشترك للعينتين ( دون اثبات ) فهو :-

$$\sigma^2 = \frac{n_1 \left(\sigma_1^2 + D_1^2\right) + n_2 \left(\sigma_2^2 + D_2^2\right)}{n_1 + n_2}$$

حيث D2, D1 هى عبارة عن انحراف وسط العينة العينة عن الوسط العام او الفرق ما بين متوسط العينة والمتوسط العام ( للعنيتن ) اى ان :

$$\mathbf{D_1 = \bar{X}_1 - \bar{X} = 400 - 431.25 = 31.25} \qquad \mathbf{(L.E)}$$

ومن ثم فأن :

$$\mathbf{D_1^2 = (-31.25)^2 = 976.5625}$$

كذلك فأن انحراف العينة الثانية عن المتوسط العام اى :

$$\mathbf{D_2 = \bar{X}_2 - \bar{X} = 450 - 431.25 = 18.75} \qquad \mathbf{(L.E\ )}$$

وعليه فأن مربع هذا الانحراف اى :

$$\mathbf{D_2^2 = (18.75)^2 = 351.5625}$$

ومن ثم فبالتعويض فى صيغة التباين المشترك فيكون :

$$\boldsymbol{\sigma^2 = \frac{n_1\left(\sigma_1^2 + D_1^2\right) + n_2\left(\sigma_2^2 + D_2^2\right)}{n_1 + n_2}}$$

$$\mathbf{= \frac{30\,(576 + 976.57625) + 50(324 + 351.5625)}{30 + 50}}$$

$$\mathbf{= 1004.4375}$$

وعليه فأن الانحراف المعيارى للعينة الكلية الجديدة هو :

$$\boldsymbol{\sigma = \sqrt{1004.4375} = 31.693} \qquad \boldsymbol{(L.E)}$$

ومن ثم فأن معامل الاختلاف النسبى للعينة الكلية الجديدة هو

$$\boldsymbol{C_V = \frac{\sigma}{\bar{\bar{X}}} \times 100 = \frac{31.693}{431.25} = 7.35\ \%}$$

**الدرجات المعياريه : Standard Scores**

تناولنا فيما سبق كيفية المقارنة ما بين تشتت مجموعتين ( او اكثر ) باستخدام مقاييس التشتت المطلقه والنسبية الا انه فى بعض الحالات قد يكون الباحث فى حاجة لاجراء عملية مقارنه فيما بين مفردتين او قيمتين فى مجموعتين مختلفتين من حيث مقاييس المتوسطات والتشتت وفى هذه الحالة فأننا نكون فى حاجة الى مقياس احصائى ياخذ فى الاعتبار الترتيب النسبى لكل من هاتين المفردتين داخل المجموعة التى تنتمى اليها. فعلى سبيل المثال لمقارنة مستوى اداء طالب واحد فى مادتين او اكثر فى مجموعة تلك المواد فلا تصح مقارنه درجة الاختبار فى مادة باخرى الا من خلال قياس مستوى اداء هذا الطالب فى تلك المواد المختلفة مقارنة باداء مجموعة زملائه فى نفس المواد فعلى سبيل المثال اذا كانت درجة طالب معين فى مادة الاحصاء هى ٧٥ درجة بينما كانت درجته فى مادة الاقتصاد هى ٨٠ درجة . فهل هذا يعنان مستوى هذا الطالب فى مادة الاقتصاد اعلى من مستواه فى مادة الاحصاء فلو. اخذنا الامور بظاهرها لقلنا ان مستوى هذا الطالب فى مادة الاقتصاد يفوق مستواه فى مادة الاحصاء حيث ان درجته فى مادة الاقتصاد ٨٠ درجة اكبر من درجته فى مادة الاحصاء ٧٥درجة. لكن فى الحقيقة فأن الامر لا يفسر هكذا حيث اننا لا نستطيع الحكم على تفوق مستوى اداء طالب فى مادة على اخرى الا اذا تعرفنا على توزيع درجات كل من هاتين المادتين على مستوى الطلاب ككل فالمسالة هنا نسبية وذلك لان مستوى الطالب فى مادة معينه يتوقف على توزيع درجات الطلاب فى هذه المادة بمعنى انه قد تكون درجة الطالب فى مادة الاقتصاد مرتفعة نتيجة لسهولة الامتحان فى تلك المادة وليس الى ارتفاع مستوى اداء الطالب فيها مما يعنى انه قد تكون درجات بقية زملائه مرتفعة ايضا بالشكل

الذى قد يصبح معة ان هذا الطالب يعتبر اقل الطلاب فى ادائه فى هذه المادة. ومن وجهة اخرى فقد تكون درجة الطالب فى مادة الاحصاء رغم انها الاقل ظاهريا الا انه قد يكون الامتحان الخاص بمادة الاحصاء على درجة عالية من الصعوبة بحيث ان يكون بقية زملائه حصلوا على درجات منخفضة فى تلك المادة وانه نسبيا يعتبر متفوقا فى الاحصاء اذا ما قورنت درجته بدرجات بقية زملائه. وعلى ذلك فأنه لمقارنة مستوى اداء هذا الطالب فى هاتين المادتين فأن يلزمنا استخدام مقياس احصائى ياخذ فى الاعتبار الاداء النسبى للطالب فى كل مادة من المادتين على حده والمقصود بالاداء النسبى هنا هو مستوى اداء الطالب فى مادة معينة مقارنا باداء بقية زملائه فى نفس المادة والمقياس الاحصائى الذى يستخدم فى هذا الشان هو ما يسمى بالدرجة او القيمة المعيارية .

ولتوضيح مفهوم الدرجة او القيمة المعيارية دعنا نفترض ان الوسط الحسابى والانحراف المعيارى لدرجات الطلاب فى مادة الاحصاء هما 67 , 4 درجات على الترتيب كما ان الوسط الحسابى والانحراف المعياربلدرجات الطلاب فى مادة الاقتصاد هما 71 , 6 درجات على الترتيب فلقياس مستوى الاداء النسبى لهذا الطالب فى هاتين المادتين فأننا نقارن بين الترتيب ( المركز ) النسبى لدرجة الطالب فى مادة معينة بالنسبة لتوزيع الدرجات فى هذه المادة ويتحقق ذلك من خلال تحديد بعد درجة الطالب فى كل مادة عن الوسط الحسابى لهذه المادة وذلك بدلالة وحدات من الانحراف المعيارى لدرجات هذه المادة ايضا ويتم ذلك من خلال طرح قيمة الوسط الحسابى لكل توزيع من قيمة الدرجة الطبيعية الخاصة بتلك المادة ثم قسمة الناتج على الانحراف المعيارى لهذا التوزيع اى ان الدرجة المعيارية هى :

$$Standard\ Score = Z = \frac{X - \bar{X}}{\sigma}$$

حيث تمثل (X) الدرجة الطبيعية للطالب اما x فهى عبارة عن الوسط الحسابى والانحراف المعيارى لتوزيع درجات الطلاب فى تلك المادة . والهدف من القسمة على الانحراف المعيارى هو تجنب الاختلاف فى وحدات القياس حيث ان الاكتفاء بحساب القيمة ( x –x) فى عمليات المقارنة قد تنطوى على نتائج مضلله ومن ثم فبايجاد الدرجة المعيارية لهذا الطالب فى كل مادة من المادتين نجد ان :

درجة الطالب المعيارية فى مادة الاحصاء هى:

$$Z_1 = S.Sc.for\ stat.degree = \frac{X_1 - \bar{X}_1}{\sigma_1} = \frac{75 - 67}{4} = 2$$

اما درجة الطالب المعيارية فى مادة الاقتصاد فهى :

$$Z_2 = S.Sc.for\ stat.degree = \frac{X_2 - \bar{X}_2}{\sigma_2} = \frac{80 - 71}{6} = 1.5$$

ومن خلال مقارنة الدرجتين المعيارتين نجد ان مستوى الاداء النسبى لهذا الطالب فى مادة الاحصاء يفوق المتوسط بما يعادل ضعف الانحراف المعيارى اما عن اداؤه النسبى فى مادة الاقتصاد فقد كان فوق المتوسط ايضا ولكن بما يعادل مره ونصف فقط من قيمة الانحراف المعيارى وهو ما يدل بدوره على ان مستوى الاداء النسبى لهذا الطالب ( اى مقارنة بالنسبة لاداء بقية زملائه ) فى مادة الاحصاء كان افضل من ادائه النسبى فى مادة الاقتصاد اى ان يمكن استنتاج ان مستوى هذا الطالب فى مادة الاحصاء افضل منه فى مادة الاقتصاد وهو عكس ما يوحى اليه درجتى الطالب الطبيعية فى هاتين المادتين اى عكس الاستنتاج المبنى على ظواهر الامور وفى ختام هذا الجزء يجب ان يؤكد على

**ان القيم المعيارية عبارة عن قيم بدلالـه وحدات مـن الانحـراف المعيـارى لـذلك فهى تكون مجردة من وحدات القيـاس وتسـتخدم بـذلك لمقارنـة قيم مختلفـة مـن مجموعات مختلفة وللقيم المعيارية عدة خصائص هى**

**١ـ الوسـط الحسـابى للقـيم المعياريـة يسـاوى الصـفر دائمـا بغض النظـر عـن الوسط الحسابى للقيم الاصلية ويمكن اثبات ذلك على النحو التالى :**

**فحيث ان الدرجة المعيارية للدرجة الطبيعية ( $x_i$) هي ($z_i$) حيث :**

$$\bar{Z} = \frac{X_i - \bar{X}}{\sigma}$$

**فيكون الوسط الحسابي للدرجات المعيارية اي ($\bar{Z}$) هو :**

$$\bar{Z} = \frac{\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\sigma}\right)}{n} = \frac{1}{n\,\sigma}\sum_{i=1}^{n}(X_i - \bar{X})$$

**وحيث ان مجموع انحرافات القيم عن وسطها الحسابي يساوي صفرا . لذا فأن**

$$\bar{Z} = \frac{1}{n\,\sigma}\ (0) = 0$$

**٢ـ التبـاين او الانحـراف المعيـاري للقـيم المعياريـه يسـاوي الواحـد الصـحيح ويمكن اثبات ذلك علي النحو التالي :**

$$\sigma_Z^2 = \frac{\sum_{i=1}^{n}(Z_i - \bar{Z})^2}{n} = \frac{\sum_{i=1}^{n} Z_i^2}{n}$$

**وذلـك لان $\bar{Z}$ تسـاوي الصـفر ( مـن الخاصـية السـابقة ) . وحيـث ان الدرجـة المعيارية (z ) هي :**

$$\bar{Z} = \frac{X_i - \bar{X}}{\sigma}$$

$$\sigma_Z^2 = \frac{\sum_{i=1}^{n} Z_i^2}{n} = \frac{\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\sigma}\right)^2}{n} = \frac{1}{\sigma^2} \cdot \frac{\sum_{i=1}^{n}(X_i - X)^2}{n}$$

$$= \frac{1}{\sigma^2}\,(\sigma^2) = 1$$

**ومن خلال الخاصيتين السابقتين يتضح لنا انه امعايره مجموعتين من المفردات فان هذا يعني اننا نردها الي مجموعتين لها نفس المتوسط الحسابي ( وهو الصفر ) ونفس الانحراف المعياري ( وهو الواحد الصحيح ) .**

***<u>مثال :</u>***

**في امتحان الشهاده الاعدادية بمحافظة قنا وجد احد الطلاب ان درجاته في الرياضيات واللغة العربية واللغة الانجليزية كانت علي الترتيب هي ٤٥ . ٤٠ . ٤٢ درجة . فاذا علمت ان متوسط درجات الطالب بالمحافظة في المواد الثلاثة السابقة هو علي الترتيب ٤٠.٤٥.٤٢ درجة . وان الانحراف المعياري للدرجات في تلك المواد الثلاثة بالمحافظة هو علي الترتيب ٥.٤.٦ فما هو الحكم الموضوعي علي مستوي هذا الطالب في المواد الثلاثة ؟**

***<u>الحل :</u>***

**لكي يكون الحكم صحيحا علي مستوي اداء الطالب في المواد الثلاث يجب حساب الدرجات المعيارية للمواد الثلاث . حيث ان :**

**الدرجة المعيارية للطالب في مادة الرياضيات ولتكن ( $z_1$) هي :**

$$\bar{Z}_1 = \frac{X_1 - \bar{X}_1}{\sigma_1} = \frac{45 - 42}{6} = 0.5$$

**والدرجة المعيارية للطالب في مادة اللغة العربية ولتكن ($z_2$)هي :**

$$\bar{Z}_2 = \frac{X_2 - \bar{X}_2}{\sigma_2} = \frac{40 - 35}{4} = 1.25$$

**اما عن الدرجة المعيارية للطالب في مادة اللغه النجليزية ولتكن ($z_3$) فهي :**

$$\bar{Z}_3 = \frac{X_3 - \bar{X}_3}{\sigma_3} = \frac{42 - 40}{5} = 0.4$$

**يتضح لنا ان مستوي اداء $z_1$ , $z_2$ , $z_3$ ومن ثم فمن خلال مقارنة القيم الثلاث للطالب في تلك المواد موضوعيا هو ان : مستوي الطالب في اللغة العربية افضل منه في الرياضات افضل منه في اللغة الانجليزية . وذها الحكم هو الحكم الصحيح علي مستوي اداء الطالب في المواد الثلاث عكس ما توحي به الارقام المجرده لدرجات الطالب في المواد الثلاث .**

## تمارين

١- الجدول التالي يبين توزيع عدد الاطفال في عدد معين من الاسرة

| No. of children | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| No. of families | 8 | 16 | 41 | 32 | 20 | 12 | 2 |

والمطلوب حساب الوسط الحسابي والانحراف المعياري لهذا التوزيع .

٢- البيانات التالية تمثل اعمار مجموعتين من الاشخاص ( بالسنوات ) :

اعمار المجموعه الاولي :

**17, 22 , 10, 18 , 27 , 15 , 13 , 30**

اعمار المجموعه الثانية :

**38 , 41 , 33 , 39 , 41 , 73, 36 , 35**

والمطلوب :-

أ – قارن بين التشتت في المجموعتين وذلك باستخدام المدس وشبيهات المدي مع التعليق علي ما تحصل عليه نتائج .

ب – احسب الانحراف المعياري والانحراف المتوسط لكل مجموعه من المجموعتين ثم بين ايهما اكثر تاثرا بالقيم المتطرفه وذلك علي ضوء نتاءجك في (ا) .

٣- الجدول التالي يبين الانفاق علي وسائل المعيشة لعدد معين من الاسر بالجنيه في مدينتين مختلفتين (a) و ( b) :-

| Expenditure classes | 25- | 35- | 40- | 50- | 60- | 70- | 80- |
|---|---|---|---|---|---|---|---|
| No. of families | 3 | 61 | 132 | 153 | 140 | 51 | 2 |
| No. of families | 2 | 14 | 20 | 27 | 28 | 7 | 2 |

**والمطلوب :**

**١-قارن بين متوسط مستوي الانفاق في المدينتين باستخدام المقياس المتوسط الملائم**

**٢-قارن بين تشتت الانفاق في المدينتين باستخدام مقياس التشتت المطلق والنسبى الملائم**

**٤- الجدول التالي يمثل مرتبات عينيه من الموظفين حجمها مائة موظفا :**

| Salary classes | 245- | 255- | 265- | 275- | 285- | 295- |
|---|---|---|---|---|---|---|
| No.of employees | 13 | 18 | 22 | 20 | 16 | 11 |

**والمطلوب حسابات مايلي :**

**أ- نصف المدي الربيعي**

**ب- الانحراف المتوسط عن الوسيط**

**ت- معامل الاختلاف المناسب**

**٥- فيما يلـي التوزيـع التكراري المتجمـع الصـاعد لاعمـار عينـه مـن الاشـخاص حجمها خمسون شخصا :**

| Less than the classes U .L | Frequencies |
|---|---|
| 10 and less 20 | 8 |
| less than 30 | 17 |
| less than 40 | 29 |
| less than 50 | 40 |
| less than 60 | 50 |

**والمطلوب :-**

**أ- ادرس تشتت هذا التوزيع باستخدام نصف المدي الربيعي .**

**ب- احسب العمر الذي يقل عند 90% من اجمالي عدد الاشخاص**

**ت- اوجد مقياس النزعه المركزية الملاءم .**

**٦- الجدول التالي يوضح توزيع اجور ١٦٠ عامل في احد المصانع**

| Wages classes | 30- | 40- | 50- | 60- | -70 | -80 | -90 | $\sum$ |
|---|---|---|---|---|---|---|---|---|
| No. of employees | 10 | 15 | 30 | 50 | 25 | 20 | 10 | 160 |

**والمطلوب حساب**

**١- الوسط الحسابي والوسيط والمنوال**

**٢- الانحراف المتوسط**

**٣- الانحراف الربيعي**

**٤- الانحراف المعياري**

**٥- معامل الاختلاف**

٦- دراسة تماثل التوزيع من خلال نتائجك في ١.

٧- الجدول التكراري التالي يمثل توزيع الاوزان بالكيلو جرام لعينه من ٨٠ شخصا :

| Weighted Classes | 62- | 66- | 70- | 74- | 78- | 82- | 86-90 |
|---|---|---|---|---|---|---|---|
| Frequency | 3 | 8 | 20 | 21 | 14 | 10 | 4 |

والمطلوب حساب :

١- الوسط الحسابي والانحراف المعياري للتوزيع

٢- الوسيط والانحراف الربيعي .

٨- الجدول التالى يبين توزيع الاطوال بالسنتيمترات لعينه من ١٠٠ شخصا :

| Length Classes | 150- | 155- | 160- | 165- | 170-175 |
|---|---|---|---|---|---|
| Frequency | 8 | 16 | 43 | 29 | 4 |

والمطلوب هو حساب الانحراف المتوسط والانحراف المعياري ،

٩- الجدول التالي يمثل توزيع درجات ١٠٠ طالب في ماده الرياضة باحدي الجامعات

| Degree classes | 40- | 45- | 50- | 55- | 60- | 65- | 70- | 75- | 80- | 85- | 90- |
|---|---|---|---|---|---|---|---|---|---|---|---|
| No. of students | 1 | 2 | 5 | 10 | 16 | 28 | 19 | 11 | 5 | 2 | 1 |

والمطلوب :

١- حساب الوسيط والمنوال .

٢- الوسط الحسابي والانحراف المعياري

٣- الانحراف الربيعي والانحراف المتوسط

٤- معامل الاختلاف الملائم .

١٠- اوجد المدي والانحراف المتوسط والانحراف المعياري للقيم التالية :

(١) ٣، ٤، ١٢، ٦، ٧، ٤، ٨، ٣، ٥

(ب) ٦، ٤، ٩، ٤، ٨، ٦، ١٠، ٦، ٦، ٥، ٨، ٨

١١- اوجد الانحراف المعياري للبيانات التالية :

ا- : ٣ ، ٦، ٢ ، ١، ٧ ، ٥

ب- : ٢ ، ٣ ، ٦ ، ٤ ،٨، ٢ ، ٢ ، ٥ ، ٤ ، ٤

ج- : صفر ، صفر ، صفر ، صفر ، ١ ، ١ ، ١

١٢- الجدول التالي يوضح توزيع تكلفة الاقامة في الغرفة الواحده لعدد ٢٠ سائح في احد فنادق .

| Cost Classes | 30- | 40- | 50- | 60- | 70- | 80- | 90- | ∑ |
|---|---|---|---|---|---|---|---|---|
| No. of tourists | 16 | 22 | 36 | 55 | 31 | 25 | 10 | 200 |

<u>والمطلوب حساب</u> :

١- الوسط الحسابي والوسيط والمنوال

٢- الانحراف المتوسط

٣- الانحراف الربيعي

**٤- الانحراف المعياري**

**٥- معامل الاختلاف**

**١٣- الجدول التكراري التالي يبين مبيعات أحد منتجات خان الخليلي حسب السعر فكانت كالتالي :**

| Pro. Price Classes | 75- | 77- | 79- | 81- | 83- | 85- | ∑ |
|---|---|---|---|---|---|---|---|
| Frequency | 3 | 23 | 52 | 15 | 7 | 2 | 102 |

**والمطلوب حساب :**

**١- الوسط الحسابي والوسيط والمنوال**

**٢- الانحراف المتوسط**

**٣- الانحراف المعياري**

**٤- معامل الاختلاف**

**١٤- فيما يلي بيان بدرجات عينية من طلاب معهد السياحه والفنادق بحلوان واخري من طلاب معهد السياحه والفنادق بمدينة السادس من اكتوبر في العام الماضي بمادة الاحصاء.**

| Degree Classes in stat. | 25- | 50- | 60- | 75- | 80-100 | ∑ |
|---|---|---|---|---|---|---|
| No. of students in Helwan | 30 | 80 | 120 | 15 | 5 | 250 |
| No. of student in $6^{th}$ of Oct. | 5 | 25 | 30 | 90 | 50 | 200 |

**المطلوب :- حساب الاتي:**

**١- الوسط الحسابي لدرجة الطلاب عينة طلاب حلوان والانحراف المعياري ومعامل الاختلاف.**

**٢- الوسيط والربيع الادني والاعلي لطلاب عينة ٦ اكتوبر حسابيا وبالرسم**

**٣- المنوال حسابيا وبالرسم لطلاب عينة حلوان.**

**٤- نصف المدي الربيعي ومعامل الاختلاف الربيعي لطلاب عينة السادس من اكتوبر.**

**٥- قارن تشتت التوزيعين .**

**٦- اوجد عدد الطلاب الذين يحصلون علي درجات اقل من ٦٥ درجه لطلاب عينة حلوان ونسبتهم .**

**٧- اوجد نسبة الطلاب اللذين تزيد درجاتهم عن ٧٨ درجة لطلاب مدينة السادس من اكتوبر .**

**٨- اوجد عدد الطلاب اللذين تتراوح درجاتهم بين ٥٨-٦٨ درجة لطلاب عينة حلوان .**

**٩- اوجد الوسط الحسابي والوسيط والمنوال لطلاب مدينة السادس من اكتوبر .**

**١٠- هل يمكن القول ان طلاب معهد السياحه والفنادق بمدينة السادس من اكتوبر اكثر كفاءة من طلاب معهد السياحه والفنادق بحلوان .**

**١١- عرض بيانات التوزيعين في شكل مدرج تكراري .**

## الباب الخامس
## العزوم والإلتواء والتفرطح

## Moments Skewness and kurtosis

**درسنا فى الابواب السابقة بعض خصائص التوزيعات التكرارية من خلال معرفة مقاييس النزعة المركزية ومقاييس التشتت الخاصة بهذه التوزيعات والتى تساعد بدرجة كبيرة فى تلخيص وصف التوزيعات التكرارية الخاصة بالظواهر محل الدراسة إلا ان هذه المقاييس ايضا وحدها لا تكفى للتعرف على كل خصائص هذه التوزيعات التكرارية والمقارنة فيما بينها. فقد يتساوى توزيعان تكراريان من حيث المتوسط والانحراف المعيارى ولكنهما يختلفان من حيث التماثل او الاعتدال لذا فأنه قد يلزم ايجاد مقاييس احصائية اخرى تقيس مدى تماثل التوزيع او التوائه وكذلك تقيس درجة تفرطح ( شكل قمة منحنى ) التوزيع او تدببه .**

**وقبل ان نتعرض لخاصيتى الالتواء والتفرطح للتوزيعات التكرارية سوف نتعرض اولا للعزوم الرياضية بانواعها المختلفة لما لها من اهمية فى دراسة مدى إلتواء التوزيع من عدمه وكذلك مدى تفرطح التوزيع من عدمه .**

**<u>اولا : العزوم</u> Moments :**

**ان كلمة عزوم بدايتها اساساً فى علم الاستاتيكا فى تحليل القوى التى تؤثر فى جسم او كتلة معينة حيث يقاس عزم القوة التى تؤثر فى جسم ما بحاصل ضرب مقدار القوة في مركز العزم ويتم حساب عزوم مجموعة من القوى بانه عبارة مجموعة حواصل ضرب كل قوة فى ذراع عزمها . والعزم فى الاحصاء ناتج عن تشبيه قيمة الظاهرة بزراع العزم اما تكرار تلك القيمة فهو بمثابة القوة وان كان العزم يقاس فى الاحصاء بطريقة مختلفة إلى حد ما عن طريقة قياسه**

فى علم الاستاتيكا . وتعتبر العزوم من اهم معالم التوزيعات الاحتمالية للظواهر المختلفة حيث يمكن معرفة اى توزيع احتمالى من خلال عزوم هذا التوزيع . وهناك انواع مختلفة من العزوم وهى :

- العزوم حول الصفر.
- العزوم حول الوسط الحسابى.
- العزوم حول وسط فرض معين.
- العزوم المعيارية .

وبصفة عامة اذا كان لدينا متغيرا وليكن (X) ياخذ مجموعة القيم $X1, X_2, \ldots.., X_n$ فان العزوم حول وسط فرضي وليكن ($\propto$) هو عبارة عن متوسط انحرافات القيم ( $X_i$) عن هذا الوسط الفرضى ($\propto$) مرفوعه للقوه(r). وهذه القوة ( r ) تفيد فى تحديد قيم العزوم المتتالية اى العزوم الاولى والعزم الثانى و ...... وهكذا لاى رتبة موجبة ومن ثم فان يمكن وضع الصيغة العامة للعزم ذو الرتبة ( r ) حول الوسط الفرضي وليكن ($\propto$)على انه:

$$M_r^{(\propto)} = \frac{\sum_{i=1}^{n}(X_i - \propto)^r}{n} \qquad (1)$$

وهذه الصيغة (١) تستخدم فى حساب العزوم المتتالية فى حالة البيانات المفردة اما فى حالة الجدول او التوزيعات التكرارية فأنه يتم ترجيع الانحرافات المحسوبة فى البسط بقيم التكرارات المقابلة لمراكز فئات التوزيع والتى تاخذ مجموعة القيم المعدة عن مراكز الفئات أي ($X_i$) ومن ثم فان الصيغة العامة للعزم ذو الرتبة ( r ) حول الوسط الفرضى ($\propto$) فى حالة البيانات المبوبة هو عبارة عن :

$$M_r^{(\propto)} = \frac{\sum_{i=1}^{n}(X_i - \propto)^r F_i}{\sum_{i=1}^{n} F_i} \qquad (2)$$

**فاذا تم التعويض عن $r = 1$ فى اى من المعادلتين (1) او (2) نحصل على العزم الاول حول الوسط الفرض (∝) واذا تم التعويض عن $r = 2$ فى هاتين المعادلتين نحصل على العزم الثانى حول الوسط الفرضى (∝)...... وهكذا .**

**وطبقا للانواع المختلفة من العزوم السابق الاشارة اليها فأنه اذا تم التعويض عن الوسط الفرضى (∝) فى المعادلتين (1) ، ( 2 ) بالقيمة صفر فاننا نحصل على العزوم المتتالية حول الصفر . اما عن العزوم حول الوسط الحسابى فأنها تنتج حينما يتم التعويض فى هاتين المعادلتين عن قيمة الوسط الفرضى على انها تساوى الوسط الحسابى للتوزيع اى $(\bar{X})$ واخيرا فالعزوم حول وسط فرضى معين وليكن (a ) فأنها تنتج من خلال التعويض المباشر فى المعادلتين (1) او ( 2 ) عن قيمة a= ∝ اما العزوم المعيارية ففيها يتم ايجاد البسط فى المعادلتين (1) ، ( 2 ) من خلال حساب متوسط الدرجات المعيارية وهو ما سيرد فيما بعد الا انه يمكن القول ان العزوم حول الصفر والعزوم حول الوسط تعتبر الاكثر استخداما فى مجال التحليل الاحصائى وفيما يلى نتناول الانواع المختلفة للعزوم وسوف نكتفى من كل نوع بالاشارة الى كيفية ايجاد العزوم الاربعة الاولى ويمكن بالطبع الحصول على العزوم من قوة اكثر اى بدءا من العزم الخامس والسادس و ..... وهكذا .**

**١- <u>العزوم حول الصفر (العزوم اللامركزية ) :-</u>**

**من خلال المعادلتين (1) ، (2) السابقتين فأنه عند التعويض عن قيمة ∝ = صفرا تصبح صيغة العزم ذو الرتبة (r) حول الصفر فى حالة البيانات الغير مبوبة على الصورة :**

$$\boldsymbol{M_r = M_r^{(0)} = \frac{\sum_{i=1}^{n}(X_i - 0)^r}{n} = \frac{\sum_{i=1}^{n} X_i^r}{n}} \qquad (3)$$

**حيث $r = 1,2,\ldots\ldots,n$**

**اما صيغة العزم ذو الرتبة (r) حول الصفر فى حالة البيانات المبوبة تكراريا فتأخذ الصورة التالية :**

$$M_r = M_r^{(0)} = \frac{\sum_{i=1}^{n}(X_i - 0)^r F_i}{\sum F_i} = \frac{\sum_{i=1}^{n} X_i^r F_i}{\sum_{i=1}^{n} F_i} \qquad (4)$$

**حيث ($X_i$) تعبر عن مركز الفئة (i) حيث $i=1.2,\ldots..,n$ كما ان (n) تعبر عن عدد فئات التوزيع فى هذه الحالة .**

**هذا ومن خلال المعادلات (3) ، (4) يمكن بيان العزوم الاربعة الاولى حول الصفر وذلك فى حالتى البيانات الغير مبوبة والبيانات المبوبة والجدول (٥-١) يبين هذه العزوم الاربعة حول الصفر :**

**جدول (٥-١)**
**العزوم الأربعة الأولى حول الصفر**

| Case<br>r $\underset{\sim}{th}$ *Moment* | Values (القيم المفردة) | Frequency distributions |
|---|---|---|
| $M_1^{(0)} = M_1$ | $\frac{\sum_{i=1}^{n} X}{n} = \bar{X}$ | $\frac{\sum_{i=1}^{n} X_i F_i}{\sum F_i} = \bar{X}$ |
| $M_2^{(0)} = M_2$ | $\frac{\sum_{i=1}^{n} X_i^2}{n}$ | $\frac{\sum_{i=1}^{n} X_i^2 \; F_i}{\sum F_i}$ |
| $M_3^{(0)} = M_3$ | $\frac{\sum X_i^3}{n}$ | $\frac{\sum_{i=1}^{n} X_i^3 \; F_i}{\sum F_i}$ |
| $M_2^{(0)} = M_2$ | $\frac{\sum X_i^4}{n}$ | $\frac{\sum_{i=1}^{n} X_i^4 \; F_i}{\sum F_i}$ |

**لاحظ انه من التعريف السابق للعزوم حول نقطة الاصل اى حول الصفر سواء فى حالة القيم**

**( المفردات ) او البيانات المبوبة تكراريا فان العزم الاول حول الصفر ما هو الا الوسط الحسابى اى ان :** $M_1^{(0)} = \bar{X}$

**٢ - *<u>العزوم حول الوسط الحسابى ( العزوم المركزية ) :</u>***

**تكمن اهمية هذا النوع من انواع العزوم فى انها تدرس العزوم حول نقطة مركز التوزيع ( وهى بمثابة قيمة الوسط الحسابى ) لذا تسمى بالعزوم المركزية .**

**هـذا ويمكـن ايجـاد العـزوم المركزيـة مـن خـلال التعـويض عـن $\bar{x}$ =∝ فـى المعـادلتين ( 1 ) ، (2) ومـن ثـم فـان صـيغة العزم ذو الرتبـة (r) حـول الوسـط الحسابى ( المركزية ) فى حالة البيانات المفردة تأخذ الصورة التالية :**

$$\grave{M}_r = M_r^{(\bar{X})} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^r}{n} \qquad (5)$$

*Where r = 1, 2, 3,………*

**اما عن صيغة العزم ذو الرتبة (r) حول الوسط الحسابى فى حالة البيانات المبوبة تكراريا فهى تأخذ الصورة التالية :**

$$\grave{M}_r = M_r^{(\bar{X})} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})\,F_i}{\sum F_i} \qquad (6)$$

*Where r = 1, 2, 3,………*

**حيـث $(X_i)$ تعبـر عـن مركـز الفئـة (i) ، $(F_i)$ يعبـر عـن تكـرار الفئـة (i) i=1,2,……n ، (n) تعبر عن عدد فئات التوزيع التكرارى .**

**والجدول (٥-٢) يلخص صيغة العزوم الاربعة الاولى حول الوسط ( المركز ) وذلك فى حالة البيانات المفردة ( الغير مبوبة ) والبيانات المبوبة .**

جدول ( ٥- ٢)

| Case<br>r $\underset{\sim}{th}$ *Moment* | Values (القيم المفردة) | Frequency distributions |
|---|---|---|
| $\grave{M}_1 = M_1^{(\bar{x})}$ | $\frac{\sum (X_i - \bar{X})}{n} = 0$ | $\frac{\sum (X_i - \bar{X}) F_i}{\sum F_i}$ |
| $\grave{M}_2 = M_2^{(\bar{x})}$ | $\frac{\sum (X_i - \bar{X})^2}{n} = \sigma^2$ | $\frac{\sum (X_i - \bar{X})^2 F_i}{\sum F_i} = \sigma^2$ |
| $\grave{M}_3 = M_3^{(\bar{x})}$ | $\frac{\sum (X_i - \bar{X})^3}{n}$ | $\frac{\sum (X_i - \bar{X})^3 F_i}{\sum F_i}$ |
| $\grave{M}_4 = M_4^{(\bar{x})}$ | $\frac{\sum (X_i - \bar{X})^4}{n}$ | $\frac{\sum (X_i - \bar{X})^4 F_i}{\sum F_i}$ |

**لاحظ ايضا انـه مـن خـلال التعريـف السـابق للعزوم حـول الوسـط الحسـابى فـأن العزم الثانى حول الوسط الحسـابى مـا هو الا تبـاين التوزيـع لـذا فهو دائمـا ابـدا مقدارا موجبا فيما عدا فى حالـة القيم الثابتـة فتباينـهـا او تشتتها صفرا بطبيعة الحال لان الثابت لا يختلف من حيث القيم .**

٣- ***العزوم حول وسط فرضى معين وليكن ( a) :***

**اذا اعتبرنـا ان $a = \propto$ حيث (a) عبـارة عن مقدار ثابـت معين فأنـه بـالتعويض فـى المعـادلتين (1) ، (2) السـابق الاشـارة لـهمـا نحصـل علـى العزوم المختلفـة**

**حول هذا الوسط الفرضى (a) ومن ثم فان صيغة العزم ذو الرتبة (r) حول الوسط الفرضى (a) فى حالة المفردات تأخذ الصورة التالية :**

$$M_r^{(a)} = \frac{\sum (X_i - a)^r}{n} \qquad \text{where } r = 1,2,3,\ldots.. \qquad (7)$$

**اما عن صيغة العزم ذو الرتبة (r) حول الوسط الفرضى (a) فى حالة البيانات المبوبة تكراريا فتأخذ الصورة التالية :**

$$M_r^{(a)} = \frac{\sum (X_i - a)^r F_i}{\sum F_i} \qquad \text{where } r = 1,2,3,\ldots.. \qquad (8)$$

**والجدول (٥-٣) يلخص صورة العزوم الاربعة الاولى حول الوسط الفرضى (a) .**

**دول (٥-٣)**

| Case<br><br>r $\underset{\sim}{th}$ *Moment* | Values (القيم المفردة) | Frequency distributions |
|---|---|---|
| $M_1^{(a)}$ | $\frac{\sum (X_i - a)}{n}$ | $\frac{\sum (X_i - a) F_i}{\sum F_i}$ |
| $M_2^{(a)}$ | $\frac{\sum (X_i - a)^2}{n}$ | $\frac{\sum (X_i - a)^2 F_i}{\sum F_i}$ |
| $M_3^{(a)}$ | $\frac{\sum (X_i - a)^3}{n}$ | $\frac{\sum (X_i - a)^3 F_i}{\sum F_i}$ |
| $M_4^{(a)}$ | $\frac{\sum (X_i - a)^4}{n}$ | $\frac{\sum (X_i - a)^4 F_i}{\sum F_i}$ |

**هذا ويمكننا تبسيط صورة العزوم حول الوسط الفرضي (a) لتأخذ الصورة التالية :**

***فى حالة المفردات***

$$M_r^{(a)} = \frac{\sum d_i^r}{n}$$

***فى حالة البيانات المبوبة***

$$= \frac{\sum d_i^r F_i}{\sum F_i}$$

**وذلك باعتبار ان $d_i$ تمثل انحرافات القيم (المفردات فى حالة البيانات الغير مبوبة او مراكز فئات التوزيع ( فى حالة البيانات المبوبة ) عن الوسط الفرضي (a) اى ان :**

$$d_i = X_i - a$$

**سواء كانت $X_i$ تعبر عن المفرده رقم (i) فى حالة المفردات او مركز الفئة (i) فى حالة البيانات المبوبة .**

**٤- <u>العزوم المعيارية</u> : Standard Moments**

**بصفة عامة تأخذ صورة العزم المعيارى ذو الرتبة (r) الصيغة التالية**

**فى حالة المفردات :**

$$M_r^{(z)} = \frac{\sum_{i=i}^{n} \left(\frac{X_i - \bar{X}}{\sigma}\right)^r}{n} = \frac{\sum Z^r}{n} \qquad where\ r = 1,2,\dots\dots\dots \quad (9)$$

**اما عن صيغة العزم المعيارى ذو الرتبة (r) فى حالة البيانات المبوبة فأنه يأخذ الصورة التالية:**

$$M_r^{(z)} = \frac{\sum_{i=i}^{n} \left(\frac{X_i - \bar{X}}{\sigma}\right) F_i}{\sum F_i} = \frac{\sum Z^r F_i}{\sum F_i} \qquad (10)$$

$$where\ r = 1,2,3,\dots\dots\dots$$

**والعزوم الاربعة الاولى المعيارية نحصل عليها من خلال المعادلتين (9)،(10) بوضع (r) مساوية 1 , 2 , 3 , 4 على التوالى .**

والعزم الاول المعيارى دائما ابدا قيمته صفرية والعزم الثانى المعيارى دائما ابدا قيمته مساوية للواحد الصحيح وذلك لان العزم الثانى المعيارى ما هو الا تباين القيم المعيارية والتى سبق وان اوضحنا ان الوسط الحسابى للقيم المعيارية هو صفرا اما التباين لها فهو الواحد الصحيح.

هذا ويمكن من خلال صيغ العزوم المختلفة فى حالة الجداول التكرارية تحويرها لكى تتناسب مع التوزيعات التكرارية النسبية . وسوف اترك للقارئ هذه الجزئية على ان يتعامل معها كما تم فى المقاييس السابقة لهذا الباب باختلاف انواعها .

***<u>مثال :</u>***

اذا كان لديك مجموعة القيم التالية : 4 , 5 , 3 , 6 , 7

<u>فالمطلوب حساب</u> :

١- العزوم الاربعة الاولى حول الصفر.

٢- العزوم الاربعة الاولى حول الوسط الحسابى .

٣- العزوم الاربعة الاولى حول المعيارية .

***<u>الحل :</u>***

١- <u>العزوم الاربعة الاولى حول الصفر تتطلب حساب المجاميع التالية :</u>

$\sum X$ , $\sum X^2$ , $\sum X^3$ , $\sum X^4$ والجدول التالى يبين الحسابات اللازمة لحساب العزوم الاربعة الاولى حول الصفر .

**جدول (٥-٤)**

**المجاميع اللازمة لحساب العزوم الاربعة الاولى حول الصفر :**

| X | $X^2$ | $X^3$ | $X^4$ |
|---|---|---|---|
| 4 | 16 | 64 | 256 |
| 5 | 25 | 125 | 625 |
| 3 | 9 | 27 | 81 |
| 6 | 36 | 216 | 1296 |
| 7 | 49 | 343 | 2401 |
| 25 | 135 | 575 | 4609 |

**وحيث ان الصيغة العامة لمعادلة العزوم ذو الرتبة (r) حول الصفر**

**فى حالة المفردات هى :**

$$M_r^{(0)} = \frac{\sum X^r}{n}$$

**لذا فان العزوم الاربعة الاولى هى :**

**العزم الاول حول الصفر:**

$$M_1^{(0)} = \frac{\sum X}{n} = \frac{25}{5} = 5 = \bar{X}$$

**العزم الثانى حول الصفر :**

$$M_2^{(0)} = \frac{\sum X^2}{n} = \frac{135}{5} = \frac{135}{5} = 27$$

**العزم الثالث حول الصفر :**

$$M_3^{(0)} = \frac{\sum X^3}{n} = \frac{575}{5} = 115$$

**اخيرا العزم الرابع حول الصفر :**

$$M_4^{(0)} = \frac{\sum X^4}{n} = \frac{4609}{5} = 921.8$$

٢ - ***العزوم الاربعة الاولى حول الوسط الحسابى ( العزوم المركزية ) :***

**من خلال العزم الاول حول الصفر والذى يعتبر بمثابة الوسط الحسابى للقيم فأنه صيغة العزوم الاربعة الاولى حول الوسط الحسابى تأخذ الصورة التالية باعتبار ان $\bar{X} = 5$ هى :**

$$\grave{M}_r = M_r^{(\bar{X})} = \frac{\sum(X_i - 5)^r}{n}$$

**والجدول التالى يوضح حسابات العزوم الاربعة الاولى لتلك المفردات**

**جدول ( ٥-٥)**

**المجاميع اللازمة لحساب العزوم الاربعة الاولى حول الوسط الحسابى**

| X | $(X_i-5)$ | $(X_i-5)^2$ | $(X_i-5)^3$ | $(X_i-5)^4$ |
|---|---|---|---|---|
| 4 | -1 | 1 | -1 | 1 |
| 5 | 0 | 0 | 0 | 0 |
| 3 | -2 | 4 | -8 | 16 |
| 6 | 1 | 1 | 1 | 1 |
| 7 | 2 | 4 | 8 | 16 |
| $\Sigma$ | 0 | 10 | 0 | 34 |

**وعليه فان العزوم الاربعة الاولى حول الوسط الحسابى هى :**

$$\grave{M}_1 = M_1^{(\bar{X})} = \frac{\sum(X-5)}{n} = \frac{0}{5} = 0 \quad ,$$

$$\grave{M}_2 = M_2^{(\bar{X})} = \frac{\sum(X - 5)^2}{n} = \frac{10}{5} = 2 \quad ,$$

$$\grave{M}_3 = M_3^{(\bar{X})} = \frac{\sum(X - 5)^3}{n} = \frac{0}{5} = 0 \quad ،$$

$$\grave{M}_4 = M_4^{(\bar{X})} = \frac{\sum(X - 5)^4}{n} = \frac{34}{5} = 6.8$$

٣- **<u>العزوم الاربعة الاولى المعيارية :</u>**

**لحساب العزوم الاربعة الاولى المعيارية يتم حساب القيم المعيارية للمفردات من خلال التحويله المعيارية التالية :**

$$Z_i = \frac{X_i - \bar{X}}{\sigma}$$

**وحيث ان العزم الاول حول الصفر($\bar{X}$) يساوى القيمة (5) والعزم الثانى حول الوسط الحسابى هو بمثابة تباين المفردات المعطاه اى ان $\sigma^2 = 2$**

**لذا فان الانحراف المعيارى هو $\sigma = \sqrt{2} \simeq 1.414$**

**ومن ثم يمكن حساب العزوم الاربعة الاولى المعيارية كما يوضحها جدول (٥-٦) وذلك على النحو المبين التالى :**

**جدول (٥-٦)**

**جدول حساب العزوم الاربعة الاولى المعيارية**

| $X_i$ | $z_i = \frac{X_i - 5}{1.414}$ | $z_i^2 = \left(\frac{X_i - 5}{1.414}\right)^2$ | $z_i^3 = \left(\frac{X_i - 5}{1.414}\right)^3$ | $z_i^4 = \left(\frac{X_i - 5}{1.414}\right)^4$ |
|---|---|---|---|---|
| 4 | $\frac{4-5}{1.414} = 0.707$ | 0.499849 | -0.353393243 | 0.249849022 |
| 5 | 0 | 0 | 0 | 0 |
| 3 | -1.414 | 1.999396 | -2.827145944 | 3.997584365 |
| 6 | 0.707 | 0.499849 | 0.353393243 | 0.249849022 |
| 7 | 1.414 | 1.999396 | 2.827145944 | 3.997584365 |
| | 0 | 4.99849 | 0 | 8.494866775 |

**وعليه فان العزوم الاربعه الاولي المعياريه هي :**

$$M_1^{(z)} = \frac{\sum\left(\frac{X_i - \bar{X}}{\sigma}\right)}{n} = \frac{0}{5} = 0$$

$$M_2^{(z)} = \frac{\sum\left(\frac{X_i - \bar{X}}{\sigma}\right)^2}{n} = \frac{4.99849}{5} = 0.999698 \simeq 1$$

$$M_3^{(z)} = \frac{\sum\left(\frac{X - \bar{X}}{\sigma}\right)^3}{n} = \frac{8.494866775}{5} = 1.698973355$$

لاحظ ان العزم الاول المعياري ما هو الا الوسط الحسابي للقيم المعياريه وهو يساوي الصفر دائما . كذلك فان العزم الثاني المعياري ما هو الا تباين القيم المعياريه وهو دائما ابدا يساوي الواحد الصحيح ان لم يكن هناك تقريب في قيمة الوسط الحسابي او الانحراف المعياري للمفردات التي تم حساب القيم المعياريه لها .

***مثال :***

احسب العزوم الاربعة الأولى حول الصفر وكذا حول الوسط الحسابي للتوزيع التكراي التالي :

| Classes | 1- | 5- | 9- | 13-17 |
|---|---|---|---|---|
| Frequency | 2 | 3 | 4 | 1 |

***الحل :-***

لحساب العزوم الابعه الاولي حول الصفر نكون الجدول التالي :-

| classes | $F_i$ | $x_i$ | $X_i F_i$ | $X_i^2F_i$ | $X_i^3 F_i$ | $X_i^4F_i$ |
|---|---|---|---|---|---|---|
| 1- | 2 | 3 | 6 | 18 | 54 | 162 |
| 5- | 3 | 7 | 21 | 147 | 1029 | 7203 |
| 9- | 4 | 11 | 44 | 484 | 5324 | 58564 |
| 13-71 | 1 | 15 | 15 | 225 | 3375 | 50625 |
| Σ | 10 | | 86 | 874 | 9782 | 116554 |

وعليه فان العزوم الاربعه الاولي حول الصفر هي :

$$M_1^{(0)} = \frac{\sum X_i F_i}{\sum F_i} = \frac{86}{10} = 8.6 = \bar{X}$$ و

$$M_2^{(0)} = \frac{\sum X_i{}^2 F_i}{\sum F_i} = \frac{874}{10} = 87.4$$ و

$$M_3^{(0)} = \frac{\sum X_i{}^3 F_i}{\sum F_i} = \frac{9782}{10} = 978.2$$ و

$$M_4^{(0)} = \frac{\sum X_i{}^4 F_i}{\sum F_i} = \frac{116554}{10} = 11655.4$$

**اما لحساب العزوم الاربعه الاولى حول الوسط الحسابي بمعلومية ان الوسط الحسابي للتوزيع ( العزم الاول حول الصفر ) يساوي القيمة 8.6 فيتم تكوين الجدول التالي :**

**جدول ( ٥ – ٧ )**

**حساب العزوم الاربعه الاولي حول الوسط الحسابي**

| classes | $F_i$ | $X_i$ | $(X_i - 8.6)$ | $(X_i - 8.6)F_i$ | $(X_i - 8.6)^2F_i$ | $(X_i-8.6)^3F_i$ | $(X_i-8.6)^4F_i$ |
|---|---|---|---|---|---|---|---|
| 1- | 2 | 3 | -5.6 | -11.22 | 62.72 | -351.232 | 1966.899 |
| 5- | 3 | 7 | 1.6 | -4.8 | 7.68 | -12.288 | 19.6608 |
| 9- | 4 | 11 | 2.4 | 9.6 | 23.04 | 55.296 | 132.7104 |
| 13-17 | 1 | 15 | 6.4 | 6.4 | 40.96 | 262.144 | 1677.7216 |
| Σ | 10 | | | 0 | 124.4 | -46.08 | 3796.992 |

**وعليه فان العزوم الاربعه الاولي حول الوسط الحسابي هي :**

$$M_1^{(\bar{X})} = \frac{\sum (X_i - \bar{X}) F_i}{\sum F_i} = \frac{0}{10} = 0$$ (بالتعريف)

$$M_2^{(\overline{X})} = \frac{\sum (X_i - \overline{X})^2 F_i}{\sum F_i} = \frac{134.4}{10} = 13.44 = \sigma^2$$

و

$$M_3^{(\overline{X})} = \frac{\sum (X_i - \overline{X})^3 F_i}{\sum F_i} = \frac{-46.08}{10} = -4.608$$

,

$$M_4^{(\overline{X})} = \frac{\sum (X_i - \overline{X})^4 F_i}{\sum F_i} = \frac{3796.992}{10} = 379.6992$$

- ***العلاقة بين الانواع المختلفة من العزوم :***

**تفيد دراسة العزوم كما سبق وأن ذكرنا في معرفة خصائص التوزيعات التكراريه. فيمكن عن طريقة العزوم معرفه أهم مقاييس النزعة المركزية وهو الوسط الحسابي حيث راينا ان العزم الاول حول نقطة الاصل (الصفر) ما هو الا عباره عن الوسط الحسابي . كما تمكننا العزوم ايضا من معرفة اهم مقياس من مقاييس التشتت المطلق وهو الانحراف المعياري ومن ثم التباين حيث راينا ان العزم الثاني حول الوسط الحسابي ( العزم الثاني المركزي ) ما هو الا التباين .**

**والاهم من ذلك فأن العزوم حول الوسط اي العزوم المركزية تمكننا من دراسة مدي تماثل التوزيع من عدمه . الا ان طريقة حساب العزوم المختلفة حول الوسط تحمل في فحواها حسابات صعبة ومعقده وخصوصا عندما يزداد عدد المفردات والقيم وبصفه خاصة اذا احتوت قيمة الوسط الحسابي ($\overline{X}$) علي قيم كسرية فنضطر الي تقريبها مما ينجم عنه تقريب كبير في النتائج النهائية وتراكم عملية التقريب هذه تؤدي الي نتائج قد تكون مضلله الي حدا ما . لذا يفضل حساب العزم الثاني والثالث والرابع حول الوسط بدلالة العزوم حول الصفر . وسوف نقتصر في علاقتنا هذه علي العلاقه ما بين العزوم حول الوسط**

**الحسابي والعزوم حول نقطة الاصل وسوف نورد هذه العلاقات دون برهان للتبسيط علي القارئ .**

**وهذه العلاقه هي :**

$$M_3^{(\bar{X})} = M_3^{(0)} - 3\,M_2^{(0)}\,M_1^{(0)} + 2\,(M_1^{(0)})^2$$

$$M_4^{(\bar{X})} = M_4^{(0)} - 4\,M_3^{(0)}\,M_1^{(0)} + 6\,M_2^{(0)}\,(M_1^{(0)})^2 - 3\,(M_1^{(0)})^4$$

- **العزوم المصصحه لشبرد :**

**سبق وان ذكرنا انه عند تبويب المفردات او البيانات في صورة جدول تكراري فاننا نعتمد علي مراكز الفئات عند حساب الكثير من المقاييس الاخصائية . والعزوم الرياضية شأنها في ذلك شأن باقي هذه المقاييس الاحصائية التي تبني علي نفس الفروض بأن التكرارات تتوزع بأنتظام علي مدار طول الفئة وهذا الافتراض وأن كان يشوبه عدم الدقه لذا قام شيرد بعمل تصحيح للعزوم المركزية حيث اوجد تصحيحا للعزم الثاني حول الوسط بحيث باخذ الصورة التالية :**

$$\text{correction of } \mathbf{M}_2^{(\bar{X})} = \mathbf{M}_2^{(\bar{X})} - \frac{{L_i}^2}{12}$$

**حيث $(L_i)$ تعبر عن طول الفئة في التوزيع التكراري**

**كما قام شبرد بتصحيح العزم الرابع حول الوسط الحسابي ليأخذ الصورة المصححة التالية :**

$$\text{correction of } \mathbf{M}_4^{(\bar{X})} = \mathbf{M}_4^{(\bar{X})} - \frac{{L_i}^2}{12}\,\mathbf{M}_2^{(\bar{X})} + \frac{7}{240}\,{L_i}^4$$

**ويشترط لاجراء هذه التصحيحات ان يكون منحني التوزيع التكراري متصلا وأن يقابل طرفي المنحني المحور الافقي وان تكون فئات التوزيع متساوية الطول . وعلي المستوي الواقع العملي فأنه نادرا ما تستخدم العزوم المصححه ولن تقوم بحسابها الا اذا نص المطلوب صراحه علي حسابها .**

## *ثانياً :- الالتواء skewness*

**بعد ان تعرضنا في دراستنا السابقه الي مقاييس النزعة المركزية ومقاييس التشتت بنوعيها المطلقة والنسبية والعزوم الرياضية للبيانات الغير مبوبة او المبوبة تكراريا لظاهرة ما تأتي الان دراسة ما اذا كان التوزيع الخاص بظاهرة ما متماثلا او ملتويا وما هو شكل ودرجة التواء التوزيع دون ان تقوم برسم منحني التوزيع . حيث يفيد ذلك في وصف وتحليل توزيع الظاهرة وذلك لان التوزيع المتماثل يعني ان المنحني التكراري للظاهرة موضع الدراسه تنقسم عند القيمة المتوسطه لها إلى قسمين متطابقين تماما ويكون تزايد او تناقص التكرارات منتظما او متشابها او بطريقة متماثله علي جانبي المحور الراسي المقام عند القيمة المتوسطة للتوزيع .**

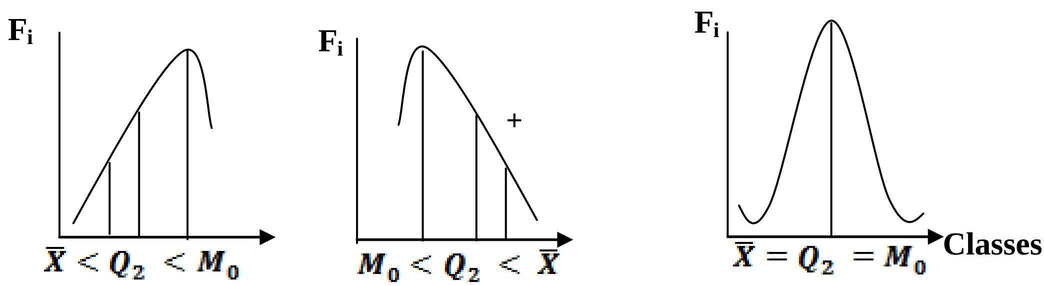
**اما التواء التوزيع فنقصد به بعد المنحني عن التماثل اي تركز التكرارات في طرف وقلتها في طرف اخر . حيث يكون التوزيع موجب الالتواء Positive skewness (أي ملتوي جهة اليمين ) إذا تركزت التكرارات الكبيرة عند الفئات الدنيا (اي الفئات الاولي ) للتوزيع اما التكرارات الصغيرة فنتشر عند الفئات العليا من التوزيع ويكون ذيل التوزيع جهة اليمين اطول من جهة اليسار العكس صحيح حيث يكون التوزيع سالب الالتواء Negative skewness (اي ملتوي جهة اليسار) نتيجة لتركيز عدد كبير من التكرارات عند الفئات**

العليا للتوزيع بينما تنتشر قليل من التكرارات عند الحدود الدنيا للتوزيع ويكون حينئذ التوزيع ذو زيل اطول جهة اليسار (توزيع ذو ذيل ايسر) .

وخلاصة ما سبق هو ان التوزيع التكراري من حيث التماثل او الالتواء يكون واحدا من الصور الثلاث : فاما ان يكون التوزيع متماثلا او ملتوي جهة اليمين او ملتوي جهة اليسار .

وفي دراستنا لمقاييس الالتواء سوف نميز بين نوعين من المقاييس حيث يعتمد النوع الاول في ايجادة علي مقاييس النزعة المركزية ومقاييس التشتت التي تعرضنا لها سابقا . اما النوع الثاني من مقاييس الالتواء فيعتمد في حسابه علي العزوم .

هذا ويمكن من خلال مقارنة مقاييس المتوسطات (الوسط – الوسيط – والمنوال ) دراسة ماهية تماثل التوزيع من عدم كما يوضحه شكل (٥-١)




Negative skewness
توزيع سالب الالتواء
(C)

Possitive skewness
توزيع موجب الالتواء
(B)

Symmetric distribution
توزيع متماثل
(A)

* فاذا كان التوزيع متماثلا فان الوسط الحسابي يتساوى مع الوسيط يتساوى مع المنوال اى ان :

كما هو واضح فى شكل (A-1-5) $\bar{X} = Q_2 = M_0$

اما اذا ابتعدت التوزيعات عن التماثل فان قيم مقاييس المتوسطات ستختلف عن بعضها. وحيث ان الوسط الحسابى يتاثر بالقيم المتطرفه او الشاذه لذلك نراه دائما مائلا فى اتجاه ذيل التوزيع بينما تكون قيمة المنوال دائما تقابل نقطة قمة منحنى التوزيع فى حين تقع قيمة الوسيط دائما بين كل من الوسط الحسابى والمنوال وبصفة عامة اذا كان:

$$\bar{X} > Q_2 > M_0$$

فان التوزيع يكون موجب الالتواء كما هو فى شكل (B – 1 – 5)

أما إذا كان : $\bar{X} > Q_2 > M_0$

فان التوزيع يكون سالب الالتواء كما هو فى شكل ( c – 1 – 5)

هذا ويزداد قرب او بعد التوزيع عن التماثل بمدى قرب او بعد مقاييس المتوسطات الثلاثه عن بعضها البعض. إلا أن هذه الدراسة تعطى اشارة لنوع الالتواء فقط دون قياس لدرجة الالتواء. لكن لتحديد درجة الالتواء بالإضافة لنوعة ايضا فيجب حسب اى من نوعى معاملات الالتواء السابق الاشارة لها والتى سترد حالا فى هذا القسم .

***١ - معاملات الالتواء بدلاله مقاييس المتوسطات و التشتت : -***

ذكرنا حالا أنه كلما بعدت مقاييس المتوسطات عن بعضها البعض كلما بعد التوزيع عن التماثل ومن ثم فأنه يمكن استخدام الفروق بين قيم هذه المتوسطات فى ايجاد مقاييس الالتواء لكن فى الحقيقة فان الفروق بين فيمابين مقاييس المتوسطات وحدها فى حد ذاتها لا تقيس درجة او حدة الالتواء بشكل صحيح اذا انه من الممكن ان يكون الالتواءكبيرا فى الوقت الذى تكون فيه قيمة الفروق صغيرة بين مقاييس المتوسطات ويرجع ذلك لصغر قيمة الانحراف المعيارى للتوزيع ومن جهة اخرى فقد يكون هذا الفرق كبير وفى نفس الوقت

**يكون الالتواء صغير وذلك نظرا لكبر قيمة الانحراف المعيارى للتوزيع. ولذلك فإنه لاستخدام تلك الفروق فيما بين مقاييس المتوسطات فى دراسة الالتواء يجب ان ننسب هذه الفروق الى مقياس التشتت المناظر ( اى من نفس طبيعة مقياس النزعة المركزية المستخدم وله نفس طريقة الحساب ) ويسمى الناتج بمعامل الالتواء وبصفة عامة فان معامل الالتواء يجب ان يحقق الشروط التالية:**

*** ان تكون قيمته صفرية فى حالة التوزيعات المتماثلة.**

*** ان تكون قيمة معامل الالتواءمجردة من وحدات القياس .**

**وفيما يلى اكثر معاملات الالتواء شيوعا :**

$$B_1 = \frac{\bar{X} - M_0}{\sigma}$$

**أ - معامل الالتواء الاول لبيرسون ($B_1$) وهو عبارة عن :**

**اى نسبة الفرق ما بين بين الوسط الحسابى والمنوال الى الانحراف المعيارى للتوزيع**

**ب – معامل الالتواء الثانى لبيرسون ($B_2$) وهوعبارة عن :**

$$B_2 = \frac{3\,(\bar{X} - Q_2)}{\sigma}$$

**اى النسبة بين ثلاثه امثال الفرق بين الوسط والوسيط الى الانحراف المعيارى للتوزيع.**

**وتتراوح قيمتى معاملى الالتواء الاول والثانى لبيرسون اى $B_1$ , $B_2$ ما بين +3 ,3- إلا أن ما يعاب على هذين المعاملين ان كلاهما لا يمكن حسابه من**

**الجداول التكرارية المفتوحة حيث يصعب حساب كل من الوسط الحسابى والانحراف المعيارى كما يتاثر هذين المعاملين بالقيم الشاذة او المتطرفة .**

**وهذين المعاملين تتساوى قيمتهم بالصفر فى حالة التوزيعات المتماثلة اما اذا كان الوسط الحسابى اكبر من كل من المنوال والوسيط فان التوزيع يكون موجب الالتواء والعكس صحيح .**

**وللتغلب على عيوب معاملى بيرسون للالتواء $B_1$ , $B_2$ من حيث عدم امكانية حسابهم فى حالة التوزيعات التكرارية المفتوحة وكذلك تاثرهم بالقيم الشاذة فقد امكن استنتاج معامل ثالث للالتواء وهو ما يسمى بمعامل بولى Bowley للالتواء او ما يسمى بمعامل الالتواء الربيعى والذى ياخذ الصوره التالية :**

$$B_3 = \frac{3\,(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1}$$

**وأهم ما يميز هذا المعامل بالاضافة لعدم تأثره بالقيم الشاذة او المتطرفه فهو المعامل الوحيد الذى يمكن ايجاده من خلال الرسم البيانى حيث يمكن من خلال رسم المنحنى التكرارى المجتمع الصاعد او الهابط ايجاد كل من الربيعين الادنى والاعلى والوسيط ثم التعويض فى صيغة المعامل.**

**هذا وتتساوى قيمة معامل بولى ($B_3$) بالصفر فى حالة التوزيعات المتماثلة وذلك لان الربيعين الادنى ($Q_1$) والاعلى ($Q_3$ ) يقعان على بعدين متساويين من الوسيط ($Q_2$) فى هذه الحالة وتتراوح قيمة معامل بولى للالتواء ما بين 1- , 1+ لكن يعاب على هذا المعامل انه لا يأخذ فى الاعتبار مجموعة القيم التى تسبق الربيع الادنى او التالية للربيع الاعلى ولذلك يعتبر اقل دقة من معاملى بيرسون للالتواء حيث لا ياخذ كل القيم فى الحساب .**

د – هناك معامل رابع يقيس نوع ودرجة الالتواء وان كان اقل استخداما من معاملات السابقة وهو ما يسمى بمعامل الالتواء الميئنى نظرا لاعتماده على حساب الميئنيات وهو ياخذ الصورة التالية :

$$B_4 = \frac{(P_{90} - Q_2) - (Q_2 - P_{10})}{P_{90} - P_{10}}$$

حيث $P_{10}$ , $P_{90}$ تمثل المئين العاشر والمئين التسعون على الترتيب.

٢- <u>قياس الالتواء بإستخدام العزوم المركزية ( حول الوسط ) :</u>

قياس الالتواء باستخدام العزوم يعتمد على قيمة العزم الثالث حول الوسط الحسابى اى ( $M_3^{(\bar{X})}$ ) فاذا كانت قيمة العزم الثالث حول الوسط مساويه للصفر يكون التوزيع متماثلا حيث تكون الانحرافات السالبة مساوية للانحرافات الموجبة ويكون المجموع الجبرى لمكعبات تلك الانحرافات مساويا للصفر اما اذا كانت قيمة العزم الثالث حول الوسط تختلف عن الصفر فيكون التوزيع متلوياً وله نفس اتجاه اشارة هذا العزم بمعنى اذا كانت :

$$M_3^{(\bar{X})} > 0$$

فيكون التوزيع موجب الالتواء اما اذا كان :

$$M_3^{(\bar{X})} < 0$$

فيكون التوزيع سالب الالتواء اما عن قياس درجة الالتواء فتكون من خلال النسبة مابين العزم الثالث حول الوسط والجذر التربيعى لمكعب العزم الثانى حول الوسط اى ان معامل الالتواء بدلاله العزوم ($B_5$) ياخذ الصورة التالية :

$$B_5 = \frac{M_3^{(\bar{X})}}{\sqrt{\left(M_2^{(\bar{X})}\right)^3}} = \frac{M_3^{(\bar{X})}}{\left(M_2^{(\bar{X})}\right)^{3/2}}$$

**وبصفه عامة فان معاملات الالتواءالخمسة السابقة تشترك فى دراسة درجة تماثل او التواء التوزيع كما يلى :**

- **فاذا كانت قيمة معامل الالتواء تساوى الصفر يكون التوزيع متماثلا .**
- **واذا كانت قيمة معامل الالتواء موجبا يكون التوزيع ملتويا جهة اليمين أى أن له التواء موجب.**
- **واذا كانت قيمة معامل الالتواء سالبا يكون التوزيع ملتويا جهة اليسار أى ان له التواء سالب.**

### **٣- *التفرطح او التفلطح* kurtosis**

**توجد خاصية اخرى للتوزيعات وحيدة القمة بالاضافة الى خاصية التماثل او الالتواء وهى خاصية الاعتدال والتفرطح فقد تتساوى منحنيات التوزيعات التكرارية فى مقاييس المتوسطات والتشتت والالتواء ولكنها قد تختلف من حيث شكل القمة اى ان الالتواء وحده غير كاف لتحديد معالم التوزيع التكرارى او لاجراء المقارنات فيما بين التوزيعات التكرارية المختلفة فقد تكون احداها ذات قمه عريضة ( مفرطح ) او قد تكون اكثر تدببا لذلك فان لابد من وجود مقياس احصائى يدرس شكل قمه التوزيع وهو ما يسمى بمعامل التفرطح والذى يقيس درجة تفرطح التوزيع اى شكل قمته وهناك معاملين اساسين يمكن استخدامهم فى دراسة شكل قمه المنحنيات التكرارية يسميان بمعاملات التفرطح وهما :**

## أ – معامل التفرطح بدلاله العزوم :

وهذا المعامل هو خارج قسمة العزم الرابع المعيارى حول الوسط على مربع العزم الثانى حول الوسط ويرمز له بالرمز ($\propto_1$) حيث ان :

$$\propto_1 = \frac{M_4^{(\bar{X})}}{\left(M_2^{(\bar{X})}\right)^2}$$

وتمت القسمة على مربع العزم الثانى حول الوسط ( مربع التباين ) وذلك للتخلص من وحدات القياس ونحصل منه على مقياس نسبى .

وقد وُجد ان قيمة هذا المعامل ($\propto_1$) تساوي ( 3) لمنحني التوزيع الطبيعي (المعتدل). وقد استقر الاحصائيون في الراي علي اعتبار ان المنحني الطبيعي متوسط التفرطح Mesokurtic ويستخدم بالتالي كاساس او معيار لمقارنة درجة تفرطح منحنيات التوزيعات الاخري. لذا فأنه بعد حساب معامل التفرطح بدلالة العزوم ($\propto_1$) فتتم مقارنة معامل تفرطح التوزيع بمعامل التفرطح الخاص بالتوزيع المعتدل ( اي بالقيمة3). وهنا نكون بصدد احد الحالات الثلاث التالية :

* اذا كانت قيمة $\propto_1 = 3$: فان هذا يعني ان المنحني طبيعي او معتدل اي متوسط التفرطح Mesokurtic كما موضح في شكل ( ٥-٢-A) .

* واذا كانت $\propto_1$ للمنحني اقل من (٣): فان هذا يعني ان مح\نحني التوزيع مفرطح Platykutric اي ذو قيمة عريضة ( اي يكون اكثر اتساعا من وسطه وتنخفض قيمته عن قيمة المنحنى المعتدل ) كما هو موضح في شكل ( ٥- ٢- B ) .

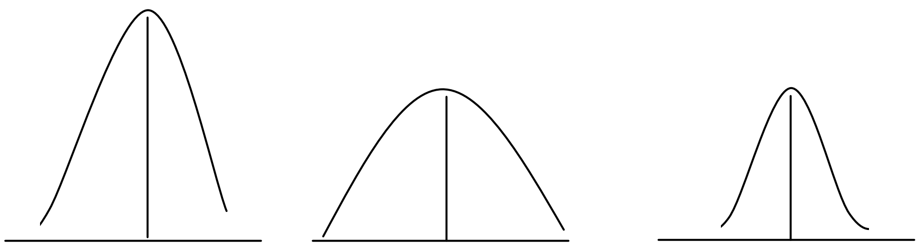
**اما اذا كانت قيمة $\propto_4$ اكبر من (٣): فان هذا يعني ان منحني التوزيع مدبب Leptokurtic اي يكون اكثر ضيقا من وسطه وترتفع قمته عن المنحني المعتدل .**

ب – معامل التفرطح بدلالة الربيعيين والئينيات :

**وهذا المعامل عبارة عن النسبه ما بين نصف المدي الربيعي وبين الفرق ما بين المئين التسعين والمئين العاشر ويرمز له بالرمز ($\propto_2$ ) اي ان**

$$\propto_2 = \frac{Q_3 - Q_1}{2\,(P_{90} - P_{10})}$$

**وتبلغ قيمة هذا المعامل 0.263 في حالة التوزيع الطبيعي في حين يكون منحني التوزيع مفرطحاً إذا كانت قيمة معامل التفرطح ($\propto_4$ ) أكبر من (0.263) ويكون منحنى التوزيع مدببا اذا كانت قيمة معامل التفرطح ( $\propto_4$ ) أقل من القيمة 0.263 هذا ويفضل استخدام هذا المعامل في حالة الجداول التكرارية المفتوحه حيث يصعب ايجاد المعامل الاول للتفرطح والمعتمد في قياسه علي العزوم . كما يفضل استخدامه في حالة وجود قيم شاذه او متطرفه في البيانات.**



Leptokurtic Dist.
توزيع مدبب

Platykurtic Dist.
توزيع مفرطح

Mesokurtic Distributio.
توزيع معتدل ( طبيعي)

**شكل (٥-٢)**

## ملاحظات حول الالتواء والتفرطح :-

١- معاملات الالتواء والتفرطح كلها مقاييس نسبيه ليس لها تمييز . ولذلك فهذه المعاملات تصلح لمقارنة توزيعات سواء متحدة او مختلفة في وحدات قياسها .

٢- يلاحظ ان قيم معاملات الالتواء السابقة قد تختلف بالنسبة للتوزيع الواحد وذلك بسبب اختلاف تعريف كل منها وهذا يعني ضرورة استخدام المعامل نفسه عند مقارنة درجة الالتواء لتوزيعين او اكثر .

٣- تتوقف اشارة اياً من معاملات الالتواء علي اشارة البسط وذلك لان اشارة المقام دائما ابدا موجبه باستثناء حالة وحيدة وهي حينما تكون قيمه($\sigma$) او الفرق بين الربيعيين الاعلي والادني مساوية للصفر وهي تمثل حالة التجانس الكامل للبيانات (ظاهرة تأخذ قيمة ثابتة دون تغير او اختلاف )

***مثال :***

ادرس خاصيتي الالتواء والتفرطح للتوزيع التكراي التالي :

| classes | 20- | 22- | 24- | 26- | 28- | 30-32 |
|---|---|---|---|---|---|---|
| frequently | 3 | 8 | 10 | 18 | 4 | 4 |

***الحل***:-

حيث ان التوزيع التكراري المعطي مغلق الطرفين لذا يفضل قياس الالتواء بدلالة العزوم وقياس التفرطح كذلك بدلالة العزوم .وهو ما يستلزم حساب العزوم الأربعة الأولى حول الحسابي. والجدول التالي يبين الحسابات اللازمة لهذه العزوم. حيث يتم حساب الوسط الحسابي أولاً ثم العزوم الأربعة حول الوسط. فكانت قيمة الوسط الحسابي هي :

| Classes | $F_i$ | $X_i$ | $X_i F_i$ | $(X_i - 26.2)$ | $(X_i - 26.2)F_i$ | $(X_i - 26.2)^2F_i$ | $(X_i - 26.2)^3F_i$ | $(X_i - 26.2)^4F_i$ |
|---|---|---|---|---|---|---|---|---|
| 20 – | 3 | 21 | 63 | - 5.2 | -15.6 | 81.12 | - 421.824 | 2193.4848 |
| 22 – | 8 | 23 | 184 | - 3.2 | -25.6 | 81.92 | -262.144 | 838.8608 |
| 24 – | 10 | 25 | 250 | - 1.2 | -12 | 14.4 | -17.28 | 20.736 |
| 26 – | 18 | 27 | 486 | 0.8 | 14.4 | 11.52 | 9.216 | 7.3728 |
| 28 – | 7 | 29 | 203 | 2.8 | 19.6 | 54.88 | 153.664 | 430.2592 |
| 30 – 32 | 4 | 31 | 124 | 4.8 | 19.2 | 92.16 | 442.368 | 2123.3664 |
| | 50 | | 1310 | | 0 | 336 | -96 | 5614.08 |

$$\bar{X} = \frac{\sum X_i F_i}{\sum F_i} = \frac{1310}{50} = 26.2$$

وعليه فإن العزوم الأربعة الأولي حول الوسط هي :

$$M_1^{(\bar{X})} = \frac{\sum(X_i - \bar{X}) F_i}{\sum F_i} = \frac{\sum(X_i - 26.2) F_i}{\sum F_i} = \frac{0}{50} = 0$$

،

$$M_2^{(\bar{X})} = \frac{\sum(X_i - \bar{X})^2 F_i}{\sum F_i} = \frac{\sum(X_i - 26.2)^2 F_i}{\sum F_i} = \frac{336}{50} = 6.72$$

،

$$M_3^{(\bar{X})} = \frac{\sum(X_i - \bar{X})^3 F_i}{\sum F_i} = \frac{\sum(X_i - 26.2)^3 F_i}{\sum F_i} = \frac{-96}{50} = -1.92$$

$$M_4^{(\bar{X})} = \frac{\sum(X_i - \bar{X})^4 F_i}{\sum F_i} = \frac{\sum(X_i - 26.2)^4 F_i}{\sum F_i} = \frac{5614.08}{50}$$
$$= 112.2816$$

وعليه فأن معامل الألتواء بدلالة العزوم هو :

$$B_4 = \frac{M_3^{(\bar{X})}}{\sqrt{\left(M_2^{(\bar{X})}\right)^3}} = \frac{-1.92}{\sqrt{(6.72)^3}} = \frac{-1.92}{17.420231} = -0.1102$$

**وهو ما يعني أن هذا التوزيع وإن كان يقرب من التماثل إلا أنه يميل للألتواء السالب بدرجة ضعيفة.**

**أما عن معامل التفرطح بدلاله العزوم ( $\propto_1$ ) فهو عبارة عن**

$$\propto_1 = \frac{M_4^{(\bar{X})}}{\left(M_2^{(\bar{X})}\right)^2} = \frac{112.2816}{(6.72)^2} = 2.4864$$

**وحيث أن قيمه معامل التفرطح أقل من ( 3) فهذا يفيد أن التوزيع مفرطح وإن كانت القيمه تقترب من القيمة (3) أى أن التوزيع يقترب من الاعتدال .**

***<u>مثال:</u>***

**إذا كان العزم الاول حول الرقم (3) هو (8) فما هى قيمة الوسط الحسابى للتوزيع .**

***<u>الحل :</u>***

**حيث أن صيغة العزم ذو الرتبة (r) حول الوسط الفرضى (a) ياخذ الصورة التالية (فى حالة المفردات )**

$$M_r^{(a)} = \frac{\sum(X_i - a)^r}{n}$$

**والقيمة المعطاه فى المثال هى العزم الاول ( أى r=1) حول المقدار (3) أى يتم التعويض عن a = 3 فى المعادلة السابقة وبالتعويض عن قيمه العزم الاول حول الرقم ( 3) أنه مساويا القيمة (8) فأن :**

$$M_1^{(3)} = \frac{\sum(X_i - 3)^1}{n} = 8$$

**ومن ثم فإن :**

$$\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} 3 = 8\,n$$

$$\sum X_i - 3n = 8\,n$$

$$\sum X_i = 8\,n + 3n = 11\,n$$

أي أن:

$$\frac{\sum X_i}{n} = 11$$

$$i.\,e., \ \bar{X} = 11$$

***مثال :***

اذا كان لديك التوزيع التكرارى المتجمع التالى لتوزيع منتظم لأمر عينة من 50 عاملا ( بالجنيه ) فى احد المصانع فى الاسبوع :

| Less that the Classes U.L | ASC. C. F |
|---|---|
| Less than 55 | 2 |
| Less than 65 | 8 |
| Less than 75 | 15 |
| Less than 85 | 35 |
| Less than 95 | 43 |
| Less than 105 | 47 |
| Less than 115 | 50 |

والمطلوب : -

أ – ايجاد معامل الاختلاف المناسب .

ب – ايجاد معامل الالتواء الثانى لبيرسون $(B_2)$

**جـ ـ حدد نسبه العمال الذين تزيد اجورهم عن ٩٠ جنيها اسبوعيا**

***الحل*** **: -**

**معطى فى المثال أن التوزيع التكرارى منتظما لذا فيمكن معرفة الحد الادنى للفئه الاولى وذلك من خلال طرح طول الفئة من الحد الاعلى للفئة الاولى فينتج أن الحد الادنى للفئة الاولى هو 45 = 10 – 55 ومن ثم فالجدول او التوزيع التكرارى المتجمع المعطى هو توزيعا متجمعاً لجدول اصلى مغلق من اعلى وحيث أن العينة حجمها 50 عاملا وأن التكرار المتجمع الصاعد للفئة الاخيرة هو 50 عاملا ايضا لذا فأن الرقم المقابل من قيم الفئات لهذا التكرار المتجمع يعتبر بمثابة الحد الاعلى للفئة الاخيرة فى التوزيع الاصلى. أى أن التوزيع التكرارى المتجمع المعطى هو توزيعا تكراريا من جدول اصلى مغلق من اسفل كذلك. والخلاصة أن هذا التوزيع المتجمع مشتق من توزيع تكرارى مغلق الطرفين لذا فأن معامل الاختلاف المناسب هو معامل الاختلاف النسبى. ولكى يتم حساب هذا المعامل يجب رد التوزيع المتجمع الصاعد لصورة التوزيع التكراري الاصلى حتى يتم حساب كل من الوسط الحسابي والانحراف المعيارى. والجدول التالى يوضح الحسابات اللازمة لهذا المعامل حيث تم استنتاج التكرارات بطريقة عكسية أى من خلال طرح التكرار المتجمع التالى من التركرار المتجمع الحالى ليعطى تكرار الفئة الحالية كما هو موضح بالجدول التالى :**

| Classes | $F_i$ | $X_i$ | $D_i = \frac{X_i - 80}{10}$ | $D_i F_i$ | $D_i^2 F_i$ |
|---|---|---|---|---|---|
| 45- | 2-0=2 | 50 | -3 | -6 | 18 |
| 55- | 8-2=6 | 60 | -2 | -12 | 24 |
| 65- | 15-8=7 | 70 | -1 | -7 | 7 |
| 75- | 35-15=20 | 80 | 0 | 0 | 0 |
| 85- | 43-35=8 | 90 | 1 | 8 | 8 |
| 95- | 47-4=3 | 100 | 2 | 8 | 16 |
| 105-115 | 50-47=3 | 110 | 3 | 7 | 27 |
| $\sum$ | 50 | | | 0 | 100 |

وعليه فأن :

$$\bar{X} = a + b\,\frac{\sum D_i F_i}{\sum Fi} = 80 + 10\left(\frac{0}{50}\right) = 80 \qquad (L.E)$$

و

$$\sigma = b\sqrt{\frac{\sum D_i^2 F_i}{\sum F_i} - \left(\frac{\sum D_i F_i}{\sum F_i}\right)^2}$$

$$= 10\sqrt{\frac{100}{50} - \left(\frac{0}{50}\right)^2} = 10\sqrt{2} = 14.14 \qquad (L.E)$$

ومن ثم فأن معامل الاختلاف النسبى هو :

$$C_v = \frac{\sigma}{\bar{X}} \times 100$$

$$= \frac{14.14}{80} \times 100 = 17.675\,\%$$

ب- لإيجاد معامل الالتواء الثانى لبيرسون أى :

$$B_2 = \frac{3\,(\bar{X} - Q_2)}{\sigma}$$

وهذا المعامل يتطلب حساب الوسيط اما الوسط والانحراف المعيارى فتم حسابهم فى المطلوب السابق فحيث أن

$$Rank\ of\ Q_2 = \frac{\sum F}{2} = \frac{50}{2} = 25$$

وعلية فأنه من خلال الجدول التكرارى المتجمع الصاعد فان قيمة الوسيط هى:

$$Q_2 = 75 + 10 = \left(\frac{25-15}{35-15}\right) = 80 \qquad (\mathrm{L.E})$$

ومن ثم فأن معامل الالتواء الثانى ($B_2$) عبارة عن:

$$B_2 = \frac{3\,(80-80)}{14.14} = 0$$

وهو ما يعنى أن التوزيع متماثل.

جـ - لإيجاد نسبة العمال الذين تزيد اجورهم عن ٩٠ جنيها فأن يمكن من الصاعد معرفة نسبة الاسر الذين تقل أجورهم عن ٩٠جنيها ثم بطرح النسبة الناتجة من 100% لينتج المطلوب مباشرة بدلا من ايجاد التوزيع التكرارى المجتمع الهابط واستنتاج النسبة المطلوبة

*(على الطالب إستنتاج المطلوب والاجابة 22%)*

## تمارين على الباب الخامس

١- فيمـايلى التوزيـع العمـرى لافـراد عينـة حجمهـا ٦٠ مـن العـاملين فـى كـل مـن المصنعين ( أ، ب ) لصناعة المنسوجات الحريرية:

| Age classes | 20- | 25- | 30- | 35- | 40- | 45- | 50- |
|---|---|---|---|---|---|---|---|
| No.of employees(A) | 4 | 5 | 12 | 18 | 11 | 8 | 2 |
| No.of Employees (B) | 2 | 4 | 11 | 15 | 12 | 10 | 6 |

والمطلوب :

أولا: احسب المقاييس التالية لكل من التوزيعين :

الوسـط الحسـابى – الوسـيط - المنـوال – الانحـراف المعيـارى – الانحـراف الربيعى .

ثانياً : قارن بين درجة تماثل كل من التوزيعين وذلك باستخدام الطرق التالية :

أ- إيجاد معاملات الالتواء المختلفة وذلك باستخدام المقاييس المحسوبة فى (أولا).

ب- تطبيق العلاقـات مـابين الانحـراف المعيـارى وبـين كـل مـن الانحـراف المتوسط ونصف المدى الربيعى والمتعلقـة بـالتوزيع المعتـاد مـع ابـداء التعليـق المنـاسب .

ج- حساب نسبة الموظفين الذين تتراوح مرتباتهم مايبن :

$\bar{X} \mp \sigma , \bar{X} \mp 2\sigma , \bar{X} \mp 3\sigma$

حيث :$\bar{X}$ ، $\sigma$ هما على الترتيب المسط الحسابى والانحراف المعيارى للمرتبات.

قارن بين قيم هذه النسب بما يجب ان تكون عليه فى حالة التوزيع المعتاد وهى 68.27% 95.45% 99.73%،مبديا ما تراه مناسب من تعلق .

٢- فيما يلى التوزيع التكرارى للدخل الشهرى لمائة اسرة:

| Income classes | 340- | 350- | 360- | 370- | 380- | 390- | 400- | 410- |
|---|---|---|---|---|---|---|---|---|
| No.06 families | 2 | 5 | 9 | 17 | 35 | 16 | 10 | 6 |

<u>والمطلوب</u> :

أ- رسم المنحنى المتجمع الصاعد ثم اشتقاق الوسيط والربيعين الادنى والاعلى ومن ثم حساب معامل الاختلاف الربيعى .

ب- قياس الالتواء والتفلطح لهذا التوزيع.

٣- استخدام بيانات أجور العمال الاسبوعية فى المصنعين ( أ، ب) لصناعة المنسوجات القطنية والوادة فى التمرين الأول من تمارين هذا الباب .

<u>المطلوب:</u>

<u>اولا :</u> باستخدام بيانات المصنع ( أ ) أحسب العزمين الثانى والثالث حول الوسط الحسابى .

<u>ثانيا :</u> قارن بين التواء توزيع الاجور فى كل من المصنعين وذلك باستخدام الحسابى مايلي :

- معاملا بيرسون الاول والثانى للا لتواء.
- معاملا الالتواء الربيعى.
- معاملا الالتواء المئينى .

***<u>ثالثا :</u>*** قارن بين التوزيعين من حيث الاختلاف النسبى .علق ما تحصل عليه من نتائج .

***<u>رابعا:</u>*** قياس التفلطح فى كل من التوزيعين .

٤- باستخدام بيانات درجات الطلاب فى كل من كليتى التجارة والتربية والواردة فى تمارين الباب الثالث .

**المطلوب :**

أ- قارن بين تشتت التوزيعين باستخدم المقياسين التاليين:

- نصف المدى الربيعى -الانحراف المعيارى

ب- باستخدام العزوم ، قارن بين تفلطح التوزيعين .

ج- قدر التواء كل توزيع واختلافه النسبى.

٥- البيانات التالية تمثل درجات ٨٠ طالبا فى مادة الاقتصاد:

| Degree classes | 40- | 50- | 60- | 70- | 80- | 90- |
|---|---|---|---|---|---|---|
| No.0f stndents | 5 | 12 | 32 | 21 | 7 | 3 |

**المطلوب :**

أ – حساب المقاييس التالية :

- الربيع الادنى والربع الاعلى ، الوسيط
- المئين العاشر والمئين التسعون.

ب –من نتائجك فى (أ) احسب المقاييس التالية :

- معانل الاختلاف - معامل الالتواء - معامل التفلطح

ج- احسب نسبة الطلاب الذين تتراوح دراجاتهم مابين : $\bar{X} \mp \sigma$

حيث $\bar{X}$ ، $\sigma$ هما على الترتيب الوسط الحسابى والانحراف المعيارى لهذا التوزيع .ثم بين كيف يمكن استخدام قيمة هذه النسبة كمؤشر على تماثل هذا التوزيع من عدمه.

**٦- فيما يلى التوزيع التكرارى لدرجات طلاب كليات التجارة موزعين حسب نظام الدراسة ،وذلك فى احد الامتحانات .**

| نظام الدراسة / فئات الدرجات | انتظام | انتساب موجه | المجموع |
| --- | --- | --- | --- |
| صفر- | ٣٨ | ٤٠ | ٧٨ |
| ١٤- | ١٢٦ | ٧٧ | ٢٠٣ |
| ٢٠- | ٢٠٤ | ١٠٧ | ٣١١ |
| ٢٦- | ٢٣٩ | ٥٧ | ٢٩٦ |
| ٣٢- | ٧٣ | ٥ | ٧٨ |
| ٣٦-٤٠ | ٢٤ | صفر | ٢٤ |
| المجموع | ٧٠٤ | ٢٨٦ | ٩٩٠ |

**<u>والمطلوب :</u>**

**أ- المقارنة بين تشتت درجات الطلاب فى كل من نظامى الدراسة مستخدما الانحراف المعيارى .**

**ب- احسب معامل الالتواء لكل من التوزيعين التكراريين ،ثم قارن بينهما**

**ج- قارن بين نسبة الطلاب الذين تتراوح درجاتهم مابين ٣٠،٢٠ درجة وذلك فى كل من تاتوزيعين (انتظام ،انتساب موجه ).**

**٩- الجدول التالى يوضح التوزيع التكرارى لدرجات الطلاب المنتظمين بكلية التجارة باحدى الجامعات فى دور مايو ١٩٨٧ م:**

| Degree classes | 0- | 10- | 20- | 30- | 40- | 60-50 | المجموع |
|---|---|---|---|---|---|---|---|
| No.of stndents | 4 | 98 | 169 | 192 | 163 | 33 | 702 |

**<u>والمطلوب :</u>**

**حساب مايلى :**

**أ- مقاييس التشتت التالية :**

**-الانحراف المعيارى. - نصف المدى الربيعى.**

**ب – نسبة رسوب الطلاب.**

**ج- عدد الطلاب الذين تتراوح درجاتهم مابين ٥٢،٣٥ درجة .**

**د – درجة ونوع التواء توةزيع الدرجات .**

**هـ - اذاكانت الدرجة النهائية لامتحان الاحصاء فى هذا الدور هى ٦٠ درجة ،احسب عدد الطلاب الذين حصلوا على تقدير جيد على الاقل إذا علمت ان عدد الطلاب حصلوا على تقدير جيد على الاقل هو ٢١٨ طالبا ،فسر ماقد تجده من اختلاف فى النتيجة التى حصلت عليها.**

## الباب السادس

## *الارتباط الخطى البسيط*

## *Simple linear correlation*

**فى دراستنا السابقة اقتصرت الدراسة حتى الان على سلوك متغير واحد او ظاهرة واحدة حيث تناولنا كيفية عرض وتبويب البيانات المتعلقة بهذا المتغير او تلك الظاهرة ثم حساب بعض المقاييس الاحصائية مثل مقاييس النزعة المركزية او التشتت بالاضافة للمقاييس الخاصة بتحديد درجة تماثل وتفرطح التوزيع الخاص بهذا المتغير او تلك الظاهرة محل الدراسة . الا انه فى كثير من الدراشات الطبيعية او الاقتصادية او الاجتماعية يكون الاهتمام منصبا على دراسة العلاقة بين ظاهرتين او اكثر بعرض الوقوف على طبيعة هذه العلاقة سواء من حيث معرفة درجة قوتها او ضعفها من ناحية او من حيث تعيين اتجاهها من ناحية اخرى .**

**فالباحث الاقتصادى قد تعنيه دراسة العلاقة ما بين حجم المبيعات من احدى السلع وتكاليف الدعاية والاعلان على هذه السلعة والباحث الزراعى قد يهتم بدراسة العلاقة ما بين كمية المنتج من محصول معين والانواع المختلفة من السماد وكمياته والباحث الاجتماعى قد يحتاج لدراسة العلاقة بين ادمان المخدرات بمرض السرطان او الايدز ....الى غير ذلك من الامثلة العديدة فى جميع فروع العلوم والمعرفة لمختلف الانشطة مما يؤدى بدوه الى اهتمام متزايد بدراسة الارتباط والانحدار .**

**هذا ويطلق على عملية دراسة العلاقة بين متغيرين او اكثر بعلاقة الارتباط حيث يهتم الباحث بدراسة ما اذا كانت هذه المتغيرات ترتبط بعلاقة خطية كما يهتم بعملية تحديد اتجاه تلك العلاقة اى تحديد ما اذا كانت علاقة الارتباط**

طردية ام عكسية ، اما عن تعيين شكل الصورة الرياضية التى تربط متغيرات الدراسة ببعضها البعض لاستخدامها فى التنبؤ بقيم متغير فى المستقبل فى ضوء مجموعة من القيم التى ياخذها متغير او مجموعة متغيرات اخرى تؤثر بشكل كبير على المتغير الاصلى او الظاهرة موضع الدراسة والتحليل فهو ما يعرف بتحليل الانحدار والذى سوف نقوم بدراسته فى الباب التالى . وتجدر الاشارة الى ان الارتباط والانحدار يختلفان فى الفرض القائم وقت الدراسة. فالارتباط هو عبارة عن دراسة للتغيير المشترك فيما بين متغيرين (او اكثر) كل منها ياخذ قيما مختلفة دون تدخل من الباحث. اى ان ازواج المشاهدات المتناظرة من المتغيرين هى بمثابة قيما عشوائية لا يعنينا عند دراسة الارتباط التمييز بين المتغيرين اى منها يعتبر متغيرا مستقلا وايهما يعتبر متغيرا تابعا اما الانحدار فيفرض وجود متغيراً واحداً تابعا يتاثر بواحدا او اكثر من المتغيرات المستقلة حيث يتدخل الباحث بتثبيت قيم المتغير او المتغيرات المستقلة عن مستويات معينة ( اى انها تقاس بدون اخطاء عشوائية ) ثم يشاهد او يسجل القيم التى ياخذها المتغير التابع عند تلك المستويات المحددة للمتغير او المتغيرات المستقلة والمحددة سلفا . ومن ازواج القيم المتناظرة فى الحالة الاولى يمكن تقدير معامل الأرتباط والذي يُستدل منه على إتجاه ودرجة قوة العلاقة بين المتغيرات موضع الدراسة . امافى الحالة الثانية فيمكن تعيين شكل العلاقة (المعادلة ) الرياضية التى تصف العلاقة ما بين المتغير التابع والمتغير او مجموعة المتغيرات المستقلة . وتستخدم المعادلة الناتجة فى التنبؤ بالقيم المستقبلية للمتغير التابع بدلالة قيم جديدة للمتغير او مجموعة المتغيرات المستقلة .

**وسوف نهتم فى هذا الباب بدراسة الارتباط الخطى البسيط بين متغيرين والارتباط ما هو الا ميل متغيرين ( او اكثر ) الى التغيير معا فهو يبين درجة قوة العلاقة بين متغيرين كما يبين اتجاه تلك العلاقة ( طردية كانت او عكسية ) ايضا.**

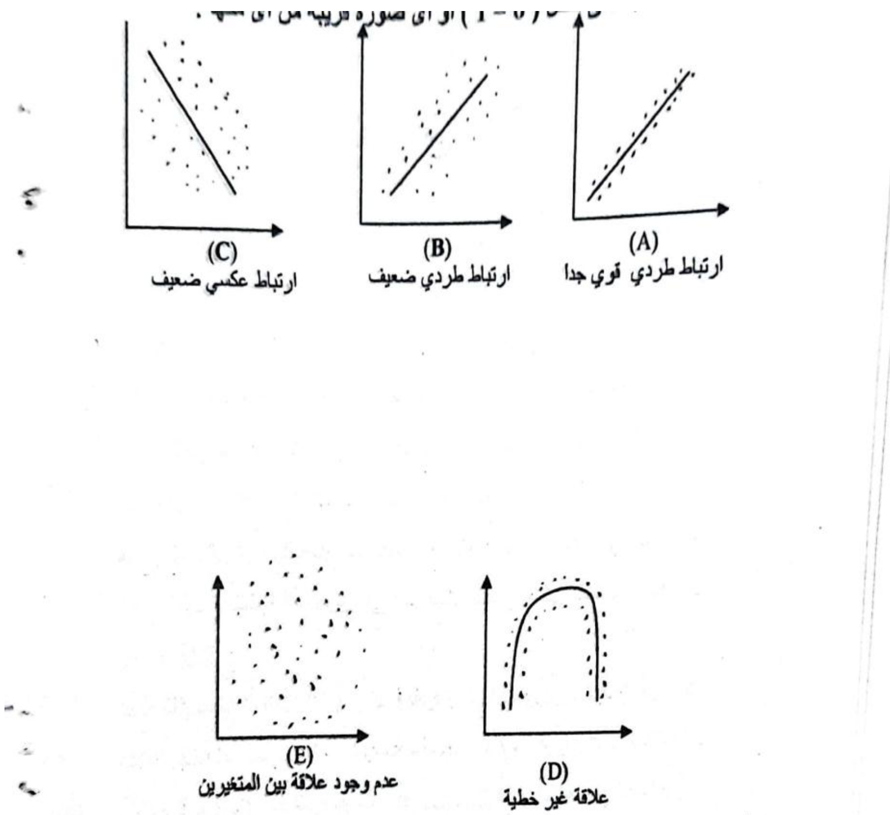
**فاذا كان لدينا متغيرين y, x وكان التغيير فى احد المتغيرين يتبعه تغير فى المتغير الاخر فى نفس الاتجاه بمعنى ان زيادة احدهما يؤدى الى زيادة فى قيمة المتغير الاخر او ان نقص احدهما يؤدى الى نقص فى قيمة المتغير الاخر فيقال فى تلك الحالة ان الارتباط بين المتغيرين طرديا ( اى موجب ) والعكس صحيح اى انه اذا كان التغيير فى احد المتغيرين يتبعه تغير فى قيمة المتغير الاخر فى الاتجاه المضاد بمعنى ان زيادة احدهما يؤدى الى نقص فى قيمة المتغير الاخر او ان نقص احدهما يؤدى الى زيادة فى قيمة المتغير الاخر فيقال فى هذه الحالة ان الارتباط بين المتغيرين عكسيا ( اى سالب ) .**

**هذا وقبل التعرض لكيفية حساب معامل الارتباط فانه من المفيد ان يتم عرض ازواج القيم المتنظرة للمتغيرين فى شكل بيانى يسمى بشكل االنتشار Scatter Diagram .**

**فاذا كان لدينا المتغيرين y , x وبفرض ان المتغيرين يرتبطان من خلال علاقة . فانه اذا سحبنا عينة عشوائية حجمها ( n ) من ازواج المشاهدات لقيم (Xi) و قيم (Yi) المناظرة حيث j=1,2,…….,n اى يصبح لدينا ازواج المشاهدات التالية :**

$$(X_1,Y_1)، (X_2,Y_2)، (X_3\ ,Y_3)، \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots X_n\ ,Y_n)$$

**فاذا تم تمثيل هذه الازواج من المشاهدات بيانيا فاننا نحصل على شكل الانتشار والذى يعطى للباحث صورة مبدئية واضحة عن العلاقة ما بين المتغيرين محل الدراسة ويساعده على اختيار النموذج القياسي الذى يناسب طبيعة البيانات موضع الدراسة. وهناك صوراً مختلفة لاشكال الانتشار نرصد منها الاشكال التالية والمبينة فى شكل ( ٦ – ١ ) او اى صورة قريبة من اى منها .**




**صورة مختلفة لاشكال الانتشار**

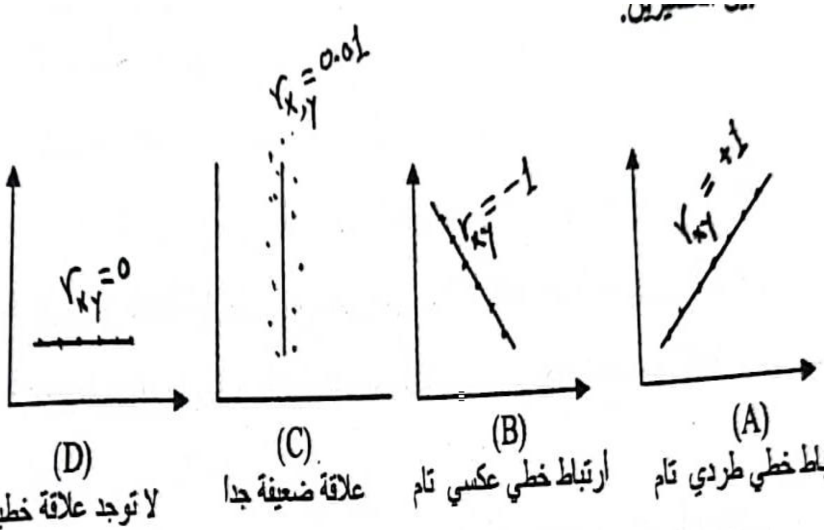
**(٦-١)**

**حيث يتضح لنا من خلال اشكال الانتشار الوارده فى الشكل (٦-١) ما يلى :**

- بالنسبة للشكل (A) فى شكل (٦-١) فانه يوضح لنا وجود علاقة طردية وقوية جدا ما بين المتغيرين Y,X بمعنى ان اى تغير فى احد المتغيرين وليكن مثلا ( X ) يصاحبة تغير فى قيمة المتغير(Y)وفى نفس الاتجاه . وكلما كانت نقاط شكل الانتشار التى تعبر عن ازواج المشاهدات($Y_i$ , Xi) قريبة من الخط المستقيم والذى يتوسط هذه النقاط فكلما دل ذلك على قوة الارتباط بين المتغيرين . والعكس صحيح فكلما ابتعدت النقاط عن الخط المستقيم الذى يتوسط تلك النقاط فان هذا يعنى ان الارتباط ضعيف كما هو فى الشكل ( B ) والذى يوضح وجود علاقة طردية ضعيفة بين المتغيرين. اما الشكل (C) فهو يوضح ان التغير فى اى من قيم المتغيرين يصاحبة تغير فى قيمة المتغير الاخر ولكن فى الاتجاه المضاد وهو ما يفيد وجود علاقة عكسية ضعيفة نظرا لبعد ازواج المشاهدات عن الخط المستقيم . والاشكال (C), (B),(A) الواردة فى شكل (٦-١) تمثل وجود علاقة خطية بين المتغيرين ويطلق على علاقة الارتباط فى هذه الحالة بانها ارتباطا خطيا حيث يمكن توفيق خط مستقيم يمر باكبر عدد ممكن من نقاط شكل الانتشار ويتوسط الباقى منها .
- اما الشكل (D) فهو يمثل علاقة غير خطية وكما يتضح من خلال هذا الشكل ان العلاقة تأخذ شكل منحنى ( درجة العلاقة بين المتغيرين اعلى من الدرجة الاولى ) ويكون الارتباط بين المتغيرين فى هذه الحالة ارتباطا غير خطيا .
- كما يوضح الشكل (E) من خلال نقاط شكل الانتشار ان النقاط لا تنتشر حول خط مستقيم او حول منحنة بل تنتشر هذه النقاط بشكل عشوائى (غير منتظم) حيث تنتشر النقاط تقريبا بالتساوى فى كل الاتجاهات مما يعنى عدم وجود علاقة بين المتغيرين . هذا وتوجد صورا اخرى لاشكال الانتشار منها التى تقع فيها كافة نقاط شكل الانتشار على خط مستقيم واحد او منحنى دون ان تنحرف

عنه اى نقطة فيعنى ذلك ان الارتباط تاما بين المتغيرين وهو ما يفيد أن كل التغيرات التي تنشأ في أي من المتغيرين يتم تفسيرها بواسطة التغيرات التى تنشأ فى المتغير الاخر كما يتضح فى شكل (٦-٢) فاذا وقعت جميع نقاط شكل الانتشار على خط مستقيم موجب الميل ( اى يصنع زوايه حادة مع الاتجاه الموجب للمحور الافقى ) كما هو فى الشكل ( ٦-٢-A ) فيعنى ذلك ان الارتباط بين المتغيرين فى هذه الحالة ارتباطا خطيا طرديا تاما .

اما اذا وقعت النقاط جميعا على خط مستقيم سالب الميل ( اى يصنع زوايه منفرجة مع الاتجاه الموجب للمحور الافقى ) كما هو موضح فى الشكل (٦-٢-B ) فان هذا يدل على ان الارتباط بين المتغيرين هو ارتباطا خطيا عكسيا تاماً بين المتغيرين.




صورة اخرى لاشكال الانتشار

شكل ( ٦-٢ )

**اما اذا تركزت نقاط شكل الانتشار حول خط مستقيم كما هو فى شكل (٦-٢-C) فان هذا يعنى وجود علاقة ضعيفة جدا بين المتغيرين .**

**وفى حالة وقوع جميع النقاط على خط مستقيم يوازى احد المحورين كما هو الحال فى شكل (٦-٢-D) فان هذا معناه انه لا توجد علاقة خطية بين المتغيرين اى ان معامل الارتباط الخطى بين المتغيرين فى هذه الحالة يكون مساويا للصفر .**

**واخيرا فانه بالرغم من ان اشكال الانتشار توضح ما اذا كانت هناك علاقة ما بين المتغيرين محل الدراسة وعما اذا كانت تلك العلاقة قوية ام ضعيفة الا اننا نحتاج دائما الى معيار او مقياس كمى يعبر عن درجة الارتباط بين المتغيرين وهو ما يسمى بمعامل الارتباط Correlation Cofficient وتتراوح قيمة معامل الارتباط ما بين 1+,1- فاذا كان الارتباط طرديا (موجبا) تاما بين المتغيرين Y,X حيث تقع جميع نقاط ازواج المشاهدات على خط مستقيم واحد ذو ميل موجب دون انحراف اى نقطة عن هذا الخط فان قيمة معامل الارتباط تكون مساوية للقيمة (1+) كما هو فى شكل (٦-٢-A). والعكس صحيح اذا كان الارتباط بين المتغيرين عكسيا (سالبا) تاما بحيث ان كافة نقاط ازواج المشاهدات تقع على خط مستقيم واحد ذو ميل سالب دون اى انحراف عنه كما هو واضح فى الشكل (٢-٦-B) فان قيمة معامل الارتباط تكون مساوية للقيمة (-1) .**

**اما ان لم تكن العلاقة بين المتغيرين تامة فان قيمة معامل الارتباط تحدد مدى قوة او ضعف العلاقة بين المتغيرين فكلما اقتربت قيمة معامل الارتباط من الواحد الصيح سواء موجبا او سالبا فان هذا يعنى قوة العلاقة بين المتغيرين . والعكس صحيح فكلما اقتربت قيمة معامل الارتباط من الصفر كلما دل ذلك على**

**ضعف العلاقة بين المتغيرين . واذا تساوت قيمة معامل الارتباط بالصفر معنى ذلك عدم وجود علاقة ارتباط او نزعة او تأثير بين قيم احد المتغيرات وقيم المتغير الاخر . اما عن اشارة معامل الارتباط فهى تبين اتجاه العلاقة بين المتغيرين حيث تكون الإشارة موجبة إذا كانت العلاقة طردية والعكس صحيح تكون الاشارة سالبة اذا كانت العلاقة عكسية بين المتغيرين .**

**هذا وعند دراستنا للارتباط فى هذا الباب فاننا سوف نقتصر فى هذه الدراسة على الارتباط الخطى البسيط بين متغيرين فقط .**

***(٦-١) العلاقة بين الظواهر ( المتغيرات ) :***

**تأخذ العلاقة بين المتغيرات صورا مختلفة ويتقف ذلك على طبيعة او نوعية البيانات وكذا عدد المتغيرات المطلوب دراسة العلاقة ما بينها وتتلخص هذه الصور فيما يلى :**

**١ - *علاقة الرتباط :***

**وهى العلاقة بين المتغيرات لبيانات يمكن قياسها رقميا ( كمية ) وينقسم الارتباط فى هذه الحالة الى عدة اقسام يتوقف كل منها على عدد المتغيرات المطلوب دراسة علاقة الارتباط فيما بينها :**

**وتتلخص أقسام علاقة الإرتباط فيما يلى :-**

أ- الارتباط الخطي البسيط : **وهو الذى يستخدم فى قياس العلاقة بين متغيرين فقط .**

ب- الارتباط المتعددة : **وهو يتناول العلاقة بين اكثر من متغيرين .**

**ج-** الارتباط الجزئى : **وهو يهتم بدراسة العلاقة بين متغيرين مع افتراض تثبيت ( او استبعاد) تأثير المتغيرات الاخرى .**

٢- ***الارتباط بين الصفات*** : اذا كانت احدى الظاهرتين المطلوب دراسة العلاقة ما بينها عبارة عن صفات ويعتذر قياسها كميا فان قياس العلاقة بين الظاهرتين يكون من خلال ما يلى :

أ- **علاقة الإقتران** :ويقيس الاقتران العلاقة بين ظاهرتين (متغيرين) احداهما على الاقل وصفيا على ان يتكون كل من المتغيرين من قسمين فقط .

ب- **علاقة التوافق** : ويقيس التوافق العلاقة بين ظاهرتين (متغيرين) احدهما على الاقل وصفيا وتحتوى احدى هاتين الظاهرتبن على أكثر من قسمين.

هذا وسوف نتناول بالتفصيل دراسة كل نوع من هذه العلاقات وذلك على النحو التالى فى هذا الباب .

***اولا : دراسة علاقة الارتباط :***

سبق وان ذكرنا حالا ان الارتباط البسيط يدرس العلاقة بين متغيرين وذلك من لال ما يسمى بمعامل الارتباط **Correlation Cofficient** وهناك صورا عديدة لمعاملات الارتباط ومع اختلاف هذه الصور الا انها تتفق جميعا فى الخصائص التالية :

***١ - معامل الارتباط بين y , x يساوى معامل الارتباط بين X , Y اى ان:***

$$r_{x,y} = r_{y,x}$$

وهذه الخاصية ناشئة عن الافتراض القائل بان القراءات المسجلة لقيم المتغيرين هى قراءات جمعت بطريقة عشوائية لا يتحكم الفرد فى اى منها.

٢- تتراوح قيمة معامل الارتباط ما بين 1+,1- اى ان $-1 \leq r_{xy} \leq +1$ .

وتزداد قوة العلاقة بين المتغيرين كلما اقتربت قيمة المعامل من 1+ أو 1- وتكون العلاقة طردية تامة حينما تكون قيمة $r_{x,y} = 1$ وعكسية تامة حينما تكون قيمة $r_{x,y} = -1$

٣- معامل الارتباط لايتاثر باى من العمليات الجبرية الاربعة ( الطرح او الجمع او الصرب او القسمة ) اى ان :

$r_{x,y} = r_{dx\,,\,dy} = r_{Dx\,,\,Dy}$

حيث :-

$d_{xi} = x_i - a$ ، $d_{yi} = y_i - a$ حيث (a) وسط فرضي بمثابة مقدار ثابت يتم طرحه من كافة المفردات سواء X أو Y كما سبق وان عرفنا الانحرافات المباشرة فى دراستنا للوسط الحسابى والانحراف المعيارى اما:

$D_{X_i} = \frac{X_i - a}{b}$ ، $D_{yi} = \frac{y_i - a}{b}$ حيث $D_i$ تعبر عن الانحرافات المعدلة السابق الاشارة اليها .

هذا وسوف نقوم بدراسة معاملات الارتباط التالية :

١- معامل الارتباط الخطى البسيط لبيرسون.

٢- معامل إرتباط الرتب لسبيرمان.

٣- معامل ارتباط كندال.

٤- معامل ارتباط فهنر.

١- <u>معامل الارتباط الخطى البسيط</u>

**Simple Linear Correlation Coefficient**

وينسب هذا المعامل الى كارل بيرسون **K. Pearson** وهناك اكثر من صورة رياضية لهذا المعامل وهو يقيس العلاقة الخطية ما بين متغيرين كميين سواء كانت البيانات مبوبة او غير مبوبة وذلك على النحو التالى :

- فى حالة البيانات الغير ميوية ( المفردات ) :

فاذا كان لدينا عدد (n) من ازواج المشاهدات للمتغيرين X , Y اى لدينا ازواج المشاهدات :

$(X_1,Y_1)$ ، $(X_2,Y_2)$، ... ... ... ... ... ... ... ... ... ... ... ... ... ... $(X_n\ ,Y_n)$

وكـان الوسـط الحسـابى للمتغيـر X هـو $\bar{X}$ والوسـط الحسـابى للمتغيـر Y هـو $\bar{Y}$وكان الانحـراف المعيـارى للمتغيـرين Y , X هـما $\sigma_x$، $\sigma_y$على الترتيب فـان معامـل الارتبـاط الخطـى البسـيط لبيرسـون $(r_{X_y})$ هـو عبـارة عن متوسط حاصل . ضرب القيم المعياريه للمتغير ين y.xأى أن :

$$r_{x,y} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma_x}\right) \left(\frac{y_i - \bar{y}}{\sigma_y}\right)$$

$$= \frac{\sum(X_i - \bar{X})\ (y_i - \bar{y})}{n\ \sigma_x\ \sigma_y}$$

$$= \frac{\sum(X_i - \bar{X})\ (y_i - \bar{y})}{n \sqrt{\frac{\sum(X_i - \bar{X})^2}{n} . \frac{\sum(y_i - \bar{y})^2}{n}}} \quad (1)$$

$$= \frac{\sum(X_i - \bar{X})\ (y_i - \bar{y})}{\sqrt{\sum(X_i - \bar{X})^2 . \sum(y_i - \bar{y})^2}} \quad (2)$$

والمقدار $\frac{\sum(X_i - \bar{X})\ (y_i - \bar{y})}{n}$ يسمى بالتغاير **Covariance** بين المتغيرين y.x ومن ثم فان كان معامل الارتباط الخطى البسيط هو عبارة عن خارج قسمة التغاير بين المتغيرين على حاصلا ضرب الإنحرافات المعياريه للمتغير.

هذا وعادة ما تكون قيم الاوساط الحسابية للمتغيرين قيا كسرية ممـا يـؤدى الـى صـعوبة استخدام صيغة المعادلـة (١) فـى حسـاب قيمـة معامـل الأرتبـاط بـين المتغيرين y. xوللتغلب على هذه المشكلة فإن يمكن استنتاج (دون الخوض فى

**الإثبات) الصيغة التاليه والتى يسهل استخدامها لحساب معامل الارتباط الخطى البسيط لبيرسون :**

$$r_{x,y} = \frac{\frac{\sum xy}{n} - \frac{\sum X}{n} \cdot \frac{\sum Y}{n}}{\sqrt{\left[\frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2\right]\left[\frac{\sum Y^2}{n} - \left(\frac{\sum Y}{n}\right)^2\right]}} \qquad (3)$$

**ووفقا لصيغة المعادلة (٣) فان يلزم لحساب معامل الارتباط الخطى البسيط المجاميع التالية:**

$$\sum x, \sum y, \sum x^2, \sum y^2, \sum xy$$

**وإنطلاقا من الخاصية الثالثة الخصائص معامل الارتباط والتى تفيد بان معامل الارتباط لايتتاثر بان ى من العمليات الجبريه الاربعة. فإنه يمكن بالاضافة للطريقة المباشرة والمستخدم فيها المعادلة (٣)السابقة فانه يمكن حساب معامل الارنباط الخطى البسيط سواء باستخدام طريقة الإنحرافات البسيطة او المعدله حيث تكون الصيغة على الصورة التالية :**

**صيغة معامل الارتباط الخطى البسيط لبيرسون فى حالة استخدام طريقة الإنحرافات البسيطة :-**

$$r_{x,y} = \frac{\frac{\sum d_x d_y}{n} - \frac{\sum d_x}{n} \cdot \frac{d_y}{n}}{\sqrt{\left[\frac{\sum d_x^2}{n} - \left(\frac{\sum dx}{n}\right)^2\right]\left[\frac{\sum d_y^2}{n} - \left(\frac{\sum dy}{n}\right)^2\right]}}$$

**حيث $d_y$. $d_x$ تعبر عن إنحرافات قيم ازواج المشاهدات عن وسط فرض معين سواء كان هذاالوسط الفرضى متساويا للمتغير ين او مختلفا.**

- **صيغة معامل الارتباط الخطى البسيط لبيرسون فى حالة استخدام طريقة الإنحرافات المعدلة:**

$$r_{x,y} = r_{D_x,D_y} = \frac{\frac{\sum D_x D_y}{n} - \frac{\sum D_x}{n} \cdot \frac{\sum D_y}{n}}{\sqrt{\left[\frac{\sum D_x^2}{n} - \left(\frac{\sum Dx}{n}\right)^2\right]\left[\frac{\sum D_y^2}{n} - \left(\frac{\sum Dy}{n}\right)^2\right]}}$$

***مثال :***

**فيما يلى بيان باسعار الفائده على راسى المال $(x_i)$وحجم المدخرات بالمليون جنيه $(y_i)$ باحد البنوك التجارية خلال سبع سنوات :**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $x_i$ | 8 | 9 | 8 | 11 | 10 | 13 | 18 |
| $y_i$ | 30 | 33 | 28 | 37 | 27 | 40 | 57 |

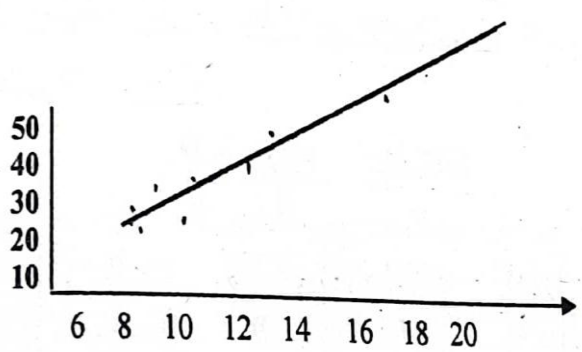
**والمطلوب :-**

**١- ارسم شكل الانتشار وعلق على نتائجك .**

**٢- احسب معامل الارتباط الخطى البسيط لبيرسون وذلك باستخدام الإنحرافات عن الوسط الحسابى ثم باستخدام القيم الاصلية وذلك باستخدام الطريقة المباشره وطريقة الانحرافات البسيطة حول وسط فرض خلاف قيمة الوسط الحسابى للمتغيرين .**

***الحل :***

**١- رسم شكل الأنتشار لأزواج المشاهدات وذلك على النحو التالى:-**




**ومـن خـلال هـذا الشـكل يتضـح لنـا ان نقـاط ازواج المشـاهدات تتركـز بدرجـه كبيرةحـول خـط مسـتقيم ذوميـل موجـب اى ان العلاقـة طرديـة وقويـة جـدا بـين المتغيرين (x) ، (y)**

**٢ـ حساب معامل الارتباط الخطى البسيط :**

**أ باسـتخدام إنحرافـات قـيم المتغيـرين عـن اوسـاطها الحسـابية اى باسـتخدام الصيغة التالية :**

$$r_{x,y} = \frac{\sum(X_i - \bar{X})\,(y_i - \bar{y})}{\sqrt{\sum(X_i - \bar{X})^2 . (y_i - \bar{y})^2}}$$

**لذا يلزم بداية حساب قيمةكل من $\bar{y}.\bar{x}$ فحيث ان :**

$$\bar{X} = \frac{\sum X}{n} = \frac{8+9+8+11+.........+18}{7} = \frac{77}{7} = 11$$

$$\bar{Y} = \frac{\sum Y}{n} = \frac{30+33+.........+57}{7} = \frac{252}{7} = 36$$

**ثم يتم تكوين الجدول التالى لاجراء الحسابات اللازمة لحساب معامل الارتباط:**

| $X_i$ | $Y_i$ | $(X_i - \bar{X})$ | $(Y_i - \bar{Y})$ | $(X_i - \bar{X})(Y_i - \bar{Y})$ | $(X_i - \bar{X})^2$ | $(Y_i - \bar{Y})^2$ |
|---|---|---|---|---|---|---|
| 8 | 30 | - 3 | -6 | 18 | 9 | 36 |
| 9 | 33 | -2 | - 3 | 6 | 4 | 9 |
| 8 | 28 | -3 | - 8 | 24 | 9 | 64 |
| 11 | 37 | 0 | 1 | 0 | 0 | 1 |
| 10 | 27 | -1 | -9 | 9 | 1 | 81 |
| 13 | 40 | 2 | 4 | 8 | 4 | 16 |
| 18 | 57 | 7 | 21 | 147 | 49 | 441 |
| 77 | 252 | 0 | 0 | 212 | 76 | 348 |

**وعليه فان معامل الارتباط الخطى البسيط هو :**

$$r_{x,y} = \frac{212}{\sqrt{(76)(648)}} \simeq 0.955$$

**وهو ما يفيد ان العلاقة طردية قوية جدا بين x ، y**

<u>(ب) حساب معامل الارتباط الخطى البسيط باستخدام القيم الاصليه -:</u>

**<u>١ ـ باستخدام الطريقة المباشره :-</u>**

**والجدول التالى يبين الحسابات اللازمة لحساب معامل الارتباط الخطى البسيط لبيرسون باستخدام الطريقة المباشره :-**

| $X_i$ | $Y_i$ | $X_i\ Y_i$ | $X_i$ | $Y_i$ |
|---|---|---|---|---|
| 8 | 36 | 240 | 64 | 900 |
| 9 | 33 | 297 | 81 | 1089 |
| 8 | 28 | 224 | 64 | 784 |
| 11 | 37 | 407 | 121 | 1369 |
| 10 | 27 | 270 | 100 | 729 |
| 13 | 40 | 520 | 169 | 1600 |
| 18 | 57 | 1026 | 324 | 3249 |
| 77 | 252 | 2984 | 923 | 9720 |

**وحيث ان :**

$$r_{x,y} = \frac{\frac{\sum xy}{n} - \frac{\sum x}{n} \cdot \frac{\sum y}{n}}{\sqrt{\left[\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2\right]\left[\frac{\sum y^2}{n} - \left(\frac{\sum y}{n}\right)^2\right]}}$$

**لذا فإن :**

$$r_{x,y} = \frac{\frac{2984}{7} - \frac{77}{7} \cdot \frac{252}{7}}{\sqrt{\left[\frac{923}{7} - \left(\frac{77}{7}\right)^2\right]\left[\frac{9720}{7} - \left(\frac{252}{7}\right)^2\right]}}$$

$$= \frac{30.2857}{\sqrt{[10.8714]\,[92.5714]}} = \frac{30.2857}{31.7235} \simeq 0.955$$

**وهى نفسى النتيجة السابقة**

**٢- باستخدام طريقة الإنحرافات البسيطة :**

**دعنا نقوم بحساب انحرافات قيم المتغير ( x )عن الوسط الفرضي وليكن القيمة (10)ونقوم بحساب انحرافات قيم المتغير (y) عن الوسط الفرضى (30)**

**اى ان** $d_x = x_i - 10$ **where** $i=1,2,\ldots,7$

$d_y = y_i - 30$ **where** $i=1,2,\ldots,7$

**ويتم تكون الجدول التالى لايجاد المجاميع اللازمه لحساب معامل الارتباط الخطى البسيط لبيرسون :**

| $x_i$ | $y_i$ | $d_x = x_i - 10$ | $d_y = y_i - 30$ | $d_x d_y$ | $d_x^2$ | $d_y^2$ |
|---|---|---|---|---|---|---|
| 8 | 30 | -2 | 0 | 0 | 4 | 0 |
| 9 | 33 | -1 | 3 | 3- | 1 | 9 |
| 8 | 28 | -2 | -2 | 4 | 4 | 4 |
| 11 | 37 | 1 | 7 | 7 | 1 | 49 |
| 10 | 27 | 0 | -3 | 0 | 0 | 9 |
| 13 | 40 | 3 | 10 | 30 | 9 | 100 |
| 18 | 57 | 8 | 27 | 216 | 64 | 729 |
| 77 | 252 | 7 | 42 | 254 | 83 | 900 |

**وحيث أن**

$$r_{x,y} = r_{D_x,D_y} = \frac{\frac{\sum D_x D_y}{n} - \frac{\sum D_x}{n} \cdot \frac{\sum D_y}{n}}{\sqrt{\left[\frac{\sum D_x^2}{n} - \left(\frac{\sum D_x}{n}\right)^2\right]\left[\frac{\sum D_y^2}{n} - \left(\frac{\sum D_y}{n}\right)^2\right]}}$$

**لذا فأن**

$$r_{x,y} = \frac{\frac{254}{7} - \frac{7}{7}\cdot\frac{42}{7}}{\sqrt{\left[\frac{83}{7} - \left(\frac{7}{7}\right)^2\right]\left[\frac{900}{7} - \left(\frac{42}{7}\right)^2\right]}}$$

$$= \frac{27.42857}{\sqrt{(10.85714)\,(92.57143)}} = \frac{30.28571}{31.7027} \simeq 0.955$$

وهو نفس ناتج الطريقتين السابقتين.

***مثال :***

فيايلى بيان عن الدخل الاسبوعى (x)والانفاق الاستهلاكى الاسبوعى(y) بالجنيه لعينه من خمسة اسر:-

| Weekly Income (x) | 50 | 60 | 80 | 100 | 120 |
|---|---|---|---|---|---|
| Weekly Expenditure(y) | 20 | 40 | 50 | 40 | 80 |

والمطلوب حساب معامل الارتباط الخطى البسيط مستخدما طريقة الإنحرافات المعدلة فى حسابه ؟.

***الحل :-***

لحساب معامل الارتباط الخطى البسيط لبيرسون باستخدام طريقة الإنحرافات المعدلة يتم تكوين الجدول التالي :

| $x_i$ | $y_i$ | $d_x = x_i - 80$ | $d_y = y_i - 40$ | $D_x = \frac{d_x}{10}$ | $D_y = \frac{d_y}{10}$ | $D_x^2$ | $D_y^2$ | $D_x D_y$ |
|---|---|---|---|---|---|---|---|---|
| 50 | 20 | -30 | -20 | -3 | -2 | 9 | 4 | 6 |
| 60 | 40 | -20 | 0 | -2 | 0 | 4 | 0 | 0 |
| 80 | 50 | 0 | 10 | 0 | 1 | 0 | 1 | 0 |
| 100 | 40 | 20 | 0 | 2 | 0 | 4 | 0 | 0 |
| 120 | 80 | 40 | 40 | 4 | 4 | 16 | 15 | 16 |
| | | | | 1 | 3 | 33 | 21 | 22 |

لاحظ انه :-

$$D_x = \frac{Xi - 80}{10} \quad , D_y = \frac{yi - 40}{10}$$

ومن خلال المجاميع المستنتجة فى الجدول السابق فان قيمة معامل الارتباط الخطى البسيط لبيرسون هى :

$$r_{x,y} = r_{D_x,D_y} = \frac{\frac{\sum D_x D_y}{n} - \frac{\sum D_x}{n} \cdot \frac{\sum D_y}{n}}{\sqrt{\left[\frac{\sum D_x^2}{n} - \left(\frac{\sum Dx}{n}\right)^2\right]\left[\frac{\sum D_y^2}{n} - \left(\frac{\sum Dy}{n}\right)^2\right]}}$$

$$= \frac{\frac{22}{7} - \frac{1}{7} \cdot \frac{3}{7}}{\sqrt{\left[\frac{33}{7} - \left(\frac{1}{7}\right)^2\right]\left[\frac{21}{7} - \left(\frac{3}{7}\right)^2\right]}} = \frac{3.081633}{\sqrt{(4.693878)(2.816326)}}$$

$$\simeq 0.848$$

وهذه النتيجة تفيد أن هناك علاقة طردية قوية بين المتغرين y,x.

## ***ملاحظات على معامل الأرتباط الخطى البسيط لبيرسون :***

١- معامل الارتباط الخطى البسيط لبيرسون يستخدم القيم الفعلية لبيانات المتغيرين فى الحساب ولاتختلف قيمتة باختلاف الطريقة المستخدمة فى حسابة . فكما اتضح لنا من الامثلة السابقة فإن معامل الارتباط لايتأثر بالعمليات الجبرية الاربعة .

٢- معامل الارتباط الخطى البسيط يقيس قوة العلاقة الخطية البسيطة بين المتغرين وهو على هذا الاساسى يحدد نسبة التغير فى احد المتغرين والتى يمكن تفسيرها على اساس وجود هذه العلاقة الخطية .ونسبة التغير هذه يتم قياسها عن طريق ما تسمى بمعامل التحديد Coefficient of Determination وهو عبارة عن مربع معامل الارتباط اى( $r^2$) ففى مثالنا السابق مثلا كانت قيمة معامل الارتباط فيابين الدخل الاسبوعى والانتقاق الاستهلاكى الاسبوعى هى 0.848 وهنا تبلغ قيمة معامل التحديد ($r^2$) القيمة 0.719104 (لاحظ ان $0.719104=(0.8480)^2$)) وهوما يعنى م ان% 71.9104 من التغيرات التى تنشافى قيمة الانقاق الاستهلاكى الاسبوعى يرجع فى أساسة للتفيرات التى تنشأ على قيمة الدخل الاسبوعى وذلك بموجب العلاقة الخطية التي تربط بينهما وان %28.0896 فقط من هذه التغيرات انما يرجع لعوامل اخرى متعلقة بالعلاقة الخطية فيما بين الدخل الاسبوعى و الانفاق الاستهلاكى الاسبوعى. وبعبارة اخرى يمكن القول أن التغير فى قيمة الانفاق الاستهلاكى الاسبوعى يرجع %71.91 منه الى التغير فى الدخل الاسبوعى وذلك بموجب العلاقة الخطية التى تربط مابين الدخل والانفاق الاستهلاكي الاسبوعى وان النسبة الباقية (%28.09)ترجع الى عوامل اخرى تؤثر على الانفاق الاستهلاكى الاسبوعى بخلاف الدخل الاسبوعى للاسرى.

٣- تتوقف اشارة معامل الارتباط (اى اتجاه العلاقة )على قيمة البسط فى صيغة هذا المعامل وذلك لان المقام والذى يعبر عنه بالإنحرافات المعيارية للمتغيرين موجب دائما .

٤- ***حساب معامل الارتباط الخطى لبسيط لبيرسون فى حالة البيانات المبوبه :-***

اذا كانت قيم ازواج المشاهدات الخاصة بالمتغيرين او الظاهرتين بين محل الدراسة كبيره بحيث يصعب معها اجراء العمليات الحسابية اللازمه لتحليلها احصائياً فأنه يجب فى هذه الحالة عرض البيانات الخاصة بالمتغيرين في صورة جدول تكراري مزدوج وهو ما سبق فى مطلع دراستنا. والجدول التكراري المزدوج ما هو إلا جدول تكراري إتجاهين حيث يتم فية تخصيص فئات احد المتغيرين فى الصفوف وفئات المتغير الثانى فى الاعمدة وتمثل القيم الموجودة بباقى خلايا الجدول التكرارات المشتركة المناظرة لفئات احد فى مقابل أحد فئات المتغيرالاخر . وانتشار التكرارات المشتركة داخل الجدول التكرارى المزدوج يعطى فكرة مبدئية عن طبيعة العلاقة بين المتغيرين (x،y) من حيث كونها طردية ام عكسية ، فتركيز التكرارات حول قطرى الجدول التكرارى المزدوج اما عن صيغة معامل الارتباط الخطى البسيط فى حالة البيانات المبوبة فهى نفس صيغ المعامل فى حالة البيانات المفردة مع ترجيح مراكز فئات كل توزيع من التوزيعين بالتكرارات المزدوجة.

وبصفة عامه فإنه لحساب معامل الإرتباط من البيانات المبوبة يتم اشتقاق التوزيعين التكراريين الهامشيين من الجدول التكراري المزدوج. ثم حساب المجاميع اللازمة لحساب الانحراف المعياري لكل توزيع هامشي ثم من خلال الجدول التكراي المزدوج واستبدال فئاته بمراكز الفئات او الانحرافات البسيطه او المعدلة عن اوساط فرضيه معينه يتم ايجاد المجموع المشترك ثم يتم

التعويض في احد الصيغ الثلاثة التالية لايجاد معامل الارتباط الخطي البسيط لبيرسون :

$$r_{x,y}=\frac{\frac{\sum\sum x_i y_j\ F_{xiyj}}{\sum F}-\left(\frac{\sum X_i F_x}{\sum F_x}\right)\left(\frac{\sum y_j F_y}{\sum F_y}\right)}{\sqrt{\left[\frac{\sum x_{x^2}\ F_x}{\sum F_x}-\left(\frac{\sum x_i f_i}{\sum x}\right)^2\right]\left[\frac{\sum y_j F_y}{\sum F_y}-\left(\frac{\sum y_j F_y}{\sum F_y}\right)^2\right]}}$$

حيث تمثل $y_i$, $x_i$ مراكز الفئتين (j),(i) للمتغيرين X،Y علي الترتيب.

أما $F_y$,$F_x$ فهي تعبر عن التكرارات الهامشيه لكل من y,x

أما $Fx_iy_j$ فتعبر عن التكرارات المزدوج لخلية تقاطع الفئتين i،j للمتغير x،y علي الترتيب .

وتمثل الصيغه السابق استخدام القيم الاصلية دون طرح او قسم اي صيغه الطريقة المباشرة لحساب معامل الارتباط . اما عن صيغة المعامل باستخدام طريقة الانحرافات البسيطة فهي نفس الصيغه السابقة مع الفارق الوحيد هو استبدال قيم المراكز ( Xi ) بالانحرافات ( $d_{xi}$ ) وكذلك استبدال قيم المراكز yj بالانحرافات البسيطة $d_{yj}$ . وكذلك صيغة طريقة الانحرافات المعدله هي نفس الصيغه لكن مع استبدال المراكز $x_i$ , $y_j$ بالانحرافات المعدلة $D_x$ , $D_y$ علي الترتيب .

***مثال :***

الجدول التالي يبين عدد سنوات الخبرة ( x ) والمرتب الشهري ( y ) لعينه من مائة عامل تم اختيارهم بطريقة عشوائية من احدي الشركات:

| Y / X | 150- | 180- | 210- | 240- | 270 - 300 | ∑ |
|---|---|---|---|---|---|---|
| 0- | 8 | 7 | | | | 15 |
| 5- | 2 | 7 | 10 | 6 | | 25 |
| 10- | | 1 | 19 | 10 | | 30 |
| 15- | | | 13 | 2 | | 15 |
| 20- | | | 8 | | 2 | 10 |
| 25- 30 | | | | | 5 | 5 |
| ∑ | 10 | 15 | 50 | 18 | 7 | 100 |

**والمطلـوب حسـاب معامـل الارتبـاط الخطـي البسـيط وفسـر مـا تـراه مناسـبا باستخدام معامل التحديد**

***<u>الحل :</u>***

**اذا نظرنـا للجدول التكراري المـزدوج نجد ان التكـرارات المزدوجـه تتركـز حـول القطر الرئيسي للجدول (من اعلي جهة اليسار حتي اسفل جهة اليسار) مما يدل بشكل مبدئي علـي وجود علاقـة طرديـة بـين المتغيـرين y , x وحيث ان طـوال الفئـات متسـاوية لكـل مـن المتغيـرين (اي ان التوزيعـات الـهاشـميه منتظمـه) لـذا يفضل استخدام طريقة الانحرافات المعدلة وذلك علي النحو المبين التالي :**

**<u>١ - ايجـاد التوزيعـات التكراريـة الـهامشـية وحسـاب المجـاميع اللازمـة لحسـاب الانحراف المعياري لكل توزيع :</u>**

**التوزيع التكراري الـهامش لعدد سنوات الخبره (x)**

| Classes | $F_i$ | $X_i$ | $D_x = x_i - 12.5$ | $D_X = \frac{d_X}{5}$ | $D_x F_x$ | $D_x^2 F_x$ |
|---|---|---|---|---|---|---|
| 0- | 15 | 12.5 | -10 | -2 | -30 | 60 |
| 5- | 25 | 7.5 | -5 | -1 | -25 | 25 |
| 10- | 30 | 12.5 | 0 | 0 | 0 | 0 |
| 15- | 15 | 17.5 | 5 | 1 | 15 | 15 |
| 20- | 10 | 22.5 | 10 | 2 | 20 | 40 |
| 25- 30 | 5 | 24.5 | 15 | 3 | 15 | 45 |
| | 100 | | | | -5 | 185 |

ومن هذا الجدول فان $\sum D_x F_x = -5$ ، $\sum d_x^2 F_x = 185$ اما عن التوزيع التكراري الهامشي للمرتب (y) فلحساب المجاميع اللازمة يتم تكوين الجدول التالي :

| Classes | $F_i$ | $Y_j$ | $d_y = Y_j - 225$ | $Dy = \frac{dy}{30}$ | $D_y F_{yj}$ | $D^2 y F y_j$ |
|---|---|---|---|---|---|---|
| 150- | 10 | 165 | -60 | -2 | -20 | 40 |
| 180- | 15 | 195 | 30- | -1 | -15 | 15 |
| 210- | 50 | 225 | 0 | 0 | 0 | 0 |
| 240- | 18 | 225 | 30 | 1 | 18 | 18 |
| 270-300 | 7 | 285 | 60 | 2 | 14 | 28 |
| $\sum$ | 100 | | | | -3 | 101 |

ومن هذا الجدول فاننا نجد ان : $\sum d^2 y F_y = 101$ . $d_y F_y = -3$

٣- اما لايجاد المجموع المشترك $\sum D_x D_y F_{xy}$ فيتم إستبدال فئات المتغيرين فى الجدول التكراري المزدوج بالإنحرافات المعدلة لهذه المتغيرات ثم حساب إيجاد حواصل ضرب إنحراف الصف (i) أي $D_{xi}$ فى تكرار الخلية المزدوجة $F_{xy}$ في العمود (j) اي Dyj ثم الجمع لكل هذه الحواصل وهو ما يوضحه الجدول التالي:

| Dx \ Dy | -2 | -1 | 0 | 1 | 2 | $\sum$ |
|---|---|---|---|---|---|---|
| -2 | 8 (32 | 7 (14 | | | | 46 |
| -1 | 2 (4 | 7 (7 | 10 (0 | 6 (-6 | | 5 |
| 0 | | 1 (0 | 19 (0 | 10 (0 | | 0 |
| 1 | | | 13 (0 | 2 (2 | | 2 |
| 2 | | | 8 (0 | | 2 (8 | 8 |
| 3 | | | | | 5 (30 | 30 |
| $\sum$ | 10 (36 | 15 (21 | 50 (0 | 18 (-4 | 7 (38 | 100 (91 |

ومن الجدول فإن المجموع المشترك $\sum D_x D_y F_{xy}$ هو (٩١) . معامل الارتباط الخطي البسيط لبيرسون هي :-

$$r_{x,y} = r_{D_x,Dy} = \frac{\frac{\sum D_x D_y F_{xy}}{\sum F} - \left(\frac{\sum D_x F_x}{\sum F}\right)\cdot\left(\frac{\sum D_y F_y}{\sum F}\right)}{\sqrt{\left[\frac{\sum D_x^2 F_x}{\sum F_x} - \left(\frac{\sum Dx\, F_x}{\sum F_x}\right)^2\right]\left[\frac{\sum D_y^2 F_y}{\sum F_y} - \left(\frac{\sum Dy\, F_y}{\sum F_y}\right)^2\right]}}$$

$$= \frac{\frac{91}{100} - \left(\frac{-5}{100}\right)\cdot\left(\frac{-3}{100}\right)}{\sqrt{\left[\frac{185}{100} - \left(\frac{-5}{100}\right)^2\right]\left[\frac{101}{100} - \left(\frac{-3}{100}\right)^2\right]}}$$

$$= \frac{0.9085}{\sqrt{[1.8475]\,[1.0091]}} = 0.665$$

اي ان العلاقة طردية متوسطة الي حد ما فيما عن عدد سنوات الخبرة والراتب الشـهري وهـو مـا يؤكـد صـحة التنبـؤ باتجـاه العلاقـه مـن خـلال شـكل تركيـز التكـرارات المزدوجـة علـي جـانبي القطـر الرئيسـي . امـا عـن معامـل التحديـد 0.4422=(0.665)وهو ما يفيد ان %44.22 من التغيرات التـي تنشأ فـي اي من المتغيرين يرجع لتغير الاخر والنسبة الباقية ترجع لعوامل اخري .

**ملاحظات على معامل الارتباط البسيط:**

١- عنـد حسـاب معامـل الارتبـاط الخطـي البسـيط لبيرسـون فـي التطبيقـات العمليـة يجـب ان يكـون عـدد القـيم المسـتخدمة علـي الاقـل (30 ) مفـرده لانـه فـي هـذه الحالة فـأن وجـود اي قيمـة شـاذه ( او حتـي اكثـر من قيمـة ) او غيـر دقيقـة لـن يكـون لهـا تـأثير كبيـر علـي قيمـة معامـل الارتبـاط الخطـي البسيط لبيرسـون لـذا ينصح اذا كان عدد القيم المستخدمه في حساب معامل الارتبـاط اقل من ثلاثـون مفرده ( والمفرده هـي زوج من المشـاهدات (y, x) ) فـأن وجـود قيمـة شـاذه سوف يكون لها تاثير كبيرا علي دقة معامل بيرسون للارتبـاط وفـي هذه الحالـة يفضـل البحـث عـن مقيـاس اخـر للارتبـاط يكـون اكثـر دقـه لان معامـل الارتبـاط الخطي البسيط بيرسون سوف يعطي نتائج مضللة في هذه الحالـة نتيجـة لتـأثره بالقيم الشاذه او المتطرفه. فعلى سبيل المثال لنوضح تلك النقطـة دعنـا نفرض ان لدينا البيانات – التالية عن المتغيرين y،x :

| | | | | | |
|---|---|---|---|---|---|
| x: | 1 | 2 | 3 | 4 | 5 |
| y: | 1 | 2 | 3 | 4 | 5 |

٢- يستخدم معامل الارتباط الخطي البسيط لقياس درجة قوة العلاقة الخطية بين متغيرين ، اي درجة تركز النقاط حول خط الانحدار Regression Line فكلما زادت القيمة المطلقة لمعامل الارتباط ($|r_{xy}|$) كلما زادت قوة العلاقة بين المتغيرين وبالتالي زادت درجة تركز النقاط حول خط الانحدار . فعلي سبيل المثال ، تكون قيمة ($|r_{xy}|$) مساوية الواحد الصحيح في الشكلين ٦-٢ / A،B بينما هي في الشكل (٦-١ / A ) اكبر منها في الشكل (٥-١ / C,B)

٣- تتوقف اشارة معامل الارتباط علي اشارة البسط والذي يمثل التغاير بين المتغيرين ، وذلك لان المقام دائما موجب حيث انه حاصل ضرب الانحرافات المعيارية $\sigma_x$ ، $\sigma_y$ في وكلاهما مقدار موجب .

٤- عند ايجاد معامل الارتباط لبيانات جدول تكراري مزدوج فأنه ليست هناك اية قيود او اشتراطات فيما يتعلق بطول الفئة او عدد الفئات في المتغيرين. فعلي سبيل المثال يمكن ان يكون عدد الفئات لأحد المتغيرين مختلفا عنه للمتغير الاخر ، كما يمكن ان تكون الفئات متساوية الطول لأحد المتغيرين وغير متساوية الطول للمتغير الاخر .

٥- أن إقتراب قيمة معامل الإرتباط الخطي البسيط من الصفر أو حتي مساواتها للصفر وأن إفاد بعدم وجود علاقة خطية بين المتغيرين ، إلا أنه لا يعني إستبعاد إمكانية وجود علاقة غير خطية بينهما . فعلي سبيل المثال فأن قيمة معامل الارتباط الخطي البسيط بين المتغيرينx .y في الشكل (D/1-5) تقترب من الصفر في حين ان هناك علاقة غير خطية قوية بين المتغيرين .

٦- عند حساب معامل الارتباط للبيانات المفردة ( غير المبوبة ) فانه اذا كان الوسط الحسابي لكل من المتغيرين x،y عددين صحيحين وكانت القيم كبيرة. حينئذ يفضل اتباع طريقة الانحرافات البسيطة باستخدام الوسطين الحسابين

كوسطين فرضين . وهذا يؤدى الى تبسيط العمليات الحسابية ،بالاضافة الى انه يمكننا التاكد من صحة قيم الانحرافات البسيطة حيث ان مجموعها لابد وان يساوى الصفر . وأما اذا كانت القيم كبيرة وكان الوسط الحسابى يحتوى على كسور فانه يفضل اتباع طريقة الانحرافات البسيطة باستخدام وسط فرضى ملائم . واخيرا فانه اذا كانت قيم المتغيرين صغيرة فانه يفضل إستخدام الطريقة المباشرة وذلك اختصارا للعمليات الحسابية .

٧- عند حساب معامل الارتباط لبيانات مبوبة فى جدول تكرارى مزدوج فانه اذا كانت فئات كل من المتغيرين متساوية الطول يفضل استخدام طريقة الانحرافات المعدلة .

٨- فى حالة حساب معامل الارتباطات لبيانات المتغيرين y،x والمبوية فى جدول تكرارى مزدوج ، فانه قبل اجراء أية عملية حسابية يمكن الاستدلال على طبيعة العلاقة بين المتغيرين وذلك على النحو التالى :

*اذا كانت التكرارات المشتركة داخل خلايا الجدول تتركزة حول قطر الجدول الممتد من اعلى ركن من جهة اليسار الى اسفل الجدول جهة اليمين فأن هذا يعنى ان هناك علاقة طردية بين المتغيرين وتزداد قوة هذه العلاقة بزيادة تركز التكرارات حول هذا القطر ، والعكس صحيح .

* اذا كانت التكرارات المشتركة داخل خلايا الجدول تتركزة حول قطر الجدول الممتد من اعلى اليمين الى ادنى اليسار ان هذا يعنى ان هناك علاقة عكسية بين المتغيرين وتتوقف درجة قوة او ضعف هذه العلاقة على مدى تركز هذه التكرارات حول هذا القطر .

٩- تلزم الاشارة الى نقطة هامة وهى : انه فى دراستنا للارتباط ،يجب علينا ان نفرق بين العلاقة الاحصائية – متمثلة فى معامل الارتباط بين المتغيرين –

وبين العلاقة السببية . فحين نجد ان هناك ارتباطا قويا بين متغيرين (اى قيمة $|r_{xy}|$ تقترب من الواحد الصحيح ) . فان ذلك لايعنى بالضرورة ان العلاقة بينهما هى بالضرورة علاقة سببية بمعنى ان التغير فى احد المتغيرين هو السبب فى التغير في قيمة الاخر . فعلى سبيل المثال : لو كنا بصدد ايجاد معامل الارتباط بين درجات مجموعة من الطلاب فى مادة الاحصاء وبين درجاتهم فى مادة القانون ، ووجدنا ان معامل الارتباط يساوى ٠.٩٥ فان هذا يعنى ان هناك ارتباطا قويا بين درجات الطلاب فى مادتى الاحصاء والقانون ، ولكن هل يعنى ذلك ان ارتفاع (او انخفاض) درجات الطلاب فى مادة القانون انما يرجع الى ارتفاع (او انخفاض ) درجاتهم فى مادة الاحصاء او العكس ؟ الاجابة بكل تاكيد هى : لا . وهنا يمكن القبول ان هناك علاقة احصائية قوية بين المتغيرين رغم انعدام العلاقة السببية بينها . ومن جهة اخرى ،قد تنعدم العلاقة الاحصائية – ممثلة فى معامل الارتباط – بين متغيرين فى الوقت الذى يوجد فيما بينهما علاقة سببية . فعلى سبيل المثال : لو كنا بصدد دراسة العلاقة بين سعر سلعة ما وكمية المباع منها ، ووجدنا ان سعر هذه السلعة ثابت لا يتغير فى حين ان كمية المباع منها فى زيادة مضطردة . فاننا لو رسمنا الشكل الانتشارى لهذين المتغيرين لحصلنا على خط مستقيم مواز للمحور الافقى ( راجع الشكل (٦-٢ ، D )) وذلك بافتراض ان المحور الافقى يمثل كمية المباع من السلعة وان المحور الراسى يمثل سعر هذه السلعة . وهذا فى حد ذاته يدل على ان معامل الارتباط بين سعر السلعة وكمية المباع منها يساوى الصفر . مما يعنى بدوره انعدام العلاقة الخطية بين المتغيرين . وفيما يتعلق بالعلاقة السببية بينهما . الا يمكننا ان نعتقد بان السبب فى ارتفاع كمية المباع من السلعة انما يرجع الى

ثبات سعرها . وهنا نستطيع القول ان هناك علاقة سببية بين المتغيرين رغم انعدام العلاقة الاحصائية بينهما .

واخيرا فان هذا لايجب ان يدعونا الى الاعتقاد بعدم امكانية وجود علاقة احصائية . بين متغيرين هى فى الواقع انعكاس لعلاقة سببية بينهما . فظاهرتا الطول والوزن على سبيل المثال قد يوجد بينهما علاقة احصائية وسببية ايضا . ولكن ما نريد ان نؤكد عليه هنا هو ان العلاقة الاحصائية بين متغيرين ليست بالضرورة علاقة سببية ، وان تحديد طبيعة العلاقة السببية – ان وجدت – انما هوامر يخرج عز نطاق علم الاحصاء والذى يهتم فقط بمجرد قياس العلاقة بين المتغيرات.

١٠- أنه أحياناً ما نصل إلى نتائج غير منطقية بمعامل الإرتباط الخطي البسيط لقياس العلاقة الخطية بين المتغيرين. بمعني أنه قد يكون من المعروف أن العلاقة الخطية بين متغيرين هى علاقة عكسية ، ولكن قيمة هذا المعامل أظهرت لنا أنها علاقة طردية أو العكس. وفي هذه الحالة يمكن تفسير هذا بأحد أمرين :-

<u>أولهما</u> : أن يكون هناك متغير ثالث يؤثر على المتغيرين (y،x ) موضع الإعتبار وذلك بالشكل الذي يؤدي إلى النتيجة غير المنطقية ولنضرب مثالاً لذلك.

ما هو تفسيرك عزيزي القارئ إذا ما حدث أن العلاقة بين كمية الإنتاج فى أحد المصانع (x) وعدد ساعات تعطل الآلات (y) وذلك خلال عشر سنوات هو 0.7 ؟ هل تستطيع القول حينئذ بأن هذه القيمة تفيد بأن هناك علاقة طردية ما بين كمية الإنتاج وعدد ساعات تعطل الآلات ؟ إن قلنا ذلك فإنه يعتبر منطقاً معكوساً ومخالفاً للواقع. لأنه إن صح أمر كهذا فما علينا إلا أن نوقف جميع الآلات حتى

نستطيع أن نحقق أكبر كمية ممكنة من الإنتاج . وهذا يبدو أمراً لا يقبلـه منطق أو يُقرُّه عاقل. إذ أن المنطقي أن العلاقة بين كميـة الإنتـاج وعدد سـاعات تعطل الآلات يجب أن تكون عكسية. إذا ما هو تفسير ما حصلناه عليه من نتيجـة غيـر منطقية هنا؟ وفي حالة كتلك ، يجب أن نفكر فى عامل(متغير) ثالـث يـؤثر علـى هذين المتغيرين (كمية الإنتاج ، عدد ساعات تعطل الآلات) وفـي نفس الاتجـاه. ومن بين التفسيرات المنطقية لذلك هو أن عدد الآلات يزداد عامـاً بعد آخـر فـى المصنع بشكل يؤدي إلى زيادة كمية الإنتاج كمـا يـؤدي أيضـاً وفـي نفس الوقـت إلى زيادة عدد ساعات تعطل الآلات. وبالتالي فإننا نجد أنه خـلال العشـر سنوات هناك – بشكل عام – زيادة فى كل من كمية الانتاج وعدد ساعات الأعطال ، مما يؤدي فى النهايـة إلـى أن معامـل الإرتبـاط الخطـي بينـهمـا يكـون موجبـاً وهـو مـا يخـالف المنطق. وفـي حقيقـة الأمـر فـإن العلاقـة العكسـية بـين المتغيـرين يمكن الحصول عليها إذا ما كان عدد الآلات في المصنع شبه ثابت على مدار السنوات العشر.

ثانيهما : أن العلاقة بين المتغيرين (y،x ) موضع الإعتبار هي علاقة ضعيفة ، بمعنـي أن معامـل الإرتبـاط الخطـي بينـهمـا يقتـرب فـي قيمتـه مـن الصـفر . وفـي حالات كتلك ، فإنه ليس من المستبعد أن نحصـل علـى مـا يفيـد بـأن هنـاك علاقـة عكسيـة بـين المتغيـرين فـي حـين أنـهـا يجـب أن تكـون طرديـة. وهـذا أمـر ممكـن ولكـن بشـرط أن تكـون العلاقـة الخطيـة بينـهمـا هـي علاقـة ضـعيفة فـى جميـع الحالات. حيث أنـه لا يضر أن نجد أن معامل الإرتبـاط الخطـي يسـاوي 0.05 فـي حين أنه يجب أن يكون 0.06- والعكس صحيح . ولكن الأمر غيـر المنطقـي هنـا أن تكون العلاقة بين المتغيرين طردية ضعيفة – علـى سـبيل المثـال – ونجـد أن $r_{xy} = -0.07$ أو العكس.

**١١- أنه فى حالة الحصول على علاقة خطية ضعيفة بين متغيرين (x . y) فإنه يجب عدم التسليم بذلك حيث يتم التأكد مما يلى:**

- **أنه لا يوجد قيم شاذه بين قيم المتغيرين ، حيث أن وجود مثل هذه القيم يؤدي إلى إضعاف هذه العلاقة ويمكن اكتشاف ذلك برسم الشكل الإنتشاري لقيم المتغيرين.**
- **أن ضعف العلاقة الخطية بين المتغيرين (y،x) قد يكون أمراً طبيعياً . ومع ذلك فإن هذا لا ينفي إحتمال أن تكون هناك علاقة قوية (بل قوية جداً) أو ربما تامة بين المتغيرين ولكنها علاقة غير خطية. ويمكن التعرف على ذلك من خلال الشكل الانتشاري أيضاً. حيث نجد أن انتشار نقاط المتغيرين لا تأخذ شكل الخط المستقيم ولكنها تأخذ صوراً أخري تعبر عن وجود علاقة غير خطية قوية بين المتغيرين.**

<u>ب – معامل فهنر للارتباط:</u>

**يبنى هذا المعامل على اساس ان العلاقة ما بين متغيرين كميين لا تعتمد على القيم الاصلية للمتغيرين وانما تعتمد اساساً على انحرافات قيم كل متغير عن وسطه الحسابى على حده .**

**فاذا كانت لدينا عدد (n) من ازواج المشاهدات لمتغيرين كميين Y,X ولتكن هذه الازواج من القيم هى ، $(y_1 - x_1)$ ، $(y_2 - x_2)$ ، .... $(y_n - x_n)$ فإنه لحساب معامل فهنر للأرتباط يتم أتباع الخطوات التالية :-**

- **حساب الوسط الحسابى لكل متغير اى$\bar{x}$ ، $\bar{y}$**
- **حساب انحرافات قيم كل متغير عن وسطه الحسابى اى يتم حساب قيم:**

$(x_i - \bar{x}), (y_{i-\bar{y}})$ For i = 1,2,…,n

- يتم حصر عدد الحالات التى تكون فيها ازواج الانحرافات للمتغيرين ذات اشارات متشابهة (موجبة كانت ام سالبة) ويمكن اجراء ذلك من خلال اعطاء رمز وليكن (S) لكل حالة من حالات تشابه تلك الانحرافات فى الاشارة فيكون مجموع حالات التشابه هو $\sum S$ .
- يتم حصر عدد الحالات التى تكون فيها ازواج الانحرافات للمتغيرين ذات اشارات مختلفة. ونرمز لكل حالة من تلك الحالات بالرمز (d). ومن ثم لاجمالي تلك الحالات بالمقدار $\sum d$ .
- يتم حساب معامل فهنر للارتباط $(r_f)$ والذى يأخذ الصورة التالية :

$$r_f = \frac{\sum s - \sum d}{n} \qquad where \quad n = \sum s + \sum d$$

حيث :

$(\sum S)$ تعبر عن عدد ازواج الانحرافات ذات الاشارات المتشايهة ،

$(\sum d)$ تعبر عن عدد ازواج الانحرافات ذات الاشارات المختلفة .

هذا وعند حساب معامل فهنر للارتباط فان قيم الانحرافات لا تعنينا فى حد ذاتها فى اى شيئ وانما يهمنا فى المقام الاول الاشارة التى يأخذها كل انحراف من هذه الانحرافات. ولذلك فانه لحساب هذا المعامل فاننا نقوم بحساب الوسط الحسابى لكل متغير ثم نحدد الاشارة التى يأخذها كل انحراف دون الحاجة الى ايجاد قيمة هذا الانحراف . حيث تكون اشارة الانحراف موجبة اذا كانت قيمته المفردة اكبر من الوسط الحسابى لهذا المتغير بينما تكون سالية اذا كانت قيمة المتغير تقل عن قيمة الوسط الحسابى لهذا المتغير والمثال التالى يوضح كيفية حساب معامل فهنر للارتباط .

***مثال:***

**البيانات التالية تمثل درجات عشرة طلاب فى مادتى الاحصاء والمحاسبة:**

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Statistic degree(x)** | **82** | **98** | **67** | **40** | **75** | **61** | **31** | **72** | **53** | **83** |
| **Accounting degree(y)** | **80** | **83** | **52** | **63** | **96** | **71** | **71** | **75** | **35** | **85** |

**والمطلوب حساب معامل فهنر للارتباط .**

***الحل :***

**لحساب معامل فهنر للارتباط $(r_f)$ فانه يجب فى البداية حساب الوسط الحسابى لكل من المتغير Y,X حيث نجد ان :**

$$\bar{x} = \frac{\sum x}{n} = \frac{82+98+\cdots+83}{10} = \frac{662}{10} = 66.2 \qquad degree$$

$$\bar{y} = \frac{\sum y}{n} = \frac{80+83+\cdots+58}{10} = \frac{711}{10} = 71.1$$

**ثم يتم بعد ذلك تحديد عدد الانحرافات المتشابهة وكذا المختلفة عن الوسط الحسابى لكل متغير كما يوضحها الجدول التالى :**

| X | Y | Difference Sign | | Case Type |
|---|---|---|---|---|
| | | x-$\bar{x} = x - 62.2$ | Y- $\bar{y} = y - 71.1$ | |
| 82 | 80 | + | + | S |
| 98 | 83 | + | + | S |
| 67 | 52 | + | - | D |
| 40 | 63 | - | - | S |
| 75 | 96 | + | + | S |
| 61 | 71 | - | - | S |
| 31 | 71 | - | - | S |
| 72 | 75 | + | + | S |
| 53 | 35 | - | - | S |
| 83 | 85 | + | + | S |

ومن خلال الجدول السابق فان :

$\sum S = 9 \ , \ \sum d = 1 \ , \ n = 10$

لذا فان معامل فهنر للارتباط هو :

$$r_f = \frac{\sum s - \sum d}{n} = \frac{9-1}{10} = 0.8$$

وحيث أن قيمة معامل فهز تقترب من الواحد الصحيح. لذا فأن هناك علاقة طردية قوية بين درجات الطلاب في مادتي الإحصاء والرياضيات

**٣- ارتباط الرتب : Rank Correlation**

قد تكون الدراسات الكمية فى بعض الاحيان وسيلة للحصول على ترتيب المفردات كما يحدث عند ترتيب اندية الدورى العام لكرة القدم حسب عدد النقاط التى يحصل عليها كل ناد . وفى احيان اخرى قد يتعذر قياس احدى الظاهرتين

على الاقل كما هو الحال فى الظواهر الوصفية او قد تكون الدقة العالية فى النتائج غير مطلوبة وفى مثل هذه الحالات يستعاض عن القيم العددية الحقيقية لكل من المتغيرين بالرتب التى يأخذها كل منها على اساس موقعه فى المجموعة التى ينتمى اليها .

فاذا اتفقت رتب قيم المتغير (x) مع رتب قيم المتغير (y) اتفاقا تاما اى لكافة قيم ازواج المشاهدات ( بمعنى ان اصغر قيمة فى قيم (x) تقابلها اصغر قيمة لـ (y) نكون بصدد حالة ارتباط طردى تام . والعكس صحيح فاذا كانت رتب (x) ، رتب (y) رتب عكسية تماما ( بمعنى ان اصغر قيمة لـ(x) تقابلها اكبر قيمة لـ(y) والقيمة البعد الصغرى لـ(x) تقابلها القيمة قبل الكبرى لـ(y) وهكذا الى ان نصل بان تكون اكبر قيمة من قيم (x) تقابلها اصغر قيمة فى قيم (y) فيكون الارتباط عكسيا وتاما. الا ان معظم التوزيعات تقع بين هذا وذاك اى قد لا يكون لا هو ارتباطا طرديا تاما ولا هو عكسيا تاما ولقياس أرتباط الرتب فهناك معامل ارتباط الرتب لسبيرمان ومعامل كندال وسوف نعرض هاذين المعاملين على النحو المبين التالى :

أ- معامل ارتباط الرتب لسبيرمان :

يعتمد معامل ارتباط الرتب الذى وضعه سبيرمان على ايجاد فروق الرتب للقيم المتاظرة للمتغيرين X , Y وهذه الفروق تتوقف قيمتها على مدى الاختلاف او الاتفاق فيما بين الرتب المتناظرة . فكلما كان الفرق بين هذه الرتب المتناظرة كبيرا دل ذلك على ضعف العلاقة بين المتغيرين والعكس صحيح كلما كانت فروق الرتب للقيم المتناظرة صغيرا دل ذلك على قوة العلاقة بين المتغيرين او الظاهرتين محل الدراسة. اما اذا كانت الرتب المتناظرة متساوية تماما ومن ثم تنعدم كافة عناصر فروق الرتب فعندئذ نكون بصدد حالة ارتباط

طردى تام . وعلى غرار ما تم بالنسبة لدراسة معامل الارتباط الخطى البسيط لبيرسون سوف نتعرض لدراسة معامل ارتباط الرتب لسبيرمان فى حالتى المفردات والبيانات المبوبة تكراريا .

- **<u>حساب معامل ارتباط الرتب لسبيرمان فى حالة البيانات الغير مبوبة :</u>**

اذا كان لدينا عدد (n) من ازواج المشاهدات للمتغيرين X , Y اى لدينا مجموعة القيم :( $x_1$ ، $y_1$)، ( $x_2$ ، $y_2$)،....( $x_n$ ، $y_n$)فأنه لحساب معامل الارتباط لسبيرمان يتم الاتي :

- صنع رتب تصاعدية او تنازلية لقيم كل متغير على حده على اساس موقع القيمة بالنسبة لجميع قيم المتغير الذى ينتمى له . فاذا تم وضع رتب تصاعدية والقيمة التى تليها فى الكبر تأخذ الرتبة (2) ، وهكذا الى ان نصل الى المفردة ذات القيمة القصوى لقيم المتغير لتأخذ الرتبة (n) وفى حالة تساوي قيمتين على الأقل من قيمة أحد المتغيرين يتم اعتبار الحسابى لتلك الرتب المتتالية لتوضع امام القيم المتساوية من قيم هذا المتغير ولتكن فى النهاية رتب (X) يرمز لها بالرمز $(R_x)$، رتب (Y) يرمز لها بالرمز $(R_y)$.
- يتم حساب فروق الرتب المتناظرة للمتغيرين ولتكن هذه الفرروق هى مجموعة القيم $D_i$ حيث i =1,2,.........,n حيث ان :

$$D_i = r_i - r_y$$

ويلاحظ ان بعض قيم ($d_i$) ستكون موجبة والبعض الاخر ستكون سالبة وقد تأخذ قيم صفرية . الا ان معيار صحة الحسابات هو ان يكون المجموع الجبرى لهذه الفروق مساوية للصغر . اى ان :

$$\sum_{i=1}^{n} D_i = 0$$

وحتى يتسنى التعبير عن مدى الاختلاف بين الرتب باستخدام الفروق ذاتها فانه لابد من التخلص من اشارات هذه الفروق وذلك من خلال ايجاد مربعاتها ثم الجمع . لذا يتم تربيع قيم $(D_i)$ای نوجد $\left(D_i^2\right)$ ثم بالجمع نحصل على $\sum D_i^2$ والذى يستخدم لبيان مدى الاتفاق والاختلاف بين الرتب المتناظرة .

- يتم حساب معامل ارتباط الرتب لسبيرمان من خلال الصيغة التالية :

$$r = 1 - \frac{6\sum D_i^2}{n(n^2 - 1)}$$

وهذه الصيغة لمعامل ارتباط الرتب مشتقاه من خلال حساب معامل الارتباط الخطى البسيط لرتب المتغيرين اى$R_x$، $R_y$

هذا وتنحصر قيمة معامل إرتباط الرتب لسبيرمان بين 1- , 1+ وعند حسابنا لمعامل ارتباط الرتب لسبيرمان بين متغيرين y , x نكون بصدد احد الصور الثلاثة التالية :

<u>الاولى</u>: عندما تكون الرتب المتناظرة لقيم المتغيرين متساوية تماما. وهو ما يؤدى الى تطابق الرتب ومن ثم تكون فروق الرتب صفريه لكافة قيم ازواج المشاهدات اى تصبح $D_i$0= لكافة قيم (i) , بالتالى فان $\sum D_i^2 = 0$ . وبذلك فان r=1 وهو ما يعنى ان العلاقة طردية تامة بين المتغيرين .

<u>الثانية</u> : اذا كانت الرتب المتناظرة لازواج المشاهدات عكسية تماما حيث تكون رتب قيم المتغير (x) هى : 2,1,........ ,(n-1),n تقابلها تماما رتب المتغير (y) والتى تأخذ الصورة 1,2,.......,(n-1),n وهنا تكون العلاقة بين المتغيرين عكسية تامة .

<u>الثالثة</u>: فى حالة عدم وجود علاقة ارتباط الرتب بين المتغيرين y , x فان قيمة معامل ارتباط الرتب لسبيرمان تكون صفرية .

**ملحوظة : ان المؤشر الحقيقى لتمام العلاقة الخطية ( سواء طردية تامة او عكسية تامة ) هو قيمة معامل الارتباط الخطى البسيط لبيرسون .**

**بمعنى اذا كانت هناك علاقة خطية بين متغيرين y , x فانه يمكن القول بانه اذا كانت العلاقة بينهما علاقة تامة. فان معامل الارتباط الخطى البسيط سيكون مساويا 1+ , 1- وفى هذه الحالة فانه معامل أرتباط الرتب لسبيرمان لابد ان يؤكد على تمام العلاقة كذلك . اى ستكون قيمة معامل ارتباط الرتب لسبيرمان مساوية ايضا 1-او 1+ . الا ان العكس غير صحيح بمعنى انه اذا كان معامل معامل ارتباط الرتب لسبيرمان مساويا 1+ أو 1- فان ذلك لا يعنى بالضرورة ان تكون العلاقة الخطية بينها تامة اى ليس بالضرورة ان تكون قيمة معامل الارتباط الخطى البسيط لسبيرمان مساوية ايضا ±1 .**

**هذا بالاضافة الى ان الاختلاف فى قيم المعاملين ( بيرسون وسبيرمان ) يرجع اصلا لاختلاف طريقة الحساب حيث يعتمد معامل الارتباط الخطى البسيط لبيرسون فى حساباته على قيم ازواج المشاهدات فى حين يعتمد معامل ارتباط الرتب لسبيرمان على رتب ازواج المشاهدات . وليس قيمها فاختلاف طرق الحساب من شأنه ان يؤدى الى اختلاف النتائج .**

***مثال:***

**أجرى احد اختبارات الذكاء لمجموع مكون من ثمانية افراد فى اعمار مختلفة فكانت نتيجة الاختبار على النحو التالى :**

| Age | 20 | 17 | 25 | 18 | 12 | 15 | 17 | 22 |
|---|---|---|---|---|---|---|---|---|
| Grade | V.Good | Good | Excellent | Good | Weak | Pass | V.Good | Good |

**والمطلوب دراسة ماهية وجود علاقة بين عمر الفرد ودرجة ذكائه فى هذا الاختبار باستخدام المقياس الاحصائى الملائم .**

***الحل :***

لاحظ ان احد الظاهرتين ( ظاهرة التقدير ) وصفية لذا لا يمكن حساب اى من معاملى الارتباط الخطى البسيط لبيرسون او فهنر للارتباط . لكن يمكن وضع رتب للظواهر سواء منها الكمية او الوصفية على حد سواء . لذا فان المقياس الاحصائى الملائم الذى يقيس ماهية وجود علاقة من عدمه فيما بين عمر الفرد والتقدير فى اختبار الذكاء هو معامل ارتباط الرتب لسبيرمان. ولحسابه دعنا نفترض ان العمر هو المتغير (x) والتقدير هو المتغير (y) ولحساب معامل سبيرمان يتم تصوير جدول الحسابات التالى:

| Age (x) | Grade(Y) | $R_X$ | $R_Y$ | $D_i = R_X - R_Y$ | $D_i^2$ |
|---|---|---|---|---|---|
| 20 | Very Good | 6 | 7 | 6-7 =-1 | 1 |
| 17 | Good | 3.5 | 5 | 3.5-5 = -1.5 | 2.25 |
| 25 | Excellent | 8 | 8 | 0 | 0 |
| 18 | Good | 5 | 5 | 0 | 0 |
| 12 | Weak | 1 | 2 | -1 | 1 |
| 15 | Pass | 2 | 3 | -1 | 1 |
| 17 | Very Weak | 3.5 | 1 | 2.5 | 6.25 |
| 22 | Good | 7 | 5 | 2 | 4 |
| $\Sigma$ | | | | 0 | 15.5 |

لاحظ ان هناك قيم مكرره فى قيم المتغيرين x , y فبالنسبة للمتغير (x) الذى يعبر عن العمر نجد ان القيمة 17 مكرره مرتان فلو لم تكن هاتين القيمتان مكررتان لكانت رتبهم هى 3 , 4 لذا تم اعتبار الوسط الحسابى للرتبتين

المتتـاليين اى $\frac{3+4}{2}$ وكـذلك بالنسـبة للمتغيـر (y) الـذى يعبـر عـن التقـدير فـان التقدير good مكرر ثلاث مرات . فلولا هذا التكرار لكانت رتب هذه المفردات هى 4 , 5 , 6 ومن ثم يتم اعتبار الوسط الحسابى لتلك الرتب الثلاث اى

$$\frac{4+5+6}{3} = 5$$

لذا يتم وضع الرتبة 5 امام كل تقدير "Good" كما بالجدول.

$$r = 1 - \frac{6\sum D_i^2}{n(n^2 -1)}$$

ومن خلال حسابات الجدول السابق فان معامل ارتباط الرتب لسبيرمان هو:

$$r = 1 - \frac{6\sum D_i^2}{n(n^2 -1)}$$

$$= 1- \frac{6\,(15.5)}{8(64-1)} = 1 - 0.1848 = 0.8155$$

وتدل قيمة معامل ارتباط الرتب لسبيرمان بـأن العلاقـة طرديـة وقويـة بـين عمـر الفرد ودرجة ذكائه .

- **حساب معامل ارتباط الرتب لسبيرمان فى حالة البيانات المبوبة :**

يتم حسـاب معامـل ارتبـاط الرتـب فـى حالـة البيانـات المبوبـة فـى جـدول تكـرارى مزدوج للمتغيرين x , y سواء كانت البيانات كمية أو وصفية حيث يـتم إعطـاء الفئـات رتبـاً تصـاعدية (أو تنازليـة) ويشـترط أن يكـون الجـدولين التكـرارين منتظمنين لكل من $X$ ، $\overline{Y}$ بمعنى ان كـون كل فئـات x متسـاوية الطول وكـذلك كـل فئـات y متسـاوية الطـول ، ويعتـذر اسـتخدام هـذه الطريقـة فـى حالـة عـدم تساوى اطوال الفئات لاحد المتغيرين او كليهما ويمتاذ هذا المعامل بانـه يصـلح لقياس الارتباط فى حالة الجداول التكرارية المفتوحة .

وجدير بالذكر ان حساب معامل ارتباط الرتب فى حالة البيانات المبوبة بالجدول التكرارى المزدوج يعطى نفس النتيجة التى تم التوصل اليها باستخدام معامل ارتباط بيرسون بالرغم من اننا نستخدم فى المعامل الاول رتب x , y بينما فى المعامل الثانى قيم x , y .

ويتم حساب معامل ارتباط الرتب لسبيرمان فى حالة البيانات المبوبة جدول تكرارى مزدوج باتباع الخطوات الاتية :

١- نعطى فئات المتغير x رتبا تصاعدية تبدأ بالوحدة اى 1,2....,n وكذلك نعطى فئات المتغير y رتبا تصاعدية بالارقام 1,2....,m حيث n, m يمثلان عدد فئات توزيع المتغير x وعدد فئات توزيع المتغير y على الترتيب .

٢- نرسم اقطار متوازية على الجدول التكرارى المزدوج نازلة من اعلى اليسار الى ادنى اليمين موازية للقطر الرئيسى بحيث تمر هذه الاقطار بجميع الخلايا التى بها تكرارات واذا كان القطر سوف يمر بمجموعة من الخلايا ليس بها تكرارات فلا داعى لرسمه . ثم يحسب الفرق بين رتبتى x , y لهذا الاقطار المتوازية حيث يكون هذا الفرق مساويا للصفر بالنسبة للقطر الرئيسى ويأخذ قيما مختلفة عن الصفر يالنسبة لباقى الاقطار ويلاحظ ان هذا الفرق بين رتبتى x , y يكون متساويا على جميع اجزاء كل قطر من هذه الاقطار لذلك تسمى هذه الطريقة احياناً بطريقة الاقطار ذات الفروق المتساوية .

٣- نكون جدول تكرارى بسيط لرتب x ونحسب منه الانحراف المعيارى لرتب x اى $\sigma_x$ وبنفس الطريقة نكون جدول تكرارى بسيط لرتب y ونحسب منه الانحراف المعيارى لرتب y اى $\sigma_y$.

٤- نكون جدول تكرارى لفروق الرتب وتكون تكرارية هى مجاميع تكرارات الاقطار السابقة تقابلها الفروق المختلفة بين رتب x ورتب y ونحسب منه تباين هذه الفروق ونرمز له بالرمز $\sigma_{d^2}$ ثم يحسب معامل ارتباط الرتب لسبيرمان بالصيغة التالية :

$$r = \frac{\sigma_x^2 + \sigma_{y-}^2 \sigma_{d^2}}{2\sigma_x\sigma_y}$$

حيث $\sigma_x^2, \sigma_y^2, \sigma_d^2$ هو تباين رتب x ، تباين رتب y تباين فروق الرتب بين x,y علي الترتيب. هذا ولتوضيح خطوات حساب معامل ارتباط الرتب لسبيرمان فى حالة البيانات المبوبة دعنا نقدم المثال التالى :

***مثال :***

الجدول التالى يمثل الانفاق الاستهلاكى (x) بالجنيهات والادخار (y) بالجنيهات فى احد الشهور لعينة مكونة من 200 اسره موزعين على النحو التالى :

| Saving CSlasses (y) / Exp. Classes (x) | 120- | 125- | 130- | 135- | 140-145 | $\sum$ |
|---|---|---|---|---|---|---|
| 280- | | | 5 | 25 | 10 | 40 |
| 290- | | | 40 | 20 | | 60 |
| 300- | | | 25 | 5 | 20 | 50 |
| 310- | | 30 | | | | 30 |
| 220-330 | 10 | 10 | | | | 20 |
| $\sum$ | 10 | 40 | 70 | 50 | 30 | 200 |

**والمطلوب حساب معامل ارتباط الرتب لسبيرمان بين الانفاق الاستهلاكى وبين الادخار للاسرة ؟**

***الحل :***

**بالنظر الى شكل توزيع التكرارات بالجدول التكرارى المزدوج نجد ان التكرارات تتركز حول القطر العكس ( من اعلى جه اليمين الى ادنى اليسار ) مما يكشف بصورة مبدئية عن وجود ارتباط عكسى بين المتغيرين محل الدراسة ولحساب معامل ارتباط الرتب لسبيرمان يتم حساب تباين الرتب لكل من توزيع المتغيرين كلا حده اولا : وذلك على النحو التالي:**

**فبالنسبة لجدول المتغير (x) فان :**

| Rank of (x) classes = $R_x$ | $F_x$ | $R_x F_x$ | $R_x^2 F_x$ |
|---|---|---|---|
| 1 | 40 | 40 | 40 |
| 2 | 60 | 120 | 240 |
| 3 | 50 | 150 | 450 |
| 4 | 30 | 120 | 480 |
| 5 | 20 | 100 | 500 |
| $\sum$ | 200 | 530 | 1710 |

**فيكون تباين فئات المتغير (x) هو :**

$$\sigma_x^2 = \frac{\sum R_x^2 F_x}{\sum F} - \left(\frac{\sum R_x F_x}{\sum F}\right)^2$$

$$= \frac{1710}{200} - \left(\frac{530}{200}\right)^2 = 1.5275 \qquad (L.E)$$

كذلك يتم حساب تباين رتب المتغير (y) والجدول التالى يعطى المجاميع اللازمة لحساب $\sigma_y^2$.

| Rank of (y) classes = $R_y$ | $F_y$ | $R_y$ $F_y$ | $R_y^2$ $F_y$ |
|---|---|---|---|
| 1 | 10 | 10 | 10 |
| 2 | 40 | 80 | 160 |
| 3 | 70 | 210 | 630 |
| 4 | 50 | 200 | 800 |
| 5 | 30 | 150 | 750 |
| $\sum$ | 200 | 650 | 2350 |

وعليه فان تباين رتب فئات المتغير (y) هو :

$$\sigma_y^2 = \frac{\sum R_y^2 F_y}{\sum F} - \left(\frac{\sum R_y F_y}{\sum F}\right)^2$$

$$= \frac{2350}{200} - \left(\frac{650}{200}\right)^2 = 1.1875 \qquad (L.E)$$

ثم استبدال فئات كل متغير فى الجدول التكرارى المزدوج بمجموعة الارقام التى تعبر عن رتب كل متغير على حده ثم نوجد الفروق ما بين تلك الرتب ولتكن $d_i = R_x - R_y$ لكل قطر من الاقطار المتوازية ونوجد مجموع التكرارات التى توجد على كل قطر على انها تكرارات لتلك الفروق المتناظرة والجدول المزدوج التالى يساهم فى حساب ذلك :

| Ry / RX | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | | | 5 | 25 | 10 |
| 2 | | | 40 | 20 | |
| 3 | | | 25 | 5 | 20 |
| 4 | | 30 | | | |
| 5 | 10 | 10 | | | |

| فروق الرتب $d_i = R_x - R_y$ | التكرار $F_d$ | $d_i F_d$ | $d_i^2 F_d$ |
|---|---|---|---|
| - 4 | 10 | - 40 | 160 |
| - 3 | 25 | -75 | 225 |
| - 2 | 45 | -90 | 180 |
| - 1 | 45 | -45 | 45 |
| 0 | 25 | 0 | 0 |
| 1 | 0 | 0 | 0 |
| 2 | 30 | 60 | 120 |
| 3 | 10 | 30 | 90 |
| 4 | 10 | 40 | 160 |
| $\sum$ 0 | 200 | -120 | 980 |

ومن ثم فمن المجاميع المبينية فان تبـأين فروق رتـب فئـات التوزيـع التكرارى المزدوج هو :

$$\sigma_d^2 = \frac{\sum d_i^2 F_d}{\sum F_i} - \left(\frac{\sum d_i F_d}{\sum F_i}\right)^2$$

$$= \frac{980}{200} - \left(\frac{-120}{200}\right)^2 = 4.54 \qquad (L.E)$$

وعليه فان معامل ارتباط الرتب لسبيرمان هو :

$$r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_d^2}{2\,\sigma_x\,\sigma_y} = \frac{1.5275 + 1.1875 - 4.54}{2\sqrt{(1.5275)(1.1875)}}$$

$$= -0.6775 \simeq -0.68$$

وهذه النتيجة تدل على ان هناك علاقة ارتباط عكسية قوية الى حد ما او متوسط فيما بين الانفاق الاستهلاكى (x) والادخار (y) للاسرة فى تلك العينة.

**ملاحظات حول معامل ارتباط الرتب لسبيرمان :**

١- يستخدم معامل ارتباط الرتب لسبيرمان لقياس العلاقة بين متغيرين سواء كانت القياسات ترتيبية اصلا او كمية ثم تم تحويلها الى قياسات ترتيبية بينما معامل ارتباط الخطى البسيط لبيرسون لا يصلح للاستخدام الا اذا كانت القياسات كمية للمتغيرين او الظاهرتين محل الدراسة .

٢- يتميز معامل ارتباط الرتب لسبيرمان بسهولة وسرعة حسابه الا انه يعتبر اقل دقة من معامل بيرسون للارتباط وذلك نظرا لاهمال معامل سبيرمان للقيم الحقيقية للمتغيرين واعتماده على رتب المتغيرين. حيث لا يوجد معنى طبيعى لحساب الفروق بين رتبتين ثم تربيع هذا الفرق . لذلك لا يفضل استخدام هذا المعامل الا فى الحالات التى لا تتطلب دقة عالية فى النتائج أي فى حالة تعذر حساب معامل بيرسون نظرا لان البيانات وصفية ويتعذر قياسها كميا بل يمكن ترتيبها فيما بينها ووفقا لخاصية معينة .

٣- عند حساب معامل ارتباط الرتب لسبيرمان فانه يلزم ترتيب المتغيرين y , x بنفس الطريقة بمعنى اذا رتبنا المتغير (x) ترتيبا تنازليا مثلا فانه يلزم حينئذ ترتيب المتغير (y) تنازليا ايضا والعكس صحيح فلا يصح ترتيب احد المتغيرين تنازليا مثلا وترتيب المتغير الاخر تصاعديا فى نفس الوقت .

**ب – معامل كندال للارتباط :**

راينا ان معامل ارتباط الرتب لسبيرمان بأنه يعاب عليه عدم وجود معنى طبيعى لحساب الفرق بين رتبتي المتغيرين ثم يتم ايجاد مربع هذه الفروق . لذا يوجد

**معامل لقياس ارتباط الرتب لكن يعتبر افضل من المعامل السابق يعرف بمعامل كندال لارتباط الرتب ومن ثم فان معامل كندال شأنه شأن معامل سبيرمان يعتمد على رتب قيم المتغير وليس على قيم المتغيرين الاصلية ويحسب معامل كندال لارتباط الرتب اذا ما توافرت لدينا ازواج من المشاهدات ولتكن :**

$(X_1,Y_1)$ , $(X_2,Y_2)$ , ............ , $(X_n,Y_n)$

**فانه يتم اتباع الخطوات التالية لحساب هذا المعامل :**

**١- ضع رتبا تصاعدية لكل من المتغير x , y .**

**٢- يتم كتابة رتب القيم المتناظرة للمتغيرين على ان تكون رتب احد المتغيرين وليكن (x) مكتوبة فى ترتيبها الطبيعى اى 1,2,3,....,n .**

**٣- باعتبار رتب قيم المتغير (y) رتبة بعد الاخرى ( يلاحظ ان رتب قيم المتغير (y) لا تأخذ الترتيب الطبيعى اللهم الا اذا كانت رتب القيم المتناظرة للمتغيرين اصلا متفقة تماما فيما بينها ) حيث يتم حساب ما يلى بالنسبة لكل رتبة من رتب (y):**

- **عدد الرتب التى تليها وتكون اكبر منها وسنرمز لهذا العدد بالرمز $N_t$ وذلك بالنسبة للرتبة (t) .**
- **عدد الرتب التى تليها وتكون اصغر منها وسنرمز لهذا العدد بالرمز Mt وذلك لنفس الرتبة (t) .**

**٤-يتم حساب الفرق ما بين $N_t$ وبين $M_t$ حيث ( t = 1,2,.....n) ثم نجمع ذلك بالنسبة لجميع رتب قيم (y) وسنرمز لهذا المجموع بالرمز (C) اى ان :**

$$C = \sum_{t=1}^{n}(N_t - M_t)$$

**٥-يتم حساب معامل كندال للارتباط ($r_k$) والذى يأخذ الصورة الرياضية التالية :**

$$r_k = \frac{2C}{n(n-1)}$$

هذا وتتراوح قيمة هذا المعامل بين 1+ , 1- اى ان :

$$-1 \leq r_k \leq 1$$

والمثال: التالى يوضح كيفية حساب معامل كندل لارتباط الرتب .

***مثال:*** احسب معامل كندال للارتباط وذلك لبيانات درجات الطلاب العشرة فى مادتى الاحصاء والمحاسبة والذى تم حساب معامل الارتباط لفهنر سابقا .

***الحل :*** الجدول التالى يوضح الحسابات اللازمة لايجاد معامل كندال للارتباط :

| Stat . degree(x) | Accounting degree(y) | $R_x$ | $R_y$ | Normal Ranking for($R_x$) and its corresponding ($R_y$) | | $N_t$ | $M_t$ | $C=N_t-M_t$ |
|---|---|---|---|---|---|---|---|---|
| | | | | $R_x$ | $R_y$ | | | |
| 82 | 80 | 8 | 7 | 1 | 4.5 | 5 | 3 | 5-3=2 |
| 98 | 83 | 10 | 8 | 2 | 3 | 6 | 2 | 4 |
| 67 | 52 | 5 | 2 | 3 | 1 | 7 | 0 | 7 |
| 40 | 63 | 2 | 3 | 4 | 4.5 | 5 | 1 | 4 |
| 75 | 96 | 7 | 10 | 5 | 2 | 5 | 0 | 5 |
| 61 | 71 | 4 | 4.5 | 6 | 6 | 4 | 0 | 4 |
| 31 | 71 | 1 | 4.5 | 7 | 10 | 0 | 3 | -3 |
| 72 | 75 | 6 | 6 | 8 | 7 | 2 | 0 | 2 |
| 53 | 35 | 3 | 1 | 9 | 9 | 0 | 1 | -1 |
| 83 | 85 | 9 | 9 | 10 | 8 | 0 | 0 | 0 |
| $\sum$ | | 55 | 55 | 55 | 55 | | | 24 |

وعليه فان معامل كندال للارتباط هو

$$r_k = \frac{2C}{n(n-1)} = \frac{2\,(24)}{10\,(10-1)} = \frac{48}{90} = 0.533$$

اى ان هناك علاقة طردية بين المتغيرين ولكنها ليست قوية .

**٤ - الارتباط بين الصفات (المبوبة تكراريا فى جداول تكرارية مزدوجة ):**

درسنا فى الاجزاء السابقة من هذا الباب كيفية قياس العلاقة بين الظواهر الكمية والتى يمكن قياسها رقميا ولكن قد يحدث أن تكون أحد الظاهرتين على الأقل عن صفات ويتعذر قاسها كمياً ( بمعنى ان احد الظاهرتين على الاقل وصفية ) ففى هذه الحالة لا نستطيع استخدام معامل الارتباط لقياس العلاقة بين الظاهرتبن وسندرس اثنين من هذه المقاييس هما معامل الاقتران ومعامل التوافق .

**أ- معامل الاقتران Coefficient of Association**

قدم يول **Yule** مقياسا لدراسة العلاقة بين ظاهرتين احداهما على الاقل وصفية يسمى بمعامل الاقتران ليول . وهذا المعامل يقيس العلاقة بين ظاهرتين احداهما على الاقل وصفية على ان تكون كل منها تنقسم الى قسمين فقط ( اى ان الجدول التكرارى المزدوج فى هذه الحالة يتكون من صفين فقط وعموديين فقط اى يتكون من اربعة خلايا فقط ) وفيما يلى نوضح كيفية حساب معامل يول للاقتران :لنفرض ان لدينا ظاهرتين تنقسم كل منها الى قسمين فقط فانه يمكن عرض النتائج المتعلقة بها على النحو المبين فى الجدول التالى :

| Second Phenomenon (y)<br>First Phenomenon (x) | First Part | Second Part |
|---|---|---|
| First Part | $F_{11}$ | $F_{12}$ |
| Second Part | $F_{21}$ | $F_{22}$ |

حيث $F_{11}, F_{12}, F_{21}, F_{22}$ تشير الى تكرارات خلايا الجدول التكرارى المزدوج حيث $F_{iJ}$ تشير الى تكرار الخلية التى تقع فى الصف (i) والعمود (j)حيث
$J=1,2$، $i=1,2$

فى هذه الحالة فان افضل مقياس لقياس الارتباط بين الظاهرتين x , y هو مامل يول للاقتران والذى سنرمز له بالرمز ($C_{Ass}$) والذى يأخذ الصورة التالية:

$$C_{ASS.} = \frac{F_{11}\,F_{22} - F_{21}\,F_{12}}{F_{11}\,F_{22} + F_{21}\,F_{12}}$$

وهذا المعامل (معامل يول للاقتران) شأنه فى ذلك شأن كافة قيم معاملات الارتباط حيث تنحصر قيمته بين 1 + , 1- فإذا كانت قيمة هذا المعامل تساوى (+1) دل ذلك على ان الاقتران بين متيرين x , y اقترانا تاما . اما اذا كانت قيمة هذا المعامل قيمة سالبة فانه يجب اهمال الاشارة السالبة لانه لا يصح ان يقال ان هناك علاقة طردية او عكسية بين ظاهرتين وصفيتين وعندما تنقدم قيمة هذا المعامل اى حينمما تساوى الصفر فان هذا يعنى عدم وجود علاقة بين المتغيرين او الظاهرتين محل الدراسة .

***مثال:***

فى بحث لدراسة العلاقة ما بين لون الشعر (x) ولون العينين (y) والتى شملت ٢٠٠ فردا تم اختيارهم عشوائيا تم الحصول على البيانات لتالية :

| Y / X | Black | Coloured | ∑ |
|---|---|---|---|
| Black | 110 | 40 | 150 |
| Yellow | 20 | 30 | 50 |
| ∑ | 130 | 70 | 200 |

والمطلوب دراسة ماهية وجود علاقة بين لون الشعر (x) ولون العينين (y) وذلك باستخدام المقياس الاحصائى الملائم .

***الحل :***

حيث أن الجدول التكرارى المزدوج للمتغيرين x , y مكون من صفين وعمودين اي يضم اربعة خلايا فقط ) لظاهرتين احدهما على الاقل وصفية لذا فان انسب معامل لقياس درجة العلاقة بين الظاهرتين هو معامل يول للاقتران ( $C_{ASS.}$ ) وهو :

$$C_{ASS.} = \frac{F_{11}\,F_{22} - F_{12}\,F_{21}}{F_{11}\,F_{22} + F_{12}\,F_{21}}$$

$$= \frac{110\,(30) - 20(40)}{110\,(30) + 20(40)} = 0.61$$

وهو ما يدل على ان هناك علاقة طردية وقوية الى حد ما ( متوسطة ) فيما بين لون شعر الفرد ولون عينيه .

<u>هذا ومن اهم اهم خصائص معامل يول للاقتران ما يلى :</u>

* اذا كانت احد التكرارات الاربعة $F_{11}$و$F_{12}$ او $F_{21}$ او $F_{22}$ على الاقل مساوية لصفر فان معامل يول للاقتران يكون حينئذ مساويا 1+ وتكون علاقة الاقتران تامة اما اذا كانت كل من التكرارات الاربعة اكبر من الصفر فان معامل الاقتران تتراوح قيمته بين 1+ , 1- وكلما اقتربت قيمة هذا المعامل من الصفر دل ذلك على ضعف العلاقة بين الظاهرتين والعكس صحيح .

* اذا كان : $F_{12}\,F_{21} = F_{11}\,F_{22}$ فان معامل الاقتران سيكون مساويا للصفر ويدل ذلك على عدم وجود علاقة بين الظاهرتين محل الدراسة والامثلة التالية توضح ذلك .

***<u>مثال:</u>***

**فى دراسة لايجاد العلاقة بين الاصابة بمرض القلب (x) والتدخين (y) تم جمع بيانات عن ١٢٠ شخصا فكانت النتائج كما هى موضحة فى الجدول التالى:**

| Smoking(y) / Illness disease | Smoke | Not smoke |
|---|---|---|
| Illness | 50 | 15 |
| Not Illness | 35 | 20 |

**والمطلوب قياس علاقة الاقتران بين التدخين والاصابة بالمرض مع التعليق على ما تحصل عليه من نتائج .**

***الحل:***

**حيث ان الجدول التكرارى يحتوى على ظاهرتين احدهما على الاقل وصفية ويتكون كل منها من قسمين فقط ( اى أن الجدول المزدوج من درجة 2x2 ) لذا فانه لقياس العلاقة بين الظاهرتين يكون ذلك من خلال معامل يول للاقتران وهو :**

$$C_{ASS.} = \frac{F_{11}\,F_{22} - F_{12}\,F_{21}}{F_{11}\,F_{22} + F_{12}\,F_{21}} = \frac{50\;(20) - 15\;(35)}{50\;(20) + 15\;(35)} = 0.311$$

**وهذا يدل على ان هناك علاقة ضعيفة ما بين التدخين والاصابة بمرض القلب ومن بيانات الجدول المعطى نستطيع القول ان التدخين يؤدى الى حد ما الى الاصابة بمرض القلب حيث نجد ان نسبة المصابين بالمرض من بين المدخنين تبلغ حوالى 59% بينما تبلغ تلك النسبة حوالى 43% بين غير المدخنين.**

***مثال:***

**الجدول التالى يوضح البيانات الخاصة بتوزيع 250 طفلا حسب الاصابة بمرض الحصبة (x) وعدم التحصين ضد هذا المرض (y) .**

| Vaccination(y) / Illness disease(X) | Not Vaccinate | Vaccinate |
|---|---|---|
| Illness | 40 | 18 |
| Not Illness | 12 | 180 |

**والمطلوب: حساب معامل الاقتران بين الاصابة بالمرض وعدم التحصين ضده مع التعليق على نتائجك .**

***الحل:***

**حيث ان معامل الاقتران :**

$$C_{ASS.} = \frac{F_{11}\,F_{22} - F_{12}\,F_{21}}{F_{11}\,F_{22} + F_{12}\,F_{21}} = \frac{40\,(180) - \;18\,(12)}{40\,(180) + \;15\,(12)} = 0.94$$

**اى ان هناك علاقة قوية ما بين الاصاب بمرض الحصبة وعدم التحصين ضده ومن بيانات الجدول المعطى نستطيع ان نستنتج ان الاطفال الذين لم يحصنوا ضد هذ المرض يكونون اكثر عرضه للاصابة به من اولئك الذين تم تحصينهم ضده حيث ان نسبة المصابين بالمرض من بين الذين تم تحصينهم قد بلغت 0.9% فقط فى حين ان هذه النسبة تبلغ 77% من بين الذين لم يحصنوا ضد هذا المرض .**

**هذا ويجب التنوية عن ما سبق تناوله من نقطة هامة وهى انه يؤخذ دائما بالقيمة الموجبة لمعامل الاقتران مهملين الاشارة السالبة للنتائج السالبة ان وُجدت . إذ أنه لا يصح ان يقال ان هناك علاقة طردية او عكسية بين ظاهرتين وصفيتين ( او احداهما على الاقل وصفية ) وذلك لان الاشارة تنتج عن جدولة**

البيانات فى الجدول التكرارى المزدوج . بمعنى انه اذا اعدنا ترتيب الجدولين السابقين فانه فىحالة التبديل لصف على آخر او عمود على آخر لنتج لنا معاملى الاقتران مساويا للقيم 0.311- , 0.94- للجدولين على التريب ) وهذا يعنى ان الاشارة السالبة فى قيمة معامل الاقتران انما تأتى نتيجة لطريقة عرض البيانات فى الجدول ولكنها لا تؤثر على طبيعة العلاقة ما بين الظاهرتين محل الدراسة .

**ب- معامل التوافق Contingency Coefficient**

درسنا فى المقياس السابق قياس علاقة الاقتران ما بين ظاهرتين احدهما على الاقل وصفية وكل منها يتكون من قسمين فقط . اما عن علاقة التوافق التى نحن بصددها الان فهى تبحث ايضا عن العلاقة بين ظاهرتين ايضا احدهما على الاقل وصفية ولكن تتكون احدهما ايضا على الاقل على اكثر من قسمين . فقد قدم لنا كارل بيرسون K.Pearson مقياسا احصائيا يقيس العلاقة بين هاتين الظاهرتين فى تلك الحالة يعرف باسم معامل التوافق لبيرسون ( اى ان هذا المعامل يستخدم حينما يكون الجدول التكرارى المزدوج يتكون من اكثر من اربعة خلايا ) .

فاذا كان لدينا ظاهرتين ولتكن الاولى (x) تنقسم الى عدد من (n) من الصفوف والظاهرة الثانية ولتكن (y) تتكون من عدد من الاقسام وليكن (m) فان جدول التوافق يكون على الصورة التالية :

| y / x | 1 | 2 | .......... | m | $\sum$ |
|---|---|---|---|---|---|
| 1 | $F_{11}$ | $F_{12}$ | ...... | $F_{1m}$ | $F_{1.}$ |
| 2 | $F_{21}$ | $F_{22}$ | ...... | $F_{2m}$ | $F_{2.}$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| n | $F_{n1}$ | $F_{n2}$ | ...... | $F_{nm}$ | $F_{n.}$ |
| $\sum$ | $F_{.1}$ | $F_{.2}$ | ...... | $F_{.m}$ | $F_{..}$ |

**ولحساب معامل التوافق يتم اتباع الخطوات التالية :**

**١- يتم تربيع تكرار كل خلية اى $(F_{iJ})^2$ حيث $i= 1,2,\ldots..,n$ ، $j=1,2,\ldots.,m$**

**٢- يتم قسمة مربع تكرار كل خلية على حاصل ضرب مجموع الصف والعمود الذى تنتمى اليها الخلية ثم يتم تجميع الناتج لجميع الخلايا وليكن هذا المجموع هو (C) . فيكون معامل التوافق هو :**

$$C_{Cont.} = \sqrt{\frac{C-1}{C}}$$

**حيث :**

$$C = \sum_{i=1}^{n}\sum_{J=1}^{m}\left(F_{iJ}^2 / F_{i}.F_{.J}\right)$$

**هذا ويأخذ معامل التوافق دائما ابدا قيمة موجبة ( ان لم تكن قيمة مساوية للصفر) كما ان الحد الاقصى ( اكبر قيمة يمكن ان يأخذها ) لمعامل التوافق يمكن تحديده من خلال العلاقة التالية :**

$$Maximum\ Value\ of\ C_{cont.} = \sqrt[4]{\frac{(n-1)(m-1)}{n.m}}$$

حيث (n) , (m) هى عدد اقسام الظاهرتين الاولى (الممثلة فى الصفوف ) والثانية (الممثلة فى الاعمدة) على الترتيب هذا وفى حالة تساوى عدد اقسام الظاهرتين ولتكن (n) لكل ظاهرة على حده فان الحد الاقصى لمعامل التوافق تكون قيمته هى :

$$Maximum\ Value\ of\ C_{cont.} = \sqrt{\frac{n-1}{n}}$$

ومن ثم فان الحد الاقصى لهذا المعامل يعتمد على عدد اقسام الظاهرتين محل الدراسة . وكلما اقتربت قيمة هذا المعامل من الصفر دل ذلك على ضعف العلاقة وكلما اقتربت قيمة هذا المعامل من قيمة الحد الاقصى له كلما دل ذلك على قوة العلاقة بين الظاهرتين. كما أن معامل التوافق لا يعطى شيئا على اتجاه العلاقة بين الظاهرتين حيث ان قيمته دائما ابدا موجبة (ان اختلفت عن الصفر) . الا ان طبيعة العلاقة امر يمكن استنتاجه من توزيع البيانات داخل جدول التوافق .

***مثال:***

الجدول التالى يوضح توزيع 80 اسرة حسب درجة تعليم الام وعدد الاطفال فى الاسرة وذلك بعد مدة زواج قدرها ستة سنوات :

| No. children (y) / Mother Education(x) | Less than 3 children | 3 childern and More | $\sum$ |
|---|---|---|---|
| Not Qualified | 2 | 10 | 12 |
| Preparatory | 3 | 6 | 9 |
| Secondary | 20 | 3 | 23 |
| University Qualification and more | 35 | 1 | 36 |
| $\sum$ | 60 | 20 | 80 |

**والمطلوب: قياس العلاقة ما بين درجة تعليم الام وعدد الاطفال فى الاسرة باستخدام المقياس الاحصائى الملائم وعلق على نتائجك .**

***الحل:***

**حيث ان الجدول التكرارى المزدوج يحتوى على ظاهرتين احدهما على الاقل وصفية وتتكون من اكثر من قسمين . لذا فان المقياس الاحصائى الملائم لدراسة العلاقة ما بين الظاهرتين محل الدراسة هو معامل التوافق لبيرسون ولحسابه يتم الاتى :**

$$C = \sum_i \sum_J \left(F_{ij^2} / F_i . F_{.J}\right)$$

$$= \frac{(2)^2}{12 \times 60} + \frac{(10)^2}{12 \times 20} + \frac{(3)^2}{9 \times 60} + \cdots$$

$$\dots . + \frac{(1)^2}{36 \times (20)} = 1.519$$

$$C_{Cont.} = \sqrt{\frac{c-1}{c}} = \sqrt{\frac{1.519}{1-1.519}} = 0.58$$

**وحيث ان الحد الاقصى لمعامل التوافق فى هذا المثال هو :**

$$maximum\ Value\ of\ C_{cont.} = \sqrt[4]{\frac{(n-1)(m-1)}{n.m}}$$

$$= \sqrt[4]{\frac{(2-1)(4-1)}{2(4)}}$$

**والنسبة ما بين قيمة معامل التوافق والحد الاقصى له :**

$$\frac{0.58}{0.78} = 0.74$$

**وهو ما يعنى ان هناك علاقة قوية بين درجة تعليم الام (الزوجة) وبين عدد الاطفال فى الاسرة. واما عن اتجاه العلاقة فهو امر يمكن استنتاجه من بيانات الجدول المعطى . حيث يمكن ملاحظة ان عدد الاطفال فى الاسرة يقل كلما زادت درجة تعليم الزوجة والعكس صحيح .**

## تمارين الباب السادس

١- البيانات التالية تمثل اعمار مجموعة من ١٠ أزواج وزوجات .

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Male Age:** | **23** | **27** | **28** | **29** | **30** | **31** | **33** | **35** | **36** | **39** |
| **Female Age:** | **18** | **22** | **23** | **24** | **25** | **26** | **28** | **22** | **30** | **32** |

والمطلوب ايجاد معامل الارتباط بين العمرين .

٢- البيانات الاتية تمثل الكميات المطلوبه من سلعة معينة وسعر هذه السلعة :

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Demand** | **16** | **10** | **8** | **5** | **4** | **2** | **1** |
| **Price** | **3** | **4** | **6** | **5** | **7** | **8** | **11** |

والمطلوب ايجاد معامل الارتباط بين الطلب والسعر .

٣- البيانات التالية تمثل الكميات المعروضة من سلعة معينة وسعر هذه السلعة:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Supply** | **4** | **5** | **8** | **9** | **10** | **12** | **15** |
| **Price** | **3** | **4** | **5** | **6** | **7** | **8** | **12** |

والمطلوب ايجاد معامل الارتباط بين العرض والسعر .

٤- اذا كانت $x$ ترمز الى سعر السلعة، $y$ ترمز الى الكمية المعروضة عند كل ثمن فالمطلوب ايجاد العلاقة بين هاتين الظاهرتين :

| X | 2 | 5 | 7 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|
| Y | 100 | 130 | 150 | 200 | 220 | 210 | 250 | 240 | 300 | 400 |

**٥- فيمـا يلـى الـدرجات التـى يحصـل عليهـا ١٠ طـلاب فـى مـادتى الرياضـة والاحصاء :**

| Math. | 12 | 15 | 8 | 10 | 11 | 10 | 7 | 17 | 6 | 5 |
|---|---|---|---|---|---|---|---|---|---|---|
| Stat. | 10 | 11 | 7 | 12 | 9 | 8 | 7 | 14 | 2 | 4 |

**والمطلوب ايجاد معامل سبيرمان لارتباط رتب المادتين .**

**٦- فيمـا يلـى الدرجات التى يحصـل عليهـا ١٠ طـلاب فـى امتحـان مـادتى اللغـة الانجليزيـة واللغة الفرنسية .**

| Engl. | 50 | 24 | 60 | 68 | 66 | 39 | 44 | 34 | 49 | 76 |
|---|---|---|---|---|---|---|---|---|---|---|
| Fren. | 59 | 38 | 65 | 71 | 68 | 50 | 53 | 44 | 52 | 77 |

**والمطلوب ايجاد معامل ارتباط الرتب .**

**٧- احسب باستخدام البيانات التالية معامل سبيرمان للارتبـاط بـين الظـاهرتين y ، x**

| X | 17 | 14.3 | 11.7 | 10.3 | 11.3 | 12.5 | 9.7 | 10.8 | 10.4 | 16.2 |
|---|---|---|---|---|---|---|---|---|---|---|
| Y | 703 | 702 | 767 | 801 | 773 | 653 | 709 | 724 | 720 | 571 |

٨- فى احدى الدراسات الاحصائية عن رأى مجموعتين من السائحين فى ستة فنادق بهدف معرفه الافضليه من حيث حسن التعامل مع السائح فكانت نتجة الدراسة كما يلى :

| الفندق | هيلتون | شيراتون | سونستا | موفنبيك | ميريديان | ماريوت |
|---|---|---|---|---|---|---|
| المجموعة الاولى | ٢ | ٥ | ١ | ٤ | ٣ | ٦ |
| المجموعة الثانية | ٥ | ٤ | ٣ | ٢ | ١ | ٦ |

والمطلوب ايجاد الارتباط ( معامل الرتب ) بين افضلية المجموعتان .

٩- فيما يلى اسعار السلع الجلدية والخزف ( للوحدة الواحدة ) خلال عشرة ايام فى منطقة خان الخليلى

السلع الجلدية ٥ ٧ ٨ ٤ ٩ ٣ ٢ ٥ ٤ ٣

السلع الجلدية ٢ ٤ ٥ ٥ ٦ ٥ ٤ ٤ ٣ ٢

والمطلوب ايجاد معامل بيرسون وايضا معامل ارتباط سبيرمان ( الرتب ) ومقارنة قيمة معاملى الارتباط وبيان سبب الاختلاف فى قيمتها أن وجد .

١٠- يوضح الجدول التالى توزيع عينة بحسب فئات دخولهم الشهرية x وإيجار المسكن الذى تقطنه بالجنيهات (y):

| x \ y | 10- | 20- | 30- | 40- | 50-60 |
|---|---|---|---|---|---|
| 50- | 4 | 3 | 3 | | |
| 75- | 2 | 5 | 9 | 4 | |
| 100- | | 6 | 15 | 10 | 4 |
| 125- | | 1 | 7 | 12 | 2 |
| 150-175 | | | 5 | 7 | |

والمطلوب : حساب معامل الارتباط بين المتغيرين : X ، Y .

اذا كان لدينا البيانات الاتية عن كلا من الظاهرة X ، Y .

$\sum XF_X = 39000$ ، $\sum X^2F_X = 1605000$

$\sum y_{F_y} = 29950$ ، $\sum_{y^2F} = 919750$

$\sum XYF_{XY} = 1197250$ ، $\sum F_X = \sum F_Y = \sum F_{XY}$

والمطلوب : ايجاد:

١- الوسط الحسابى والانحراف المعيارى ومعامل الاختلاف للظاهرتين

قارن تشتت الظاهرتين X ، Y .

٢- حساب معامل الارتباط بين X ،Y .

١١- فيما يلى بيانات عن عينة من نزلاء فندق سونستا بالقاهرة حسب أعمارهم (X) وعدد الليالى السياحية التى قضوها بالقاهرة y

| X \ Y | 2 | 3 | 5 | 6 | $\sum$ |
|---|---|---|---|---|---|
| 20- | 1 | 2 | 2 | - | 5 |
| 25- | 2 | 3 | 3 | - | 8 |
| 30- | - | 10 | 5 | 5 | 20 |
| 40- | 2 | 3 | 6 | 1 | 12 |
| 50-60 | - | 2 | 2 | 1 | 5 |
| $\sum$ | 5 | 19 | 18 | 7 | 50 |

والمطلوب :

**ايجاد قوة العلاقة ونوعها بين اعمار نزلاء الفندق و عدد الليالى السياحة وعلق على النتيجة .**

**١٢- ارسم الشكل الانتشارى ثم أوجد معامل الارتباط الخطى البسيط بين المتغيرين x , y وذلك من واقع البيانات التالية :**

**X: ١٢، ١٨، ٢١، ٢٤، ٢٨، ٣٢، ٤٥، ٥٦، ٦٢، ٧١**

**Y: ١١، ١٣، ٢٢، ٢٣، ٣٠، ٣٥، ٤٢، ٥٠، ٦٠، ٦٥**

**١٣- فيما يلى الدخل الاسبوعى (x) والانفاق الاسبوعى (y) لعينة مكون من ٨ أسر**

**X: ٤٣، ٥٨، ٥٥، ٦٢، ٨١ ، ٧٢، ٩٨، ٧٦**

**Y : ٣٥، ٥٥ ، ٥٣ ، ٦١ ، ٧٩، ٤٥، ٥٠، ٦٢**

**والمطلوب : حساب معاملات الارتباط التالية :**

- **معامل الارتباط الخطى البسيط .**
- **معامل ارتباط الرتب ( معامل سبيرمان ).**
- **معامل ارتباط كندال .**
- **معامل ارتباط فهنر .**

**١٤- البيانات التالية تبين تقديرات نجاح عشر طلاب فى مادتى الاحصاء (x) والرياضيات (y).**

| الطالب | تقديرات (X) | تقديرات(Y) |
|---|---|---|
| ١ | مقبول | جيد |
| ٢ | جيد | جيد |
| ٣ | مقبول | جيد جداً |
| ٤ | ممتاز | مقبول |
| ٥ | جيد جداً | جيد جداً |
| ٦ | جيد جداً | جيد |
| ٧ | جيد | مقبول |
| ٨ | ممتاز | جيد جداً |
| ٩ | مقبول | ضعيف |
| ١٠ | ضعيف | ضعيف جداً |

**والمطلــوب : قيــاس العلاقــة بــين تقــديرات الطــلاب فــى مــادتى الاحصــاء والرياضيات وذلك باستخدام المقياس الاحصائى الملائم.**

**١٥- البيانات التالية تمثل الدخل الاسبوعى (x) والانفاق الاسبوعى (y) وعدد أفراد الاسرة (z) وذلك بالنسبة لعينة مكونة من ١٠أسر :**

**X : ٨٠ ، ١١٠ ، ١٣٠، ٧٥ ، ٨٢ ، ٦٥ ، ٧٦ ، ٩١ ، ٩٥ ، ١٠٠**

**Y : ٧٥ ، ٩٣ ، ١٠٢ ، ٦٠ ، ٥٥ ، ٥٠ ، ٦٣ ، ٨٢ ، ٨٠ ، ٩٤**

**Z : ٥ ، ٦ ، ٤ ، ٣، ٣ ، ٢ ، ٣ ، ٢ ، ٥ ، ٦**

**والمطلوب :**

**أ- ايجاد معاملات الارتباط الخطى بين أزواج المتغيرات المختلفة.**

**ب- احسب معاملات التحديد المختلفة مع توضيح مدلول كل منهما.**

**جـ - على ضوء نتائجك فى أ ، ب علق بما تراه مناسبا .**

**١٦- من بيانات التمرين السابق ، احسب معاملات الارتباط التالية :**

**- معامل ارتباط سيبرمان.**

**- معامل ارتباط فهنر .**

**- معامل ارتباط كندال.**

**وذلك لازواج المتغيرات : (س ، ص ) ، (س ،ع ) ، ( ص ،ع ) . مع ابداء التعليق المناسب .**

**١٧- فيما يلى بيان بعدد ساعات التدريب (x) التى تلقتها مجموعة من العاملين باحدى الشركات ونتيجة تقييم أدائهم فى نهاية الملائم (y) .**

**X : ٢٠ ٤٢ ٥٠ ٣٥ ٤٠ ٦٠ ٦٥**

**Y : مقبول ممتاز جيد جدا جيد جيد جيد جدا ممتاز**

**والمطلوب : قياس مدى فاعلية البرنامج التدريبى على مستوى أداء العاملين وذلك باستخدام المقياس الاحصائى الملائم .**

**١٨- الجدول التالى يوضح التوزيع التكرارى المزدوج للمتغيرين x , y:**

| X / Y | ٢٥- | ٣٥- | ٤٥- | ٥٥- | ٦٥- | ٧٥- | ٨٥- | المجموع |
|---|---|---|---|---|---|---|---|---|
| ٤٠- | ٣ | ٥ | ٣ | | | | | ١١ |
| ٥٠- | ٥ | ٦ | ١ | ١ | | | | ١٣ |
| ٦٠- | | ١ | ٧ | ١٠ | ١ | | | ١٩ |
| ٧٠- | | | ٦ | ٩ | ٧ | ٣ | | ٢٥ |
| ٨٠- | | | ١ | ٣ | ٨ | ٥ | ١ | ١٨ |
| ٩٠- | | | | | | ٥ | ٩ | ١٤ |
| **المجموع** | ٨ | ١٢ | ١٨ | ٢٣ | ١٦ | ١٣ | ١٠ | ١٠٠ |

**والمطلوب :**

- **ايجاد معامل الارتباط الخطى البسيط بين X ،Y .**
- **باستخدام معامل التحديد فسر ما تحصل عليه من نتائج.**

**١٩- فيما يلى التوزيع التكرارى المزدوج لدرجات ٥٠ طالبا فى كل من مادتى الاحصاء (x) والرياضيات (y) :**

| (x) / (y) | ٥٠- | ٦٠- | ٧٠- | ٨٠- | ٩٠- | المجموع |
|---|---|---|---|---|---|---|
| ٦٠- | ٢ | ٤ | ٣ | ٢ | | ١١ |
| ٦٥- | ٤ | ٨ | ٩ | ٥ | | ٢٦ |
| ٧٠- | ٢ | ٤ | ١١ | ٣ | ٢ | ٢٢ |
| ٧٥- | | ٢ | ١٥ | ٩ | ٤ | ٣٠ |
| ٨٠- | | | ٣ | ٢ | ٢ | ٧ |
| ٨٥- | | | | ٢ | ١ | ٣ |
| ٩٠- | | | | | ١ | ١ |
| **المجموع** | ٨ | ١٨ | ٤١ | ٢٣ | ١٠ | ٥٠ |

**والمطلوب :**

**أ – حساب معامل الارتباط الخطى بين x, y . ثم تفسير ماتحصل عليه من نتائج باستخدام معامل التحديد .**

**ب- قارن بين الاختلاف النسبى لكل من x, y.**

**جـ - ادرس تماثل توزيع الدرجات فى كل مادة من المادتين .**

**٢٠- الجدول التالى يوضح توزيع عينة من الاشخاص حسب لون العينين فى كل من الاباء x والابناء y:**

| X / Y | سوداء Black | غير ذلك Other | المجموع |
|---|---|---|---|
| سوداء Black | ٢٥ | ٨ | ٢٣ |
| غير ذلك Other | ١٢ | ١٥ | ٢٧ |
| المجموع | ٣٧ | ٢٣ | ٦٠ |

**والمطلوب : دراسة اثر العوامل الوراثية فى لون العينين وذلك باستخدام المقياس الاحصائى المناسب .**

**٢١- فيما يلى توزيع عينة من ١٠٠ شخص حسب درجة تعليمهم (x) وارائهم تجاه عمل المراة (y):**

| x / y | Univ. | Second | Prep And Less | المجموع |
|---|---|---|---|---|
| مؤيد | ٥ | ٢٨ | ٩ | ٤٢ |
| معارض | ٥٠ | ٢ | ٦ | ٥٨ |
| المجموع | ٥٥ | ٣٠ | ١٥ | ١٠٠ |

**والمطلوب وضح الى اى حد تعتمد وجهة النظر تجاه عمل المراة على درجة التعليم مستخدما فى ذلك المقياس الاحصائى الملائم.**

**٢٢- فيما يلى نتيجة امتحان رياضيات الاستثمار(X) فى احدى كليات التجارة (منتظمين و منتسبين ) (y) وذلك فى دور يناير ١٩٨٧ :**

| X / Y | Failed | Pass | المجموع |
|---|---|---|---|
| إنتظام | ١٩٢ | ٧١٦ | ٩٠٨ |
| إنتساب | ١٤٤ | ٢٠٨ | ٣٥٢ |
| المجموع | ٣٣٦ | ٩٢٤ | ١٢٦٠ |

**والمطلوب : وضح الى اى حد تتوقف نتيجة الامتحان على نظام الدراسة مستخدما فى ذلك المقايس الاحصائى الملائم ؟**

**٢٣- الجدول التالى يوضح نوزيع الطلاب المنتظمين و المنتسبين فى احدى كليات التجارة وذلك حسب تقديراتهم فى نتيجة امتحان مادة الاحصاء لدور مايو ١٩٨٧**

| التقدير / نظام الدراسة | ضعيف جداً | ضعيف | مقبول | جيد | جيد جداً | ممتاز | المجموع |
|---|---|---|---|---|---|---|---|
| إنتظام | ١٦٦ | ١٤٨ | ١٧٠ | ١٦٧ | ٣٦ | ١٥ | ٧٠٢ |
| إنتساب | ١٠٨ | ٨٨ | ٥٩ | ٢٤ | ٥ | ١ | ٢٨٥ |
| المجموع | ٢٧٤ | ٢٣٦ | ٢٢٩ | ١٩١ | ٤١ | ١٦ | ٩٨٧ |

**<u>والمطلوب</u> : باستخدام المقياس الاحصائى الملائم ،حدد درجة قوة وطبيعة العلاقة مابين نظام الدراسة وتقديرات الطلاب فى مادة الاحصاء.**

**٢٤- البيانات التالية تمثل درجات عشرة طلاب فى كل من مادتى الفلسفة وعلم النفس:**

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| X | ١٦ | ١٠ | ١٧ | ١٣ | ١٢ | ١٥ | ١٨ | ١٦ | ١٤ | ١٩ |
| Y | ١٤ | ١٢ | ١٥ | ١٤ | ١٦ | ١٣ | ١٧ | ١٦ | ١٢ | ١٥ |

**<u>والمطلوب:</u> حساب معامل الارتباط الخطى البسيط .ثم علق على النتيجة التى تحصل عليها باستخدام معامل التحديد.**

**٢٥- فيما يلى التوزيع التكرارى المزدوج للمتغيرين :-**

| Y \ X | ١٠- | ٢٠- | ٣٠- | ٤٠- | ٥٠-٦٠ | المجموع |
|---|---|---|---|---|---|---|
| ٢٥- | ٢ | ١ | ١ | | | ٤ |
| ٣٠- | ١ | ٤ | ٢ | | | ٧ |
| ٣٥- | ٢ | ٣ | ٥ | ٣ | | ١٣ |
| ٤٠- | | ٢ | ٦ | ٦ | | ١٤ |
| ٤٥-٥٠ | | | ٥ | ٤ | ٣ | ١٢ |
| المجموع | ٥ | ١٠ | ١٩ | ١٣ | ٣ | ٥٠ |

**<u>والمطلوب</u>**

**أ- حساب معامل الارتباط الخطى ثم فسر النتيجة التى تحصل عليها وذلك باستخدام معامل التحديد.**

**ب- قارن بين النشتت النسبى لكل من المتغيرين**

**ج- ياستخدام معامل بيرسون الاول قارن بين الالتواء فى كل من التوزيعين.**

**٢٦- فى دراسة لتحديد العلاقة مابين درحة تعليم الرجل ووجهة نظره تجاه عمل المراة تم استطلاع اراء مائةرجل وكانت آراؤهم كمايلى**

| درجة التعليم<br>وجهة نظر تجاه عمل المراة | جامعى فما فوق | ثانوى | اعدادى فأقل |
|---|---|---|---|
| مؤيد | ٢٨ | ٢٢ | ٥ |
| معارض | ٥ | ١٠ | ٣٠ |

**المطلوب :**

**أ- تحديد طبيعة ودرجة قوة العلاقة مابين درجة التعليم للرجل ووجهة النظر تجاه عمل المراة وذلك باستخدام المقياس الاحصائى الملائم .**

**ب- لو لفترضنا ان هناك شكوكا حول صحة آراء الرجال الحاصلين على مؤهل اعدادى فاقل وانع تم استبعادهم نهائيا من الدراسة. قس درجة قوة العلاقة مابين درجة التعليم وتأييد عمل المرأة فى هذه الحالة مستخدما المقياس الاحصائى الملائم**

**ج- قارن بين النتيجنين فى كل من (أ) ، (ب) مع إبداء التعليق المناسب**

# الباب السابع

## *الانحدار الخطي البسيط*

## *Simple linear regression*

**رأينا في الباب السابق ان الإرتباط يقوم بدراسة درجة او قوة العلاقة بين ظاهرتين او متغيرين كلاهما يأخذ قيما أو عناصر مختلفة دون اي تدخل من الباحث في تقدير احدهما اي دون تحديد ما هو المتغير المستقل وما هو المتغير التابع . اما الانحدار فهو يمثل دراسة لشكل العلاقة بين متغيرين كميين احدهما مقيد عند مستويات معينه اي انه مشاهد بدون اي اخطاء ويسمي بالمتغير المستقل Independent Variable وليكن (x) والمتغير الاخر ويأخذ قيما مختلفة عند مستويات أو قيم المتغير المستقل ويسمي بالمتغير التابع وليكن (y) Dependent Variable ولقد عُرف الانحدار عام ١٨٨٥ علي يد فرانسيس جالتون حينما اهتم بدراسة الصفات الوراثية للابناء والاباء. فقد لاحظ ان الابناء لاباء طوال القامه يميلون في المتوسط الي ان يكونوا قصارا عن ابائهم في حين ان الابناء لاباء قصار القامه يميلون في المتوسط الي ان يكونوا طوالا عن ابائهم فاستخدم جالتون لفظ انحدار للتعبير عن هذا الاتجاه نحو المتوسط او الدرجة المتوسطة في هذه العلاقه . ومنذ ذلك الحين ظل هذا اللفظ متداولا في سائر العلوم الادارية والاقتصادية والبيولوجية .... والخ**

**ويقصد بتحليل الانحدار عرض شكل العلاقة بين المتغيريين ( او المتغيرات ) في صورة معادلات رياضية بهدف التنبؤ بقيمة المتغير التباع ( y ) بمعلومية قيمة ( قيم ) المتغير المستقل او (المتغيرات المستقلة ) الذي (التي ) يؤثر (تؤثر) علي المتغير التابع . فعملية التنبؤ Prediction بالقيم المستقبلية للظواهر في غاية الاهمية وخصوصا في مجال الاقتصاد والادارة والسياحه**

والتعليم والصحة والقوي العامله و...الخ . فتخصيص الميزانية لحظة خمسية قادمة يتطلب التنبؤ بعدد السكان المتوقع خلال سنوات الخطة وتوزيعهم العمري والنوعي والتنبؤ باعداد الطلاب المتوقع في المراحل التعلمية المختلفة واعداد الخريجين وقوة العمل وحجم المباني والالات والاجهزة المطلوبة لتنفيذ المستهدف من الخطة . وتلعب معادلات الانحدار الدور الرئيسي في الوصول الي مثل هذه التنبؤات كما سنري فيما بعد .

وكمثال مبسط لتعريف فكرة المتغير التابع والمتغير المستقل في علاقة الانحدار نشير الي انه من الطبيعي ان تكون هناك علاقة بين الدخل الشهري للاسرة وبين الانفاق الشهري لها . وهنا لايجد القارئ عناء في تحديد اي من المتغير هو التابع وايهما هو المستقبل . وإذ أن الانفاق بطبيعة الحال هو الذي يتأثر بمستوي الدخل .. ومن ثم فأن الدخل يعتبر بمثابة المتغير المستقل بينما يمثل الانفاق المتغير التابع . ولتوضيح الفرق ما بين مفهوم الارتباط ومفهوم الانحدار مستخدمين الدخل الشهري والانفاق الشهري كمتغيرين موضع الاعتبار . يمكننا في هذا الصدد ان نقول انه بحساب معامل الارتباط بين الدخل والانفاق الشهري لعدد وليكن خمسين اسرة وبإفتراض أن هذا المعامل هو 0.844 وهو ما يفيد أن هناك علاقة طردية قوية ما بين الدخل والانفاق . بمعني أن التغير في احد المتغيرين ( سواء زيادة أو نقصانا ) سوف يتتبعه تغير وفي نفس الاتجاه في المتغير الاخر . واما الانحدار فهو يعبر عن شكل تلك العلاقة في شكل معادلة يمكن بواستطها تحديد قيمة الانفاق الشهري المتوقعه عند المستويات المختلفة من الدخل الشهري وذلك بأعتبار الدخل بمثابة متغيرا مستقلاً والإنفاق الشهري بمثابة متغيراً تابعاً.

هذا ويوجد نوعان من الانحدار الخطي هما الانحدار الخطي البسيط والانحدار الخطي المتعدد .ففي حالة الانحدار الخطي البسيط يكون هناك متغير مستقل واحد يؤثر علي المتغير التابع . اما في حالة الانحدار الخطي المتعدد فيوجد متغيريين مستقليين علي الاقل والتي تؤثر علي المتغير التابع . ويعد الانحدار الخطي البسيط في أغلب الحالات تبسيطا شديدا للواقع بل واغفالا لجانب كبير مما هو كائن في الحياه العملية حيث نجد في الواقع ان المتغير التابع يعتمد في تفسيره وتغييره علي عدد كبير من المتغيرات المستقلة . ولذلك نجد ان الانحدار المتعدد هو الاكثر شيوعا واستخداما في الحياه العلمية فهو يفيد في حل الكثير من مشاكل التحليل الاقتصادي والتجاري والادارة الصناعية وسوف نقتصر في هذا الباب علي دراسة نموذج الانحدار الخطي البسيط بشيئ من التفصيل تاركين الانحدار المتعدد لدراسات احصائية فيما بعد أن شاء الله .

**الانحدار الخطي البسيط :**

بينا في المقدمة السابقة انه عند دراستنا للانحدار الخطي البسيط لابد من تحديد كل من المتغير المستقل والمتغير التابع . فاذا كان لدينا المتغيران y . x فأن المتغير ( y ) يمكن ان يكون المتغير التابع و المتغير (X) هو المستقل او العكس صحيح ومن ثم نجد ان هناك خطين مستقيمين يمثلان معادلة خطي الانحدار وكلا منها يشرح العلاقة بين المتغيرين X . y ويمكن توضيح ذلك علي النحو التالي :

الخط المستقيم الاول والذي يكون فيه المتغير ( y ) هو المتغير التابع ،(x ) هو المتغير المستقل ويأخذ الصورة التالية :

$$\hat{Y} = a + bx$$

ويسمي هذا النموذج بخط انحدار y علي X حيث ($\hat{y}$ ) هي القيم المستنتجه ( المتوقعه ) من معادلة خط انحدار y علي x وذلك عند مجموعة القيم المختلفة للمتغير المستقل (x ) .

واما الخط الثاني فيكون فيه المتغير (x ) هو المتغير التابع والمتغير ( y ) هو المتغير المستقل ويأخذ الصوره الخطية التالية :

$$\hat{X} = \grave{a} + \grave{b}\, y$$

ويسمي هذا النموذج بخط انحدار x علي y حيث $\hat{x}$ هي القيم المستنتجه ( المتوقعه ) من معادلة خط انحدار X علي y وذلك عند قيم y المختلفة .

هذا وكلما زادت تركز نقاط الشكل الانتشاري حول الخط المستقيم المقرر كلما زادت قوة العلاقة الخطية فيما بين المتغيرين محل الدراسة والعكس صحيح . وكما سبق وان وضحنا عند دراستنا للارتباط فانه اذا كانت العلاقه بين المتغيرين X،Y تامة( اي ان:1± = r) فأن جميع ازواج المشاهدات(Y،x ) تحقق ايا من هذين الخطين y علي x او x علي y . وفي تلك الحالة فأن كافة القيم الفعلية تتساوي مع كافة القيم المقررة المقابلة لها اي أن $y = \hat{Y}$ ، $X=\hat{X}$ وذلك لجميع قيم y, x وهنا يجب التمييز بين كل من $Yi$ ، $\hat{Y}_i$ او بين $Xi$ ، $\hat{X}_i$ حيث $Y_i$ ، $X_i$ هي مجموعة القيم المشاهدة ( الفعلية ) للمتغيرين y , x علي الترتيب $\hat{Y}$ ، $\hat{X}$ فانهما كما سبق تعريفها يمثلان القيم المتوقعه لكل من x . y وذلك من معادلتي خطي الانحدار y علي x و xعلي y علي الترتيب . هذا وكلما كان الفارق بين القيم المشاهدة والقيم المتوقعة صغيرا كلما زادت قوة العلاقة الارتباطية بين المتغيرين محل الدراسة والعكس صحيح وسوف نتعرض الان لدراسة معادلة خطي الانحدار آخذين في الاعتبار البيانات الغير مبوبة والبيانات المبوبة وذلك علي النحو المبين التالي :

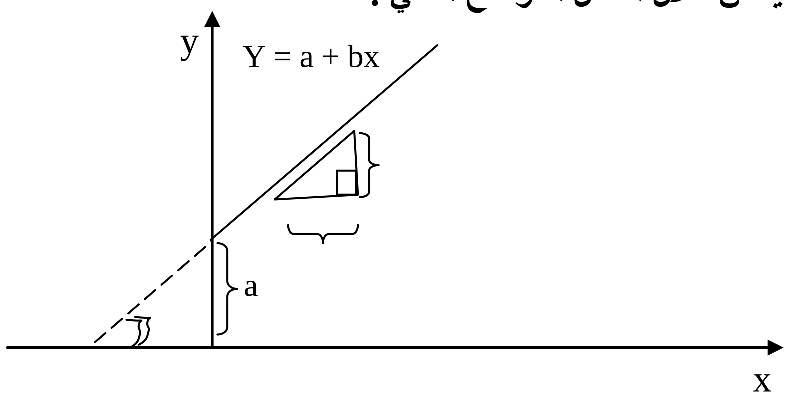
***١ - معادلة خط انحدار y علـى x :***

**معادلة خط انحدار y علي x تاخذ الصورة التالية**

**$\hat{Y} = a + b$ (1- 7 )**

**حيث :**

**$\hat{Y}$ تعبر عن القيمة المتوقعه للمتغير (y ) عند قيم ( x ) المختلفة b, a ثابتان . وهذان الثابتان يحددان الخط المستقيم تماما لذا يسمي هذان الثابتان بمعلمتي Parometers خط الانحدار . ومن خلال العلاقة (٧-١) فان الثابتa هو عبارة عن قيمة غ عندما 0= x اما عن الثابت (b ) فهو عبارة معدل التغير في قيمة المتغير (y ) بالنسبة الي ( x) ويسمي الثابت (a ) بثابت معادلة الانحدار اما (b) فتسمي بمعامل خط انحدار y علي x. أما X فهي عبارة عن القيم المشاهدة (الفعلية ) للمتغير x ويمكن تمثيل معادلة خط انحدار(y) علي ( x ) بيانيا من خلال الشكل الموضح التالي :**




**شكل (٧-١)**

**حط إنحدار (y) على (X)**

**وفي حالة تمام العلاقة بين المتغيرين ( y . x ) (اي حينما تكون $r_{xy} = \pm 1$) فان كافة النقاط الفعلية لازواج المشاهدات تقع علي خط الانحدار المقدر اي ان**

$$\hat{Y}_i = Y_i = a + b\,X_i \qquad where\ i = 1, 2, \dots\dots, n$$

وهو ما معني انعدار ما يسمي بأخطاء التقدير وهي عبارة عن الفرق ما بين القيم الفعلية والمتوقعه اي ان :

$$Y_i - \hat{Y}_i = 0 \qquad for\ i = 1, 2, \dots\dots, n$$

الا انه من الناحية العملية فانه عادة ما تختلف القيم الفعلية (y) عن المتوقعه ($\hat{Y}$) والمحسوبة من معادلة خط الانحدار المقدر . ويقل هذا الاختلاف لا شك كلما زادت قوة العلاقة بين المتغيرين والعكس صحيح تماما . وهذا يعني انه عادة ما تختلف القيمة ($y_i \hat{Y}_i$ – ) عن الصفر وذلك نظرا لأن القيمة المطلقة لمعامل الارتباط الخطي عادة ما تختلف عن الواحد الصحيح . فاذا رمزنا لهذه القيمة بالرز ( $e_i$ ) اي ان :

$$e_i = Y_i - \hat{Y}_i \qquad for\ i = 1, 2, \dots\dots, n$$

فان ($e_i$) وهي عبارة عن الفرق ما بين القيمة المشاهدة والقيمة المتوقعه أو ما يسمي بالخطأ في تقدير ($y_i$) وهذا الفرق يكون سالبا اذا كانت قيمة المشاهدة اسفل خط الانحدار ويكون موجبا اذا كانت قيمة (y) المشاهدة اعلي خط الانحدار .

وبأستخدام المعادلة (٧-١) السابقة فانه يمكن التعبير عن الخطأ في اتقدير علي الصورة :

$$e_i = Y_i - \hat{Y}_i = Y_i - a - bx_i$$

**ايجاد معادلة خط انحدار (y) على (x) للبيانات الغير مبوبة :**

اذا كان لدينا عدد (n) من ازواج المشاهدات للمتغيرين x . y ولتكن :

$$(x_1, y_1), (x_2, y_2), (x_3, y_3) \ \dots\dots\dots\dots\dots\dots\dots,\ (X_n, Y_n)$$

فأن يمكن استخدام هذه الازواج من المشاهدات في توفيق افضل معادلة خط مستقيم يمثل العلاقة بين المتغيرين x . y ويتمثل توفيق افضل خط مستقيم في تحديد افضل تقدير لمعلمتي النموذج (٧-١) وهما a , b. وما نعنيه بكلمة افضل في شأن المعلمتين a , b كما سبق وان ذكرنا في ان الفروق ما بين القيم المشاهدة والقيم المتوقعه للمتغير (y) يمكن علي اساسها الحكم علي مدي جودة تمثيل خط الانحدار للعلاقة بين المتغيرين Y،X ولذلك فان المعيار الذي يمكن من خلاله الحكم علي افضلية ودقة التقدير لتلك المعلمات سيكون من خلال تقليل الفروق بين القيم الفعلية ( المشاهدة ) والقيم المتوقعه للمتغير (y) . ويمكن تحقيق ذلك اعتمادا علي اي من الاسس الثلاث التالية :

* توفيق خط الانحدار بحيث يكون مجموع قيم ($e_i$) المطلقة اصغر ما يمكن اي ان :

$$\sum|\acute{e}_i| = \sum|y_i - \hat{y}_i| \qquad \textit{in its minimum value}$$

* توفيق خط الانحدار بحيث يكون أكبر الفروق المطلقة أصغر ما يمكن .

* توفيق خط الانحدار الذي يجعل مجموع مربعات الفروق بين القيم المشاهدة والقيم المتوقعه لـ y اقل ما يمكن اي ان :

$$\sum e_{i^2} \quad \textit{in its minimum value}$$

هذا وتسمي الطريقة الثالثة بطريقة المربعات الصغري Least Squares Method وهي تعتبر افضل الطرق الثلاث من حيث دقة تقديرها لثوابت او معلمتي معادلة خط الانحدار a , b حيث تتميز هذه الطريقة في تقديراتها بخصائص احصائية مفضلة من بينها خاصية عدم التحيز والكفاءة وهو ما يخرج عن نطاق دراستنا في هذا الكتاب .

- **تقدير معلمتي نموزج خط انجدار Y على X باستخدام طريقة المربعات الصغرى :**

وفقا لطريقة المربعات الصغري فانه يمكن تقدير معلمتي معادلة خط انحدار Yعلي X علي النحو التالي :

حيث ان معادلة خط انحدار Yعلي X هي : $Y = a + bx$

فأنه يمكن ايجاد قيمتي a , b والتي تجعل مجموع مربعات الفروق( $ei^2$ ) اقل ما يمكن وذلك من خلال ايجاد النهاية الصغري للدالة (٧ – ٣ ) مع استخدام الشرطين اللازمين لكي تكون هذه الدالة نهاية صغري عند نقطة وذلك علي النحو التالي :

$$\sum_{i=1}^{n} e_{i^2} = \sum_{i=1}^{n} (y_i - a - bx_i)^2 \qquad (3-7)$$

وهذان الشرطان اللازمان هما :-

- ✓ التفاضل الاول للدالة عند النقطة يساوي الصفر .
- ✓ التفاضل الثاني عند هذه النقطة يساوي مقدارا موجبا .

ومن ثم فبتطبيق الشرط الاول علي الدالة ( ٧-٣) وذلك بأعتبار التفاضل الجزئي لهذه الداله بالنسبة الي كل من a،b مساويا للصفر. فأننا نحصل علي ما يلي :-

$$\frac{\sigma}{\sigma a} \sum_{i=1}^{n} (y_i - a - bx_i)^2 = -2 \sum_{i=1}^{n} (y_i - a - bx_i) = 0$$

وعليه فأن

$$\sum_{i=1}^{n} (y_i - a - bx_i) = 0$$

أي أن (٧-٤)

$$\sum y_i = na + b \sum x_i$$

وبالمثل فإنه :

$$\frac{\sigma}{\sigma a}\sum_{i=1}^{n}(y_i - a - bx_i)^2 = -2\sum_{i=1}^{n}(y_i - a - bx_i) = 0$$

وعليه فإن :

$$\sum_{i=1}^{n}(y_i - a - bx_i) = 0$$

أي أن

$$\sum x_i\, y_i = a\sum x_i \;\; + b\sum x_{i^2} \qquad (٧-٥)$$

هذا وتسمي المعادلتان (٧ ـ ٤ ) و (٧ ـ ٥) بالمعادلات الطبيعية للنموذج وبحل هاتين المعادلتين معا يمكن الحصول علي تقدير لمعلمتي نموذج خط الانحدار B , A . او من هاتين المعادلتين يمكن الوصول الي الشكل التالي :-

$$\hat{b} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$

اي ان

$$= \frac{\frac{\sum xy}{n} - \frac{\sum x}{n} \cdot \frac{\sum y}{n}}{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2} \qquad (٧-٦)$$

**وقد فضلنا الصيغه (٧- ٦ ) عن سابقتها وذلك لوجود نوع من التشابه بينها وبين صيغة معامل الارتباط الخطي البسيط لبيرسون . اما قيمة ثابت معادلة فيتحدد من خلال العلاقة التالية :**

$$\hat{a} = \bar{y} - \hat{b}\,\bar{x}$$

**هذا ويمكن اثبات ان المشتقه الاولي عند النقطة( b , a ) تحقق الشرط الثاني لايجاد النهاية الصغري للدالة (٧ – ٣) حيث تتحدد قيمة (a) بالصيغة (٧– ٧) بينما تتحدد قيمة معامل الانحدار b من خلال الصيغة (٧-٦)**

**هذا ويمكن ان تيسر علي القارئ كيفية تذكر المعادلتين الطبيعتين (٧-٤) و (٧-٥) وذلك علي النحو التالي :**

**حيث ان معادلة خط انحدار Y علي X تأخذ الشكل التالي :**

$$\hat{Y} = a + bx$$

**فأنه يمكن الحصول علي المعادلة الطبيعية الاولي (٧-٤) وذلك من خلال اعتبا مجموع طرفي معادلة خط احدار Y علي X مع اعتبار ( $\hat{Y}$ ) بدلا من حيث أن مجموع القيم الفعلية يساوي مجموع القيم المتوقعة اي أن $\sum y = \sum \hat{y}$**

**اما المعادلة الطبيعية الثانية (٧ – ٥) فانه يمكن الحصول عليها من خلال ضرب طرفي معادلة خط انحدار Y علي X في قيمة المتغير المستقل (x) ثم اعتبارا او اخذ مجموع الطرفين ( وكذلك مع اعتبار y بدلا من $\hat{Y}$ )**

**هذا ووفقا لطريقة المربعات الصغري فأن :**

$$\sum_{i=1}^{n} e_i = 0 \;\Rightarrow\; \sum_{i=1}^{n} (y_i - \hat{y}) = 0$$

**ومن ثم فإن :-**

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i \qquad (٧\text{-}٨)$$

ومن ثم فإن (٧-٨)

اي ان مجموع القيم الفعلية يساوي مجموع القيم المتوقعه . ومن ثم فأن الوسط الحسابي للقيم الفعلية او المشاهدة يساوي الوسط الحسابي للقيم المتوقعه . ومن خلال استخدامنا لهذه الخاصية لطريقة المربعات الصغري يمكن استنتاج صورة اخري لمعادلة خط انحدار (y )علي (x ) وذلك كما يلي :

فحيث أن :

$$\hat{y} = a + bx$$

لذا فأن

$$\bar{y} = a + b\bar{x}$$

$$\hat{y} - \bar{y} + b(x - \bar{x}) \qquad (٧-٩)$$

ومن المعادلة (٧ – ٩) يمكن التوصل ألي صورة ثالثة لمعادلة خط انحدار (y ) علي X وهي :

$$\hat{y} = \bar{y} + b(x - \bar{x})$$

(٧-١٠)

هذا وسوف نقتصر في دراستنا علي استخدامنا للطريقة المباشرة في تقدير معلمتي خط الانحدار .

***مثال***

(١)الجدول التالي يعطي بيانات عن سعر الفائدة $(x_i)$ في المائة وحجم المدخرات $(y_i)$ بالمليون جنيه بأحد البنوك .

| $X_i$ | 7 | 9 | 11 | 10 | 13 | 12 | 15 |
|---|---|---|---|---|---|---|---|
| $Y_i$ | 10 | 12 | 15 | 12 | 18 | 16 | 22 |

والمطلوب:-

١- تقدير معادلة خط انحدار Y علي X بأفتراض انها خطية .

٢- ايجاد حجم المدخرات المتوقع عندما يصل سعر الفائدة الى 20 (في المائة)

٣- ايجاد معادلة خط الانحدار حينما تقل قيمة اسعار الفائدة بمقدار 5 ( في المائة ) .

**الحل:**

١ – بأفتراض ان نموذج خط انحدار Y علي X يأخذ الصورة الخطية التالية

$$\hat{y}_i = \hat{a} + \hat{b}x_i$$

ولتقدير معلمتي خط الانحدار أي $\hat{a}\hat{b}$ , يجب الحصول علي المجاميع الموضحه في الجدول التالي :

| $X_i$ | $y_i$ | $X_i\ y_i$ | $X_i^2$ |
|---|---|---|---|
| 7 | 10 | 70 | 49 |
| 9 | 12 | 108 | 81 |
| 11 | 15 | 165 | 121 |
| 10 | 12 | 120 | 100 |
| 13 | 18 | 234 | 169 |
| 12 | 16 | 192 | 144 |
| 15 | 22 | 330 | 225 |
| 77 | 105 | 1219 | 889 |

$$\hat{a} = \frac{\frac{\sum xy}{n} - \frac{\sum x}{n} \cdot \frac{\sum y}{n}}{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

$$= \frac{\frac{1219}{7} - \frac{77}{7} \cdot \frac{105}{7}}{\frac{889}{7} - \left(\frac{77}{7}\right)^2}$$

$$= \frac{9.143}{6} = 1.524$$

أما عن ثابت معادلة انحدار y علي x فهو :

$$\hat{b} = \bar{y} - a\bar{x}$$

$$= \frac{105}{7} - (1.524)\left(\frac{77}{7}\right) = -1.769$$

**ومن ثم فان معادلة خط انحدار Y علي X تأخذ الصورة الخطية التالية**

$$\hat{y} = \hat{a} + \hat{b}x = 1.764 + 1.524x$$

**٢- لايجاد حجم المدخرات المتوقع اي ( $\hat{y}$ ) عندما يصل سعر الفائدة اي (x) الـي 20في المائة فأنـه يتم التعويض عن قيمـة المتغير المستقل ( x ) فـي معادلـة خط الانحدار المقدرة بالقيمة (20)**

**اي ان :**

$$\hat{y} = 10764 + 1.524(20) = 28.716 \qquad (milion\ l.E)$$

**٣- حينما تقل اسعار الفائدة بمقدار ( 5 ) في المائة لكافة النقـاط المعطـاة فأنـه بـدلا من أيجاد المعادلة المقدرة من خلال طرح 5 في المائة من كافة قيم ( x ) المعطاة ثم تقدير المعالم يمكن أستنتاج المعادلـة الجديدة حينمـا تـنخفض أسـعار الفائـدة بمقـدار (5 ) فـي المائـة بـأن يـتم التعـويض فـي معادلـة النمـوذج المقدر سابقا (*) عن قيمة ( x) تساوي x مطروحا منها (5 ) اي ان :**

$$\hat{y}\downarrow = -1.764 + 1.524\,(x - 5)$$

**X= X - 5**

**= -1.764 + 1.524x – 7.62**

$$\hat{y} = -9.384 + 1.524x$$

**اي ان اثـر انخفـاض سـعر الفائـدة بمقدار (5 ) فـي المائـة يقـع فقـط علـي ثابـت معادلة الانحدار أي (a) أما ميل خط الانحدار (b ) فهو لم يتأثر بهذا التغير .**

**<u>أخطاء التقدير :-</u>**

**بعد تحديـد معادلـة خـط انحـدار Y علـي X فأنـه يمكـن ايجـاد القـيم المتوقعـه للمتغير (y ) المختلفة وذلك كمـا هو موضـح فـي الجـدول التـالي بـالتطبيق علـي المثال السابق :-**

**حساب القيم المتوقعه لــ  yاي  y**

**وحساب اخطاء التقدير في علاقة إنحدار Y على X : فيتم تكوين الجدول التالي بعد تقدير نموذج خط إنحدار Y على X :**

| Xi | Yi | $\hat{y}_i$= -1.764+1.524$x_i$ | $e_i = y_i - \hat{y}_i$ |
|---|---|---|---|
| 7 | 10 | -1.764 + 1.524(7) = 8.904 | 1.096 |
| 9 | 12 | 11.952 | 0.048 |
| 11 | 15 | 15.0 | 0 |
| 10 | 12 | 13.47 | -1.476 |
| 13 | 18 | 18.048 | 0.048 |
| 12 | 16 | 16.524 | -0.524 |
| 15 | 22 | 21.094 | 0.904 |
| 77 | 105 | 105 | 0 |

**لاحظ  اننا حصلنا علي قيم ($\hat{y}$) والمبينة في العموم الثالث من الجدول السابق وذلك من خلال التعويض المباشر عن قيم (x ) الموجودة في العموم الاول ومن خلال الجدول السابق لاحظ أن :**

$\sum y_i = \sum \hat{y}_i$ ,

$\sum e_i = 0$

**وهاتان الخاصيتان تأتيان كنتيجة طبيعية لطريقة المربعات الصغري والتي سبق الاشارة اليها . ويمكنك عزيزي الطالب ايجاد مجموع مربعات الاخطاء وهذا المجموع سيكون بمثابة اقل مجموع مربعات يمكن الحصول عليه بمعني لا يمكن الحصول من خلال اي طريقة اخري من طرق تقدير معالم النموذج**

**الخطي ( بخلاف طريقة المربعات الصغري ) علي مجموع مربعات للاخطاء يقل عن المجموع الناتج من طريقة المربعات الصغري .**

**<u>ايجاد معادلة خط انحدار Y على X للبيانات المبوبة :</u>**

**لايجاد معادلة خط انحدار Yعلي X للبيانات المبوبة من خلال جدول تكراري مزدوج نقوم بأتباع الخطوات التالية :-**

**- اعتبر (x ) المتغير الاول ( المستقبل ) حيث x هي اقسام هذا المتغير في حالة كونه متغيرا منفصلا او مراكز فئاته اذا كان متغيرا متصلا .**

**- احسب متوسطات المتغير الثاني ( التابع ) المقابلة لاقسام هذا المتغير المنفصل او مراكز فئات المتغير (x ) المتصل . وهذه المتوسطات تعتبر بمثابة قيم المتغير (y ) .**

**- اوجد معادلة انحدار Yعلي X والتي تأخذ الصيغه (٧ – ١) وذلك من خلال تقدير معالمها من خلال المعادلتين (٧-٦) و (٧ – ٧) السابقتين .**

**والمثال التالي يوضح كيفية تطبيق الخطوات السابقة :-**

***<u>مثال٢ :-</u>***

**فيما يلي الجدول التالي يبين توزيع 50اسرة حسب فئات الدخل (x ) والانفاق (x ) الاسبوعي بالجنيه .**

| x / y | 60- | 70- | 80- | 90- | 100- | 110 | 120- | ∑ |
|---|---|---|---|---|---|---|---|---|
| 40- | 2 | 2 | | | | | | 4 |
| 50- | 1 | 3 | 2 | 1 | | | | 7 |
| 60- | | 2 | 5 | 3 | | | | 10 |
| 70- | | 1 | 6 | 4 | 1 | | | 12 |
| 80- | | | 1 | 3 | 5 | | | 9 |
| 90- | | | | 1 | 2 | 2 | | 5 |
| 100- | | | | | | 2 | 1 | 3 |
| ∑ | 3 | 8 | 14 | 12 | 8 | 4 | 1 | 50 |

**والمطلوب: اوجد معادلة خد انحدار الانفاق ( y ) علي الدخل( x ) . اي معادلة خط انحدار Yعلي X .**

***الحل :***

**فيما يلي حساب مراكز فئات الدخل المختلفة (x) وقيم ( y) المقابلة لها مركز الفئة الاولي لــ x =65 فأن قيمة y المقابلة هي :**

$$Y=\frac{2(45)+1(55)}{3}=48.3$$

**مركز الفئة الثانية لـ x = 75 فان قيمة ( y) المقابلة هي :**

$$y=\frac{2(45)+3\,(55)+2(65)+\,1(75)}{8}=57.5$$

- **مركز الفئة الثالثة لـ x = 85 فأن قيمة (y) المقابلة لها هي :**

$$y=\frac{2(55)+5\,(65)+6(75)+\ 1(85)}{8}=69.3$$

- مركز الفئة الرابعة لـ x = 95 فأن قيمة (y ) المقابلة لها هي :

$$y = \frac{1(55) + 3\,(65) + 4(75) + \ 3(85) + 1(95)}{8} = 75$$

- مركز الفئة الخامسة لـ x = 105 فأن قيمة (y ) المقابلة لها هي :

$$y = \frac{1(75) + 5\,(85) + 4(95)}{8} = 86.3$$

- مركز الفئة السادسة لـ x =115 فأن قيمةْ (y ) المقابلة لها هي :

$$y = \frac{2(95) + \ 2\,(105)}{4} = 10$$

مركز الفئة السابع لـ x = 125 فأن قيمة المقابلة لها هي :

$$y = \frac{1(105)}{1} = 105$$

وبعد حساب كل من مراكز الفئات المختلفة لـــ ( X ) وما يقابلها من قيم ( Y) فأنه يتم اجراء نفس الحسابات الخاصه بمعادلة خط انحدار Y علي X للبيانات الغير مبوبه لذا يلزم الحصول علي المجاميع الموضحه بالجدول التالي

| X | Y | X Y | X2 |
|---|---|---|---|
| 65 | 48.3 | 313905 | 4225 |
| 57 | 57.5 | 4312.5 | 5625 |
| 58 | 69.3 | 5890.5 | 7225 |
| 59 | 75 | 7125 | 9025 |
| 105 | 86.3 | 9061.5 | 11025 |
| 115 | 100 | 11500 | 13225 |
| 125 | 105 | 13125 | 15625 |
| 665 | 54104 | 54154 | 65975 |

وبأفتراض الصورة الخطية لنموذج خط انحدار y علي ( x) هي :

$$\hat{y} = \hat{a} + \hat{b}x$$

حيث أن :

$$\hat{b} = \frac{\frac{\sum xy}{n} - \frac{\sum x}{n} . \frac{\sum y}{n}}{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

$$\hat{b} = \frac{\frac{54154}{7} - \left(\frac{665}{7}\right)\left(\frac{541.4}{7}\right)}{\frac{65975}{7} - \left(\frac{665}{7}\right)^2}$$

اي ان :

$$\hat{b} = \frac{388.714}{400} = 0.9718$$

أما عن ثابت معادلة الانحدار فهو :-

$$\hat{a} = \bar{y} - \hat{b}\bar{x} = \frac{\sum y}{n} - \hat{b}\frac{\sum x}{n}$$

$$= \left(\frac{541.4}{7}\right) - 0.9718\left(\frac{665}{7}\right) = -140978 \qquad (l.E)$$

أي أن معادلة خط انحدار (y ) علي (x ) تأخذ الصورة التالية :

$$\hat{y} = -14.978 + 0.9718x$$

**<u>العلاقة بين معامل انحدار( y ) على ( x ) وبين معامل الارتباط الخطي البسط : ( $R_{XY}$ )</u>**

من خلال معادلة ميل خط انحدار Yعلي X الواردة بالمعادلة ( ٧- ٦) ومعادلة معامل الارتباط الخطي البسيط لبيرسون سنورد العلاقة ما بين معامل خط الانحدارYعلي X ومعامل الارتباط الخطي البسيط دون اثبات تسهيلا علي القارئ : حيث أن :-

$$\hat{b} = r_{xy}\left(\frac{\sigma y}{\sigma x}\right) \qquad (١١ – ٧)$$

هذا ومن خلال المعادلة او العلاقة ( ٧ – ١١ ) يمكننا استنتاج أن معامل إنحدار yعلي x يأخذ نفس إشارة معامل الارتباط الخطي البسيط لبيرسون بين المتغيرين حيث أن النسبة ما بين الانحرافات المعيارية للمتغيرين هي دائما ابدا موجبه . هذا ويمكن اعادة كتابة المعادلة (٧ – ١٠) لمعادلة انحدار Yعلي X بدلالة معامل الارتباط الخطي البسيط بين المتغيرين وذلك علي النحو التالي :

فحيث ان

$$\hat{y} = \bar{y} + b(x - \bar{x})$$

لذا فأن :

$$= \bar{y} + r_{xy}\frac{\sigma y}{\sigma x}(x - \bar{x})$$

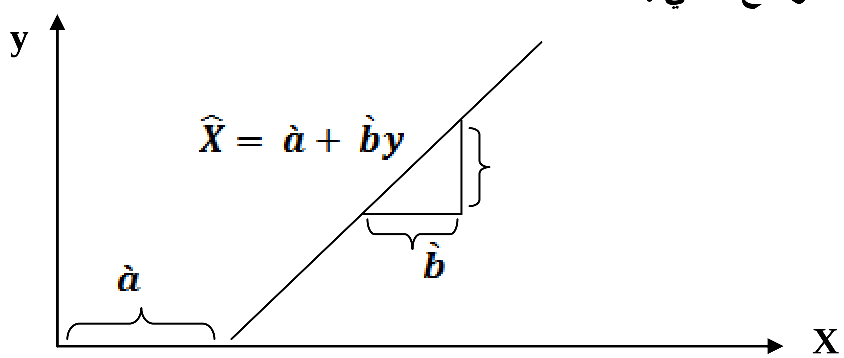
**٢ـ معادلة خط انحدار ( x) علي ( y ) :-**

**تاخذ معادلة خط انحدار x علي y الصورة الخطية**

$$\hat{x} = \grave{a} + \grave{b}y \qquad (١٢ - ٧)$$

**حيث ( $\hat{x}$ ) هي عبارة عن القيم المستنتجة ( المتوقعه ) للمتغير ( x) عند القيم المختلفة للمتغير ( y ). أما $\grave{b}.\grave{a}$ فهما ثابتان ويسميان بمعلتي Parameters خط انحدار X علي y حيث $\grave{a}$ تسمي بثابت معادلة الانحدار وهي عبارة عن قيمة x عندما تكون قيمة y مساوية للصفر . اما $\grave{b}$ فتسمي بميل خط الانحدار وهي عبارة عن معدل التغير في ( x ) بالنسبة الي ( y) .**

**هذا ويمكن ايضا تمثيل معادلة خط انحدار x علي ( y ) بيانياً وذلك من خلال الشكل الموضح التالي :**




**خط إنحدار X على Y**

هذا ويمكن تعميم كل ما سبق عرضه في حالة انحدار (Y)علي(X) علي حالتنا هذه ( اي حالة خط انحدار xعلي y وذلك من خلال استبدال yبـــ x والعكس صحيح . ووفقا لذلك فأن (x) في معادلة خط انحدار xعلي y تمثل المتغيرالتابع بينما تمثل ( y) المتغير المستقل . وهنا سوف نتعرض بدون أثبات الي العلاقات المتعلقه بمعادلة خط انحدار xعلي y وذلك علي نحو مماثل لما سبق عرضه في حالة خط انحدار Yعلي X فخطأ التقدير في قيم ( x ) يمكن التعبير عنه بالعلاقه

$$\grave{e}_i = x_i - \hat{x}_i \qquad for \quad i = 1, 2, \dots\dots, n$$

ولإيجاد معادلة خط إنحدار $(x)$ علي $(y)$ فأنه يمكن تحديد معلمتي النموذج أي قيمتي $\grave{b}, \grave{a}$ وذلك من خلال إستنتاج المعادلتين الطبيعيتين المتعلقتين خط أنحدار x علي y وذلك علي النحو التالي :

$$\left.\begin{aligned} \sum x &= na + \grave{b}\sum y \\ \sum xy &= \grave{a}\sum y^2 + \grave{b}\sum y \end{aligned}\right\} \quad (١٣ - ٧)$$

وهاتين المعادلتين تحققان ان مجموع مربعات اخطاء الفروق $\sum {}^{12}\grave{e}_i$ يكون اصغر ما يمكن ولعله من المفيد أن نسهل علي القارئ كيفية تذكر المعادلتين الطبيعيتين الواردتين في العلاقه ( ٧ -١٣ ) وذلك علي النحو التالي :

حيث يمكن الحصول علي المعادلة الاولي بأخذ مجموع الطرفين للمعادلة (٧-١٢) وذلك بأعتبار ان (x) بدلالة من $(\hat{x})$. وأما المعادلة الثانية فيمكن الحصول عليها من خلال ضرب طرفي المعادلة (٧-١٢) فى المتغير المستقل (Y) فى هذه الحالة ثم إعتبا رالمجموع بعد الضرب ( مع أعتبار x بدلالة من $\hat{x}$ أيضا) .

هذا ويمكن من خلال هاتين المعادلتين أستنتاج معلمتي خط الانحدار وذلك علي النحو التالي :

$$\hat{b} = = \frac{\frac{\sum xy}{n} - \frac{\sum x}{n} \cdot \frac{\sum y}{n}}{\frac{\sum y^2}{n} - \left(\frac{\sum y}{n}\right)^2}$$

$$\hat{a} = \bar{x} - \hat{b}\bar{y}$$

وبعد تقدير معلمتي خط انحدار x علي y تصبح صورة نموزج خط الانحدار المقدر هي :

$$\hat{x}_i = \hat{a} + \hat{b}y_i$$

هذا ويمكن كتابة معادلة خط انحدار x علي y علي الصورة الثانية التالية :

$$(x - \bar{x}) = \hat{b}\,(y - \bar{y}) \qquad (١٥\text{-}٧)$$

أو

$$x = \bar{x} + \hat{b}\,(y - \bar{y})$$

(٧ – ١٦ )

كما أن العلاقة ما بين خط انحدار x علي y ومعامل الارتباط الخطي البسيط لبيرسون تاخذ الصورة التالية :

$$\hat{b} = r_{xy}\left(\frac{\sigma\bar{x}}{\sigma y}\right) \qquad (١٧\text{-}٧)$$

ومن ثم يمكن كتابة المعادلة (٧-١٦) علي الصورة :

(٧-١٨)

$$X = \bar{x} + r_{xy}\left(\frac{\sigma\bar{x}}{\sigma y}\right)(y - \bar{y})$$

كما أن مجموع القيم الفعلية للمتغير التابع x تساوي مجموع القيم المتوقعه $\hat{x}$

اي ان $\sum x = \sum \hat{x}$

***مثال (٣) :***

البيانـات التاليـة تمثل تكـاليف الاعـلان (x) وقيمـة المبيعات (y) وذلك بـآلاف الجنيهات خلال عشر سنوات متتالية لإحدي الشركات :

| X | 10 | 12 | 14 | 16 | 16 | 18 | 22 | 22 | 24 | 26 |
|---|---|---|---|---|---|---|---|---|---|---|
| Y | 110 | 115 | 120 | 130 | 125 | 135 | 140 | 150 | 140 | 180 |

والمطلوب :-

أ- إيجاد معادلة خط انحدار تكاليف الاعلام (x) علي قيمة المبيعات (y)

ب- إذا بلغت قيمة المبيعات في أحد السنوات ١٦٠ الف جنيـه فمـا هـي القيمـة المتوقعه للمنصرف علي الاعلام في هذا العام .

ج- حساب اخطاء التقدير لقيم (x) المختلفة مع التعليق علـي مـا تحصل عليـه من نتـائج .

***الحل :***

أ- بافتراض أن معادلة خط انحدار تكاليف الاعلام (x) علي قيمة المبيعات (y) تأخذ الصورة الخطية التالية .

$$\hat{x} = \hat{a} + \hat{b}\, y$$

ولتقدير معلمتي خط الانحدار a , b يتم إيجاد المجاميع الموضحه في الجدول التالي :

| x | y | X y | 2 |
|---|---|---|---|
| 10 | 110 | 1100 | 12100 |
| 12 | 115 | 1380 | 13225 |
| 14 | 120 | 1680 | 14400 |
| 16 | 130 | 2080 | 16900 |
| 16 | 125 | 2000 | 15625 |
| 18 | 135 | 2430 | 18225 |
| 22 | 140 | 3080 | 19600 |
| 22 | 150 | 3300 | 22500 |
| 24 | 170 | 4080 | 28900 |
| 26 | 180 | 4550 | 30625 |
| 180 | 1370 | 25680 | 192100 |

فحيث أن :

$$\hat{b} = \frac{\frac{\sum xy}{n} - \frac{\sum x}{n} \cdot \frac{\sum y}{n}}{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

$$= \frac{\frac{25680}{10} - \left(\frac{180}{10}\right)\left(\frac{1370}{10}\right)}{\frac{192100}{10} - \left(\frac{1370}{10}\right)^2}$$

$$= \bar{X} - \hat{b}\,\bar{y} \left(\frac{102}{441}\right) = 0.23 \quad \left(\frac{1370}{10}\right) \quad = \left(\frac{180}{10}\right) - 0.23 \left(\frac{1370}{10}\right) \hat{a}$$

$$= -13.51$$

ومن ثم فأن خط انحدار x علي y تأخذ الصورة التالية

$$\hat{x} = -13.51 + 0.23y$$

ب – بعد إيجاد معادلة خط انحدار x علي y فأن يمكن حساب القيمـة المتوقعه لــ x عندما تساوي y القيمة 160 وذلك علي النحو التالي :

$$\hat{x} = -13.51 + 0.23(160) = 23.29$$

أي أنـه يمكن القول أنـه إذا بلغت قيمة المبيعات في أحد الاعوام 160ألـف جنيـه فأنه من المتوقع أن يكون المنصرف علـي الاعلام فـي هذا العام هو ٢٣.٢٩ الف جنيه .

ث- الجدول التالي يوضح القيم المتوقعه لــ x وأخطاء التقدير المناظرة لها :

| $x_i$ | $x_i$ | $\widehat{x_i}$ = -13.51 + 0.23 $y_i$ | $x_i$ - $x_i$ |
|---|---|---|---|
| 10 | 110 | -13.51 + 0.23 ( 110 ) = 11079 | - 1.79 |
| 12 | 115 | 12.94 | - 0.94 |
| 14 | 120 | 14.09 | - 0.09 |
| 16 | 130 | 16.39 | - 0.39 |
| 16 | 125 | 15.24 | 0.76 |
| 18 | 135 | 17.24 | 0.46 |
| 22 | 140 | 18.69 | 3.31 |
| 22 | 150 | 20.99 | 1.01 |
| 24 | 170 | 25.59 | - 1.59 |
| 26 | 175 | 26.74 | |
| 180 | 1370 | 180 | 0 |

ومن الجدول السابق لاحظ ان:

$$\sum x_i = \sum \widehat{x_i} = 180 \ \sum e_i = 0$$

- ***إيجاد معادلة خط انحدار x على y للبيانات المبوبة :***

لايجاد معادلة خط انحدار x علي y للبيانات المبوبة في جدول تكراري مزدوج يمكننا أتباع الخطوات التالية :

- نعتبر ( y ) هي قيم المتغير الاول ( المستقل ) حيث ( y ) هي أقسام هذا المتغير في حالة كونه متغيرا منفصلا أو مراكز فئاته أذا كان متغيرا متصلا

- يتم حساب متوسطات قيم المتغير الثاني ( التابع) المقابلة لمراكز فئات المتغير ( y) أو المقابلة لاقسام هذا المتغير . وهذه المتوسطات تمثل قيم المتغير ( x )

- نوجد معادلة انحدار x علي y وذلك من خلال أيجاد المجاميع اللازمة لمعلمتي خط الانحدار . والمثال التالي يوضح ذلك .

***مثال :-***

بإستخدام التوزيع التكراري المزدوج الموضح في مثال (٢) من هذا الباب اوجد معادلة خط انحدار الدخل (x ) علي الانفاق (y ) .

***الحل :***

- فيما يلي حساب مراكز فئات الانفاق (y ) المختلفة وقيم (x ) المقابلة لها :

- مركز الفئة الاولي لــ y = 45 يقابله قيمة x المقابلة لها هي

$$x = \frac{2(65)+2(75)}{4}$$

- مركز الفئة الثاني لـــ y = 55 يقابله قيمة x المقابلة لها هي :

$$X = \frac{1\,(65)+3(75)+(85)+1(95)}{7} = 79.3$$

مركز الفئه الثالثة لــ y = 65 يقابله قيمة x المقابلة لها هي :

$$x = \frac{2\,(75) + 5(75) + (85) + 3(95)}{10} = 86$$

مركز الفئة الرابعة لـ y = 75 يقابله : قيمة x المقابلة لها هي

$$x = \frac{1\,(75) + 6(85) + 4\,(95) + 1(105)}{12} = 89.2$$

مركز الفئة الخامسه لـ y = 85 يقابله : قيمة x المقابلة لها هي :

$$x = \frac{1\,(85) + 3(95) + 5(105)}{9} = 99.4$$

مركز الفئة السادسة لـ y = 95 يقابله : قيمة x المقابلة لها هي :

$$x = \frac{1\,(95) + 2(105) + 2(115)}{5} = 107$$

مركز الفئة السابعه لـ y = 105 يقابله قيمة x المقابلة لها هي :

$$x = \frac{2\,(115) + 1(125) + 1(125)}{3} = 118.3$$

وبعد حساب كل من مراكز الفئات المختلفة لـ y وقيم ) المقابلة لها فأنه يتم إيجاد المجاميع اللازمة لحساب معلمتي خط الانحدار xعلي y والتي يوضحها الجدول التالي :

| X | y | xy | $y^2$ |
|---|---|---|---|
| 70.0 | 45 | 3150.0 | 2025 |
| 79.3 | 55 | 4361.5 | 3025 |
| 86.0 | 65 | 5590.0 | 4225 |
| 89.2 | 75 | 6690.0 | 5625 |
| 99.4 | 85 | 8449.0 | 7225 |
| 107.0 | 95 | 10165.0 | 9025 |
| 118.3 | 105 | 12421.0 | 11025 |
| 649.2 | 525 | 508270 | 42175 |

والان يمكننا إيجاد قيمة $\hat{b}, \hat{a}$
علي النحو التالي :

$$\hat{\hat{b}} = \frac{\frac{\sum xy}{n} - \frac{\sum x}{n} \cdot \frac{\sum y}{n}}{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

$$= \frac{\frac{50827}{7} - \left(\frac{649.2}{7}\right)\left(\frac{525}{7}\right)}{\frac{42175}{7} - \left(\frac{525}{7}\right)^2}$$

$$\frac{305.29}{400} = 0.76$$

وكذلك فأن :

$$\hat{\hat{a}} = \overline{x} - \hat{\hat{b}}\overline{y}$$

$$= \frac{649.2}{7} - (0.76)\left(\frac{525}{7}\right) = 92.24 - 54$$

$$= 35.74$$

وعليه تأخذ صورة معادلة خط إنحدار x علي y الصورة التالية :

$$\hat{x} = 35.74 + 0.764$$

العلاقة ما بين ميلي خطي الانحدار اي $\hat{b}$ , $\hat{\hat{b}}$ ومعامل الارتباط الخطي البسيط لبيرسون $r_{xy}$ وذلك علي النحو التالي :

فحيث أن :

$$\hat{b} = r_{xy} \frac{\sigma\overline{x}}{\sigma y}$$

$$\hat{\hat{b}} = r_{xy} \frac{\sigma\overline{x}}{\sigma y}$$

وعليه فإن: $\hat{b}\hat{\hat{b}}\, r_{xy}^2$

ومن ثم فإن:                .

وعليه فإن :   (٧- ١٩)   $r_{xy} = \sqrt{\hat{b}\hat{\hat{b}}}$

ويأخذ معامل الارتباط الخطي البسيط لبيرسون إشارة أي من معاملي الانحدار ومن خلال العلاقة (٧- ١٩) فإن يمكن القول أن:

(٢٠ - ٧)

$$\hat{b} = \frac{r_{xy}^2}{\hat{\hat{b}}}$$

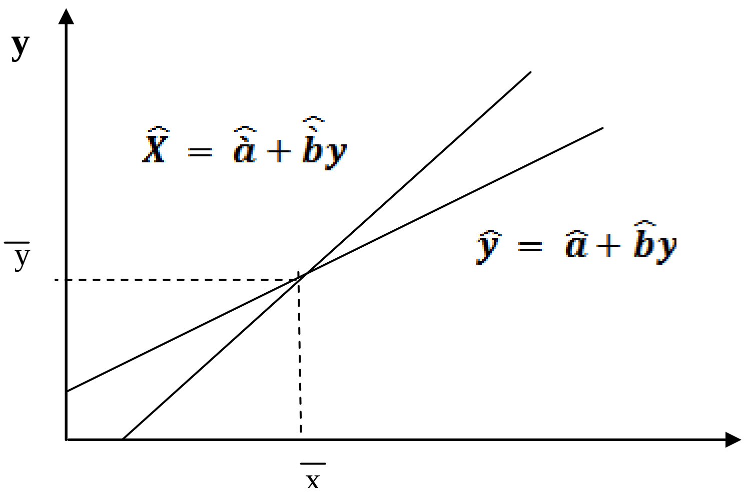
$$\hat{\hat{b}} = \frac{r_{xy}^2}{\hat{b}}$$

وأخيرا فأنه يمكن القول أن خطي الانحدار y علي x , x علي y ينطبقان علي بعضهما تمام الانطباق في حالة تمام العلاقة بين المتغيرين x , y اي حينما تكون قيمة $r_{xy}$ = ±1 وفيما عدا ذلك فأن الخطان يتقاطعان عند نقطة الاوساط الحسابية للمتغيرين أي $(\bar{x}, \bar{y})$ واللذان يحققان معادلتي خطي الانحدار بطبيعة الحال . لذا فأن جميع القيم المتوقعه للمتغير ( y) أي $(\hat{y})$ تساوي نظيراتها من القيم المشاهدة ( y) لهذا المتغير . وكذلك الامر بالنسبة للمتغير x وعليه فأن أنحدار y علي x يمكن كتابتها علي الصورة :

**Y = a + bx**

ويلاحظ هنا أننا كتبنا في الطرف الايمن y بدلا من $\hat{y}$ يحث أن القيم الفعليه تتساوي مع القيم المشاهدة ( أي أن $y_i = \hat{y}_i$ لجميع قيم i ) لوجود علاقة تامة بين x ,y ومن هذه المعادلة يمكن استنتاج معادلة خط إنحدار x علي y

وهما في واقع الامر معادلة واحدة في تلك الحالة ويمكن بيان ذلك من خلال الشكل البياني التالي :




هذا وتتوقف درجة اقتراب الخطية من بعضها علي قيمة ١ rxy ١ فكلما اقتربت هذه القمة من الواحد الصحيح كلما زادت درجة اقتراب الخطية من بعضها والعكس صحيح . وينطبق الخطان علي بعضهما تمام الانطباق ليصبحا معادله واحده وذلك في حالة وجود علاقة تامة بين المتغيرين أو الظاهرتين محل الدراسة.

# تمارين الباب السابع

**(١)- اذا كانت :**

$$X = 52 \ , \ y = 7 \quad , r_{xy} = 0.4 \quad , \sigma_x = 2 \quad , \sigma_y = 3$$

**اوجد معادلة انحدار y علي x**

**(٢)- في تحليل ٦٠ مفرده وجد أن معادلتي خطي الإنحدار هي كالاتي :**

$$1000y = 768x - 3608$$

$$5x = 64 + 24$$

**والمطلوب تحديد معامل الارتباط من هذه البيانات .**

**(٣)- الجدول التالي يبين الارقام القياسية للانتاج ( x ) والدخل القومي (y) بالملايين في الفترة من ١٩٤١حتي ١٩٤٧ في احدي الدول .**

| السنةYear | الرقم القياسي (x ) | الدخل القومي ( y ) |
|---|---|---|
| ١٩٤١ | ١٠٠ | ٥٠٠ |
| ١٩٤٢ | ١٠٥ | ٥٢٠ |
| ١٩٤٣ | ١٠٧ | ٥٨٠ |
| ١٩٤٤ | ٩٨ | ٦٠٠ |
| ١٩٤٥ | ١٠٩ | ٦١٠ |
| ١٩٤٦ | ١١٠ | ٦٠٠ |
| ١٩٤٧ | ١١٤ | ٦٥٠ |

**<u>والمطلوب</u> :**

١- ايجاد معامل الارتباط بين الظاهرتين .

٢- ايجاد معادلتي الانحدار ( Yنحدار y علي x وانحدار x علي y )

٣- الحصول علي القيمة المتنبؤ بها للدخل القومي عندما يكون الانتاج ١٢٠

٤- الحصول علي القيمة المتنبؤ بها للانتاج عندما يكون الدخل القومي ٧٠٠

(٤)- البيانات التالية تمثل الكميات المعروضة (x ) عند الاسعار المناظرة (y )

| | | | | | | |
|---|---|---|---|---|---|---|
| سعر السلعة (y ) | ٥ | ٢٨ | ١٠ | ٩ | ٢٠ | ١٥ |
| الكميات المعروضة ( x ) | ٩٠ | ٤٠٩ | ١٢٠ | ١٥٥ | ٣٧ | ١٥٣ |
| سعر السلعة (y ) | ١٢ | ٢٦ | ١٨ | ١٢ | | |
| الكميات المعروضة( x ) | ١٣٠ | ٤٥٢ | ٢٥٠ | ١٩٩ | | |

والمطلوب ايجاد معادلتي خطي الإنحدار .

(٥)- اذا فرضنا أن البيانات التالية : حيث ( x) مثل أعمار الرجال المتزوجين ، (y ) تمثل عدد مالديهم من أطفال

$\sigma y = 1.3535$ , $\sigma x= 5.5605$, $y= 2.09$ ,$x =$ ، $.484= r_{xy}$ , 33.175

والمطلوب

- حساب خط انحدار y علي x.
- حساب خط إنحدار xعلي y.

- **متوسط عدد الاطفال لرجل عمره ٣٠ سنة.**
- **متوسط العمر لرجل لديه طفلان .**

**(٦)- الجدول التالي يبين دخل مجموعة من ٨ أسر في الاسبوع (x ) وما تنفقه من خلال نفس الفترة (y ):**

| X | ٦٤ | ٥٢ | ٨٤ | ٦٤ | ٧٦ | ٥٦ | ٦٨ | ٥٤ |
|---|---|---|---|---|---|---|---|---|
| Y | ٥٢ | ٤٠ | ٦٠ | ٥٢ | ٦٠ | ٤٢ | ٥٠ | ٢٥ |

**والمطلوب :**

**١- معادلة انحدار الدخل علي الانفاق .**

**٢- معادلة انحدار الانفاق علي الدخل**

**٣- قدر إنفاق الاسر التي يبلغ دخلها ٧٠ جنيه في الاسبوع**

**٤- قدر دخل الأسرة التي تنفق ٤٨ جنبه في اسبوع**

**٥- معامل الارتباط بالطريقة العادية والتحقق من النتائج عن طريق معاملي خطي الإنحدار .**

**٦- إيجاد معادلتي الإنحدار بمعلومية معامل الارتباط وقارن النتائج التي حصلت عليها في (١) ، (٢ ).**

**(٧)- عرف كل من الارتباط والانحدار موضحا الفرق بينهما .**

**(٨)- من بيانات الانتاجية في العمل والخبرة فتح الجدول الاتي لعدد من ٧ عمال.**

| الانتاجية y | الخبرة بالسنوات X |
|---|---|
| ٧٠و٠ | ٣,٩ |
| ٤٩و٠ | ٣,١ |
| ٩٧و٠ | ٤,٩ |
| ٧١و٠ | ٣,٢ |
| ٤٥و٠ | ١,٨ |
| ٦١و٠ | ٢,٩ |
| ٣٩و٠ | ٢,١ |

**من البيانات السابقة احسب :**

**١- الوسط الحسابي لكل من المغتربين**

**٢- معامل الارتباط بينهما**

**٣- ماذا تستخلص من قيمة معامل الاتباط هل تحتاج لحساب معادلة خط الانحدار**

**(٩) -الجدول التالي يمثل توزيع عينية من ٥٠٠ شخص تبعا للطول (y) والوزن (x) :**

| Y \ X | اقل من ١٦٠ | ١٦٠- | ١٧٠- | ١٨٠- | ١٩٠ فأكثر | مجموع |
|---|---|---|---|---|---|---|
| اقل من ٦٠ | ١٥ | ٢٠ | ١٥ | - | - | ٥٠ |
| ٦٠- | ١٠ | ٥٠ | ٤٠ | - | - | ١٠٠ |
| ٦٥- | - | ٥٠ | ٤٠ | ١٠ | - | ١٠٠ |
| ٧٠- | - | ٣٠ | ٤٠ | ٣٠ | - | ١٠٠ |
| ٧٥- | - | - | ٤٠ | ٤٠ | ٢٠ | ١٠٠ |
| ٨٠- فاكثر | - | - | ١٠ | ١٥ | ٢٥ | ٥٠ |
| مجموع | ٢٥ | ١٥٠ | ١٨٥ | ٩٥ | ٤٥ | ٥٠٠ |

**والمطلوب :**

**١- إحسب معامل الارتباط الخط البسيط بين الوزن والطول من هذه البيانات**

**٢- استنتج معادلة الانحدار الخطي للطول علي الوزن بدلالة معامل الارتباط**

**٣- إستخدم المعدل التي حصلت عليها في تقدير طول أحد الاشخاص وزنه ٨٥كيلو جرام**

**٤- إستنتج معادلة إنحدار الوزن علي الطول بدلالة معامل الارتباط للبيانات المبوبة .**

**(١٠)- البيانات التالية توضح العلاقة بين اطوال وأوزان عشرة أشخاص :**

**Weight (x) : ٧٢ ٦٢ ٧٤ ٥٩ ٦٨ ٦٩ ٧٦ ٦٤ ٦٤ ٦٦**

**Length (y) : ١٥٧ ١٤٩ ١٨ ١٣٤ ١٥٦ ١٦٨ ١٧٩ ١٥٩ ١٣٤ ١٤٣**

**والمطلوب :**

**أ – إيجاد معادلة خط انحدار الطول علي الوزن.**

**ب- القيمة المتوقعة لــy عند x = ٨٠,٦٩**

**(١١)- البيانات التالية توضح الدخل الاسبوعي (x ) والاتفاق الاسبوعي علي الطعام (y ) وحجم المدخرات**
**الاسبوعية ( z ) وذلك لعينة مكونه من عشرة اسر :**

**X : ٩٨ ٨٢ ٧١ ٥٣ ٤٤ ٦٧ ٨٣ ٩٠ ٦٥ ٧٨**

**Y : ٥٣ ٤٢ ٣٨ ٢٥ ٢٨ ٣٢ ٤٩ ٧٠ ٤٥ ٥٢**

**Z : ١٠ ٧ ٣ ١٢ ٢ ٨ ١٣ ٥ ١٠ ٨**

**<u>والمطلوب :</u>**

**أ- حساب الانحراف المعياري لكل من x , y , z.**

**ب- حساب معاملات الارتباط الخطية البسيطة الاتية : $r_{xy}$ , $r_{xz}$ , $r_{yz}$.**

**ج - من نتائجك (ا) ، (ب) اوجد :**

**- معادلة خط انحدار حجم المدخرات الاسبوعي للاسرة علي الدخل الاسبوعي لها( أي معادلة خط انحدار z علي x ).**

**- معادلة خط انحدار الاتفاق الاسبوعي للاسرة علي الطعام علي الدخل الاسبوعي( أي معادلة خط انحدار z علي x )**

**- بين ايهما أكثر تأثرا بالدخل الاسبوعي ، الاتفاق الاسبوعي علي الطعام أم حجم المدخرات الاسبوعيه .**

**- من نتائجك في (ج) إحسب ما يلي :**

**- أخطاء التقدير في حساب كل من y , z وذلك عند القيم التالية لــ x : ٨٢ ، ٤٤ ، ٢٠**

- إحسب القيم المتوقعه لكل من y , z وذلك عند اليم التاليـة لــ x : ٧٥، ١٢٠، ٤٠

- ما هو الاثر الناتج علي حجم المدخرات الاسبوعية إذا نقص دخل الاسرة بمقدار عشرة جنيهات اسبوعيا ؟

(١٢)- البيانات التالية توضح تكاليف الاعلان وقيم المبيعات (بآلاف الجنيهات ) وذلك لإحدي الشركات في ثماني شهور متتالية :

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| الاعلان (x) : | ٢٨ | ٥١ | ٢٠ | ٤٦ | ٥٣ | ٣٥ | ٢٠ | ٢٥ |
| المبيعات( y ) : | ١٠٤ | ١٣٠ | ١٣٦ | ١١٣ | ١٥٠ | ١٧٥ | ١٢٧ | ١٠٨ |

والمطلوب :

أ- إيجاد معادلتي خطي إنحدار y علي x وانحدار x علي y

ب- ما هو الاثر الناتج علي المبيعات إذا نقصت تكاليف الاعلان بمقدار خمسة آلاف جنيه شهريا .

(١٣)- استخدام البيانات التالية عن الظاهرتين : x ، y وذلك لايجاد معادلتيي خط انحدارy علي x وانحدار xعلي y:

$\Sigma x = 50$ , $\Sigma y = 20$ , $r_{xy} = ٠,1037$ , $\sigma x = 1.392$ , $\sigma y = .866$ , $n = 8$

(١٤)- استخدام البيانات التالية عن الظاهرتين x، y لاجراء مقارنة بين التشتت النسبي في كل منها :

b (معامل انحدار yعلي x ) = ١,٨ ، $\frac{\Sigma y}{\Sigma x} = 2,5$

$r_{xy} = .72$ ,

(١٥)- اثبت أن معامل الارتباط الخطي بين x ، y ( $r_{xy}$ ) يمكن كتابته علي الصورة التالية

$$r_{xy} = b \ \frac{\sigma y}{\sigma x}$$

حيث b هي معامل انحدار y علي x و σx و σy هما الانحرافات المعيارية للمتغيرين x و y علي الترتيب .

(١٦)- بافتراض أن العلاقة بين المتغيرين بين Xو y علاقة خطية أوجد باستخدام البيانات المعطاه في الجدول التالي معامل الانحدار y علي x و معادلة انحدار x علي y . ثم احسب معامل الارتباط الخطي بين المتغيرين x و y

| X \ y | ٩٥ | ١٠٥ | ١١٥ | ١٢٥ |
|---|---|---|---|---|
| ٥٢.٥ | ٤ | ٧ | ٥ | ٢ |
| ٥٧.٥ | ٦ | ١٠ | ٧ | ٤ |
| ٦٢.٥ | ٦ | ١٢ | ١٠ | ٧ |
| ٦٧.٥ | ٣ | ٨ | ٦ | ٣ |

(١٧)- البيانات التالية تمثل درجات عشرة طلاب في كل مادتي الرياضيات (x ) والمحاسبة (y ) :

| Student | ١ | ٢ | ٣ | ٤ | ٥ | ٦ | ٧ | ٨ | ٩ | ١٠ |
|---|---|---|---|---|---|---|---|---|---|---|
| X | ١٨ | ١١ | ١٠ | ١٦ | ١٥ | ٢٠ | ١٣ | ١٦ | ١٤ | ١٢ |
| Y | ١٣ | ٨ | ١٤ | ١٠ | ١٣ | ١٨ | ١٧ | ١٢ | ١٢ | ١٥ |

**والمطلوب :**

**أ- حساب معامل الارتباط الخطي x ، y**

**ب- إيجاد معادلة انحدار yعلي x ثم حساب خطأ تقدير y عند القيمة x=15**

**جـ- من نتائجك في ( أ و ب) استنتج معادلة خط انحدار x علي y**

**د- هل تتفق مع الراي القائل بأن مستوي أداء الطالب السادس في مادة الرياضيات ( x ) كان أفضل منه في مادة المحاسبة ( y ) ؟**

**(١٨)- البيانات التالية توضح الدخل اليومي ( x ) والانفاق اليومي ( y ) وذلك لعينة مكونه من سبع أسر :**

| Incomes(x) | ١٧ | ١٣ | ٩ | ١١ | ١٥ | ٢١ | ١٩ |
|---|---|---|---|---|---|---|---|
| Expedition (y) | ٨ | ٦ | ٤ | ٥ | ٧ | ١٠ | ٩ |

**والمطلوب**

**أ – أوجد معادلة خط انحدار yعلي x**

**ب- إجسب القيم المتوقعه لـ ( y ) وذلك باستخدام المعادلة المحسوبة في (أ)**

**ج- علي ضوء نتائجك في (أ) ، (ب) إستنتج – دون حساب – معامل الارتباط الخطي البسيط x ، y**

**د- من نتائجك في أ ، ب ، ج استنتج أي اي من المتغيرين x ، y هو الاكبر تشتتا**

**هـ- من نتائجك في أ ، ج أوجد معادلة خط انحدار x علي y**

**(١٨)- ( عند اجراء العمليات الحسابية لهذا التمرين قرب نتائجك إلي أقرب اربعة ارقام عشرية ) .**

**في دراسة لتحديد العلاقة ما بين درجات محموعة من الطلاب في مادة الاحصاء ( x ) ودرجاتهم في مادة المحاسبة ( y ) ، وكانت معادلتا خط انحدار y علي x و x علي y علي النحو التالي**

$$Y = -5.9 + 1.35x$$

$$X = 4.51 + 0.73y$$

**المطلوب :**

**أ – احسب معامل الارتباط الخطي بين x ، y**

**ب – مستعينا بنتائجك في ( أ ) ، حدد أيهما الأكبر من حيث التشتت النسبي درجات الاحصاء x أم أم درجات المحاسبة ( y ) ؟**

**ج – إذا حصل طالب معين من بين محموعة الطلاب علي الدرجة ١٥ في كل من مادتي الاحصاء والمحاسبة ، هل تنفق مع الراي القائل بأن مستوي أداء الطالب في المادتين كان واحدا .**

**د – فيما يتعلق بالطالب المشار إليه في (جـ) خطأ تقدير درجته في كل من مادتي الاحصاء والمحاسبة ، ثم علق علي ضوء نتائجك في (أ)**

**(١٩)- البيانات التالية تعطي التوزيع التكراري المزدوج لخمسين من الازواج موزعين حسب عمر الزوج ( x ) وعمر الزوجه (y) :**

| X / Y | 25- | 35- | 45- | 55- | 65-75 | ∑ |
|---|---|---|---|---|---|---|
| 25- | 5 | 4 | 3 | | | 12 |
| 35- | 3 | 6 | 5 | | | 14 |
| 45- | | | 6 | 8 | 2 | 16 |
| 55-65 | | | | 2 | 6 | 8 |
| ∑ | 8 | 10 | 14 | 10 | 8 | 50 |

**المطلوب :**

**أ- حساب معامل الارتباط بين x و y**

**ب- اوجد معادلة خط انحدار عمر الزوجة y علي عمر الزوج x ثم احسب العمر المتوقع للزوجة عندما يكون عمر الزوج مساوياً ٤٠سنة**

**جـ- ما هو الاثر الناتج علي عمر الزوجة إذا نقص عمر الزوج بمقدار خمس سنوات**

**د- إستخدم أي معاملي بيرسون الاول أو اثاني للالتواء وذلك لتحديد شكل كل توزيع من حيث نوع ودرجة الالتواء .**

## المراجــع

١- د / أبو بكر عبد الرحمن عبد المتعال : (مبادئ الإحصاء) ، كلية التجارة ، جامعة جنوب الوادي ، ٢٠٠٦ .

٢- د / إبراهيم موسى عبد الفتاح : (مقدمة في الإحصاء الوصفي) ، كلية التجارة ، جامعة الزقازيق ، مكتبة المدينة بالزقازيق .

٣- د / أحمد عبادة سرحان : (طرق التحليل الإحصائي) ، معهد الإحصاء ، جامعة القاهرة .

٤- د / عبد المنعم ناصر الشافعي : (مابدئ الإحصاء) ، مكتبة النهضة المصرية .

٥- د / عبد المجيد فراج : (الأسلوب الإحصائي) ، دار النهضة العربية .

٦- د / صليب رفائيل : (مبادئ علم الإحصاء للتجاريين والاقتصاديين) ، كلية الاقتصاد والعلوم السياسية .

٧- د / محمد محمود حامد خطاب : (مقدمة في الإحصاء التطبيقي) ، كلية التجارة ، جامعة سوهاج ٢٠٠٤ .

٨- د / محمد عبد السميع العناني : (تطبيقات في الإحصاء الوصفي والتحليلي) ، كلية التجارة جامعة الزقازيق .

٩- د / سمير كامل عاشور ، د/ سامية أبو الفتوح سالم : (مقدمة في الإحصاء الوصفي) ، معهد الإحصاء ، جامعة القاهرة ١٩٩٠ .

١٠- د / هشام حسن مخلوف : (الإحصاء) ، مكتبة عين شمس ، القاهرة.

١١- د / نادية مكاري جرجس : (مذكرات في مبادئ الإحصاء) ، كلية الاقتصاد والعلوم السياسية .

١٢- د / محرم وهبي محمود : (النظرية الإحصائية وتطبيقاتها) الجزء الأول ، معهد التخطيط القومي .