"الإحصاء الوصفى "

لطلاب الفرقة الثانية

إعداد وتأليف

د. محمد عبدالحميد الربع

كلية التجارة

قسم الأساليب الكمية

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 على الإحصائي 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 . ويختص هذا الفرع 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

محمدعبد الحميد الربع

الهدف من تدريس مقرر الإحصاء الوصفى :

يهدف هذا المقرر الى:

(1): اكساب الطالب بمعارف ومفاهيم علم الاحصاء وفروعه وكيفيه اتمام عمليه جمع البيانات وعرضها بيانيا وجدوليا.

(2): تنميه قدره الطالب على كيفيه تلخيص الظواهر الكميه والوصفيه من خلال حساب مؤشرات رقميه تلخص طبيعه تلك الظواهر.

(3): تنميه ورفع قدره الطالب على اتمام البحوث العلميه ومعالجه المشاكل باسلوب علمى احصائى يتمتع بدرجه عاليه من الدقة والثقة .

(4): تنميه قدره الطالب الاحصائيه على معرفة ودراسه قوة العلاقة فيمابين المتغيرات او الظواهر بانواعها المختلفه وكيفيه تحديد شكل العلاقة الخطية فيما بين متغيرين كميين.

(5): تزويد الطالب بمعرفه عناصر المشكله السكانيه وطبيعتها وكيفيه حساب معدلات النمو السكانى.

(6): تنميه قدره الطالب على كيفيه دراسه التغيرات فى اسعار وكميات وقيم السلع الاقتصاديه من خلال دراسة الارقام القياسية .

محتويات المقرر

الفصل الأول
المفاهيم الأساسية في الإحصاء

(1-1):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 حيث يعتبر هذا العلم بمثابة مدخلا كميا للعلوم فى إتخاذ القرارات . ويمكن النظر إليه باعتباره أحد فروع الرياضة التطبيقية . كما يمكن تعريف علم الإحصاء 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 وإتخاذ 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 أيدى رياضيين مشهورين أمثال جالتون وباكون ولابلاس وجاوس وبرونولي. ومع نهاية القرن التاسع عشر وإستحداث نظرية الاحتمالات وتطورها فأصبحت بمثابة الأساس العلمي للأساليب الإحصائية مما أدى إلى ظهور الإحصاء الرياضي والتحليلي أو الاستدلالي وتطبيقاته في جميع مجالات أنواع المعرفة في الحياه وهو ما ساهم في دراسة خصائص

المجتمعات والتي لابد فيها من اتخاذ قرار محدد بالرغم من عدم توافر بيانات كاملة عن هذه المجتمعات وذلك من خـلال عينـات مسـحوبة مـن تلـك المجتمعات. هذا وبعد التطور المتسارع فى علم الإحصاء وأساليبه المختلفة فقد أصبح لهذا العلم أهمية كبرى في كثير من المجالات الحيويـة المختلفـة وذلك لأن الأسلوب الإحصائي أصبح قادراً على بحث وتقدير المخاطر التـي تنجم عن اتخاذ قرار معين. ففي الدراسات الطبية والدوائية مـثلا يسـتخدم الاسلوب الإحصائي لدراسة وتحديد أسباب انتشار الأمراض المختلفة ومدى فاعلية الأدوية المختلفة في علاجها ، هذا بالإضافة إلـى دراسـة العوامـل الوراثية وتأثيرها ...

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

وفي الأبحاث الزراعية أيضا يتم استخدام الإحصاء في عملية تصميم وتحليل نتائج التجارب التي قد تحدد أفضل أنواع السـماد أو المبيـدات الحشـرية الواجب استخدامها أو أثر اختلاف نوع التربة وطريقة الـري علـى كميـة المحصول من أحد المنتجات ...

وكذلك نجد أن التحليل الإحصائي هو الأساس الكبير في الدراسات الطبيعيـة والكيميائية وأبحاث الفضاء والذرة .. إلى غير ذلك من كافة فـروع أنـواع المعرفة الحديثة . هذا ويضاف إلى تلك الزيـادة الملحوظـة لاسـتخدامات الإحصاء فـي العلـوم الاجتماعيـة كالدراسـات الاقتصـادية والسياسـية واستطلاعات الرأي العام وأبحاث الاجتماع وعلم النفس ، كمـا يجـب 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 ":

تعتبر عملية جمع البيانات الخاصة حول أى مشكلة موضع دراسة بطريقة علمية صحيحة ودقيقة بعيدة عن التحيز حتى يسهل استخدامها في عملية التحليل الإحصائي للظاهرة موضع الدراسة بمثابة هدفا رئيسيا لعلم الإحصاء.

الهدف الثاني :- " عرض البيانات الإحصائية "

حيث 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 أو من خلال حساب بعض المقاييس الإحصائية التي تستخدم لتفسير الظاهرة محل الدراسة .

الهدف الثالث : " تحليل البيانات الإحصائية : -

فبعد عملية جمع وعرض البيانات الإحصائية سواء جدوليًا أو بيانيًا أو من خلال حساب 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 تلـــك الظواهر .

هذا وهناك مجموعة من التعريفات الأساسية اللازمة لدراستنا فـــي مبـــادئ الإحصاء الوصفي لعل من أهمها ما يلي :

أولاً : المجتمع الإحصائي والعينة : Statistical Population & Sample

تختلف معنى كلمة المجتمع عند الإحصائي عن المعنى الشائع الاســتخدام . 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 والمعرفـــة بحدود زمنية أو مكانية تشترك في صفة أو مجموعة خصائص محددة يمكن ملاحظتها أو قياسها كمياً .

فمثلاً إذا كان هدف الدراسة الإحصائية هو تحديد متوسط أجر ساعة العامـــل المصري . فإن المجتمع الإحصائي هو عبارة عن أجـــور جميـــع العـــاملين المصريين في الساعة . وإذا كان هدف الدراسة هو تحديد نســـبة الوحـــدات المعيبة في هذا الخط الجديد من خطوط الإنتاج فـــالمجتمع الإحصـــائى هـــو عبارة عن كل الوحدات المنتجة فى هذا الخط الجديد. وكذلك إذا كان هـــدف 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.

فالخلاصة أنه يمكن تعريف المجتمع الإحصائي بأنه عبارة عـــن مجموعـــة المفردات أو العناصر المحددة والمعرفة بحـــدود زمانيـــة ومكانيـــة معينـــة وتشترك في صفات وخصائص محددة يمكن ملاحظتها وقياسها كمياً . ومهما

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

<u>الأول</u> : <u>مجتمع إحصائي محدود</u> : **Limited or Bounded Population**

وهو المجتمع الذي يحتوي على عدد محدود من الوحدات أو الأحداث القابلة للعد أو الحصر ومن ثم القابلة لتحديد الحجم . ومثـــال للمجتمـــع المحـــدود مجموعة الطلاب المصريين لعام دراسي معين فالمجتمع الإحصائي هنا مهما كبر حجمه ليصل للآلاف أو حتى الملايين من المفردات إلا أنه قابل للحصـــر وبالتالي للعد وتحديد الحجم . فما دام هناك إمكانية لتحديد إطار لمفردات أو عناصر المجتمع الإحصائي يتم تصنيفه على أنه مجتمع إحصائي محدود .

<u>الثانـي</u> : <u>مجتمع إحصائي غير محدود</u> **Unlimited or Unbounded Population**

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

هذا وتسمى الخصائص الدالة على أي مجتمع إحصائي والتي يمكن ملاحظتها أو قياسها كمياً باسم <u>معلمات المجتمع</u> **Parameters** . وغالباً ما تكـــون القيم الحقيقية لمعلمات المجتمع الإحصائي مجهولة (غير معلومة). إلا أنـــه يمكن تقديرها من خلال الاستدلال عنها وذلك من خلال سحب عينة عشوائية بطريقة إحصائية سليمة لكي تكون ممثلة لهذا المجتمع تمثيلاً تاماً (جيـــداً) . وتجري الدراسة الإحصائية على تلك العينة لإستخلاص مايسمى <u>بإحصـــاءات العينة</u> ثم بإستخدام طرق أو أساليب الإحصاء التحليلي (الاســـتدلالي) يـــتم

تقدير معلمات المجتمع وإجراء إختبارات إحصائية حول تلك المعلمات وهـــو ما سيرد دراسته في الفصل الدراسي الثاني بمشيئة الله تعالى .

(2-1):خطوات جمع البيانات الإحصائية : ـ

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

1-تحديد هدف ومشكلة مجتمع الدراسة : فإن تحديــد هــدف البحـــث هـــو الخطوة الأولى للقيام بأي بحث أو دراسة إحصائية. فبناءًا على هذا الهـــدف يتحدد "المجتمع" محل البحث وكذلك تتحدد "المفردة" التي يجب جمع البيانات عنها ونوع ومصدر البيانات والأسلوب الذي يجب استخدامه في جمع تلـــك البيانات .

فمثلاً إذا كان الهدف من بحث ما هو المقارنة ما بين كيفية إنفـــاق دخـــل الأسرة في ريف ج.م.ع بكيفية إنفاق دخل الأسرة في حضـــر ج.م.ع خـــلال فترة زمنية معينة . فإن المجتمعات محل الدراسة تكون هي كلاً مـــن ريـــف وحضر ج.م.ع خلال هذه الفترة الزمنية المحددة . كمـــا أن المفـــردة التـــي تتكون منها هذه المجتمعات والتي يجب أن تُجمع عنها البيانات هي "الأسرة" والبيانات المطلوبة هي دخل الأسرة في تلك الفترة الزمنية المحددة وكيفيـــة توزيعه على مصادر الإنفاق المختلفة .

فالمجتمع الإحصائي هو كما ذكرنا سابقًا عبارة عن مجموع المفردات التـــي يجب أن تُجمع عنها البيانات . كما أن المفردة محل البحث هي وحدة جمـــع البيانات والمفرده محل البحث فقد تكون سلعة منتجة بطريقة معينة أو هـــي القرية ذات المساحة المحدودة أو هي الطفل في سن معين .... إلخ . فمـــن

الواضح أن تحديد المجتمع وكذلك مفرده البحث يتوقف على نوع الدراسة أو الهدف من البحث .

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

أ- مصادر تاريخية : 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- مصادر ميدانية : 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 العينة **Sample** الإحصائية هي عبارة عن مجموعة من المفردات مسحوبة من مجتمع إحصائي حول ظاهرة معينة . ولكي يتم سحب عينة عشوائية بطريقة إحصائية سليمة تعطي لكافة مفردات المجتمع الفرصة المتساوية لكي يتم تمثيلها في العينة يجب وضع قائمة تحتوي علـــى كافـــة مفردات المجتمع محل الدراسة والتي تسمى بالإطار **Frame** والذي يجـــب أن يتوافر فيه الدقة والشمول وعدم التكرار . ثم تجري عملية سحب متتالي لمفردات العينة من مجتمع الدراسة سواء يدويًا أو آليًا أو باستخدام جـــداول الأرقام العشوائية .

3– 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 صـــحيفة الاستقصاء (**Questionnaire**) ويجب أن تراعي بعض القواعد عند وضع هذه الأسئلة . فمثلاً يجب أن تكون مختصرة بقدر الإمكـــان وأن لا تتضـــمن أسئلة محرجة . كما يجب أن تتجنب تلك الاستمارة الأســـئلة ذات الإجابـــات المعتمدة على الحكم الشخصي للفرد (فمثلاً بدلاً من سؤال الفرد المبحوث عن مستوى تعليمه يمكن أن يطلب منه وضع علامة أمام أحد الحالات التعليميـــة المعروفة : لا يقرأ ولا يكتب ، يقرأ فقط ، حاصل على شـــهادة متوســـطة ، .... إلخ...) . كما يجب وضع الأسئلة في صورة يســـهل الإجابـــة عليهـــا بطريقة واضحة .

هذا وبعد التأكد من صلاحية استمارة البحث وبعـــد أن تـــتم عمليـــة تهيئـــة المجتمع لعملية جمع البيانات (وذلك بالإعلان عن أهمية البحـــث وضـــرورة الدقة في الإجابة ومدى سرية البيانات) وبعد أن يتم تدريب الأفـــراد اللـــذين

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 .

(3-1):وسائل جمع البيانات :

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 عبر التليفون . وهذه الطريقة تتميــز بالبســاطة والســهولة والسرعة وقلة التكاليف . لكن أهم ما يعيبها أنه ربما تتضمن العينــة محــل البحث مجموعة من المفردات ليس في حوزتهم تليفونات .

أما عن وسيلة المقابلة الشخصية وهي أكثر الطرق شيوعاً واستخداماً فــي عملية جمع البيانات من الميدان . وهي تعتمد على أن يقوم جامع البيانـــات بالذهاب إلي الفرد المبحوث للحصول على اجابات للأسئلة الموجــودة فــي استمارة البحث .

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

وتعتمد هذه الوسائل عادة على الاستجابات التي يمكن تحليلهــا فيمــا بعــد لاستنتاج نتائج البحث والتى لا تعتمد على الاستجابة مباشرة . وتنتشر مثل هذه الوسائل خصوصاً في مجال الصحة والبحوث النفسية وبحوث الــدوافع حيث يلجأ الباحثون إلى العمليات العقلية اللا شعورية . فعلى ســبيل المثــال طريقة التداعي الحر ووصف عينة من خلال السلوك أو الإسقاط وغيرهــا . والهدف منها التوصل إلى حقيقة نفسية ودوافع المبحوث عن طريــق غيــر مباشر أي من غير طريقة السؤال المباشر فيما يختص بالسؤال المطلوب عن الموضوع المطلوب الإجابة عليه .

<u>(4-1):أساليب جمع البيانات:-</u>

عند جمع البيانات الاحصائية يمكن استخدام إما اسلوب الحصر الشــامل أو اسلوب العينة (المعاينة) . فاستخدام اسلوب الحصر الشامل لجمع البيانات معناه جمع البيانات من جميع مفردات المجتمع . أما اســلوب العينــة فهــو

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 قد يفضل أسلوب العينة علي أسلوب الحصر الشامل إذا كانت البيانات المطلوبة باستخدام أسلوب الحصر الشامل 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 :-

(1):<u>خطأ التحيز</u> :- وينشأ هذا النوع من الخطأ نتيجة تحيز الأفراد – لا شعورياً – لبعض القيم أو البيانات فهو ينجم عن عملية اختيار بعض مفردات المجتمع دون البعض الآخر عند جمع البيانات منها . 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 .

(2): <u>خطأ الصدفة العشوائى (الخطأ العشوائى):</u>- ويظهر هذا الخطأ سواء عند استخدام اسلوب العينة أو أسلوب الحصر الشامل وإن كان في أسلوب الحصر الشامل بدرجة أعلي . وهذا النوع من الأخطاء ليس له اتجاه ثابت

ويمكن التحكم فيه وتقليل آثاره وذلك إذا ما تم اختيار العينة بطريقة علميـــة أو إحصائية سليمة . وكذلك فإنه يمكن استخدام الأساليب الرياضية لتقـــدير مدى تأثير هذا النوع من الأخطاء على نتائج الدراسة .

وأخيراً فإنه يمكن القول أنه كون العشوائية (الصدفة) تحكم عملية اختيـــار وحدات المعاينة الإحصائية (العينة) فإن نتائج عملية التحليـــل الإحصـــائي ستختلف باختلاف العينات الاحصائية (حجماً أو نوعاً أو عدداً) والتي يـــتم 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 المعاينـــة الإحصائية Sampling Error . ويحكم حجم خطأ المعاينة مجموعة من العوامل وهي :-

<u>(1): نوع العينـة</u> : حيث يفضل استخدام العينات العشوائية الاحصائيةالبسيطة والتى يكون فيها لكل مفردة من مفردات مجتمع الدراسة لها نفس فرصـــة الاختيار والتمثيل ضمن مفردات العينة .وفيها يمكن التحكم وقياس أخطـــاء المعاينة العشوائية على خلاف العينة العمديـــة Judgement Sample والتي يختار الباحث مفرداتها بشكل شخصي (قصدي) من مجتمع مشـــكلة البحث .

(2): **<u>حجم العينـة</u>** :- حيث توجد علاقة عكسية فيما بـــين حجـــم أخطـــاء المعاينة وحجم العينة المختارة . كما أن هناك علاقة عكسية مابين درجـــة الدقة الاحصائية وبين حجم العينة من جانب وحجـــم الميزانيـــة المرصـــودة للبحث من جانب آخر .

(3): **<u>تباين المجتمع</u>** :- فكلما زاد التباين(الأختلاف)فيمـــا بـــين مفـــردات المجتمع فإن ذلك يتطلب زيادة حجم تمثيل العينة المســـحوبة مـــن مجتمـــع

مشكلة البحث وذلك لتقليل أخطاء المعاينة العشوائية . وإن كان نوع وحجم العينة المسحوبة يؤديان إلى تحسين عملية الحصول على عينة احصائية جيدة .

وفيما يلي الجدول التالي يبين الفرق بين أسلوبي جمع البيانات : الحصر الشامل والعينة :

الفرق ما بين أسلوب الحصر الشامل والعينة:

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

خطوات (منهج) البحث الاحصائي (أو العلمي) :-

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 والإثبات 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 :

الأول : تعريف المشكلة : -

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 ظاهرة 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 .

الثاني : وضع الفروض الإحصائية :

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</u> : <u>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 و الميداني) . فمن قبل أن يلجأ الباحث إلى عملية جمع البيانات من خلال المصادر الميدانية بوسائلها المتعددة فإنه يجب الحرص عند استخدام البيانات تجنب الوقوع في مصادر أخطاء البيانات وذلك 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</u> : <u>تحليل البيانات:</u>

من أساسيات تحليل البيانات هو تفهم الاختلافات فيما بينها 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 (أو المتغيرات) محل الدراسة .هذا ولقد ساعد التطوير المتسارع في الحاسبات الآلية (خاصة على مستوى البرمجيات الإحصائية الجاهزة) في تحقيق السهولة واليسر ، والدقة العالية وذلك 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>

تعتبرعملية اختبارات صحة الفروض الإحصائية الموضوعة علـــى ضـــوء نتائج مخرجات عمليات التحليل الإحصائي المستخدمة ، وهي احتماليـــة (أي إمكانية أن تكون صوابًا أو خطأ بمستويات ثقة محددة) من أهم آليات تفسير النتائج. وتتم عملية التفسير لمخرجات عمليات التحليل الإحصـــائي علـــى مستويين ، وهما :

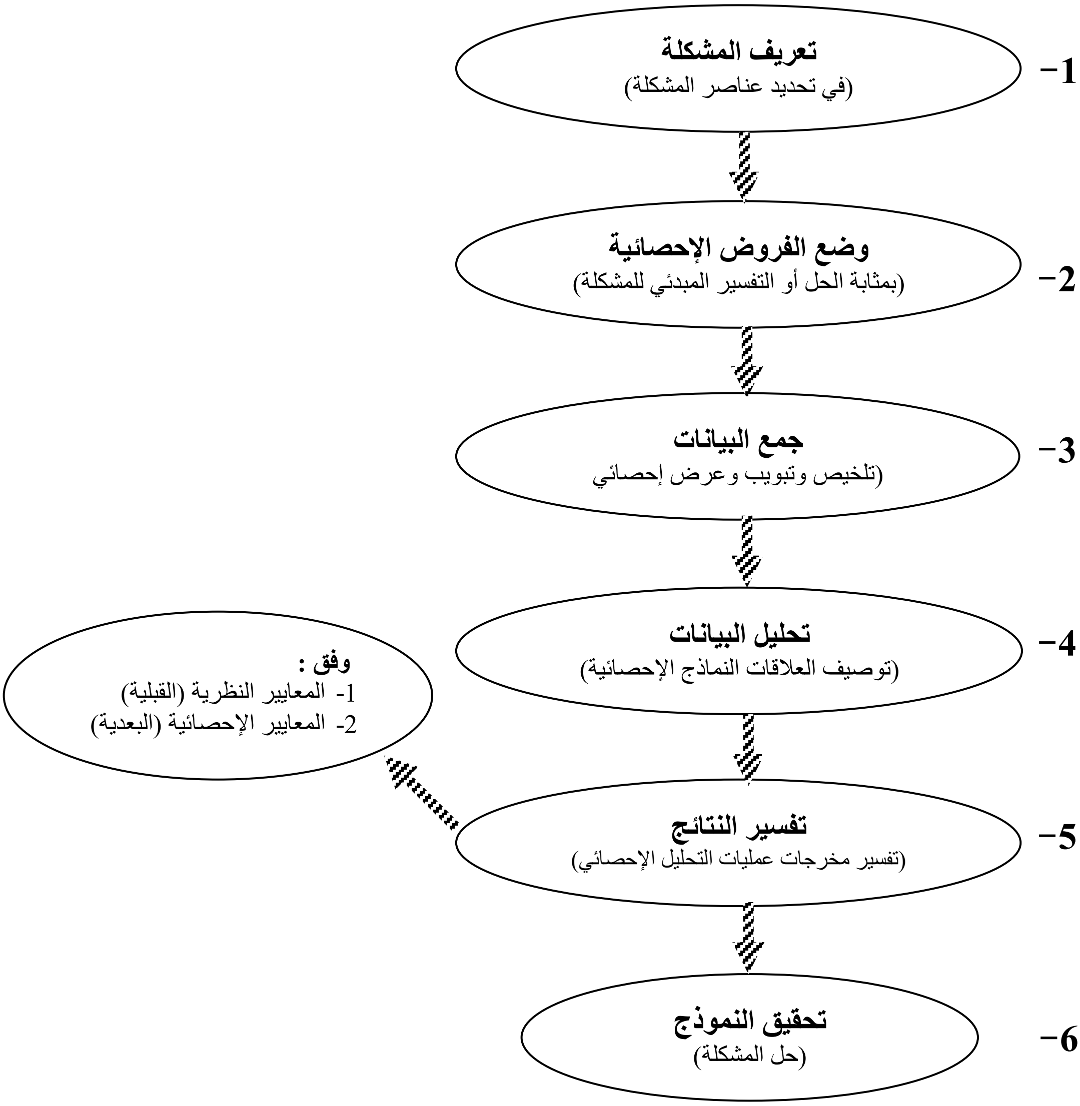
<u>(1):معيار قبلي : Prior</u> : يسترشد الإحصائي بالمعايير النظرية الحاكمـــة لعلاقات الظاهرة أو المتغيرة تحت المعالجات الإحصائية ، ويقع عـــبء هـــذا الدور أساسًا على المتخصص في مجال الدراسة ، فقد يكون : ديمـــوجغرافي أو اجتماعي أو اقتصادي أو سياسي أو نفسي أو تربوي أو طبـــي ...... أو غيره ، في مجالات العلوم وأفرع المعرفة المختلفة .

<u>(2):معيار بعدي Posterior</u> :

وهنا يجب التنويه إلى أن المعايير الإحصائية (البعدية) هي داعـــم أو مؤكـــد للمعايير النظرية القلبية ولكن لا تجبها .

<u>السادس : تحقيق النموذج:–</u>

يأتي علاج المشكلة تحت الدراسة بتحقق صحة الفروض النظرية الموضوعة باحتمال معين وذلك على ضـــوء مخرجـــات عمليـــات التحليـــل الإحصـــائي المستخدمة ، وغالباً ما تصاغ في صورة علاقات إحصائية تتوافر فيها القدرة على تفسير سلوك الظاهرة تحت التفكير الاحصائي في الماضـــي والحاضـــر 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 ، فـــي الشـــكل التوضيحي الآتي :–



شكل (1-1)

توضيحي عام لخطوات التفكير الإحصائي

(5-1):التحليل الاحصائي واستخدام الحاسب الآلي:–

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 عبأه 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 .

(1-5-1):خطوات التحليل الاحصائي باستخدام الحاسب الآلي:

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 وذلك من خلال الخطوات التالية :

الأول : تفريغ البيانات: وفيها تتم عملية تفريغ البيانات في استمارة خاصة يخصص فيها لكل متغيرة عدد من الحقول(يراعى أكبر قيمة ممكنة للمتغيرات) على أن يكون قد تمت المعالجات اللازمة لبعض مشاكل البيانات .

(1): عملية الترميز Coding : وتتم عملية الترميز عند التعامل مع الظواهر أو المتغيرات الوصفية مثال الإجابات وذلك بمستوياتها المختلفة على مستوى السؤال الواحد ، في استمارة الاستبيان .

(2): القيم المفقودة : والتي تنتج عن حالة النسيان أو الرفض أو الغير صحيحة لبعض إجابات استمارة الاستبيان . وتتم معالجة القيم المفقودة اما بالاستبعاد أو بإعطاءها رمز خاص . ويجب التعامل معها بطريقة ملائمة لموضوع الدراسة . كما يجب التأكيد على ضرورة المراجعة الدقيقة لعملية تفريغ البيانات كون درجة سلامتها تنعكس بشكل مباشر على درجة الدقة الإحصائية لمخرجات عمليات التشغيل الاحصائي المختلفة إن لم تكن خاطئة .

الثاني : إنشاء ملف البيانات :

وهي عملية إدخال البيانات على الحاسب في ملف يسمى بملف البيانات وفق قواعد البرنامج الاحصائي الجاهز المستخدم مع التأكيد على أهمية مراجعة عملية إدخال البيانات للتأكد من سلامتها خاصة البرمجيات 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 .

الثالث : وصف البيانات : تبدأ عملية التحليل الاحصائي للبيانات بخطوة وصف البيانات ، مستخدمين الجداول التكرارية والرسوم البيانية والمقاييس الاحصائية الأولية (المتوسطات ، التشتت ، الالتواء ، التفلطح) وذلك 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 يتم الآتى:

<u>(1):اختبارات الفروض الاحصائية</u> :

يتم اختبار مدى صحة الفروض الاحصائية المقابلة للفروض النظرية والتوافق فيما بينهما ضروري مستخدماً فيها الباحث الاحصائي الأساليب والطرق الاحصائية المعملية Parametric أو اللا معلمية Non-Parametric الملائمة للبيانات وهدف الدراسة للحكم باحتمالية على مدى صحة الفروض الاحصائية من عدمه (وفق المعايير الاحصائية) .

<u>(2):قياس العلاقات (أو النماذج) الاحصائية</u> :

يهدف التحليل الاحصائي في نهاية المطاف للتوصل 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 مستخدمين من الأساليب والطرق الاحصائية المتنوعة مثال : تحليل الانحدار (البسيط أو المتعدد) ، وتحليل التباين(في اتجاه واحد أو أكثر) ، وتحليل التغاير ، والتحليل العاملي ، وتصميم التجارب ، .... وغيرها ، مع ملاحظة أنه يمكن عامة تقسيم مخرجات عمليات التحليل الاحصائي إلى ثلاث مجموعات ، وهي :

أ – وصف البيانات .

ب- اختبارات الفروض الاحصائية (المعلمية أو اللامعلمية) .

جـ – قياس العلاقات (أو النماذج) الاحصائية .

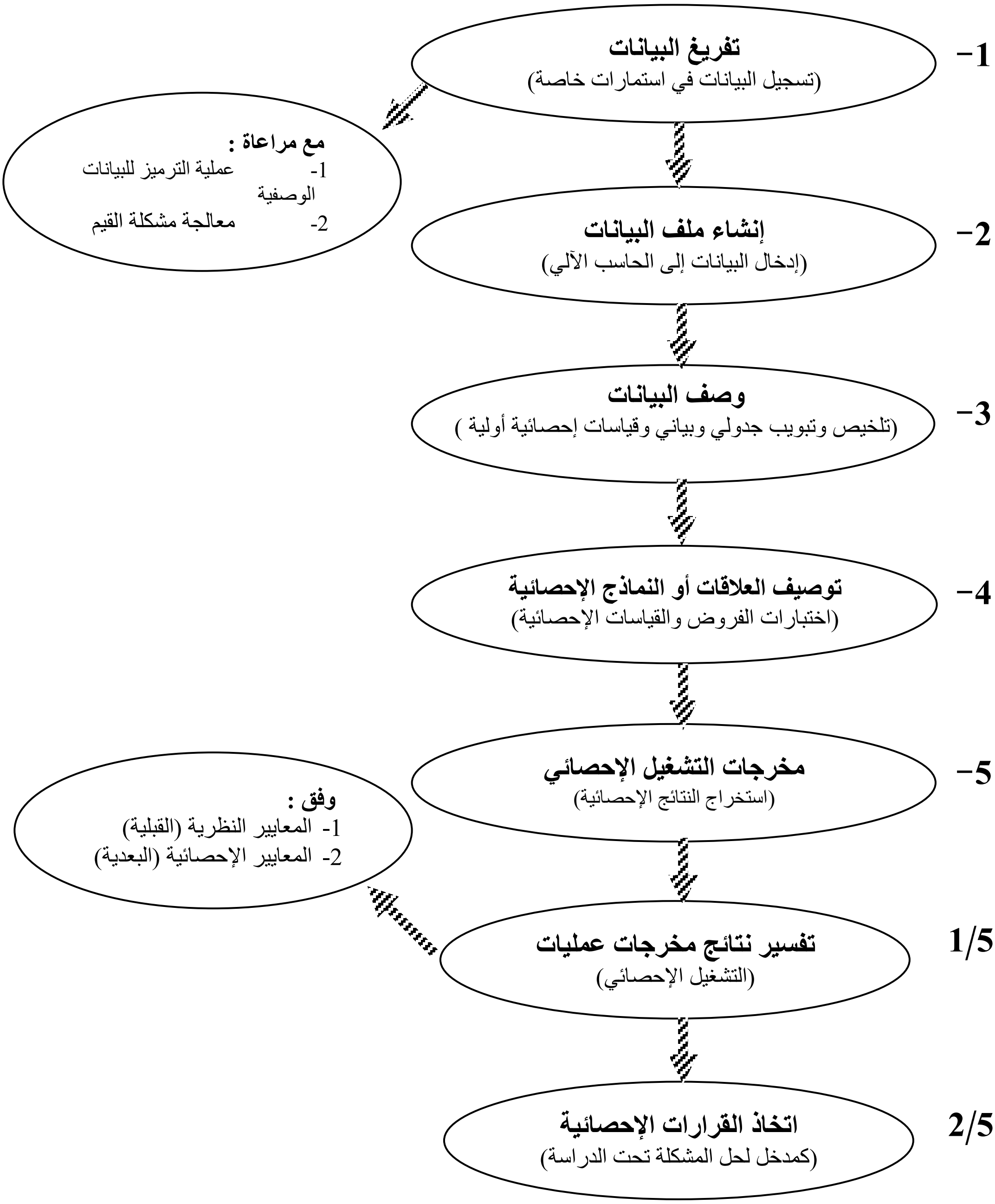
<u>الخامس : مخرجات عملية التشغيل الاحصائي :</u>

ويتم التعامل مع مخرجات التشغيل الاحصائي المختلفة على مستويين :

(1):تفسير نتائج مخرجات عمليات التشغيل الاحصائي (وصف البيانـــات ، اختبارات الفروض ، قياس العلاقات) وذلك على ضوء : المعايير النظرية (القبلية) الحاكمة لعلاقات النظرية للظــاهرة أو المتغيــرة محــل التحليــل الاحصائي ، وبدعم المعايير الاحصائية الحاكمة للطرق والأساليب الاحصائية المستخدمة .

(2):اتخاذ القرارات الاحصائية على ضوء التفسير الســابق كمــدخل لحــل المشكلة تحت التحليل الاحصائي .

ويمكن إيجاز خطوات التحليل الاحصائي باستخدام الحاسب الآلي في الشــكل التالي :



شكل (1-2)

توضيحي عام خطوات التحليل الإحصائي باستخدام الحاسب الآلي

(1-5-2):البرمجيات الاحصائية الجاهزة :

ويمكن تناول تعريف البرمجيات الاحصائية الجاهزة تحت نوعيين رئيسيين :

أ- **البرمجيات الإحصائية البسيطة :**

وهي برمجيات لا يتطلب تشغيلها كتابة أوامر معينـــة ، بـــل تعمـــل تحـــت مجموعة من الاختيارات والتي تشمل على العديد من الأســـاليب ، والطـــرق الإحصائية (الأقل تعقيدًا) ومنها البرنامج الإحصائي الجاهز Microstate .

ب- **البرمجيات الإحصائية الأكثر تعقيدًا :**

وهي برمجيات يتطلب تشغيلها كتابة مجموعة أوامر (حسب نـــوع العمليـــة الإحصائية تحت الاستخدام) ، وذلك على مستوى البرمجيات التي كانت تعمل تحت نظام التشغيل Dos وبعد ظهور نظام التشـــغيل Windows ســـارعت شركات البرمجيات الجاهزة ، إلى إعادة توفيق برامجها لتعمل تحـــت نظـــام التشغيل الجديد ، مما جعلها أكثر سهولة ومتعة لدى المســـتخدمين ، حيـــث التعامل أصبح يتم من خلال الاختيارات بالتأشير باستخدام الماوس Mouse على مجموعة قوائم (وفق مسارات محددة ، ومن خلال مجموعة من النوافذ المتتابعة الظهور على سطح المكتب Desktop) ، والمسئولة عـــن تنفيـــذ العديد من الوسائل والطرق الإحصائية (المعلمية أو اللامعلمية) بـــدءًا مـــن عمليات وصف البيانات ، إلى توصيف العلاقـــات (أو النمـــاذج) الإحصـــائية والأكثر تعقيدًا ، ومنها الحزم الإحصائية الجاهزة :

Bmdp , Minitab , Sas , Spss ..... وغيرها .

تعريف البرنامج الإحصائي الجاهز Spss :

وهو برنامج إحصائي متكامل يقوم بالعديد من عمليات التحليـــل الإحصـــائي البسيطة والمتقدمة ، وعلى الأخص في مجال العلوم الاجتماعية (الظواهر أو المتغيرات الوصفية)بداية من وصف البيانات إلى التحليل الإحصائي المتعدد .

(1-5-3):أهم العمليات الإحصائية تحت البرنامج الإحصائي Spss :

(1): Reliability Analysis (لقياس ثبات ، ومصداقية البيانات) .

(2): Store (للترتيب التصاعدي أو التنازلي للبيانات) .

(3): Frequencies (لعمل الجداول التكرارية ، العرض البياني ، حساب مقاييس إحصائية أولية (الموضع أو المتوسطات ، التشتت ، الإلتواء والتفلطح).

(4):Cross tabs (لعمل الجداول التكرارية المزدوجة).

(5): Descriptive (لحساب العديد من المقاييس الإحصائية الأولية للبيانات).

(6): Tests of Hypotheses (لإجراء إختبارات الفروض الإحصائية المعلمية) .

(7): NPAR (لإجراء اختبارات الفروض الإحصائية اللامعلمية) .

(8): Graph (لرسم أشكال البيانية للمتغيرات وكذا أشكال الأنتشار) .

(9): Correlation (لقياس قاهية قوة العلاقات فيما بين المتغيرات وذلك من خلال حساب معاملات الارتباط بين المتغيرات الكمية أو الوصفية).

(10): Regression (لتحديد الشكل الرياضي لنماذج الانحدار الخطي البسيط أو المتعدد وكذلك الأنحدار غيرالخطي بين المتغيرات) .

(11): Cluster Analysis (التحليل العنقودي) .

(12): Discriminant Analysis (تحليل التمايز) .

(13): Factor Analysis (التحليل العاملي) .

(14): Logistic Regression (الانحدار اللوجيستي) .

(15): Curve Estimation (توفيق المنحنيات ، ومنها السلاسل الزمنية Time Series) . وعمليات أخرى متنوعة .

مع ملاحظة : أهم العمليات على البيانات :

(1): 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 .

(2): يمكن قصر التحليل الإحصائي على مجموعة جزئية من المتغيرات .

(3): يمكن فصل مجموعة من المتغيرات في ملف بيانات مستقل .

(4): 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 .

(1-6):أنواع العينات :

سبق وأن ذكرنا أن العينة هي عبارة عن مجموعة مختارة من مفردات 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 عن بيانات الحصر الشامل . ويمكن بدراسة خصائص العينة (أو ما يسمى باحصاءات العينة) استنتاج الاحصائيات 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 :-

1- العينة العشوائية البسيطة Simple Random Sample

وهي أبسط أنواع العينات ويسخدم هذا النوع من العينات في حالـــة إختيـــار عينة من مفردات مجتمع تتسم بدرجة عالية جدا من التجانس(عدم التشتت). وتتميزالعينة العشوائية البسيطة بأنها تعطي فرصاً متساوية لجميع مفـــردات المجتمع للظهور أو التمثيل في العينة ، كما أن عامل الصـــدفة يكـــون هـــو العامل الوحيد المحدد للمفردات التي ستتم إختيارها بالعينـــة . ولـــذلك فـــإن اختيار مفردات العينة يتم بأي طريقة عشوائية : (أي بطريقة تضمن تساو ي فرص الظهور في العينة أمام جميع مفردات المجتمع) . فمثلاً عند اختيـــار طالبين من خمسة عشر طالباً بطريقة عشوائية يمكن اســـتخدام 15 بطاقـــة متماثلة تماماً يكتب على كل منها اسم أحد الطلبة ثم تخلط البطاقـــات جيـــداً وتسحب إحداها فيكون الاسم المدون بها هو اسم أول طالب في العينـــة ثـــم تخلط البطاقات الباقية جيداً ويسحب منها بطاقة أخرى ويكون الاسم المدون بها هو اسم الطالب الثاني في العينة ، هذا 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 يـــؤدي إلى ارتفاع نفقات جميع البيانات . ويضاف إلى ذلك أنه إذا كـــان المجتمـــع محل البحث كبيراً فإن عملية الاختيار العشوائي تكون عملية مطولة ومملـــة وللتغلب على هذه العيوب تستخدم أنواع أخرى من العينات .

2- العينـة العشوائية المنـتظمة Systematic Random Sample

والهدف من استخدام هذه العينة هو تجنب التكرار المطول الذي يظهر عنـــد استخدام العينة العشوائية البسيطة . ويعتمد اختيار مفـــردات هـــذه العينـــة

أساساًعلى تقسيم المجتمع إلى فئات متساوية (عددها يساوي عدد مفـــردات العينة) ثم يتم اختيار مفردة واحدة من الفئة الأولـــى بطريقـــة عشـــوائية واختيار باقي المفردات من الفئات المتتالية بطريقة منتظمة .

فمثلاً : إذا كان المجتمع مكون من 480 مفردة وإذا كان المطلوب اختيـــار عينة حجمها 60 مفردة ، فإننا نقسم المجتمع إلى 60 فئة يتكون كل منهـــا من 8 مفردات : (60/480 = 8) وتكون أرقام مفردات الفئة الأولى عبارة عن : (1 ، 2 0000 ، 8) وأرقام مفردات الفئة الثانية هي : (9، 10 ، .. ، 16) .... وهكذا . ثم نبدأ باختيار أحد مفردات الفئـــة الأولـــى بطريقـــة عشوائية ولنفرض أنها كانت المفردة رقم 4 ، فتكون الخطوة التاليـــة هـــي اختيار مفردة من الفئات المتتالية الثانية والثالثة والرابعة ..... وهكذا حتي الفئة الستون بطريقة منتظمة ، فتكون المفردة المسحوبة من الفئة الثانيـــة رقمها هو (4 + 8 = 12) ثم المفردة المسحوبة من الفئة الثالثة رقمهـــا بنفس الطريقة هو(12+ 8 = 20) .....وهكذا 0فنجد أن المفردات التي تظهر في العينة هي المفردات المناظرة للأرقام : 4 ، 12 ، 20 ، 28 ، 36 ، 00000 ، 468 ، 476 .

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

3– العينـة العشوائية الطبقية Stratified Random Sample

إذا كان المجتمع محل الدراسة تتسم مفرداته بعدم التجانس (التشتت) فإنـــه يمكن تقسيم المجتمع إلي مجموعة من الطبقـــات أو المجموعـــات الغيـــر

متجانسة فيما بينها (لكن كل طبقة تكون مفرداتها متجانسة) . ثم يتم إختيار عينة عشوائية بسيطة من كل طبقة . والخلاصة هنا هو أنه إذا 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 . فيكون مجموع هذه العينات يمثل العينة الكلية .

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

أ – <u>حجم الطبقة</u> : فحجم العينة 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>4– العينة العشوائية المتعددة المراحل</u> Multi-stage Random Sample

يستحدم هذا الأسلوب من أساليب المعاينة إذا كان المجتمع محل الدراسة كبيراً وبه الكثير من الاختلافات والعوامل المؤثرة على المفردات . فإذا 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 انفاق دخل الأسرة في جمهورية مصر العربية ، فإنه يمكن – كمرحلة أولى

– تقسيم الجمهورية إلى مجموعات متشابهة من المحافظات (مـــثلاً وجـــه قبلي ، وجه بحري ، الواحات ، 000) .ويتم اختيار عينة من كل مجموعة ، وتبدأ المرحلة الثانية باختيار مجموعة من مدن وقـــرى المحافظـــات التـــي ظهرت في المرحلة الأولى . ثم تأتي المرحلة الثالثة وهي اختيار مجموعـــة من أسر القرى والمدن التي ظهرت في المرحلة الثانية . وبذلك نكـــون قـــد استخدمنا عينة من ثلاث مراحل وتمتاز هذه العينة بأن الأسر التي تظهر في المرحلة الأخيرة تكون مركزة في مجموعة محدودة مـــن المـــدن والقـــرى وبالتالي فإن نفقات جمع البيانات لن تكون مرتفعة . كما أن نتيجة الطريقـــة المتبعة في الاختيار أصبحت العينة ممثلة لجميع أنحاء الجمهورية دون أن يكون حجمها مبالغاً فيه .

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

الجداول العشوائية ما هي إلا الأرقـــام 0 ، 1 ، 2 ، ..... ، 9 مرتبـــة فـــي صورة أعمدة متتالية تبعًا لترتيب الحصول عليها بإحدى الطرق العشوائية . ولتوضيح كيفية استخدام هذه الجداول فى عملية إختيـــار مفـــردات عينـــة بطريقة عشوائية فلنفرض أننا نريد اختيار عينة عشوائية بسيطة مكونة من أربع عمال من شركة إنتاجية معينة تحتوي علي عدد 650 عاملاً . فإنه فى هذه الحالة نبدأ بتكوين "الإطار" وهو قائمة تشـــمل أســـماء جميـــع العمـــال بالشركة (دون حذف أو تكرار) ونعطى لكل منهم رقمًا مسلسلاً 1 ، 2 ، ... وهكذا حتى 649 ، 650 . وهنا نعتبر أن كل عامل يمثله عدد واحد (مكـــون من ثلاث خانات أو خلايا رقمية ، أى أن الأرقام المسلســـلة للعمـــال تأخـــذ الصورة(001) ، (002) ، (003) ، (004) وهكذا حتى العامل الأخير رقم(650) . وعليه يتم إختيار ثلاث أعمدة من أعمدة الجـــداول الخاصـــة

بالأعداد العشوائية ونقرأ أربعة أعداد متتالية فيكون كل مــنهم يمثــل أحــد العمال اللذين يتم اختيارهم في العينة . فمثلاً إذا اخترنــا الأعمــدة الــثلاث الأخيرة من الجدول المعطى في نهاية هذا الكتاب نجد أن الأربعة أعداد التي نحصل عليهم هي : 503 ، 529 ، 102 ، 113 وكل منها يمثل أحد العمال في العينة (لاحظ أننا تجاهلنا الرقم 702 لأنه أكبر من 650 ، حيث لا يوجــد عامل في القائمة التي لدينا يناظر هذا العدد) .

وإذا كان المطلوب هو اختيار عينة عشوائية بسيطة من مجتمع حجمه 100 مفرده فإنه يمكن توفير الكثير من الجهد المرتبط بعملية الاختيار العشوائي. حيث يتم إعداد القائمــة باســتخدام الأرقــام 00 ، 01 ، 02 ، ... ، 99 أو الأرقام 01 ، 02 ، ... ، 99 لأنه في هذه الحالات يتم استخدام عمــودين فقط من أعمدة الجدول العشوائي بدلاً من استخدام ثلاثة . والبــديل الآخ أن يشمل الإطار الذى يحوي مفردات المجتمع مجموعة الأرقام المتسلسلة علي النحو التالي 1 ، 2 ، 00 ، 100 فإنه يجب في هذه الحالة أن نعتبر أن كل رقم مكون من ثلاث خانات أوثلاث خلايا أي علي الصورة التاليــة : 001 ، 002 ، 000 ، 099 ، 100 وذلك حتى يكون لجميع الأعداد نفس فرصــة الظهور في العينة) . كذلك إذا كان حجم المجتمع هو 170 مفردة مثلاً فيجب استخدام ثلاثة أعمدة ولكن بدلاً من إهمال جميع الأعداد من 170 أو حتــى 999 فإنه يمكن اعتبار الرقم المسلسل 1 مقابلاً للأعداد 001 ، 201 ، 401 ، 601 ، 801 . والرقم المسلسل 2 مقابلاً للأعــداد 002 ، 202 ، 402 ، 602 ، 802 ..... وهكذا فالرقم المسلسل 170 يكون مقابلاً للأعداد 170 ، 370 ، 570 ، 770 ، 970 .

الفصل الثاني
أساليب عرض البيانات الإحصائية

بعد أن تتم عملية جمع البيانات 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:

(1):تبويب البيانات وعرضها في صورة جداول تكرارية (العرض الجدولي للبيانات الإحصائية) . وهذه الطريقة تتيح لنا عملية عرض البيانات مهما كان حجمها في صورة مختصرة ومبسطة يسهل فهمها .

(2):عرض البيانات الإحصائية بيانيًا (العرض البياني) وذلك من خلال استخدام الرسوم أو الأشكال البيانية الملائمة للظاهرة موضع الدراسة والتي تساعد بمجرد النظر إليها فهم تطور تلك الظاهرة حتى لغير المتخصصين .

(3):إيجاد بعض المقاييس الإحصائية المعروفة لتلك البيانات 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 :

<u>الأول : المتغير الوصفي:</u> 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 البشري (ذكر – أنثى) أو الجنسية أو الديانة أو ...... إلخ .

هذا ويستلزم التحليل الإحصائي للبيانات أو الظواهر الوصفية القيام بعملية ترميز أو تكويد Coding لتلك البيانات وذلك من خلال تخصيص قيمًا رقمية لكل عنصر من عناصر نطاق المتغير أو الظاهرة موضع الدراسة أو لكل مستوى من 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 .

<u>الثاني : المتغير الكمي :</u> Qualitative Variable

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

<u>(1):متغير كمي متقطع (منفصل) :</u> Discrete

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 (1 ، 2 ، 3 ، ....) وكذلك المتغير الذي يعبر عن عدد

الوحدات المنتجة من سلعة لا تقبل التقسيم أو التجزئة مثل عـــدد الوحـــدات المنتجة من الثلاجات أو الغسالات أو ..... إلخ .

<u>(2):متغير كمي متصل (مستمر) :</u> Continuous

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

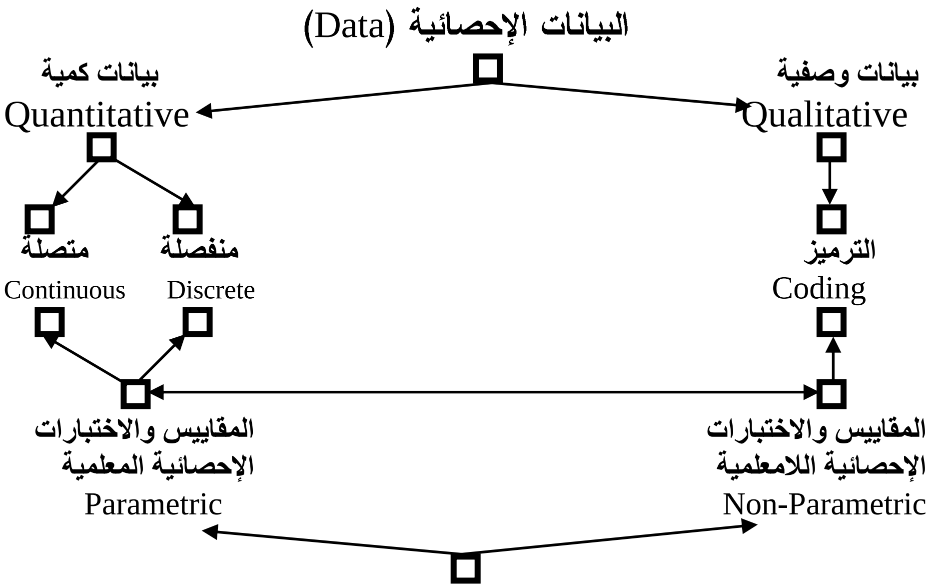
هذا ويجب التنويه إلى أن البيانات الإحصائية بصورتها المباشرة غالبًـــا مـــا تكون غير كافية بحد ذاتها بل تستخدم للتوصل إلى استنتاجات معينة وهو ما يعرف بعملية التحليل أو المعالجة الإحصائية للبيانات . ويقوم علم الإحصاء بالدور الكبير في هذا الخصوص .

وعلى الدارس أن يكون على دراية كافية بالبيانات محل الظـــاهرة موضـــوع الدراسة عند اختيار الطرق والأساليب الإحصائية المستخدمة فـــي عمليـــات التحليل الإحصائي للبيانات .

ويمكن عرض مجمل كافة المقاييس والاختبارات الإحصائية المستخدمة فـــي عمليات التحليل الإحصائي للبيانات تحت فرعين أساسيين هما : <u>المقـــاييس والاختبارات الإحصائية المعلمية</u> Parametric : ولها أولوية الاستخدام (ما دامت الشروط الإحصائية متحققة) ، وإلا استخدمت المقـــاييس <u>الإحصـــائية اللامعلمية</u> Non-Parametric عند التعامل مع البيانات الكمية وتكون لهـــا أولوية الاستخدام ما دامت الشروط الإحصـــائية غيـــر متحققـــة لإســـتخدام المقاييس المعلمية . هذا بالأضافة لإستخدامها عند التعامـــل مـــع البيانـــات الوصفية غير أن اتساع دوائر استخدامات الحاسب الآلـــي ومـــع التطـــوير الموازي والمتسارع في تطبيقات البرامج والحزم الإحصائية الجاهزة قد سهل

الأمر على كافة المستخدمين مـــن غيـــر الإحصـــائيين عنـــد التعامـــل مـــع المجموعات الكبيرة من البيانات على مستوى العـــرض الجـــدولى والبيـــانى والتحليل الإحصائي بسرعة ودقة كبيرتين (وإن كان يفضل دائمًا عند التعامل مع التحليل الإحصائي للبيانات أن تتم بمعرفة أو تحت إشراف إحصائي) .

هذا ويمكن تقديم عرض تلخيصي لمجمـــل موضـــوع البيانـــات ومعالجتهـــا الإحصائية في الشكل التوضيحي التالي :



شكل (2) : الطرق والأساليب الإحصائية

أولاً : العرض الجدولي للبيانات الإحصائية :

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 يتم التعبير عنها في صورة 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 :

(1): تحديد أقسام (أو ما تسمى بفئات) الظاهرة أو المتغير محل الدراسة والمراد تبويبها جدوليًا .

(2): عملية التفريغ في الجدول (تحديد العلاقات أو الحزم التكرارية) أمام كل قسم من أقسام الظاهرة أو المتغير .

(3):عملية العد (أي ترجمة العلامات أو الحزم التكرارية إلى أرقام) .

هذا وتختلف عملية تحديد أقسام الظاهرة أو المتغير محل الدراسة والتي تعتبر بمثابة المرحلة الأولى لعملية العرض الجدولي وذلك طبقًا لطبيعة الظاهرة أو المتغير من حيث كونه وصفيًا أو كميًا (منفصلاً أو متصلاً) . أما المرحلتين الثانية (عملية التفريغ) والثالثة (عملية العد) فهما لا تختلفان 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 .

<u>أولاً : مراحل العرض الجدولي للبيانات الوصفية :</u>

وهنا 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 (1)</u> : فيما يلي تقديرات نجاح 20 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 المراحل الثلاث لعملية تبويب هذه البيانات جدوليًا .

حيث يتم سرد أقسام الظاهرة الأربعة في العمود الأول تحت اسم فئات التقدير (أو تقديرات النجاح)ويفضل وصفها في صورة مرتبة سواء تنازليًاأو 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 علامـــة مناظرة أمام التقدير جيد تعبر عن تقدير نجاح الطالب الثاني ، ثـــم علامـــة مناظرة أمام التقدير مقبول ، ثم علامة مناظرة أمام التقدير جيد وهكذا حتـــى يتم تفريغ كل المفردات المعبرة عن التقديرات والتي تنتهي بتقدير ممتـــاز . مع مراعاة أنه لتسهيل عملية العد فإنه اُتُفِق على أن تكون العلامة المعبـــرة عن المفردة عبارة عن شرطة مائلة (/) وأن تظهر العلامات في شكل حزم تحتوي كل حزمة على خمسة علامات . ولتجنب الوقوع في أخطاء عند العد فإن العلامة الخامسة – إن وجدت – توضع بطريقة عكسية لوضع العلامات الأربعة التي تسبقها لتكون حزمة مكونة من خمس علامات . وبـــذلك تأخـــذ الحزمة الصورة(////\ ). وفي واقع الأمر فإن الحزم تساعد كثيرًا في تسهيل عملية العد النهائية حيث يكون معروفًا ان كل حزمة تحتوي علـــى خمـــس مفردات . هذا ويجب أن يراعى عند تسطير الجدول الذي ستتم فيه عمليـــة التبويب أن يكون العمود الثاني متسعًا إلى حد ما حتى يتسنى ظهور الحـــزم بشكل واضح وذلك من خلال ترك فراغ ملائم فيما بين كل حزمة وأخـــرى .

ومما لا شك فيه أن درجة اتساع هذا العمود تتوقف على عدد المفــردات أو حجم البيانات المطلوب تبويبها . والجدول التالي يلخص المراحــل الــثلاث لعملية تبويب بيانات تقديرات نجاح الطلاب جدوليًا :

جدول (2−1)

عملية تفريغ تقديرات نجاح عشرون طالبًا في مادة الإحصاء

تقديرات النجاح (الفئات)	العلامات	عدد الطلاب (التكرار)
مقبول	////	4
جيد	/// ~~////~~	8
جيد جدًا	/ ~~////~~	6
ممتاز	//	2
المجموع		20

<u>ملحوظة</u> : يجب ألا تحتوي أقسام الظاهرة (الفئات) على أي قسم لا تنتمــي إليه أي من مفردات الظاهرة أو المتغير محل البحث والدراسة سواء كانــت البيانات وصفية أو كمية . بمعنى أنه في المثال السابق لا يوجد في عمــود الفئات مجموعة تقديرات الطلاب الراسبين (ضعيف جدًا ، ضعيف) حيــث أن البيانات المعطاة هي تقديرات نجاح مجموعة من الطلاب لا تحتــوي علــى هذين التقديرين (ضعيف جدًا ، ضعيف) .

ثانياً : مراحل العرض الجدولي 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 نوعين أساسيين وهما متغير كمي متقطع (منفصل) ومتغير كمي متصل (مستمر) ، لذا سوف نفرق مابين طريقتي عرض البيانات الكمية المتقطعة والمستمرة وذلك علي النحو المبين التالي:

(1) **:مراحل العرض الجدولي للبيانات الكمية المتقطعة(المنفصلة) :**

بصفة 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 (2) :فيما يلي البيانات التالية تعبر عن أحجام 30 أسرة :

5 ، 4 ، 3 ، 5 ، 3 ، 2 ، 4 ، 6 ، 1 ، 3 ، 5 ، 4 ، 3 ، 2 ، 5 ، 4 ، 2 ،
3 ، 6 ، 5 ، 2 ، 3 ، 4 ، 1 ، 2 ، 3 ، 4 ، 2 ، 1 ، 4

والمطلوب تبويب هذه الظاهرة (تفريغها) جدوليًا .

الحل :لاحظ أن الظاهرة محل الدراسة في هذا المثال هي حجم الأسرة وهذا المتغير(حجم الأسرة)لا شك أنه يعتبر متغير كمي متقطع (لأنه يأخذ قيمًا

عددية منفصلة أي أعدادا صحيحة فقط ولا يمكن لقيم هذا المتغير أن تقبـل التقسيم أو التجزئة) لذا يتم تفريغ هذه البيانات من خلال استعراض الأحجام المختلفة لمجموعة الأسر المعطاة . أي أن أقسـام الظـاهرة عبـارة عـن مجموعة الأرقام التالية 1 ، 2 ، 3 ، 4 ، 5 ، 6 . فإذا بدأنا بقراءة أحجـام الأسر واحدًا تلو الآخر ملتزمين بترتيب المفردات كما هي واردة بالمثال (2) فأننا نبدأ بوضع علامة مناظرة للحجم (5) فـي العمـود الثـاني بالجـدول الموضح (2 – 2) ثم علامة مناظرة للحجم (4) ثم علامة مناظرة للحجم (3) .... وهكذا حتى يتم تفريغ الجدول حسب أحجام الأسـر والـذي يوضـحه الجدول التالي :

جدول (2-2)

تفريغ أحجام 30 أسرة

حجم الأسرة	العلامات	عدد الأسر
1	///	3
2	/ ~~////~~	6
3	// ~~////~~	7
4	// ~~////~~	7
5	~~////~~	5
6	//	2
المجموع		30

(2) :<u>مراحل العرض الجدولي للبيانات الكمية المتصلة(المستمرة) :</u>

أما عن مراحل العرض الجدولي 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 (or Limit) **والحد الآخر يسمى بالحد الأعلى للفئة** Upper Bound (Limit) **. وذلك مع اعتبار أن الحد الأعلى لأي فئة يعتبر بمثابة الحد الأدنى للفئة التالية لها . وفي هذه الحالة فلكي يتم تحديد عدد الفئات والحد الأدنى والأعلى لكل فئة من الفئات فإنه يتم اتباع الخطوات التالية :**

1– **<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 + 1

أي أن :

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

2– **<u>تحديد عدد الفئات</u>**: The Number of Intervals or Classes (NI) or (C)

وهناك عددًا من القواعد المختلفة التي تساعد في حساب عدد الفئات المناسب للجدول التكراري . إلا أن هذه القواعد تتفق في النهاية على أساس

واحد وهو وجوب أن يتناسب عدد فئات الجدول التكراري طرديًا مـــع عـــدد المفردات وليكن (n) مفرده . وأشهر هذه القواعد في تحديد عـــدد الفئـــات قاعدة ستورجس Sturge’s والتي تفيد بأن عدد الفئات الذي يتناسب مع عدد (n) من المفردات يستنتج من خلال القانون التالي :

$$(NI)\ or\ (C) = 1 + 3.3 \operatorname{Log}(n)$$

وهذه القاعدة تعطي نتائج مناسبة لأى عدد من المفردات .

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 صـــغيرًا 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 6 ، 10 فئات ويجب ألا يزيد عن 20 فئة .

3- **تحديد طول الفئة:** Interval Length or width (IL (or) W)

بعد تحديد عدد الفئات السابقة يتم تحديد طول الفئة ((IL (or) W)

وذلك من خلال قسمة المدى (R) على عدد الفئات((NI)or (C)) أي أن :

$$\text{طول الفئة} = \frac{\text{المدى}}{\text{عدد الفئات}} = \frac{R}{NI\ or\ C}$$

أى أن :

$$W = (\text{Range}\ (R)) /\ C = (\text{Maximum value} - \text{Minimum value} + 1) /\ C$$

وينتج في تلك الحالة طول فئة ثابت لكل فئات التوزيـــع أي يعطـــي جـــدولاً تكراريًا متساويًا من حيث أطوال فئاته . هذا ويفضل أن تكون أطوال فئـــات الجدول التكراري متساوية حتى يسهل معها عرض البيانات سواء جدوليًا أو

بيانيًا كما سنرى فيما بعد . مع مراعاة في حالة ما إذا كـــان نـــاتج عمليـــة حساب طول الفئة رقمًا كسريًا فيفضل تقريب الناتج إلى أقرب عدد صحيح . فمثلاً إذا كان المدى المحسوب في الخطوة الأولى وليكن 200 وكـــان عـــدد الفئات هو 7 فئات فإن ناتج القسمة هو 28,571 يتم تقريبه ليصبح طـــول الفئة 29 وحدة قياس تابعة لنفس وحدة قياس الظـــاهرة أو المتغيـــر محـــل الدراسة .

هذا ويفضل اختيار عدد الفئات رقمًا بحيث يكون ناتج قسمة المـــدى عليـــه يعطي طول الفئـــة مســـاويًا للـــرقم (5) أو (10) أو مضـــاعفاتهم تســـهيلا للحسابات .

هذا وتجدر الإشارة إلى أنه في بعض الحالات تستلزم طبيعة البيانات استخدام فئات غير متساوية من حيث أطوالها وذلك نظرًا لقلة التكـــرارات المقابلـــة لهذه الفئات . أو قد تكون فئات الجدول أو التوزيع التكراري محددة سلفًا في صورة فئات غير متساوية من حيث أطوالها . وبصفة عامـــة يطلـــق علـــى الجدول أو التوزيع التكراري المتساوي من حيث أطـــوال فئاتـــه بـــالتوزيع المنتظم . أما الجدول التكراري الذي تشذ فيه طول أحد فئاته على الأقل عن باقي فئات التوزيع فيسمى بالجدول أو التوزيع التكراري الغير منتظم .

4- **تحديد الحد الأدنى والحد الأعلى لكل فئة من فئات التوزيع :**

عند تحديد الحد الأدنى للفئة الأولى من فئات التوزيع التكـــراري يراعـــى 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 لـــن يـــتم تمثيلها في الجدول . كما يجب ألا يقل الحد الأعلى للفئة الأخيرة مـــن فئـــات 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 تنتهي 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 :

- 20− فتقرأ : من 20 جنيه حتى أقل من 30 جنيه .
- 30− وتقرأ : من 30 جنيه حتى أقل من 40 جنيه .
- 40− وتقرأ : من 40 جنيه حتى أقل من 50 جنيه .
- 50− 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 29.5 جنيه يقع في الفئة الأولى (20−) .

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 .

5- **مرحلة العد :** 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) .

مثال (3) :

فيما يلي البيانات التالية تمثل الأجور الأسبوعية بالجنيه التي حصل عليها 40 عامل في أحد المصانع .

36	46	43	42	35	43	49	40
35	31	41	39	44	37	36	42
32	35	53	38	51	44	48	38
40	50	43	46	52	41	43	30
42	47	52	32	47	48	41	39

والمطلوب تفريغ هذه البيانات في صورة جدول تكراري .

الحل : لتفريغ بيانات أجور العاملين المعطاه يتم الآتي :

(1):حساب المدى :

حيث أن $R = Max - Min + 1$

$= 53 - 30 + 1 = 24 \quad (L.E)$

(2):تحديد عدد الفئات :

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

$$
\begin{aligned}
NI \text{ or } C &= 1 + 3.3 \text{ Log } (n) \\
&= 1 + 3.3 \text{ Log } 40 \\
&= 1 + 3.3\,(1.6021) \\
&= 6.2868 \\
&= 6 \quad \text{(Classes)}
\end{aligned}
$$

<u>(3):تحديد طول الفئة</u> : فحيث أن:

(IL (or) W) = R / (NI or C) = 24 / 6 = 4 (L.E)

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

<u>(4):تحديد الحد الأدنى والأعلى لفئات الأجور الأسبوعية</u> :

بعد تحديد قيمة المدي وعدد الفئات ومن ثم طول الفئة فإن فئـــات التوزيـــع الخاصة بالأجر الأسبوعي تأخذ الصورة التالية :

30− ، 34− ، 38− ، 42− ، 46− ، 50− 54

وهنا يتم تفريغ 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 . فنحصـــل على الجدول التكراري (3−2) . وبقراءة المفردات المعبرة عن أجور العمال الأسبوعية المفردة تلو الأخرى وذلك بنفس الترتيب الوارد في بيانات مثـــال (3) فأننا نجد أن قيمة الأجر 40 جنيه تنتمي إلى الفئة (38−) لذا يتم وضع علامة (/) أمام تلك الفئة . بينما نجد أن قيمة الأجر الثاني وهو 49 جنيه تنتمي إلى الفئة (46−) لذا يتم وضع علامة أمام تلك الفئة . أمـــا المفـــردة الثالثة وهي قيمة الأجر 43 جنيه فهي تنتمي للفئة (42−) فيتم وضع علامة أمام تلك الفئة و ..... و هكذا حتى نصل للمفردة الأخيرة وهي قيمة الأجـــر 42 جنيه وهي تنتمي للفئة (42−) فيتم وضع علامة أمامها ومن ثم تنتهـــي عملية التفريغ . فيتم ترجمة تلك العلامات إلى أرقـــام تعبـــرعن التكـــرارات 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 لعـــدد

المفردات الإجمالي المطلوب تبويب بياناتها وهي (40) مفرده فـــي مثالنـــا. والجدول التالى (3-2) يلخص مجموعة الخطوات السابقة .

جدول (3-2)

تفريغ جدول التوزيع التكرارى للأجر الأسبوعي لـ 40 عاملاً (بالجنيه)

فئات الأجر الأسبوعي	العلامات	عدد العمال (التكرار)
30-	////	4
34-	~~////~~ /	6
38-	~~////~~ ////	9
42-	~~////~~ ////	9
46-	~~////~~ //	7
50-54	~~////~~	5
المجموع		40

وفي حالة عدم تساوي مجموع التكرارات لعدد المفردات المطلـــوب تبويـــب 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 التوزيـــع في غير موضعها . لذا يجب توخي الحذر أثناء عملية التفريغ حتى يعكـــس الجدول التكراري أو جدول التفريغ البيانات الفعلية للظاهرة موضع الدراسة .

(5):مرحلة العد:

وفي تلك المرحلة يتم الحصول على التوزيع التكراري للأجـــور الأســـبوعية 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 الأسبوعية يأخذ الصورة المبسطة التالية :

جدول (4–2)

فئات الأجر الأسبوعي	30–	34–	38–	42–	46–	50–54	المجموع
عدد العمال (التكرار)	4	6	9	9	7	5	40

وفي هذا التوزيع التكراري الوارد بالجدول (4–2) نجد أن الحد الأدنى للفئة الأولى في التوزيع معلوم (30 جنية) وأن الحد الأعلى للفئة الأخيرة (54 جنية) . كذلك يتضح من هذا التوزيع التكراري بأن هذا التوزيع التكراري توزيع مغلق (مقفل) من الطرفين. وفي حالة غياب أحد هذين الحدين يسمى التوزيع التكراري مفتوح سواء من طرف واحد أو من الطرفين . فمثلاُ 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 فًيسمى التوزيع التكراري مفتوح الطرفين .

<u>الجداول أو التوزيعات التكرارية المفتوحة</u> : 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(جدول(5 – 2)) يبين توزيعًا تكراريًا مفتوح الطرفين .

جدول (5-2)

التوزيع التكراري لدرجات 50 طالب

فئات الدرجات	أقل من 10	10-	20-	30-	40-	50-	60 فأكثر
عدد الطلاب (التكرار)	7	9	15	6	4	7	2

ففي هذا الجدول فإن الحد الأدنى للفئة الأولى (أقل من 10) والحد الأعلى للفئة الأخيرة (60 فأكثر) غير معلومين . لذا فإن هذا التوزيع التكراري يسمى توزيع تكراري مفتوح الطرفين .

ولا تنسى عزيزي القارئ أنه ليس بالضرورة تساوي أطوال فئات التوزيع فيمكن 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 (جدول(6 – 2)) يعطي صورة واضحة عن الجداول التكرارية الغير منتظمة .

جدول (6-2)

التوزيع التكراري للدخل الشهري (بالجنيه) لمائة أسرة

فئات الدخل	230-	240-	250-	260-	270-	280-	290-	310-340
عدد الأسر (التكرار)	4	6	10	12	27	28	14	9

فيلاحظ على بيانات الجدول (6-2) أن أطوال فئاته الستة الأولى هي 10 جنيه أما الفئتين الأخيرتين (290-) ، (310-340) فإن أطوالهما هما 20 ، 30 جنيه على الترتيب . لذا فإن هذا التوزيع التكراري توزيعًا غير متساوي أطوال فئاته (توزيعًا تكراريا 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 الجـــداول الغير منتظمة حيث يلزم كتابة الحد الأعلى للفئة الأخيرة صراحة .

ملحوظة : بعد إتمام عملية تفريغ البيانات والحصول على الجداول التكرارية كما هو وارد عل سبيل المثال في جـــدول (5−2) أو (6−2) فـــإن الجـــدول 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 (5−2) فإن هناك ســـبعة طـــلاب درجاتهم تقل عن عشرة درجات وكذلك فإن هنـــاك تســـعة طـــلاب تتـــراوح درجاتهم ما بين عشرة درجات وأقل من 20 درجة وهناك 15 طالبًا تتراوح درجاتهم ما بين عشرين درجة وأقل من 30 درجة وهكذا .

هذا ويمكن من خلال تلك الجداول التكرارية البسيطة الحصول علـــى نســـبة التكرارات التي تنحصر ما بين حدي كل فئة من فئات التوزيع التكراري وهو ما يسمى بالجداول التكرارية النسبية والتي سيتم تناولها في الجزء التالي . وهنا نود أن ننوه بأنه تصح مقارنة توزيعين تكراريين بسيطين لفئات واحدة متى تساوت مجموع التكرارات الأصلية لهما . أما فى حالة إختلاف مجموع التكرارات للتوزيعين فلابد من الحصول على التكرارات النسبية لكل توزيـــع حتى تصح عملية المقارنة فيما بينهما.

التوزيعات التكرارية النسبية : 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 إلا 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(i) وليكن (F_i) وبين مجموع تكرارات التوزيع التكراري (ΣF_i) : أي أن :

$R_i = (F_i / \Sigma F_i)$, where $i = 1 , 2 , ... , r$

حيث (r) تعبر عن عدد فئات التوزيع ،

أما عن ΣF_i فهي تعبر عن مجموع تكرارات التوزيع اتتكراري.

أو هناك صورة أخرى هي:

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

والفرق ما بين الصيغتين أن الأولى تعطي التكرار النسبي فى صورة 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 100% .

هذا ولقد سبق وأن ذكرنا في التوزيعات التكرارية البسيطة أنه تصح المقارنة فيما 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 تكراريين 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 . حيث أنه بحساب التكرارات النسبية نجد أن كل عمود من عمودي التكرارات النسبية يكون مساويًا للواحد الصحيح ومن هنا تم التخلص من أثر عدم التساوي مجموع التكرارات المطلقة للتوزيعين التكراريين في تلك الحالة ومن ثم تصح عملية المقارنة بعد حساب التكرارات النسبية .

<u>**مثال (4)**</u>: الجدول التالي يوضح التوزيع التكراري لدرجات مجموعة من الطلاب والطالبات .

جدول (7-2)

التوزيع التكراري لدرجات مجموع من الذكور والإناث

فئات الدرجات	أقل من 10-	10-	20-	30-	40-	50-	60 فأكثر	مجـ
عدد الطلاب	7	12	17	9	8	5	2	60
عدد الطالبات	15	17	18	18	10	9	13	100

والمطلوب إجراء المقارنة فيما بين التوزيعين باستخدام التوزيع التكراري الملائم.

<u>**الحل**</u> : لاحظ أن فئات التوزيعين واحدة لكن بالنظر لمجموع تكرارات كلاً من الطلاب والطالبات لنفس فئات التوزيع نجد أن هناك اختلاف في مجموع التكرارات الأصلية للطلاب عن الطالبات (60 ، 100 على الترتيب) . لذا لا تصح المقارنة من خلال التكرارات المطلقة أو التكرارات الأصلية المعطاة في المثال لذا يجب التخلص من أثر عدم تساوي هذا المجموع (60 ، 100)

وذلك من خلال حساب التكرارات النسبية لكل من الطلاب الذكور ولتكن هي (R_i M) وكذلك للطالبات الإناث ولتكن (R_i F) وذلك كما يوضحها الجدول (8-2) التالى :

جدول (8-2)

التوزيع التكراري الأصلي والنسبي لدرجات 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 أن مجموع عمودي التكرارات النسبية سواء للذكور أو الإناث مساويًا 100% 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 الطالبات الإناث الحاصلين على أقل من عشرة درجات .

* كذلك بالنسبة للفئة الثانية (من الدرجة 10 حتى أقل من 20 درجة) فإن نسبة الطلاب الذكور (20%) تزيد عن نسبة الطالبات (17%) . لاحظ أن مقارنة التكرارات المطلقة لهذه الفئة تعتبر مقارنة مضللة فهي عكس ما

تحدده عملية المقارنة من خلال التكرارات النسبية وهو الأصح فى هذة الحالة .

*وهكذا إلى أن نصل للفئة الأخيرة (60 درجة فأكثر) فإن نسبة الطلاب الذكور (5%) تقل عن نسبة الطالبات (13%) .

<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 كانت 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 (صفوف) ورأسيًا (أعمدة) اى إلى مجموعة من الصفوف والأعمدة . حيث يتم تمثيل فئات أحد الظاهرتين في الصفوف والأخرى في الأعمدة . والتوزيعات 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)كما سبق 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 الخطوة السابقة إلى أرقام أو تكرارات (عملية العد) . أي أن مراحل إنشاء الجدول التكراري المزدوج هي :

(1):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 التكرارية البسيطة .

(2):عملية التفريغ: (العلامات) بعد تحديد أقسام أو فئات الظاهرتين يتم تصميم الجدول التكراري المزدوج بحيث تخصص الصفوف لأحد المتغيرين ويكون عدد الصفوف مساوياً لعدد فئات هذا المتغير . بينما تخصص الأعمدة لتوزيع أو لفئات المتغير الآخر بحيث يكون عدد الأعمدة مساوياً لعدد فئات المتغير الثاني . وبهذا فإن الجدول التكراري المزدوج يتكون من عدد من الخلايا يبلغ 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 البيان أو زوج المشاهدات حيث يتم تحديد الصف والعمود الذي ينتمي إليه هذا البيان ومن ثم يتم تحديد الخلية وهي تقاطع صف وعمود هذا البيان ثم توضع

شرطة أو علامة (/) في تلك الخلية . وتتم عملية التفريغ بقراءة البيانات أو أزواج المشاهدات للبيانات أو الزوج تلو الآخر وفي كل مرة توضع شرطة (/) في الخلية التي ينتمي إليها هذا البيان مع استخدام طريقة الحزم السابق الإشارة إليها .

(3):عملية العد : وهنا تتم عملية ترجمة العلامات داخل كل خلية من خلايا الجدول المختلفة إلى أرقام فنحصل بذلك على الجدول التكراري المزدوج.

هذا ويمكن من خلال الجدول التكراري المزدوج استنتاج التوزيعات التكرارية الهامشية Marginal Distribution لكل متغير على حدة . 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 فئاته في الأعمدة ومايناظرها من مجاميع تكرارات الأعمدة في الجدول التكراري المزدوج . ومن هذه التوزيعات التكرارية الهامشية يمكن الحصول على التوزيعات التكرارات النسبية كما سبق وأن أوضحناها .

والمثال التالي يوضح كيفية تكوين الجدول التكراري المزدوج وكيفية اشتقاق التوزيعات الهامشية للظاهرتين محل الدراسة .

مثال (5):

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

y	X	Y	X	Y	X	Y	X
89	100	78	84	59	90	43	60
69	80	41	73	89	100	75	83
90	110	65	91	51	86	43	79
71	78	54	68	88	105	84	101
79	96	56	74	52	71	72	80
106	112	69	90	75	93	84	95
72	91	44	65	56	88	62	72
75	81	68	82	69	73	53	77
94	103	67	89	87	109	78	100
79	84	79	92	98	115	76	85
90	98	68	90	85	97	91	104
108	119	69	81	105	125	75	95
				68	81	80	89

والمطلوب :-

1- إنشاء جدول تكراري مزدوج لبيانات الـدخل (x) والإنفـاق (y) لتلـك الأسر .

2- من خلال نتائجك في (1) استنتج التوزيع الهامشي للدخل الأسبوعي (x) للأسروكذا التوزيع الهامشي للإنفاق الأسبوعي (y) للأسر .

3- من خلال نتائجك في (2) اوجد التوزيع التكراري النسبي لكل من الدخل الأسبوعي (x) وكذا التوزيع التكراري النسبي للإنفاق الأسبوعي (y) .

الحل :

(1):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 for (x) = Max – Min + 1

  = 125 – 60 + 1 = 66 (L.E)

- NI (or) C = 1 + 3.3 Log (n)

$$= 1 + 3.3 \text{ Log } (50)$$
$$= 6.6066 = 7 \text{ Intervals (classes)}$$

- LI (or) W = Range ÷ NI

$$= 66 \div 7 = 9.43 \approx 10 \quad \text{(L.E)}$$

ويتم تقريب طول الفئة بالزيادة فتصبح طول الفئة مساوياً عشرة جنيهات تقريبا . وحيث أن أصغر قيمة في الدخل هي ستون جنيهاً لذا فإنه 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 : 120-130 , -110 , -100 , -90 , -80 , -70 , -60

<u>ثانياً: بالنسبة لبيانات الإنفاق (y) :</u>

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

لاحظ أن عدد الفئات للمتغير (y) باستخدام قاعدة ستورجس سيكون نفس عدد الفئات المتغير (x) نظراً لثبات قيمة (n) . وعلية تصبح طول الفئة هى:

- LI (or) W = 68 ÷ 7 = 9.71 ≈ 10 (L.E)

وحيث أن أصغر قيمة في الإنفاق هي إحدى وأربعون جنيهاً . لذا فإنه 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 : 100- 110 , -90 , -80 , -70 , -60 , -50 , -40

ومن خلال نتائجنا عن فئات كل من المتغيرين x , y يتم تصوير الجدول التكراري المزدوج حيث يتم تمثيل فئات الدخل (x) وليكن في الصفوف وفئات

الإنفاق (y) في الأعمدة فيعطي الصورة الموضحة في الجدول (9 - 2) وذلك على النحو المبين التالي :

جدول (9-2)

تصميم جدول تكراري مزدوج يوضح العلاقة بين الدخل

الأسبوعي (x) والإنفاق الاسبوعي (y) لخمسين أسرة

x \ y	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) الزوج تلو الآخر . وفي كل مرة نضع علامة (/) في الخلية التي ينتمي إليها هذا الزوج . فمثلاً لو بدأنا بيانات الدخل والإنفاق متبعين الترتيب الوارد في هذا المثال فإننا نبدأ بالزوج الأول للمشاهدات وهو(x=60 تقابلها y=43) . فحيث أن قيمة الدخل 60 تنتمي للفئة الأولى من فئات الدخل وهي (- 60) وأن قيمة الإنفاق 43 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 علامة فى الخلية التي تقع في الصف الأول والعمود الأول أي وضع العلامة(/) في الخلية التي تقع فى الصف الأول وعمود الأول . أما الزوج الثاني من المشاهدات وهو (x = 83 ، y = 75) فإنه ينتمي إلى

الخلية الناشئة من تقاطع الصف الثالث الممثل للفئة (- 80) مـــع العمـــود الرابع الممثل للفئة (- 70) . لذا يتم وضع علامة (/) في الخلية الناشئة عن تقاطع الصف الثالث مع العمود الرابع .

وهكذا بالنسبة لبقية أزواج المشاهدات حتى نصل إلى الـــزوج الأخيـــر مـــن المشاهدات وهو(x= 119 ، تقابلها y=108) والـــذي ينتمـــي للخليـــة الناشئة عن تقاطع الصف السادس الممثل للفئة (- 110) مع العمود السابع الممثل للفئة (100− 110) لذا يتم وضع شرطة في الخليـــة الناشـــئة عـــن تقاطع الصف السادس مع العمود السابع كما هو موضح في جدول التفريـــغ (10− 2) والذي يبين العلامات التي تعكس ازواج المشاهدات .

جدول (10−2)

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-		//	~~////~~	~~////~~ /	/			14
90-		/	///	////	///	/		12
100-				/	~~////~~	//		8
110-						//	//	4
120-130							/	1
∑	4	7	10	12	9	5	3	50

وبترجمة العلامات الواردة بخلايا جدول التفريغ المزدوج المبين في الجدول (10−2) إلى أرقام نحصل علي الجدول التكراري المـــزدوج الـــذي تحتـــوي خلاياه على التكرارات المختلفة كما هو موضح في الجدول (11−2) التالى :

جدول (11–2)

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 (11–2) القول أن تكرار فئة معينة من فئات الدخل (x) هو عبارة عن مجموع التكرارات المشتركة فى صف هذه الفئة . فمثلاً يمكن القول من خلال بيانات الصف الأول في هذا الجدول (وهو صف فئة الدخل من 60 جنيه حتي أقل من سبعون جنيها) أن هناك ثلاث أسر يتراوح دخلها ما بين ستون جنيها وأقل من سبعون جنيهاً . وكذلك هناك ثماني أسر يتراوح دخلها ما بين سبعون جنيهاً وأقل من ثمانون جنيهاً . أما بالنسبة لفئات الإنفاق فإنه يمكن القول أن هناك أربعة أسر يتراوح إنفاقها الأسبوعي ما بين أربعون جنية وأقل من خمسون جنيهاً وكذلك هناك سبعة أسر يتراوح إنفاقها الأسبوعي ما بين خمسون جنيها وأقل من ستون جنيهاً وهكذا .

هذا ويجب أن ننوه إلى أنه يعتبر من المألوف أن نجد عدداً من الخلايا في الجدول التكراري المزدوج لا يحتوي على تكرارات وهو مايعنى أن تكرارات هذه الخلايا تكرارات صفرية . 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 .

(2): من خلال التوزيع التكراري المزدوج الوارد بالجدول (2−11) 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 (2−11) دون أخذ الإنفاق الاسبوعي (y) في الاعتبار . والجدول (2−12) يبين التوزيع التكراري الهامشي للدخل الاسبوعي (x) .

جدول (2-12)
التوزيع التكراري الهامشي للدخل الأسبوعي
(x) للعينة المكونة 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 مابين ثمانون جنيها 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 (2-11) وما يقابلها من مجاميع تكرارات لهذه الأعمدة دون الأخذ في الاعتبار فئات الدخل الأسبوعي (x) . والجدول (2-13) يبين التوزيع التكراري الهامشي للإنفاق الأسبوعي (y) .

جدول (2-13)

التوزيع التكراري الهامشي للإنفاق الأسبوعي (y)

للعينة المكونة 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 .

(3): بعد الحصول على التوزيعات الهامشية للمتغير (x) وكذلك للمتغير(y) فإنه يمكن قسمة قيم تكرارات تلك التوزيعات التكرارية الهامشية على إجمالي عدد المفردات أو التكرارات (خمسون عينة) فنحصل على التوزيع التكراري النسبي للمتغير (x) وكذلك للمتغير (y) كما هو موضح في الجدولين (14- 2) ، (15 -2) على الترتيب .

جدول (2−14)

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 24% من إجمالي عدد الأسر يتراوح دخلهم ما بين تسعون جنيها وأقل من مائة جنيه وهكذا..

جدول (2−15)

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%

ومن خلال هذا التوزيع التكراري النسبي للإنفاق يتضح لنا على سبيل المثال أن هناك 18% من إجمالي عدد الأسر يتراوح إنفاقها ما بين ثمانون جنيهــا وأقل من تسعون جنيهًا وهكذا .

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

1− ليس بالضرورة أن يتساوى عدد الفئات أو طول الفئة للمتغيرين في الجدول التكراري المزدوج . إذ أن عدد الفئات أو طــول الفئـــة لكـــل متغير يتوقف على المدى المطلق (Range) لهذا المتغير .

2− قد يكون المتغيران الممثلان في الجدول التكراري المزدوج متغيـــران 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) والإنفـــاق الشهري (متغير كمي متصل) أو جدول تكراري يمثل العلاقة ما بـــين الحالة الاجتماعية للفرد وظاهرة التدخين (متغيـــران وصـــفيان) أو جدول تكراري يمثل العلاقة ما بين النوع والدخل الشهري لعينـــة من الأســـر و ..... وهكذا .

3- التوزيع التكراري الهامشي لمتغيرًا ما هـــو عبـــارة عـــن التوزيـــع التكراري البسيط لهذا المتغير . والتوزيع التكراري البسيط يبين عـــدد التكرارات التي تقع ما بين الحد الأدنى والأعلى لكل فئة مـــن فئـــات 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 الحـــد الأدنـــى أو الأعلى لأي فئة من فئات التوزيع أو أي قيمة تقع فيما بينهما . فمـــثلاً قـــد يكون الباحث في حاجة لأن يعرف ماهو عدد الطلاب اللذين تقل درجاتهم عن 90 درجة لتوزيع تكراري بسيط معطي . وفي أحيان أخرى قد يهتم الباحـــث بمعرفة عدد المفردات أو التكرارات التي تأخذ قيمة تزيد علـــى أو تســـاوي قيمة معينة سواء كانت تلك القيمة تمثل الحد الأدنى أو الأعلى لأي فئة مـــن فئات التوزيع أو أي قيمة تقع فيما بينهما . وليكن على سبيل المثال مـــاهو

عدد الموظفين اللذين تبلغ مرتباتهم الأسبوعية 95 جنيهاً على الأقــل (95 جنيهاً فأكثر) .

ففي واقع الأمر فإن الجداول التكرارية البســيطة لا تلبــي بطريقــة مباشرة الإجابة علي مثل هذه التساؤلات . ولكي يصل الباحث إلى ما يريــد فإنه يقوم بتجميع التكرارات تصاعدياً في الحالة الأولى بينما يقوم بتجميعها تنازلياً في الحالة الثانية . وتسمى الجدول فــي الحالــة الأولــى بالجــدول التكراري المتجمع الصاعد Ascending Cumulative Frequency Distribution (Asc.C.F.D)
بينما يسمى الجدول في الحالة الثانية بالجدول التكراري المتجمع الهــابط أو النازل Descending Cumulative Frequency distribution (Dsc.C.F.D) ويتكون كلاً من الجدول التكرارى المتجمــع الصــاعد أو الهابط من خلال الجدول التكراري البســيط للظــاهرة موضــع الدراســة . ويستخدم الجدول أو التوزيع التكراري المتجمع الصاعد (Asc.C.F.D) في تحديد عدد المفردات التي تأخذ قيماً أقل من قيمة معينة . كما يستخدم فــي تحديد القيمة التي يقل عنها عدداً معيناً أو نسبة من المفردات . أما الجــدول أو التوزيع التكراري المجتمع الهابط (Dsc.C.F.D) فإنه يستخدم في تحديد عدد المفردات التي تأخذ قيماً تساوي أو تزيد عن قيمة معينة كما يســتخدم في إيجاد القيمة التي يزيد عنها أو يساويها عدداً معينا أو نسبة معينة مــن مفردات التوزيع التكراري .

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

(1): 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 أو

فيما بين حدي فئة معينة فلاحظ أن هناك مرادفات لغوية مكافئة لكلمة أقــل من حيث أن : -

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

فكل هذه المرادفات أو ما يعادلها لغوياً يفيد ضــرورة تكــوين الجــدول أو التوزيع التكراري المتجمع الصاعد (Asc.C.F.D) واستنتاج المطلــوب منه .

(2): على العكس مما سبق صحيح تماماً . بمعنى أنه إذا كان المطلوب هــو تحديد عدد التكرارات أو نسبة التكرارات التي تزيد عن(أو تبلغ قيمــة مــا فأكثر) سواء كانت هذه القيمة تمثل الحد الأدنى أو الحد الأعلى لفئة معينــة أو قيمة معينة تقع فيما بين حدي فئة من فئات التوزيع التكراري الأصــلي . فأنه يجب ملاحظة أن هناك مترادفات لغوية مكافئة لكلمة تزيد عن وهي على سبيل المثال : -

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

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

(3): 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 الأصــلي على الأقل (أو على الأكثر) فإنه يمكن استنتاج المطلوبين معاً أما من خلال التوزيع التكراري المتجمع الصاعد أو من خلال التوزيع التكراري المتجمــع الهابط كما سنرى في الأمثلة فيما بعد .

مثـال (6) :- **فيما يلي لديك التوزيع التكراري للأجــور الأســبوعية لعينــة مكونة من خمسين عاملاً (بالجنيه) كما يوضحها الجدول(16-2) التالي :**

جدول (16 – 2)

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

<u>والمطلوب</u> :–

<u>أولا</u>: إيجاد التوزيع التكراري المتجمع الصاعد وكذا التوزيع التكراري النسبي المتجمع الصاعد .

<u>ثانيا</u>: إيجاد التوزيع التكراري المتجمع الهابط وكذا التوزيع التكراري النسبي المتجمع الهابط .

ثالثا: بإستخدام التوزيع التكراري الملائم حدد ما يلي :

(أ):عدد العمال اللذين تتراوح أجورهم ما بين 80 وأقل من 90 وكذا النسبة.

(ب):عدد العمال اللذين تبلغ أجورهم 73 جنيهاً على الأكثر وكذا النسبة .

(ج):ما هي قيمة الأجر الذي يتقاضاه 65 % من إجمالي عدد العمـــال علـــى الأقل .

(د):ما هو الحد الأدنى للأجر الذي يبلغه 47 عاملاً كحد أدنى .

(هـ):ما هي نسبة العمال اللذين تتراوح أجورهم ما بين 55 ، 87 جنيه .

(و):إذا كانت نسبة العمال اللذين تتراوح أجورهم ما بين 53 ، X جنيه تبلغ 40% من إجمالي عدد العمال . فما هي قيمة الأجر x .

<u>الحل : –</u>

<u>أولا</u>: وفقاً لتعريف التوزيع التكراري المجتمـــع الصـــاعد . فـــإن التوزيـــع التكراري المتجمع الصاعد (Asc.C.F.D) المناظر لأي فئة من فئات التوزيع الأصلي ما هو إلا عبارة عن تكرار هذه الفئة مضافاً إليه تكـــرارات الفئـــات السابقة . ويتكون هذا الجدول من عمودين يـــتم كتابـــة الفئـــات التكراريـــة

المتجمعة الصاعدة فى صورة (أقل من الحد الأعلى لكـــل فئـــة فـــي فئـــات التوزيع) في العمود الأول أما التكرارات المتجمعة الصـــاعدة (Asc.C.F.) فيتم تحديدها في العمود الثاني من هذا الجدول . أمـــا التكـــرارات النســـبية المتجمعة الصاعدة (R_i .Asc.C.F) للفئة (i) فهي عبارة عن التكـــرارات المتجمعة الصاعدة للفئة (i) مقسومة على مجموع التكرارات سواء ككسر عشرى أو كنسبة مئوية . أي أن :-

$$R_i.Asc.C.F\ (i) = Asc.C.F\ (i) / \sum F_i\ ,\ \text{where}\ i = 1,2,\ldots,r$$

حيث (r) تمثل عدد فئات التوزيـــع ، كمـــا أن Asc.C.F (i) هـــو التكرار المتجمع الصاعد للفئة (i) . والجدول (2-17) يبـــين حســـابات التوزيع التكراري المتجمع الصاعد والتوزيع التكـــراري المتجمـــع النســـبي الصاعد .

جدول (2-17)

التوزيع التكراري المتجمع الصاعد والنسبي المتجمع الصاعد

لأجور عينة من خمسين عاملاً في إحدى الشركات

Classes	F_i	less than the Class Upper Limit	Asc.C.F	R_i.Asc.C.F (%)
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 (2-17) ما يلي : -

(1): قمنا بافتراض أنه توجد فئة وهمية (افتراضية) تسبق الفئـــة الأولـــى فيكون الحد الأعلى لتلك الفئة الوهمية ما هو إلا الحد الأدنى للفئـــة الأولـــى للتوزيع التكراري الأصلي (30جنيه) فيكون التكرار المتجمع الصاعد المقابل للحد الأدنى للفئة الأولى (وهو الحد الأعلى لتلك الفئة الوهمية يكون دائمـــاً أبداً مساوياً للصفر حيث لا توجد هناك تكرارات سابقة لتلك الفئة الأولـــى) . هذا ويمكن الاستغناء عن السطر الأول فـــي التوزيـــع التكـــراري المتجمـــع الصاعد . وفي هذه الحالة نبدأ حسب تمييـــز العمـــود الأول مـــن التوزيـــع التكراري المتجمع الصاعد وهو أقل من الحد الأعلى للفئة الأولى والذي يبلغ 40 جنيه . فيكون عدد العاملين اللذين يقل أجرهم عن أربعون جنيهاً هـــو عامل وحيد وهو بمثابة تكرار الفئة الأولى فى التوزيع التكرارى الأصـــلى . ويلي ذلك السطر الثاني في جدول التوزيع التكراري المتجمع الصـــاعد فـــإن عدد العاملين اللذين يقل أجرهم عن الحد الأعلى للفئة الثانية (والذي يبلـــغ 50 جنيه) هو عبارة عن 5 عمال وهذا التكرار المتجمع بمثابة تكرار الفئة الأولى مضافاً إليه تكرار الفئة الثانية فى التوزيع التكرارى الأصـــلى أي أن (4 +1 = 5) وبذلك يكون التكرار المتجمع الصاعد المقابل للحـــد الأعلـــى للفئة الثانية ما هو إلا تكرار الفئة الأولى مضافاً إليه تكرار الفئة الثانية فـــى التوزيع التكرارى الأصلى . وهكذا يتم إيجاد التكرار المتجمع الصاعد لبقية الفئات فئة تلو الأخرى حتى نصل إلى التكرار المتجمع الصاعد للحد الأدنـــى للفئة الأخيرة (وهو بمثابة الحد الأدنى لفئة وهمية تلي الفئة الأخيرة والـــذي يعتبر بمثابة الحد الأعلى للفئة الأخيرة بالجدول التكراري الأصـــلي) والـــذي يجب أن يكون مساوياً للمجموع الكلي للتكرارات أي خمسون عاملاً في هذا المثال .

(2): 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 التكــراري الأصــلي . فمثلاً للتأكد من ذلك فإن التكرار المتجمع الصاعد المقابل لأقل من 80 جنيه وهو 37 عاملاً ما هو إلا التكرار المتجمع الصاعد السابق لتلك الفئة (أقــل من 70 جنيه)مضافاً إليه تكرار الفئة من 70 حتى أقل من 80 جنيه أي أن: (37 = 11 + 26) .

(3): من التوزيع التكراري المتجمع الصاعد يمكن تحديد قيمة الأجر الــذي يقل عنه عدداً معيناً من العمال . فمثلاً قيمة الأجر الذي يقل عنه 45 عــاملاً ما هو إلا الأجر 90 جنيه . كذلك يمكن من جدول التوزيع التكراري المتجمع الصاعد تحديد أو حساب عدد العمال اللذين تقل أجورهم عن قيمة معينــة لا توجد في عمود الفئات وليكن مثلاً الأجر 64 جنيه وهي قيمة غير موجــودة في الجدول وهو ما سيتم استنتاجه جبرياً أو بيانياً كما سنرى فيمــا بعــد . وأيضاً العكس صحيح يمكن تحديد ما هو الأجر الذي يقل عنه عدداً معيناً من التكرارات المتجمعة التي لا توجد في التوزيع التكراري المتجمــع الصــاعد بنفس الآلية الجبرية(طريقة الاستكمال العددي أو النسبة والتناسب) أو بيانياً كما سنرى فيما بعد في موضع لاحق إن شاء الله .

(4): جدول التوزيع التكراري المتجمع الصاعد النسبي ما هو إلا صورة من جدول التوزيع التكراري المتجمع الصاعد لكن بعد قسمة عمــود التكــرارات المتجمعة الصاعدة بالكامل على مجموع تكرارات التوزيــع الأصــلي وهــي القيمة 50 عامل في مثالنا محل الدراسة .

(5): جدول التوزيع التكراري المتجمع الصاعد أما أن يبدأ بالصفر (وذلــك في حالة افتراض أن هناك فئة وهمية تسبق الفئة الأولى فيكون الحد الأعلى

لتلك الفئة الوهمية بمثابة الحد الأدنى للفئة الأولى) أو يبدأ بتكــرار الفئــة الأولى (وذلك في حالة بداية الفئات المتجمعة بالحد الأعلى للفئة الأولـــى) وينتهي بالمجموع الكلي لتكرارت التوزيع الأصلي .

وكذلك فإن جدول التوزيع اللتكراري المتجمع الصاعد النسبي أما أن يبـــدأ بالصفر أو بالتكرار النسبي للفئة الأولى وينتهي بالواحد الصحيح أو بالنسبة 100 % وهذا المعيار يؤكد صحة حسابات جدول التوزيع التكراري المتجمع الصاعد وكذا جدول التوزيع التكراري المتجمع الصاعد النسبي .

ثانيا: لإيجاد جدول التوزيع التكراري المتجمع الهابط (Dsc.C.F.D) وكذا جدول التوزيع التكراري المتجمع الهابط النسبي (R. Dsc.C.F.D) فإنـــه يعاد تعريف الفئات كما هو موضح في الجدول (18 – 2) حيث يتم كتابـــة فئات التوزيع التكراري المتجمع الهابط على صورة " الحـــد الأدنـــى للفئـــة الأصلية فأكثر . ومن ثم فبالاستعانة ببيانات الجـــدول (16 – 2) ســـوف نوضح كيفية تكوين جدول التوزيع التكراري المتجمع الهابط وكـــذا الهـــابط النسبي وهوما يوضحه الجدول(18–2) .

جدول (18 – 2)

التوزيع التكراري المتجمع الهابط (Dsc.C.F.D)

وكذا التوزيع التكراري المتجمع الهابط النسبى

(R.Dsc.C.F.D) لأجور عينة من 50 عامل

Classes	F_i	More than Lower Limit	DscCFD	R_iDCFD (%)
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
∑	50	More than 100	0	0

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

(1): التكرار المتجمع الهابط المقابل لفئة معينة (للحد الأدنى لفئة معينة فأكثر) من فئات التوزيع الأصلي ما هو إلا عبارة عن مجموع التكرارات مطروحاً منه تكرارات الفئات السابقة لتلك الفئة . فبداية فإن التكرار المتجمع الهابط للفئة الأولى(أى عدد العمال الذين يتقاضون أجرا 30 جنية فأكثر ما هو إلا عبارة عن المجموع الكلي للتكرارات (50عاملاً) . بينما التكرار المتجمع الهابط للفئة الثانية (40 فأكثر) فهو عبارة عن المجموع الكلي للتكرارات مطروحاً منه تكرار الفئة السابقة (أي تكرار الفئة الأولى وهو 1 عاملاً) . أي أن التكرار المتجمع الهابط للفئة الثانية هو عبارة عن(50– 1 = 49 عاملا). وهكذا بالنسبة لبقية الفئات .

(2): يمكن تكوين جدول التوزيع التكراري المتجمع الهابط بطريقة أسهل عن طريق اتباع نفس الأسلوب المستخدم في حالة جـدول التوزيـع التكـراري المتجمع الصاعد وذلك من خلال بداية عملية التجميع المتتالية مـن أسـفل لأعلى على عكس التوزيع التكراري المتجمـع الصـاعد . فمـثلاً التكـرار المتجمع الهابط المقابل للحد الأدنى للفئة الأخيرة (وهو 90 جنيه) يقابلـه التكرار المتجمع الهابط (5 عامل) وهو نفس تكرار الفئـة الأخيـرة فـي التوزيع التكراري الأصلي . بينما يكون التكرار المتجمع الهابط للفئـة قبـل الأخيرة أو الحد الأدنى للفئة قبل الأخيرة فهو عبارة عن تكرار الفئة الأخيرة مضافاً إليه تكرار الفئة قبل الأخيرة وذلك من التوزيع التكراري الأصـلي أي 5 + 8 = 13 وهكذا إلى أن نصل إلى التكرار المتجمع الهابط المقابل للحـد الأدنى للفئة الأولى والذي يجب أن يكون مساوياً للمجموع الكلي للتكرارات. أو يمكن القول أن التكرار المتجمع الهابط لفئة ما هو إلا التكـرار المتجمـع الهابط للفئة السابقة مطروحاً منه تكرار الفئة السابقة من التوزيع الأصلي . فمثلاً التكرار المتجمع الهابط للفئة الأولى هو (50) نظراً لعدم وجود فئـات سابقة للفئة الأولى . أما التكرار المتجمع الهابط للفئة الثانية فمـا هـو إلا التكرار المتجمع الهابط للفئة الأولى مطروحا منه التكـرار الأصـلي للفئـة الأولى (1 عامل) . كذلك فإن التكرار المتجمع الهابط للفئة الثالثة ما هو إلا التكرار المتجمع الهابط للفئة الثانية (49 عاملاً) مطروحـاً منـه التكـرار الأصلي للفئة الثانية (4 عامل) أي أن 49 – 4 = 45 . وهكذا إلا أن نصل إلى التكرار المتجمع الهابط للفئة الأخيرة (90 فأكثر) فيكون عبـارة عـن التكرار المتجمع الهابط للفئة قبل الأخيرة (80 فـأكثر) أي (13عـاملاً) مطروحاً منه تكرار الفئة (80–) من التوزيع الأصلي (أي 8 عامل) أي أن 13 – 8 = 5 . أما عن السطر الأخير في الجدول التكراري المتجمع الهابط

(100 فأكثر) فتعتبر كتابته اختيارية . وفي هذا الخصــوص يفتــرض أن هناك فئة وهمية تلي الفئة الأخيرة فيكون الحد الأدنى لتلك الفئة الوهمية ما هو إلا الحد الأعلى للفئة الأخيرة في التوزيع (90 – 100) أي القيمة 100 جنيه . فلا شك أن ليس هناك أي عامل يزيد أجره عن 100جنية على حسب بيانات هذا المثال .

(3): التكرارات المتجمعة الهابطة تبدأ بمجموع التكرارات وتنتهــي بتكــرار الفئة الأخيرة وذلك في حالة الوقوف بفئات التوزيع التكراري المتجمع الهابط عند الحد (90 فأكثر) في مثالنا محل الدراسة أو تنتهي بالتكرار المتجمــع الهابط صفراً وذلك في حالة افتراض أن هناك فئة وهمية تلي الفئة الأخيــرة وأن الحد الأدنى لتلك الفئة الوهمية ما هو إلا الحد الأعلى للفئــة الأخيــرة .

(4): جدول التوزيع التكراري المتجمع الهابط النسبي ما هو إلا صورة مــن جدول التوزيع التكراري المتجمع الهابط مــع الاخــتلاف الوحيــد وهــو أن التكرارات النسبية المتجمعة الهابطة النسبية (R_i.Dsc.C.F.) ما هــي إلا التكرارات المتجمعة الهابطة منسوبة إلــى (أو مقســومة علــى) مجمــوع التكرارات .

(5): من جدول التوزيع التكراري المتجمع الهابط يمكن الحصول على كثيراً من المعلومات والتي لا يمكن الحصول عليها بطريقة مباشــرة مــن خــلال استخدام التوزيع التكراري الأصلي . فمثلاً :

- هناك 38 عاملاً تبلغ أجورهم 60 جنيه فأكثر أسبوعيًا .
- أن قيمة الأجر الذي يحصل عليه أو أكثر منه 49 عاملاً هو 40 جنيه .
- أنه يمكن تحديد عدد العاملين الذين يحصلون على أجوراً تزيــد علــى أو تساوي قيمة معينة لا توجد بصورة صريحة في جدول التوزيع التكراري المتجمع . وعليه يمكن تحديد قيمة الأجر الذي يحصل عليه أو أكثر منــه

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 .

(6): من خلال قراءة الجدولين (17 – 2) ، (18 – 2) لاحظ الآتي :–
(التكرار المتجمع الصاعد + التكرار المتجمع الهابط) لأي فئة من فئات التوزيع يكون مساوياً لمجموع تكرارات التوزيع التكراري الأصلي. فمثلاً عدد العاملين الحاصلين على أقل من 30 جنيه + عدد العاملين الحاصلين على 30 جنيه فأكثر = صفر + 50 = 50 عامل وهو نفس المجموع الكلي للتكرارات . كذلك عدد العاملين الحاصلين على أقل من 40 جنيه + عدد العاملين الحاصلين على 40 جنيه فأكثر = 1 + 49 = 50 عامل وهو نفس المجموع الكلي للتكرارات . وهكذا لكل فئات التوزيعين التكراريين المتجمعين الصاعد والهابط . وكذلك على مستوى التوزيعات التكرارية النسبية المتجمعة الصاعدة والهابطة فإن مجموع التكرارين المتجمعين الصاعد والهابط النسبيان يكون مساوياً للواحد الصحيح أو 100 % .

أي أن :–

التكرار المتجمع الصاعد + التكرار المتجمع الهابط = مجموع تكرارات التوزيع الأصلي وذلك لأي فئة من فئات التوزيع . وكذلك فإن :
التكرار المتجمع الصاعد النسبي + التكرار المتجمع الهابط النسبي = 1 أو 100% لأي فئة من فئات التوزيع .

ثالثا: **باستخدام التوزيع التكراري الملائم تحديد** :-

(أ) – لتحديد عدد العمال اللذين تتراوح أجورهم ما بين 80 جنيه حتى أقل من 90 جنيه وكذلك نسبة هذا العدد من إجمالي التكرارات فأنه :

بداية مثل هذه التساؤلات يتم الإجابة عليها أولاً من خلال البحث عن ماهية وجود فئة في فئات التوزيع التكراري الأصلي كما هو معطى في المثال من عدمه . فإذا كانت هناك فئة صريحة في التوزيع التكراري الأصلي فإن الإجابة تكون من خلال التوزيع التكراري الأصلي وإلا فإن الإجابة تكون من خلال أحد التوزيعات التكرارية المتجمعة كما سنرى حالاً . فبالنسبة للفئة : (80 – 90 جنيه) فهناك فئة صريحة في التوزيع التكراري الأصلي تفيد بوجود 8 عمال يتراوح أجورهم ما بين 80 ، أقل من 90 جنيه . ومن ثم فإن النسبة المطلوبة وهي : –

نسبة العمال اللذين تتراوح أجورهم ما بين 80 وأقل من 90 جنيه

= (8 /50) × 100 = 16% .

أي أن هناك 16% من إجمالي عدد العاملين تتراوح أجورهم ما بين 80 وأقل من 90 جنيه اسبوعياً .

(ب) – عدد العمال اللذين تبلغ أجورهم الاسبوعين 73 جنيه على الأكثر وكذلك النسبة .

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

لاحظ أنه في مثل هذه الحالات فإن تمييز(وحدة قياس) العمودين اللذين يشكلان الجدول التكراري المتجمع الصاعد حيث تمثل فئات الأجور بالجنيه أما التكرارات المتجمعة الهابطة فوحدة قياسها هي العامل . والمعطى 73

جنيه تقع في عمود الفئات المتجمعة والمطلوب تحديد عدد العمال المقابل فى عمود التكرار المتجمع الصاعد . وهنا يجب التنويه أن فكرة إنشاء الجـــداول التكرارية البسيطة وكذا جداول التوزيعات التكرارية المتجمعة تبنـــى علـــى أساس أن التكرارات تتوزع بانتظام (بالنسبة والتناسب) على مـــدار طـــول الفئة الواحدة . لذا فإن قيمة الأجر الأسبوعي 73 جنيه تقع ما بين الأجر 70 جنيه ، 80 جنيه اللذان يقابلهم تكرارات متجمعة صـــاعدة كمـــا يوضـــحها الجدول(17−2) وهما 26 ، 37 عامل . لذا نفرض أن عدد العمال اللذين تقل أجورهم عن 73 جنيه وليكن (A) عامل ولاستنتاج القيمة A يتم تصـــوير الشكل الكروكي التالي :−

الفئات المتجمعة الصاعدة		التكرارات المتجمعة الصاعدة
Less than 70	$\longrightarrow$	26
Less than 73	$\longrightarrow$	A
Less than 80	$\longrightarrow$	37

وباستخدام فكرة النسبة والتناسب فإن :

(الحد الأخير− الحد الأول) / (الحد الأخير – الحد الأوسط) للفئات المتجمعـــة = (الحد الأخير− الحد الأول) / (الحد الأخير – الحـــد الأوســـط) للتكـــرارات المتجمعة. أي أن :

$(A - 37) / (26 - 37) = (73 - 80) / (70 - 80)$. أى أن :

$(A - 37) / 11 = 7 / 10$. وعلية فإن :

i.e., $10 (37-A)= 7 \times 11 = 77$

i.e., $370-77=10A$

فتكون قيمة A هى : $A = 293 \div 10$

$= 29.3 \approx 29$ (Persons)

ومن ثم فإن عدد العمال اللذين تقل أجورهم عن 73 جنبه ما هــو إلا 29 عاملاً . لاحظ أن الناتج لابد أن يقع ما بين التكرارين المتجمعين الصــاعدين أي ما بين 26 ، 37 عاملاً وهو معياراً لدقة الحسابات . ولاحــظ أيضــاً أن الناتج (29 عاملاً) يقترب من التكرار المتجمــع الصــاعد (26) حيــث أن المقابل على مستوى الفئات المتجمعة نجد أن الأجر 73 جنيه يقترب أيضــاً من 70 جنيه وليس 80 جنيه . أما عن النسبة المطلوبة فإن نسبة العمــال اللذين تقل أجورهم عن 73 جنيه (أو تبلغ أجورهم 73 جنيه على الأكثــر) فما هي إلا النسبة ما بين عدد العمال الذين تقل أجورهم عن 73 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 58 % من إجمالي عدد العمال تبلغ أجورهم 73 جنيه علــى الأكثر .

(جــ)– لتحديد قيمة الأجر الذي يتقاضاه 65 % من إجمالي عــدد العمــال على الأقل . لاحظ أولاً أن اللفظ على الأقل ≈ القيمة فأكثر . لذا فإن الإجابــة عن هذا التساؤل تكون من التوزيع التكراري المتجمع الهابط أو من التوزيع التكراري المتجمع الهابط النسبي .

فحيث أن 65 % من إجمالي عدد العمال ما هو إلا :

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

وهذا العدد الناتج (33 شخص) يقع ما بين التكرارين المتجمعين الهابطين 38 ، 24 كما هو موضح في الجدول (18-2) أو أن النسبة 65 % من الإجمالي تقع ما بين 76% ، 48% الواردة بنفس الجدول .

لذا نفرض أن قيمة الأجر الذي يتقاضاه 33 عاملاً على الأقل (65%) هو الأجر (x) . ولاستنتاج قيمة (x) يتم استخدام قاعدة النسبة والتناسب كما يلي :

More than 60 ⟶ 38
More than x ⟶ 33
More than 70 ⟶ 24

فيكون :

$$(33 - 24) / (38 - 24) = (x - 70) / (60 - 70)$$

أى أن : $(x - 70) / 10 = (-14) / (-9)$. أى أن :

i.e., $90 = 14(70 - x)$

$14x = 890$

i.e., $x = 63.571$ (L.E)

أي أن الدخل الذي يتقاضاه على الأقل 65% من إجمالي عدد العمال هو 63,571 جنيهاً .

لاحظ أن الدخل الناتج يقع ما بين 60 ، 70 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 47 عاملاً . لاحظ أن اللفظ كحد أدنى ≈ القيمة فأكثر . لذا يتم استنتاج المطلوب من جدول التوزيع التكراري

المتجمع الهابط . وبملاحظة العدد 47 عاملاً نجـد أنـه يقـع فـي عمـود التكرارات المتجمعة الهابطة (وليس الفئات) ما بين التكرارين المتجمعـين الهابطين 49 ، 45 عاملا . وبافتراض أن الأجر المطلـوب ولـيكن (Y).
فلاستنتاج قيمة (y) يتم تكوين الرسم الكروكي التالي :
استخدام قاعدة النسبة والتناسب كما يلي :

More than 40 $\longrightarrow$ **49**
More than (Y) $\longrightarrow$ **47**
More than 50 $\longrightarrow$ **45**

فيكون :-

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

i.e.,

$$\frac{10}{50-Y} = \frac{-4}{-2}$$

i.e. ,

$$20 = 4\,(50-Y)$$
$$4y = 200-20$$

i.e. ,

$$Y = 180 \div 4 = 45 \qquad \text{(L.E)}$$

أي أن قيمة الأجر الذي يتقاضاه 47 عاملاً كحد أدنى ما هو إلا 45جنية.

<u>ملحوظة</u> : لاحظ من النتائج السابقة أنه حينما كان عدد العمال 47 يقع في منتصف التكرارات المتجمعة الهابطة ما بين 49 ، 45 عاملاً بالضبط وهى بالضبط فإن الناتج وهو قيمة(Y)القيمة 45 جنية تقع فى منتصف الفئة أي في نقطة المنتصف ما بين 40 ، 50 جنيه . وبلغة أخرى : فحيث أن:

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

لذا فأن :

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

وهى نفس النتيجة السابقة .

(هـ): لتحديد نسبة العمال الذين تتراوح أجورهم ما بين 55 ، 87 جنيــه. فحيث أنه لا توجد فئة صريحة من فئات التوزيع التكراري الأصلي الــواردة بالجدول (16–2) تعطي صورة مباشرة عن تلك الفترة أو هذه الفئة مــن(55–87) لذا يجب استخدام أحد جداول التوزيعات التكرارية المتجمعة أمــا الصاعد أو الهابط . والآن دعنا نستنتج المطلوب مــن التوزيــع التكــراري المتجمع الصاعد . فحيث أن الدخل 55 جنيه يمثل نقطة المنتصف للفئة 50 – 60 جنيه . لذا فإن :

عدد العمال اللذين يقل أجرهم الأسبوعي عن 55 جنيه = (5+12) ÷ 2 أي 7,5 ≈ 8 عاملاً .

هذا وبافتراض أن عدد العمال اللذين يقل أجرهم الأسبوعي عن 87 جنية هو (B) عامل فإنه لإستنتاج القيمة (B) فإن ذلك يكون من خلال جدول التوزيع التكراري المتجمع الصاعد الوارد بالجدول (17 – 2) . فحيث أن الأجر 87 جنية يقع فى الفئة من 80 إلى 90 وعليه يكــون لــدين الشــكل الكروكى التالى :

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

فيكون :

i. e. ,

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

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

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

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

i.e. , B = 42.6 ≈ 43 (Persons)

والآن دعنا نلخص النتائج المستخلصة حتى الآن :-

هناك 9 عمال يقل أجرهم عن 55 جنيه . وهناك 43 عامل يقل أجرهم عـــن 87 جنيه .ويمكن وضع ملخص للنتائج السابقة فى الشكل الكروكي التـــالي:

Less than 30 L.E $\longrightarrow$ 0 Persons
Less than 55 L.E $\longrightarrow$ 9 Persons
Less than 87 L.E $\longrightarrow$ 43 persons

فيكون عدد العمال اللذين تتراوح أجورهم الأسبوعية ما بين 55 جنيه ، 87 جنيه هو بمثابة الفرق ما بين 43 ، 9 عمـــال أي 43 – 9 = 34 عـــاملاً .
وعليه فإن نسبة العمال اللذين تتراوح أجورهم الأسبوعية ما بين 55 ، 87 جنيه هي : (34 / 50) % أى 68 %.
أي أن هناك 68 % من إجمالي عدد العمال يتراوح أجرهم الأســبوعي مـــا بين 55 ، 87 جنيهاً .

(و): حيث أنه لا توجد فئة من فئات التوزيـــع التكـــراري الأصـــلي الـــوارد بالجدول (16−2) تبدأ بالقيمة 53 جنيه . لذا فإن هـــذا المطلـــوب لا يمكـــن استنتاجه من التوزيع التكراري الأصلي . ويمكن استنتاج هذا المطلوب مـــن أحد التوزيعين التكرارين المتجمعين أما الصاعد وأما الهابط .

والآن دعنا نلخص معطيات هذا المطلوب واستنتاجه ولـــيكن مـــن التوزيـــع التكراري المتجمع الصاعد . فحيث أن 40% من إجمالي عدد العمال هو:

50 × 40% = 20 عامل . وحيث أن عدد العمال اللذين يقل أجرهم عن 53 جنيه يمكن استنتاجه من خلال افتراض أن هذا العـــدد ولـــيكن (z) عامل . فلاستنتاج قيمة (z) يتم الآتي :−

Less than 50 L.E ⟶ 5
Less than 53 L.E ⟶ Z
Less than 60 L.E ⟶ 12

فيكون :

(60 − 50) / (60 − 53) = (12 − 5) / (12 − Z) . وعليـــــه فإن:

$10\ z = 120 - 49 = 71$

وعليه فإن: $Z = 7.1 \approx 7$ (persons)

لاحظ أن الناتج 7 أشخاص يقع ما بين التكرارين المتجمعين الصـــاعدين 5 ، 12 المقابلين للفئات المتجمعة 50 ، 60 جنيه .

والآن لاستنتاج قيمة الدخل (x) فدعنا نلخص كلا من المعطيـــات ونتائجنـــا حتى الآن وهي :−

*** هناك 7 عامل يقل أجرهم الاسبوعي عن 53 جنيه .**

*** هناك 20 عاملاً يتراوح دخلهم ما بين 53 ، x جنيه**

ولاستنتاج قيمة الدخل x دعنا نفرض الفرضين التاليين :-

الأول: وهو ان قيمة الدخل (X) < 53 جنيه .

والثانى : ان قيمة الدخل (X) > 53 جنيه.

هذا ولإحتبار مدى صحة هذين الفرضين فيكون ذلـك مـن خـلال الرسـم الكروكي التالي للحالتين السابقتين :

أولا: حالة إفتراض أن قيمة الدخل (X) تقل عن أى < 53 جنيـه . أي أن قيمة X تقع مابين الدخل 30 ، 53 جنية . وعليه فإنه يمكن تلخيص النتائج السابقة فى الشكل الكروكى التالي :

Less than 30 L.E	↓	0	Persons
Less than X L.E	7	↓	Persons
Less than 53 L.E		20	persons

ومن خلال الرسم الكروكى السابق يتضح لنا أن هذا الأتجاة مرفوض لأنه يتنافي مع المنطق الرياضـى الذى يحكم بضـرورة أن يكـون الجزء أقل من الكل .

ثانيا : حالة إفتراض أن **قيمة الدخل (X) > 53 جنيه. فإنه يمكن تلخيص النتائج السابقة فى الشكل الكروكى التالى:**

30 L.E **0** ↓ **7 person**

53 L.E

X L.E ↓ **20 person**

وهذ الأتجاة صحيحا نظرا لأستقلال السهمين فى هذه الحالة .
وعليه فمن 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 = 7 + 20 = 27 عامل. ومـــن ثـــم فالمطلوب الآن هو تحديد قيمة الأجر(x) الذي يقل عنه 27عــاملاً . وبــالرجوع للجدول التكراري المتجمع الصاعد الوارد بالجدول (17– 2) نجــد أن التكـــرار المتجمع الصاعد 27 يقع ما بين التكـــرارين المتجمعـــين الصـــاعدين 26 ، 37 المقابلين للفئتين المتجمعتين 70 ، 80 جنيه على الترتيب ومن ثم يصـــبح لـــدينا الشكل الكروكى التالي :

Less than 70 ⟶ 26
Less tha x ⟶ 27
Less than 80 ⟶ 37

فيكون :

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

$$\frac{10}{80 - x} = \frac{11}{10}$$

i.e. , $11(80 - x) = 10 \times 10 = 100$
i.e. , $880 - 100 = 11x$

فتكون قيمة x هي :

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

أي أن قيمة الدخل (x) هو 70,909 جنيه .

<u>ملحوظة</u> :- في المطلوب الأخير من هذا المثال لاحظ أنه تم رفض الفـــرص الأول الذي افترض أن قيمة x تقل عن 53 جنيه نظراً لأن هذا الفـــرض يتنـــافى مـــع المنطق الرياضي الذي يحكم بضرورة أن يكون الجزء(وهو الرقم الذى يدل على 20 عامل فى الرسم الكروكى) أقل من الكل (وهو الرقم 7 عامـــل فـــى نفـــس الرسم) . أي أن هذا الفرض سيتم رفضه في الحالات التي يتنـــافى فيهـــا مـــع المنطق الرياضي . لكن إذا كانت معطيات التمرين وما تـــم اســـتنتاجه أفـــادت أن الجزء أقل من الكل فسوف تكون النتيجة هي قبول هذا الفرض أيضاً . وفي هـــذه الحالة يتم طرح التكرارين المتجمعين اللذين تم استنتاجهم بدلاً من الجمع كما هو في الفرض الثاني من هذا المثال . وخلاصة ما سبق هو أنه فـــي حالـــة قبـــول الفرض الأول فستكون هناك إجابتين صحيحتين لقيمة الأجر (x) . حيث أن الاتجاه الثاني سيكون مقبول دائماً نظراً للاستقلال دائماً فيما بين الجـــزئين اللـــذين تـــم استخلاصهم من المعطيات ونتائج الحل .

<u>ثانياً : – العرض البياني للبيانات الاحصائية :</u>

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

وخلاصة الأمر هو أنه رغم أن عملية عرض البيانات جدولياً تيسر على القــارئ فهم هذه البيانات كما تساعد على استخلاص النتائج والعلاقات التي تــربط بــين الظواهر المختلفة إلا أن هذه العملية (العرض الجدولي) قد لا تلقي قبــولاً مــن القارئ العادي الغير متخصص حيث يجد صعوبة في عملية استنباط النتائج مــن 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 فــي إجراء عمليات المقارنة والتي يمكن إدراك نتائجها بشكل أيسر وأسرع ممــا لــو استخدمنا العرض الجدولي لهذه البيانات . هذا بالإضافة إلى أن العرض البيــاني يمكننا من تحديد أشكال التوزيعات التكرارية المختلفة . وبصفة عامة وكمــا هــو الحال في العرض الجدولي فإن العرض البياني للظواهر يعتمد على طبيعة البيانات أي طبيعة المتغيرات محل الدراسة من حيث كونها وصــفية أم كميــة بنوعيهــا المنفصل أو المتصل . أي أننا سنتناول بشيء من الإيجاز والتوضــيح العــرض البياني للظواهر أو المتغيرات المختلفة سواء كانت :-

أ – الظواهر أو المتغيرات وصفية ومبوبة تكراريا.

ب- الظواهر أو المتغيرات كمية بنوعيها المنفصل والمتصل والمبوبة تكرارياً.

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

<u>أولاً :- العرض البياني للبيانات الوصفية :-</u>

وهناك مجموعة من الأساليب البيانية التي تستخدم في عــرض الجــداول التكرارية للظواهر الوصفية بيانياً منها :

1- 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 لقياس تكرار فئات 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 (7) : فيما يلي الجدول التالي يوضح تقديرات نجاح 15 طالباً في مادة الاقتصاد :-

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

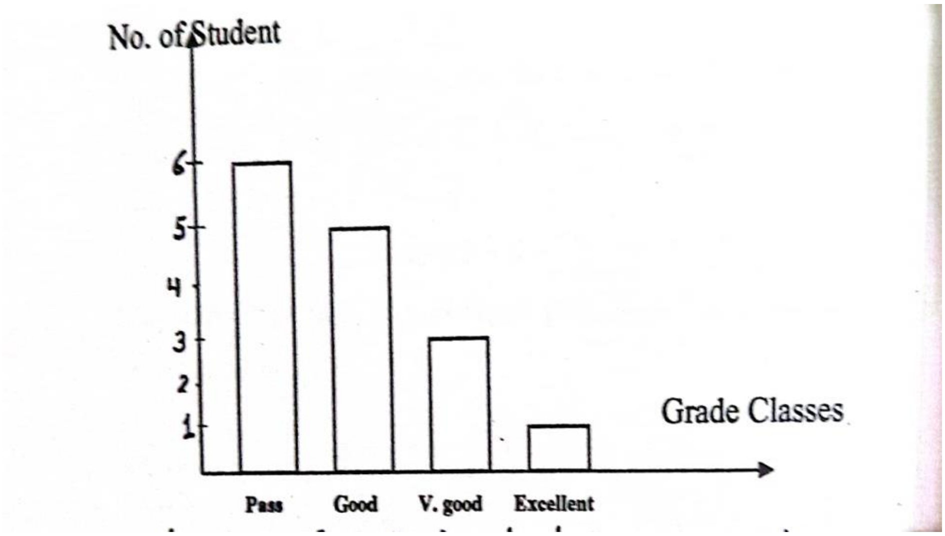
والمطلوب تمثيل بيانات هذا الجدول باستخدام الشكل البياني الملائم .

<u>الحل :-</u>

حيث أن الظاهرة محل الدراسة – تقدير الطالب – هي ظاهرة وصفية لذا يمكن عرضها بيانياً باستخدام الأعمدة البيانية (المستطيلات)كما يوضحها الشكل(2-1) .

شكل (2-1)

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 للتفرقة بينهما وفيما بين الأعمدة الملتصقة Clustered Bar Charts والأعمدة المجزأة Stacked Bar Charts

مثال (8) :- فيما يلي الجدول التالي يبين نتيجة طلاب إحدى كليات التجارة في إحدى السنوات :

جدول (2-20)

الفرقة / النتيجة	طلاب ناجحون Pass	طلاب منقولون بمواد Passwith subj.	طلاب راسبون Failled	Σ
الأولى	2000	1200	800	4000
الثانية	1300	400	300	2000
الثالثة	700	500	600	1800
الرابعة	1200	800	1000	3000

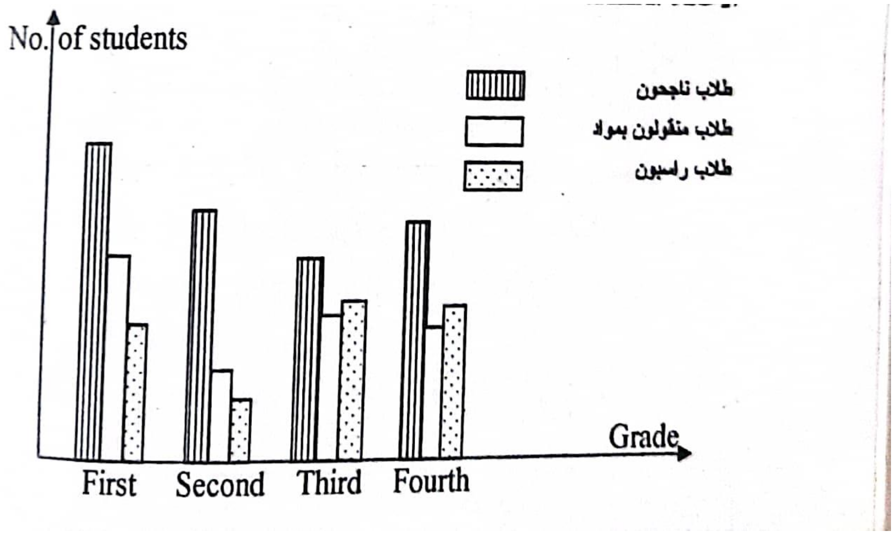
<u>والمطلوب</u>: استخدام الشكل البياني الملائم لعرض بيانات هذا الجدول .

<u>الحل</u> :-

يمكن تمثيل بيانات الطلاب (ناجحون – منقولون بمواد – راسبون) لكل فرقة على حدة إما باستخدام الأعمدة أو المستطيلات الملتصقة أو المجزأة السابق الإشارة إليها فيما قبل المثال السابق وهو ما يوضحه الشكل البياني (2-2-أ) ، (2-2-ب) .

شكل (2 – 2 – أ)

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



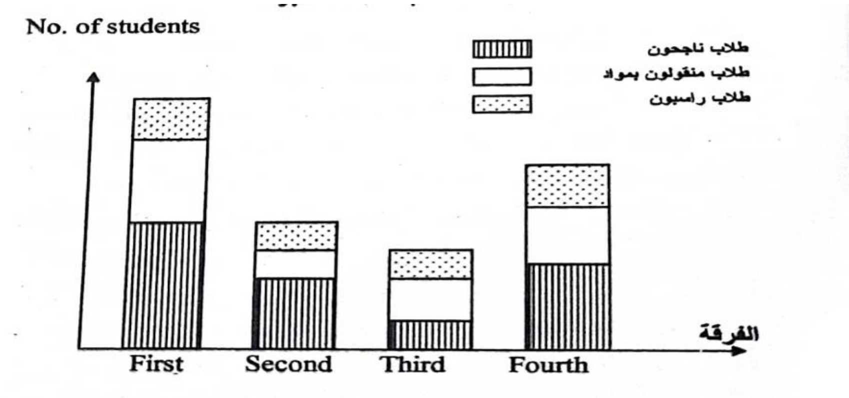
حيث

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

أما عن تمثيل كل فرقة بعمود واحد يتم تقسيمه إلى ثلاثة أجزاء تبين على الترتيب الطلاب الناجحون والمنقولون بمواد والراسبون وهو ما يوضحه الشكل (2 – 2 – ب).

الشكل (2 – 2 – ب)

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



هذا ويمكن تمثيل البيانات السابقة باستخدام سواء الأعمدة الملتصقة أو المجـزأة وذلك على أساس نسبي متبعين في ذلك نفس الأسلوب السابق . حيث يكون فـي 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 عمليـات المقارنة فيما بين أقسام الظاهرة في السنوات المختلفة وذلك على أساس نسـبي وليس مطلق . ويكون ذلك بصفه خاصة إذا ما اختلفت مجاميع التكـرارات كمـا سبق ذكره .

<u>ملاحظات</u> :عند استخدام الأعمدة البيانية كأسلوب من أساليب العـرض البيـانى فإنه يجب مراعاة ما يلى :-

1- ان تكون قواعد الأعمدة متساوية.

2- حيث ان مساحة المستطيل (العمود) = القاعدة × الارتفاع.

ولما كانت قواعد الاعمدة متساوية فإن مساحة الاعمدة تتناسـب مـع ارتفاعهـا والتى تعكس بدورها قيم تكرار اقسام 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 .

3- ان تكون المسافات فيما بين الاعمدة متساوية.

4- 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 .

<u>مثال رقم(8)</u>: فيما يلى الجدول التالى يبين قيمة المبيعات بآلاف الجنيهات فى احدى السنوات وذلك لخمسة مصانع تنتج المنسوجات القطنية.

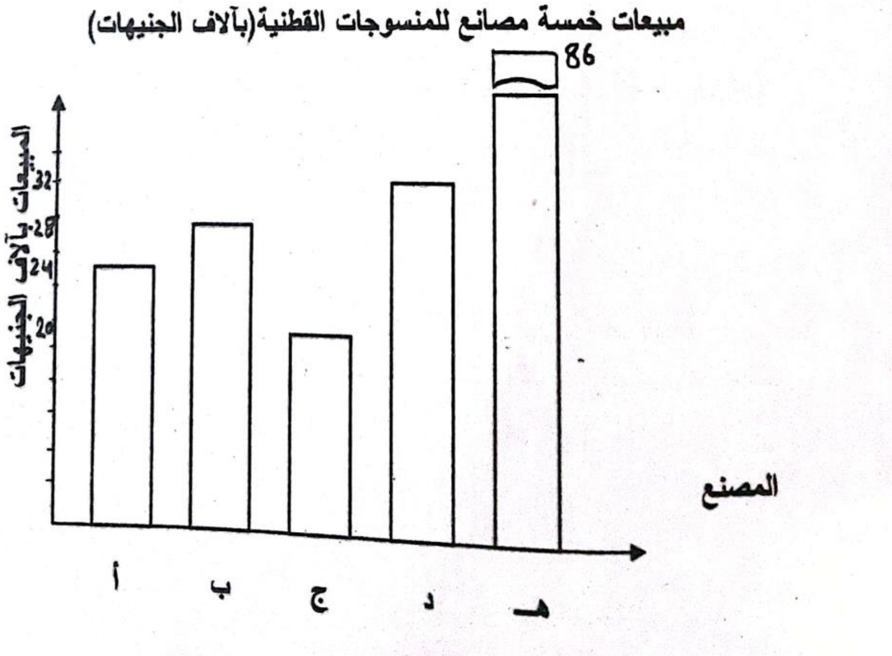
جدول(21-2)

المصنع	أ	ب	ج	د	هـ
قيمة المبيعات	24	28	20	32	86

<u>والمطلوب</u> : عرض هذه البيانات بإستخدام الاعمدة البيانيه .

<u>الحل</u> : 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(2-3) .

شكل(2–3)



2– 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 آخر زاوية مركز الدائرة الى قطاعات دائرية تتناسب مساحتها مع النسب المئوية للمكونات الجزئية للظاهرة . ولما كانت الزاوية المركزية للدائرة مساوية 360° درجة فأن الزاوية المركزية لكل قطاع جزئى تتحدد بضرب تلك النسبة المئوية المقابلة لكل جزء او قسم من اقسام الظاهرة فى 360 . ثم يتم تظليل كل قطاع بشكل مخالف.

أما اذا كنا نود اجراء المقارنة فيما بين ظاهرتين كل واحدة تحتوى على مجموعة من الأوجه، ففى هذه الحالة يتم تخصيص دائرة لتمثيل كل ظاهرة على حده . ولكن يجب فى تلك الحالة رسم مساحة دائرتين متناسبتين مع قيمة الظاهرتين ، وحيث ان:

$$\frac{\text{مساحة الدائرة الاولى}}{\text{مساحة الدائرة الثانية}} = \frac{\text{القيمة الاجمالية للظاهرة الاولى (ق1)}}{\text{القيمة الاجمالية للظاهرة الثانية(ق2)}}$$

اى ان :

$$(\text{ق}_1 / \text{ق}_2) = (\text{ط نق}_1{}^2 / \text{ط نق}_2{}^2) = (\text{نق}_1{}^2 / \text{نق}_2{}^2)$$

حيث نق_1 ، نق_2 هما نصفى قطر الدائرة الاولى والثانية على الترتيب وبإعتبـــار الجذر التربيعى للعلاقة السابقة فيكون :-

$$\frac{\text{نق}_1}{\text{نق}_2} = \sqrt{\frac{\text{القيمة الاجمالية للظاهرة الاولى (ق1)}}{\text{القيمة الاجمالية للظاهرة الثانية(ق2)}}}$$

فمثلا اذا كانت اجمالى قيم اقسام الظاهرة الاولى هو1600 واجمالى اقسام الظاهرة الثانية هو900 فأن:-

$$\frac{\text{نق}_1}{\text{نق}_2} = \left(\frac{1600}{900}\right)^{2/1} = \left(\frac{4}{3}\right)$$

فاذا افترضنا ان الدائرة الاولى نصف قطرها 6 ســـــــــــــــــــم فأن:

نصف قطر الدائرة الثانية = $6 \times (\text{نق}_2 / \text{نق}_1) = 6 \times (4/3)$

= 4.5 سم

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

زاوية اى قطاع=(قيمة الظاهرة فى هذا القطاع/مجموع قيم الظاهرة)× 360 °

<u>مثال (9)</u> : فيما يلى البيانات التالية تمثل المبالغ المنفقه (بالمليار جنيه) على القطاعات المختلفة فى احدى الدول عام 1994م :

القطاع	المبالغ المنفقة بالمليار جنيه
القطاع الصناعى	20
القطاع الزراعى	16
القطاع السياحى	4
قطاع الخدمات	8

والمطلوب استخدام الدوائر لتمثيل هذه البيانات بيانيا.

<u>الحل</u>:-لعرض هذه البيانات بإستخدام الدائرة فإننا نقوم برسم دائرة بنصف قطر مناسب . ثم يتم تقسيم المساحة الكلية للدائرة الى قطاعات دائرية حيث أنا :

الزاوية التى تمثل اى قطاع =(قيمة القطاع / اجمالى القطاعات) × 360

اى ان :

الزاوية التى تمثل القطاع الصناعى=(20 /48)× 360 =150° درجة

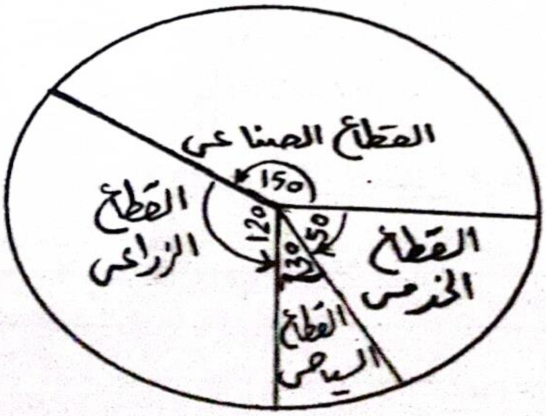
والزاوية التى تمثل القطاع الزراعى = (16 / 48) × 360 = 120°

والزاوية التى تمثل القطاع السياحى = (4 / 48) × 360 = 30°

والزاوية التى تمثل القطاع الخدمى = (8 / 48) × 360 = 60°

ومن ثم يكون شكل الدائرة على النحو المبين بالشكل (2-4) الاتى :-

شكل(2-4)



مثال(10) : الجدول التالى يوضح اعداد الطلاب (ذكورا واناثا) فى إحدى الكليات خلال السنوات(1960-1965) .

جدول(2-22)

اعداد الطلاب ذكورا واناثا فى احدى الكليات

خلال السنوات1960-1965

السنة	1960	1961	1962	1963	1964	1965
ذكور	120	150	175	210	280	320
إناث	90	100	120	150	160	175

والمطلوب مقارنة عدد الطلاب فى عامى 1960 ، 1965 بنوعيهم بيانيا بإستخدام الدوائر.

الحل :

للمقارنة فيما بين عدد الطلاب بنوعيهم فى عامى 1960، 1965 يتم الاتى:-

(1):تحديد نصف قطر كلا من الدائرتين ورسم دائرة لكل نوع على حده:-

فحيث ان :

$$\frac{نق_1}{نق_2} = \left(\frac{\text{مجموع الظاهرة خلال عام 1960}}{\text{مجموع الظاهرة خلال عام 1965}}\right)^{0.5}$$ أى أن:

$$= \left(\frac{90+120}{175+320}\right)^{0.5} = \left(\frac{210}{495}\right)^{0.5} = 0.65$$

لذلك يمكن اختيار نصف قطر الدائرة الاولى = 0.65 ونصف قطر الدائرة الثانية = 1 أو اختيار اى قيم اخرى لأنصاف الاقطار للدائرتين على ان تكون النسبة فيما بينهما مساوية دائما للقيمة0.65 فمثلا يمكن ضرب كلا من نق1 ، نق2 فى اى مقدار ثابت وليكن مثلا (5) فنجد ان:

نق $_1$ = 0.65 × 5 =3.25 سم ، نق$_2$= 1 × 5 = 5 سم

(2): تحديد زوايا كل قطاع : وهى ما يوضحها الجدول التالى:

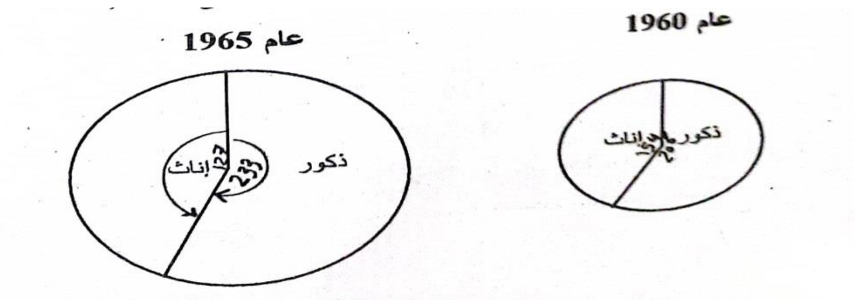
السنة النوع	1960		1965		زاوية القطاع	
	عدد الطلاب	النسبة %	عدد الطلاب	النسبة %	1960	1965
ذكور	120	%57	320	%65	=360×210/120 206	=360×495/320 233
إناث	90	%43	175	%35	=360×210/90 154	=360×495/175 127
مجـــــــــــ	210	100 %	495	%100	360	360

وبذلك يمكن رسم الدائرتين مع تحديد القطاعات داخل كل دائــرة كمــا يوضــحه الشكل(2-5).

شكل(2-5)

توزيع الطلاب(ذكورا،واناثا) فى احدى الكليات فى عامى 1965,1960

عام 1960 عام 1965



واخيرا فإن عيوب استخدام الدوائرفى 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 ابيانية. وعليه فأنــه إذا كــان عــدد اقســام الظاهرة كبيرا فإنه من الافضل عدم استخدام الدائرة وذلك لصعوبة تمييز كل قطاع فى هذه الحالة ويفضل هنا استخدام الاعمدة البيانية.

ثانيا : العرض البيانى للظواهر الكمية المبوبة تكراريا :-

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

بصفة عامة فإنه فى حالة البيانات المبوبة(الجداول التكرارية) للظاهرة التــى تــم التعبير عنها بمتغير كمى متقطع(منفصل) يتم تمثيلها بيانيا وذلك من خلال تحديد اقسام الظاهرة او المتغير على المحور الافقى ويتم رسم او اسقاط اعمدة تتناسب فى ارتفاعها مع التكرار المقابل لكل قسم اى ان ارتفاع كل عمود يكون مســاويا لقيمة تكرار المتغير المتقطع .

<u>مثال(11) :</u>

فيما يلى الجدول التالى بين توزيع 20اسرة حسب حجم الاسرة :

جدول(23-2)

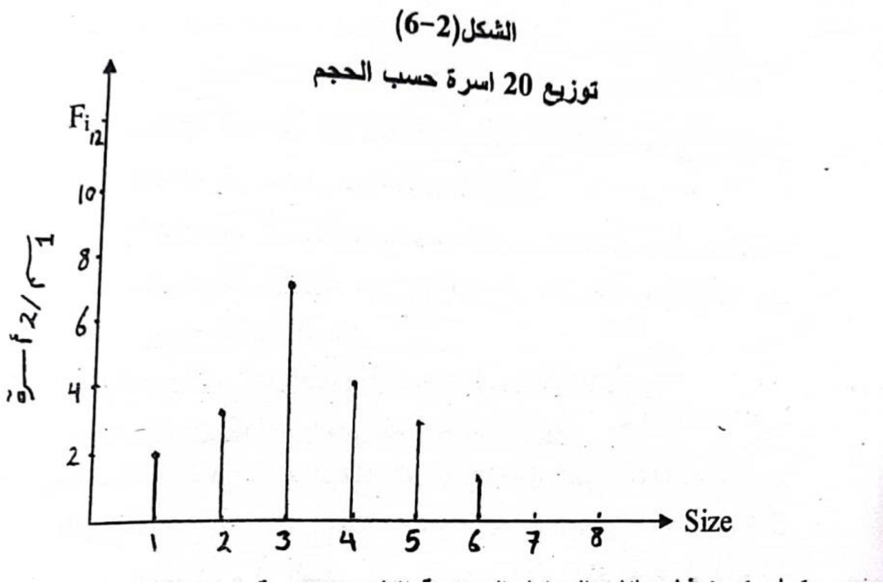
توزيع20اسرة حسب الحجم

Family size	1	2	3	4	5	6	∑
Frequency(F_i)	2	3	7	4	3	1	20

<u>الحل:-</u> يتم تمثيل حجم الاسرة على المحور الافقى والتكرارات علـــى المحـــور الرأسى بمقياس رسم مناسب(1سم / 2 وحدة تكرار). والشـــكل(2-6) يوضـــح التمثيل البيانى للجدول(23-2):

الشكل(2-6)

توزيع 20 اسرة حسب الحجم



كما يمكن تمثيل بيانات الجداول التكرارية للظواهر الكمية المنفصلة باستخدام التكرارات النسبية بدلا من التكرارات المطلقة وفى تلك الحالة فإنه يتم تمثيل التكرارات النسبية على المحور الرأسى.

ب_ التمثيل البيانى لجداول الظواهر الكمية المتصلة (المستمرة) :

هناك ثلاثة طرق رئيسية للتمثيل البيانى للجداول التكرارية التى تحوى متغيرات كمية متصلة وهذه الطرق هى :

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

هذا وعند رسم اى من هذه 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 حيث أن :

$$\text{التكرار المعدل للفئة}(MFi) = \frac{\text{التكرارالأصلي الفئة } (i)}{\text{طول الفئة} (i)}$$

اى ان :

$$MF\,i = \frac{F_i}{L\,i}$$

حيث :

where i= 1,2,……..,r

وسوف نتعرض لدراسة الاشكال البيانية المستخدمة فى التمثيل البيانى للظــواهر الكمية المتصلة بشئ من التفصيل كما يلى :-

(1):المدرج التكرارى Histogram :

والمدرج التكراراى ماهو إلا مجموعة من المستطيلات المتلاصقه كــل مســتطيل 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):رسم المدرج التكرارى للجداول المنتظمة(متساوية اطوال الفئات):

وتتلخص الخطوات فى الاتى :-

- رسم المحورين الافقى والرأسى. بحيث يتم تمثيل فئات جدول التوزيع التكراري على المحور الافقى وتكرارات تلك الفئات على المحور الرأســى بمقيــاس رســم مناسب . هذا ويراعى عند رسم المحور الافقى انه ليس من الضرورى ان يبــدأ هذا المحور فى تدريجه من الصفر ولكن يمكن أن يبدأ من فئة سابقة عن الفئـــة الاولى فى الجدول وبنفس طول هذه الفئة الاولى . كما يراعى ان يكــون هنـــاك 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 . وفى حالة ما إذا كانت تكرارات التوزيع التكراري جميعا كبيرة نسبيا وتبعد كثيرا عــن الصــفر فيمكن تصور عمل قطع (على الصورة ≈) في المحور الرأسى ثم البداية مــن بعد هذا القطع برقم قريب نسبيا من أصغر قيمـــة فـــي قــيم تكــرارات التوزيـــع التكراري.

- يتم تمثيل كل فئة بمستطيل تمثل قاعدته طول الفئة اما ارتفاعه فيمثل تكــرار الفئة. وحيث ان الجدول التكرارى الذى نحن بصدده متساوى من حيــث اطــوال الفئات فإن مساحة المستطيلات المختلفة سوف تتناســب فــى ارتفاعاتهــا مــع التكرارات المقابلة التى تمثلها هذه المستطيلات . ومن ثم فإن إجمالى مســاحات هذه المستطيلات المختلفة سوف يكون متناسبا مع المجموع الكلى للتكرارات.

والمثال التالى يوضح كيفية رسم المدرج التكرارى لجدول تكرارى منتظم .

<u>مثال(12):</u> فيما يلى لديك التوزيع التكرارى التالى لأجور 50عاملا :

جدول(23-2)

فئـــــات الاجر	30-	40-	50-	60-	70-	80-	90-	مجـــــــ
التكرارات	1	4	7	14	11	8	5	50

<u>والمطلوب</u> تمثيل هذا التوزيع بيانيا بإستخدام الشكل البيانى الملائم .

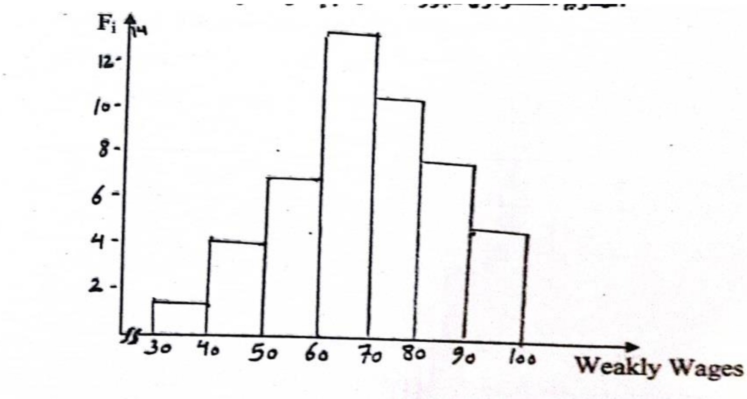
<u>الحل:</u>-

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

<u>وحيث ان</u> : الجدول التكرارى فئاته متساوية من حيث الطول لذا يتم رسم المدرج التكرارى مباشرة .حيث يتم تمثيل فئات الاجر على المحور الافقى وعدد العمــال(التكرارات) على المحور الرأسى بمقياس رسم وليكن 1سم/2 عامــل كمــا هــو موضح فى الشكل(2-7).

شكل(2-7)

المدرج التكرارى لأجور 50عاملا بإحدى الشركات



هذا ويمكن أيضا تمثيل تلك البيانات باستخدام التكرارات النسبية بدلا من التكرارات المطلقة. وفى تلك الحالة فإن المحور الرأسي يتم تمثيل عليه التكرارات النسبية للعمال .

(ب):رسم المدرج التكرارى فى حالة الجداول التكرارية الغير منتظمة: 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:

$$MF\,i = F_i / L_i$$
$$\text{where} \qquad i = 1,2,\ldots\ldots,r$$

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

مثال(13) :الجدول التالى يبين توزيع درجات مجموعة من الطلاب فى إحدى المواد.

جدول(2–24)

Classes	5-	10-	20-	30-	50-	60-	80-100	∑
No.of students	15	70	100	120	200	300	160	965

والمطلوب رسم المدرج التكرارى لتوزيع الدرجات .

الحل :–

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(2–25) الاتى :

جدول (2–25)

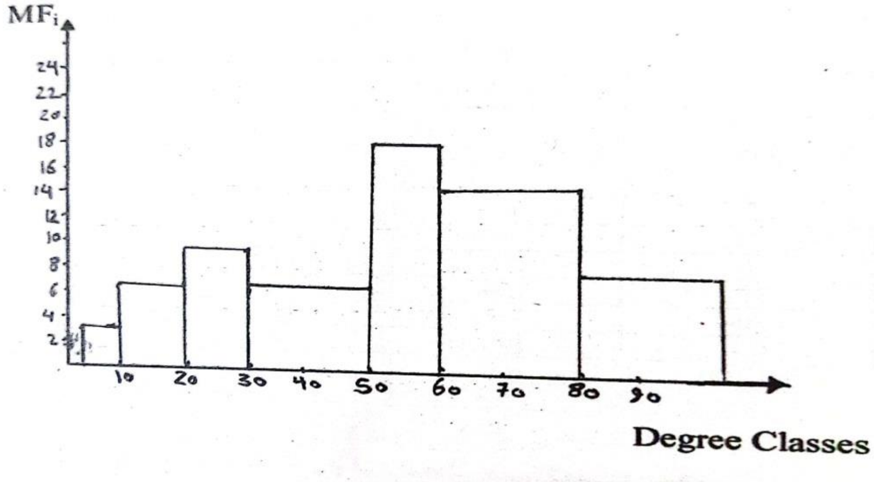
حساب التكرار المعدل (MFi) لتوزيع الدرجات

Score Interval	F_i	L_i	$MF_i = F_i / L_i$
5 -	15	5	15 / 5 = 3
10 -	70	10	70 /10 = 7
20 -	100	10	100 / 10 = 10
30 -	120	20	120 / 20 = 6
50 -	200	10	200 /10 = 20
60 -	300	20	300 / 20 = 15
80 – 100	160	10	160 / 10 = 16
∑	965		

والهدف من حساب التكرارات المعدلة هــو تحقيــق التناســب بــين مســاحات المستطيلات المكونة للمدرج التكرارى وارتفاعاتها . أى تحقيق التناســب مــابين مساحة المستطيل والتكرار الذى يمثله هذا المستطيل. هذا وبرسم الفئات كما سبق على المحور الافقى بمقياس رسم مناسب وكذا التكرارات المعدلة بمقياس رســم مناسب على المحور الرأسى وليكن 1سم/ 2 وحدة تكرار معدل نحصل على الرسم البيانى التالى الموضح فى الشكل (2–8).

شكل(2–8)

المدرج التكرارى لتوزيع درجات الطلاب



ويلاحظ على الشكل(2−8) 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) وعليه فإن ناتج مساحة كل مستطيل(الطول × العرض) ماهو إلا التكرار الاصلى للفئة والذى يمثل مساحة هذا المستطيل(وذلك لأن التكرار الأصلى = طول الفئة × التكرار المعدل للفئة). ومن ثم فإنه إذا ماتم إيجاد مجموع مساحات المستطيلات المكونة للشكل البيانى السابق فإن ناتج المجموع سيكون بمثابة مجموع التكرارات الاصلية للتوزيع وهى 965 طالب فى هذا المثال.

<u>(2):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 منتصفات أو مراكز الفئات المقابلة للتكرارت الاصلية (فى حالة الجداول التكرارية المنتظمة)أو التكرارات المعدلة(فى حالة الجداول الغير منتظمة) وبصفة عامة فإن مركز الفئة ((Mid Point X_i))يمكن استنتاجها من خلال احد الزوايا التالية:−

$$\text{مركز الفئة} = \frac{\text{الحد الأدنى للفئة} + \text{الحد الأعلى للفئة}}{2}$$

أو :

X_i or MP =(Class Lower Limit+Class Upper Limit)/ 2

i.e., X_i = (C.L.L + C.U.L) / 2

كذلك يمكن إيجاد مركز الفئة من خلال اضافة نصف طول الفئة للحد الادنى

الفئة . أى أن :

$$X_i \text{ or } MP = C.L.L + 0.5\, L_i \,(\text{ or } W_i)$$

$$\text{Where : } i = 1, 2, 3, \ldots\ldots, r$$

حيث(r) تمثل عدد فئات التوزيع.

هذا ويجب التنويه الى انه عند رسم المدرج أو المضلع وكذلك المنحنى التكرارى 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(أو المنحنى التكرارى) فيجب بداية التأكد من مدى إنتظام التوزيع التكرارى من عدمه أولا. فإذا كان التوزيع التكرارى منتظما يتم حساب مراكز الفئات والرسم مباشرة حيث يتم تمثيل مراكز الفئات على المحور الافقى والتكرارات الاصلية 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 ومايقابلها من تكرار اصلى بخط منكسر بالمسطرة فنحصل 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(14): الجدول التالى يوضح توزيع الاجور الشهرية لمجموعة من العاملين بإحدى الشركات .

جدول(25−2)

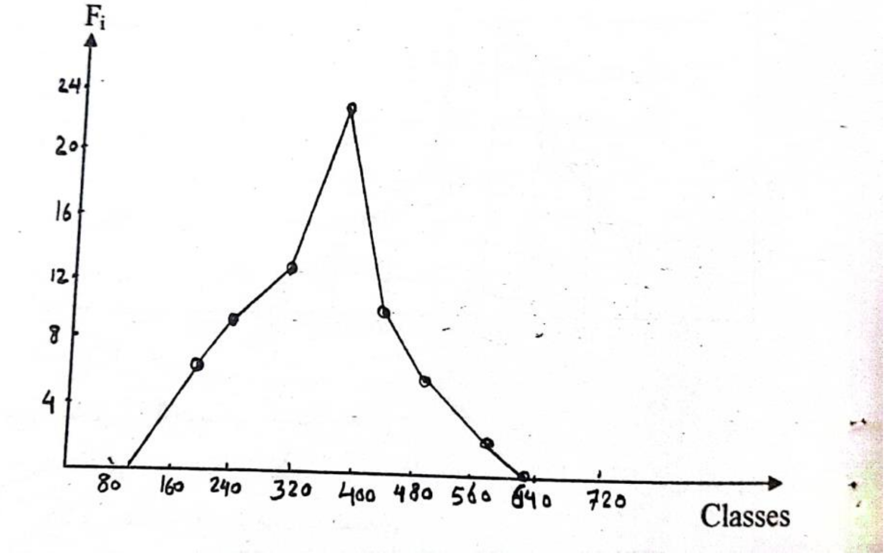
Classes	150-	220-	270-	360-	430-	500-	570-640
Fi	6	10	18	25	9	6	2

الحل :

لاحظ ان بيانات التوزيع التكرارى الموضحة فى الجدول(23-2)عبارة عن جدول تكرارى منتظم لذا يجب رسم المضلع التكراري مباشـــرة . حيـــث يـــتم تمثيـــل مراكزالفئات على المحور الافقى . كما يتم تمثيل التكرارات الاصلية على المحور الراسى بمقياس رسم مناسب كما هو موضح بالشكل(2-9).

شكل(2-9)

رسم المضلع التكرارى للأجر الشهرى للعاملين

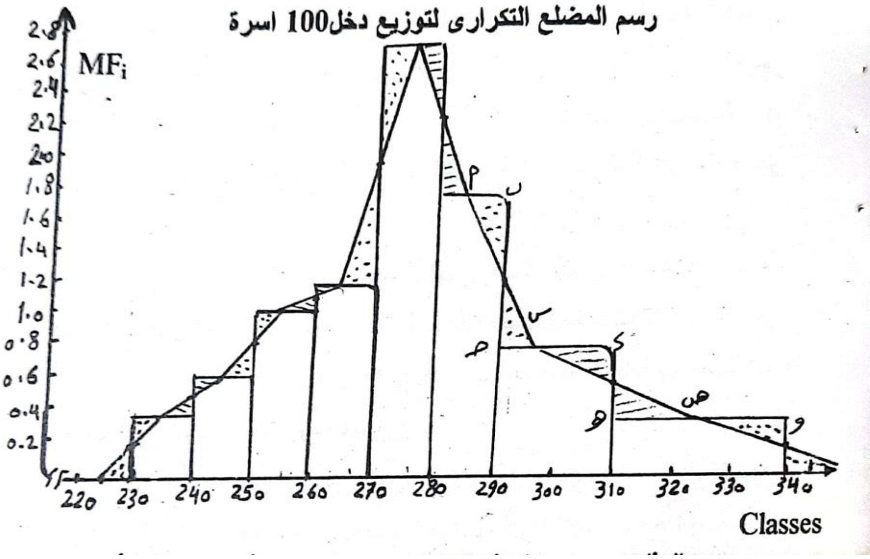


هذا ويمكن الحصول على المضلع التكرارى من خلال رسم المدرج التكرارى أولا كما سبق وان ذكرنا ، ثم من خلال تنصـــيف قمـــم المســـتطيلات(القواعد العليـــا للمستطيلات) وتوصيل هذه المنتصفات بخطوط مستقيمة مع إقفـــال الشـــكل مـــن البداية والنهاية وذلك لأن منتصفات قمم المســـتطيلات مـــاهى إلا مراكـــز فئـــات التوزيع التكرارى. ويمكن توضيح ذلك من خلال الرسم الموضح فى شكل(2-10) حيث تم رسم المدرج التكرارى أولا ثم تنصـــيف قمـــم مســـتطيلاته وتوصـــيلها

بالخطوط المستقيمة والتي تعطي شكل المضلع التكراري للأجر الشهرى لمجموعة العاملين .

شكل(2−10)

رسم المدرج والمضلع التكرارى للاجر الشهرى للعاملين



ومن الشكل(2−10) يمكن ملاحظة مايلى:−

−إن مركزى الفئتين الأفتراضيتين يقعان على المحور الافقى وذلك لأن تكرار كلا من هاتين الفئتين مساويا للصفر.

−المساحة تحت المدرج التكرارى مساوية للمساحة تحت المضلع 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(2−10) تساوى تماما مساحة المثلث المشار اليه بالرمز(ب) فى نفس الشكل.

- إن الملاحظة السابقة لاتتحقق اى لايتحقق تساوى المساحة تحت المضلع مــع المساحة تحت المدرج إلا اذا تم اقفال شكل المضلع التكرارى وذلــك مــن خــلال اضافة الفئتين قبل الاولى وبعد الاخيرة ولعل ذلك يبرر اضافة هاتين الفئتين .

وأخيرا ليس بالضرورة رسم المدرج التكرارى قبل المضلع فــى حالــة الجــداول المنتظمة حيث يمكن رسم المضلع مباشرة دون رسم المــدرج كمــا ســبق وان اوضحناه فى الشكل(2−9).

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

<u>مثال(15):</u>

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

جدول(2−26)

Income 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.

<u>الحل:−</u>

لاحظ أن التوزيع التكرارى لدخل هؤلاء الاسر غير منتظم. لذا فإنه لرسم المضلع التكرارى يجب الحصول على التكرارات المعدلة اولا ثم رسم المــدرج التكــرارى ثانيا ويلى ذلك رسم المضلع التكرارى وذلك على النحو المبين بالجدول (27−2) التالى :

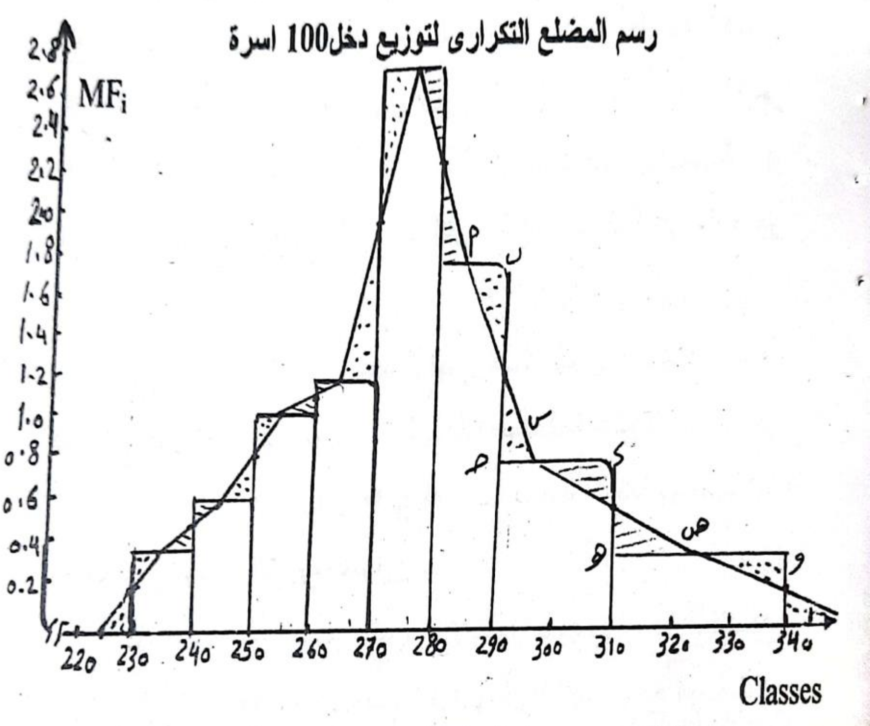
جدول (27-2)

Income Classes	F_i	L_i	$MF_i = F_i / L_i$
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(2-11). ولرسم المضلع التكرارى بإستخدام 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 10جنيه) وأما فيما يتعلق بالفئة السابعة(290 وحتى اقل من 310 جنيه) 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 يقطع قمة المستطيل الممثل للفئة (290حتى أقل من310 جنيه) وليكن فى النقطة(س) وذلك حتى يتحقق أن يكون المثلث المستبعد يساوى فى مساحته المثلث المضاف.

شكل(2−11)

رسم المضلع التكرارى لتوزيع دخل100اسرة



وبالمثل يتم رسم الخط المستقيم بدءا من النقطة(س)الممثلة أخيرا بمنتصف الارتفاع او الضلع (دهـ) ويمتد على استقامته الى ان يقطع قمة المستطيل الممثل للفئة(310 حتى أقل 340)فى نقطة ولتكن(ص) وذلك حتى يتحقق أن يكون المثلث المستبعد يساوى فى مساحته المثلث المضاف . واخيرا يتم رسم خط مستقيم بدءا من النقطة (ص) ومارا بمنتصف الارتفاع (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.

(3):<u>المنحنى التكرارى</u> : Frequency curve

إذا تم تمهيد نقاط المضلع التكرارى باليد بدلا من المسطرة فإننا نحصل على شكل منحنى يسمى بالمنحنى التكرارى . ولذلك فإنه عند رسم المنحنى التكرارى يراعى

نفس الشروط المتبعة سابقا عند رسم المدرج أو المضلع التكـــرارى مـــن حيـــث البحث في طبيعة التوزيع تكرارى من حيث كونة توزيعا تكراريا منظمـــا أم غيـــر منتظم حيث يلزم تعديل التكرارات وتؤخذ مراكز الفئـــات علـــى المحـــور الافقـــى والتكرارات الاصلية فى حالة التوزيعات التكرارية المنتظمة او التكرارات المعدلـــة فى حالة التوزيعات التكرارية الغير منتظمة على المحور الرأسى .

وبصفة عامة إذا كان تمهيد المنحنى جيدا فإن المساحة تحت المنحنى التكـــرارى سوف تساوى المساحة تحت المضلع التكرارى وكذا تساوى المساحة تحت المدرج التكرارى والتى تعبر هذه المساحات كما بينا فيما قبل عـــن مجمـــوع تكـــرارات التوزيع التكرارى .

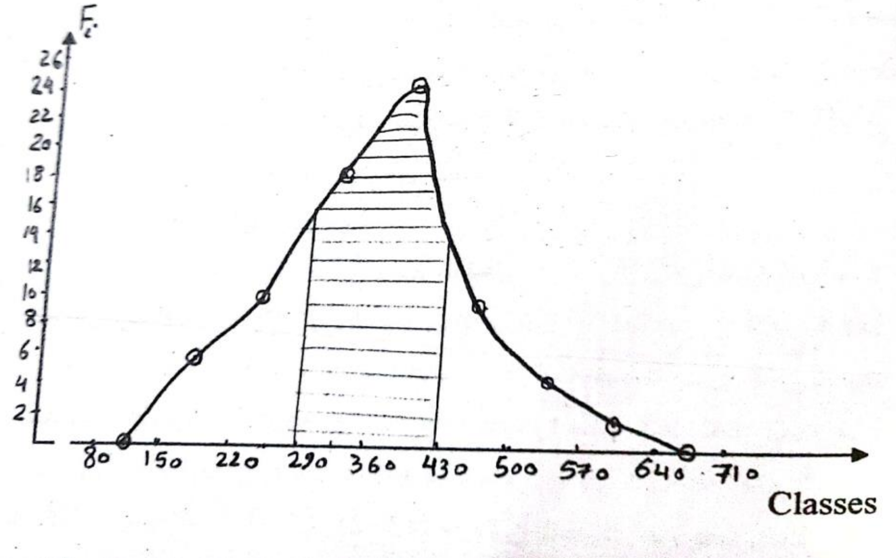
هذا ونتيجة لأن المنحنى التكرارى لايمر بجميع نقاط المضـــلع التكـــرارى فـــإن المساحة تحت المنحنى التكرارى قد لاتتســـاوى مـــع المســـاحة تحـــت المضـــلع التكرارى. وفى واقع الامر فإنه كلما كانت أطوال فئات التوزيع التكرارى قصـــيره وفى نفس الوقت زاد حجم البيانات فأن المضلع التكرارى يتحول تدريجيا الى شكل انسيابى يقترب اكثر 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 مســـاحة المضـــلع التكرارى من مساحة المنحنى التكرارى.

مثال(16):

ارسم المنحنى التكرارى لتوزيع أجور العاملين بإحدى الشركات الـــوارد بياناتهـــا بالجدول(25-2) . وهى بيانات لتوزيع تكرارى منتظم . لذا يتم رســـم المنحنـــى التكرارى مباشرة من خلال تحديد نقاط تلاقى مراكز فئات التوزيع ومايقابلها مـــن تكرارات أصلية حيث يتم تحديد الفئات او مراكز الفئات علـــى المحـــور الافقـــى والتكرارات على المحور الرأسى . ويتم تحديد نقاط المنحنى التكرارى كمـــا هـــو موضح على الرسم المبين فى شكل(2-12).

شكل(2−12)

المنحنى التكرارى لتوزيع الاجر الشهرى لمجموعة من العاملين بإحدى الشركات



وفى المنحنى التكرارى 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 290−430 مع المساحة المظللة فى الشكل (2−12). هذا ويمكن رسم المضلع او المنحنى التكرارى لتوزيع تكرارى بإستخدام التوزيعات التكرارية النسبية بدلا من التكرارات الاصلية. وتستخدم المنحنيات التكرارية فى مقارنة توزيعين تكراريين أو اكثر. إلا أنه يجب التنويه إلى أنه فى حالة تساوى المجموع الكلى للتكرارات فإن عملية المقارنة تصح مباشرة من خلال رسم المنحنى التكرارى لتلك التوزيعات بإستخدامك 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 الاهمية وذلك لأنها تستخدم فى الحكم على مدى التواء التوزيع من عدمه (أى مدى قرب أو بعد التوزيع عن التماثل) كما تستخدم

فى التعرف على شكل قمة المنحنى (دراسة التفلطح) وهو ماسـيرد فـى أبـواب لاحقة بإذن الله.

أشكال المنحنيات التكرارية:

تناولنا فيما سبق كيفية رسم المنحنى التكرارى ورأينا ان شكل المنحنى يتوقـف على التوزيع التكرارى الذى يمثله ، كما يبرز أهمية دور المنحنيات التكرارية فى استخدامها فى المقارنة فيما بين التوزيعات المختلفة . هذا ومن المتوقع ان نجـد عددا كبيرا من الاشكال المختلفة للمنحنيات التكرارية ومـع ذلـك فإننـا سـوف نتعرض هنا الى أكثر هذه الاشكال استخداما والموضحة فى شكل(2–13) مرجئين الحديث عن بعض الاشكال الاخرى حتى سـيتم دراسـتنا لمقـاييس الالتـواء او التفلطح فيما بعد فى مقررنا الدراسى بمشيئة الله.

هذا ويمكن تناول المقارنة فيما بين أشكال المنحنيات للتوزيعات التكراريـة مـن خلال الجدول التالي (جدول (28 –2)).

جدول(28–2)

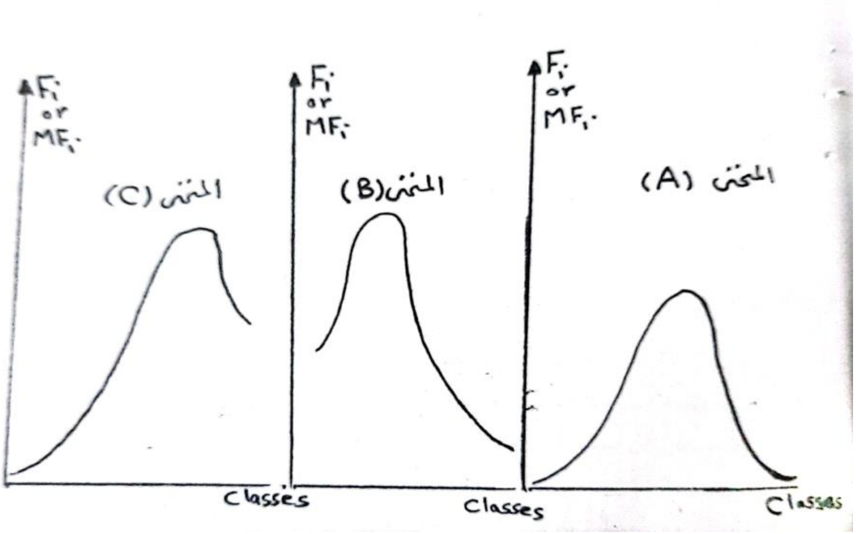
توزيعات تكرارية مختلفة

Classes	10 -	22 -	34 -	46 -	58 -	70 -	82 -	∑
Fa	4	8	13	15	13	8	4	65
Fb	3	13	25	11	7	4	2	65
Fc	2	4	7	11	25	13	3	65

لاحظ أن هذه التوزيعات الثلاثة متساوية من حيث مجموع التكرارات ولكن بتوزيع للتكرارات مختلف بالشكل الذى يخدم الغرض الذى من أجله نود التفرقة فيما بين أنواع هذه التوزيعات الثلاث . حيث تمكننا التوزيعات الثلاث من تصنيف الاشـكال المختلفة للمنحنيات التكرارية الموضحة فى الشكل(2–13) .

شكل(2−13)

بعض أشكال المنحنيات التكرارية



هذا ولكى تسهل عملية التمييز فيما بين أشكال المنحنيات الثلاثة (a,b,c) المعطاه فى الجدول (2−27) والتى تناظر علـى الترتيـب المحنيـات التكراريـة a,b,c الموضحة فى الشكل(2−13). حيث نجد أن:−

(1): التوزيع أو المنحنى(أ) أو (A) :-

وهو توزيع متماثل حول العمود الرأسى الذى يمر بنقطة النهاية العظمى للمنحنى. اى أن العمود الرأسى الذى يسقط من نقطة النهايـة العظمـى للمنحنـى يقسم المساحة تحت المنحنى الى قسمين متطابقين تماما ويسمى المنحنى التكرارى فى هذه الحالة بالمنحنى المتماثل . حيث يكون تزايد او تناقص التكـرارات متشـابها ومنتظما بطريقة متماثلة على جانبى المحورالرأسى والذى يمر بمركز التوزيـع . ويمكن ملاحظة ذلك فى الشكل (2−13) ومن الجدول(2−27) . وبعبـارة اخـرى فأن المنحنى التكرارى (أ) والموضح فى شكل (2−13) والـذى يمثلـه التوزيـع التكرارى (أ) أوِ (A) فى الجدول (2−27) يمثل هذا التوزيع صورة للتوزيعـات المتماثلة Symmetric Distribution .

(2): <u>التوزيع أو المنحنى (ب) او (B)</u>) :-

وفى هذا التوزيع تتزايد التكرارات أو تتناقص بشكل غير منـــتظم علـــى جـــانبى المحور أو العمود المنشأ عند وسط التوزيع . ولذلك فهو توزيـــع غيـــر متماثـــل تتزايد فيه التكرارات بسرعة حتى تصل الى القمة وبعدها تتناقص ببطء . ويطلق على هذه التوزيعات والتى يميل فيها التكرارات الكبيرة الى التركيز عنـــد الفئـــات الدنيا (التى تقع فى بداية التوزيع) بالتوزيعات الموجبـــة الالتـــواء Positive Skewness Distributions. والمنحنى التكرارى الـــذى يمثـــل مثـــل هـــذه التوزيعات يسمى بالمنحنى غير المتماثل حيث يكون طرفه(ذيله) الايمن أطول من طرفه الايسر (انظر الشكل(2−13) المنحنى(ب)) والتوزيع التكرارى(ب) الموضح فى الجدول(27−2) يمثل هذا النوع من المنحنيات.

<u>(3): التوزيع أو المنحنى (ج) أو (C):-</u>

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 القمـــة ثم بعدها تأخذ فى التناقص بسرعة. ومثل هذا التوزيع والذى تميل فيه التكرارات الكبيرة الى التركيز عند الفئات العليا(الأخيرة) يسمى بـــالتوزع ســـالب الالتـــواء (Negative Skewness Distribution). والمنحنى التكرارى الممثل لهذا التوزيع هو منحنى غير متماثل حيث يكون الطرف(الذيل) الايسر فيه أطول مـــن الطرف الايمن (أنظر الشكل (2−13) والمنحنى(جـــــ) يمثـــل هـــذا النـــوع مـــن المنحنيات التوزيع التكرارى(جــ) الوارد فى جدول(27−2) .

<u>مثال(17):</u>

الجدول التالى يوضح التوزيع التكرارى لدرجات مجموعة مـــن الطـــلاب (ذكـــورا وإناثا على حده) فى احدى المواد.

جدول(2-29)

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

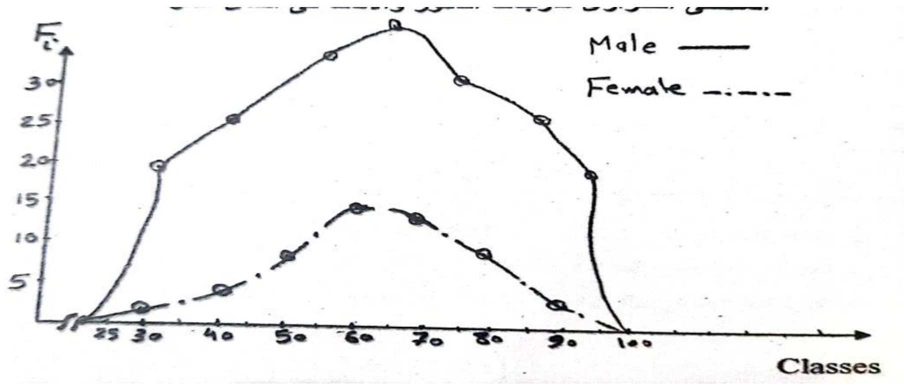
Classes	-25	-35	-45	-55	-65	-75	-85	$\sum$
No.of Males	20	26	35	43	30	27	19	200
No.of Femals	2	4	7	13	12	9	3	50

والمطلوب:- المقارنة مابين التوزيعين التكرارين 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)والآن دعنا نقوم بالمقارنة من خلال المنحنى التكرارى . فإذا تمت هذه المقارنة البيانية بتلك الصورة فإننا نحصل على الشكل البيانى الوارد بالشكل (2-14). والشكل(2-14) يوضح رسم المنحنى التكرارى للذكور والاناث مباشرة.

شكل(2-14)

المنحنى التكرارى لدرجات الذكور والاناث فى احدى المواد



فيتضح من شكل (2–14) ان المقارنة بين التوزيعين غير واضحة إذ أن المساحة تحت كلا من المنحنيين تختلف عن الاخرى نظرا لاختلاف مجموع التكرارات فــى كلا من التوزيعين عن الأخر. لاحظ عزيزى القارئ أن مجموع التكرارات فى حالة الذكور(200) مختلفا عن مجموع التكرارات فى حالة الاناث(50) ، فإذا تم إجراء المقارنة فيما بين التوزيعين مباشرة دون الحصول علــى التكــرارات النســبية، سيكون النتيجة مقارنة غير واضحة وغير موضوعية نظــرا لاخــتلاف مجمــوع التكرارات . وعليه فلكى تكون المقارنة موضوعية وعلى اساس إحصائى ســليم فأن المساحة تحت المنحنى الممثل لدرجات الذكور يجب أن تتساوى مع المساحة تحت المنحنى الممثل لدرجات الاناث ولكى يتحقق ذلك لابد من ايجاد التكراريــة النسبية لكل من الذكور والاناث حيث تتساوى المساحتين فى هذه الحالة.

والجدول (2–29) يوضح التوزيع التكرارى النسبى لدرجات الطلاب (ذكورا وإناثا) فى إحدى المواد.

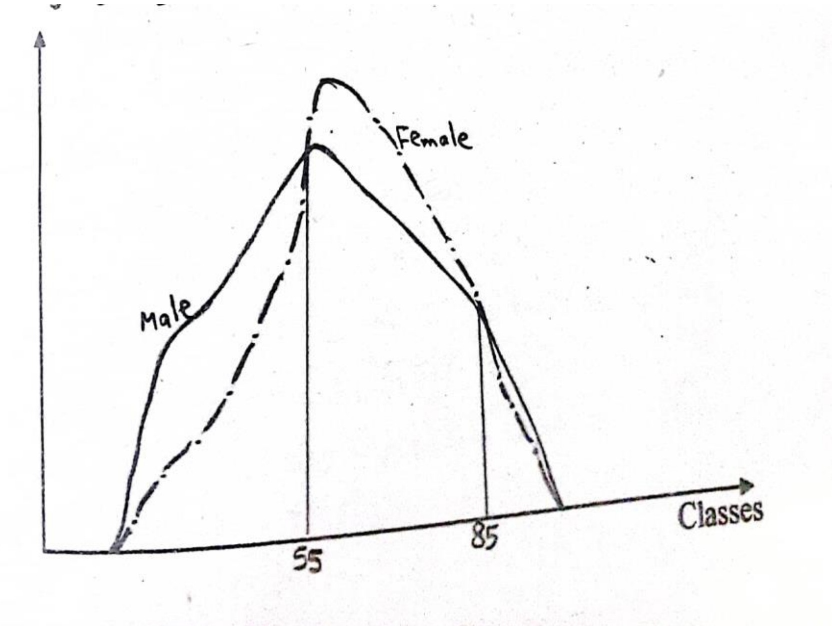
جدول (30– 2)

Score Classes	Frequences		Relative Frequences %	
	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(2–30) يمكننا رسم المنحنى الذى يمثل التوزيــع التكــرارى النسبى لكلا من درجات الذكور والأناث كلا على حده وذلك كما هو موضح بالشكل (2–15) حيث يبين هذا الشكل منحنى التكرارات النسبية لكلا من الذكور والانـــاث للبيانات الوارد بالجدول(2–30).

شكل(2−15)

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



ومن خلال شكل(2−15) يتضح لنا الاتى:−

− ترتفع نسبة الذكور عند الحدود الدنيا للتوزيع (وبالتحديد عند الدرجات التى تقل عن 55 درجة) عن مثيلتها للإناث حيث يقع منحنى التكرار النسبى للذكور أعلى من منحنى التكرار النسبى للإناث.

− ترتفع نسبة الاناث عند الحدود العليا لفئات التوزيع (وبالتحديد عند الدرجات التى تزيد عن 55 درجة وتقل عن 85 درجة)حيث يقع منحنى التكرار النسبى للإناث اعلى من منحنى التكرار النسبى للذكور . ثم تزيد نسبة الذكور بنسب طفيفة عنها فى الإناث عندما تزيد الدرجات عن 85 درجة .

وبصفة عامة عند المقارنات بإستخدام المدرج أو المضلع أو المنحنى وإن كان يفضل إستخدام المضلع او المنحنى التكرارى لكى تتضح عملية المقارنة سواء بإستخدام التكرارات الاصلية (فى حالة تساوى مجموع التكرارات) أوبإستخدام التكرارات المعدلة(فى حالة عدم تساوى أطوال الفئات أى الجداول الغير منتظمة بأفتراض تساوي مجموع التكرارات). حيث يتم تحديد نقاط تقاطع المضلعات أو

المنحنيات التكرارية ويتم إسقاط من تلك النقاط اعمدة على المحور الافقى ثم تـــتم عملية المقارنة فيما بعد ذلك كما سبق فى الشكل(2–15). أما إذا أختلفت مجاميع التكرارات فلابد من الحصول على التكرارات النسبية في حالة الجداول التكراريـــة المنتظمة أو التكرارات النسبية المعدلة فى حالة الجداول أو التوزيعات التكراريـــة الغير منتظمة .

<u>المنحنيات التكرارية المتجمعه:-</u> Cumulative Frequency Curves

بصفة عامة تستخدم المنحنيات التكرارية المتجمعة لتمثيل الجـــداول التكراريـــة المتجمعة(الصاعدة أو الهابطة) بيانيا وذلك لاستخدامها فى تحديد عدد التكـــرارات التى تقابل قيمة اقل من (أو أكبر من) قيمة ما فى فئات التوزيع التكرارى أو فيما بين فئتين من فئات التوزيع او العكس صحيح.

ولما كان هناك نوعان من الجداول التكرارية المتجمعة وهما الجداول التكراريـــة المتجمعة الصاعدة والجداول التكرارية المتجمعة الهابطة لذا فإنه من الطبيعى ان يكون هناك نوعين للمنحنيات التكرارية المتجمعة وهى:–

***المنحنى التكرارى المتجمع الصاعد** Ascending Cumulative Frequency Curve.

***المنحنى التكرارى المتجمع الهابط** Descending Cumulative Frequency Curve.

حيث يستخدم المنحنى التكرارى المتجمع الصاعد لرسم الجدول التكرارى المتجمع الصاعد بيانيا كما يفيد فى حساب عدد التكرارات (أو نسبة التكرارت) التى تأخـــذ قيمة اقل من قيمة ما(سواء فئة معينة او فيما بين فئتين) من قيم فئات التوزيـــع التكراري. أو العكس تفيد فى حساب ماهى قيمة الفئة (أو القية التى تقع مـــابين فئتين من الفئات المتجمعة التى يقل عنها عدد معين مـــن التكـــرارات (او نســـبة معينة من التكرارات).

ولرسم المنحنلى التكرارى المتجمع الصاعد فأننا نختار المحور الرأســـى لتمثيـــل التكرارت المتجمعة الصاعدة بينما يتم تمثيل الفئات على المحور الافقى . حيث يتم تمثيل كل فئة ومايقابلها من تكرار متجمع صاعد بنقطة إحداثياها هما الحد الاعلى

للفئة والتكرار المتجمع الصاعد لهذه الفئة . ثم نقوم بتوصيل هذه النقـــاط بخـــط منحنى ممهد باليد فنحصل بذلك على المنحنى التكرارى المتجمع الصـــاعد. امـــا المنحنى التكرارى المتجمع الهابط فيستخدم لرسم الجـــدول التكـــرارى المتجمـــع الهابط بيانيا كما يفيد فى حساب عدد التكرارات(اونسبة التكرارات) التى تبلغ قيمة مافى الفئات (أو بين فئتين) فأكثر او يفيد فى تحديد ماهى القيمة التى يبلغها عددا ما من التكرارات (أونسبة معينة من التكرارات) فأكثر .وبأسلوب مماثل يمكن رسم المنحنى التكرارى المتجمع الهابط حيث يتم تمثيل التكرارات المتجمعـــة الهابطـــة على المحور الرأسى بينما يمثل المحور الافقى فئات التوزيع التكرارى المتجمـــع الهابط. ويتم تمثيل كل فئة ومايقابلها 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 بـــذلك علـــى المنحنى التكرارى المتجمع الهابط والمثال التـــالى يوضـــح أهميـــة المنحنيـــات التكرارية المتجمعة.

مثال(18-2) :- فيما يلى لديك التوزيع التكرارى التالى يبـــين توزيـــع الاجـــر الاسبوعى لخمسين عاملا بإحدى الشركات :

جدول(31-2)

Classes	-30	-40	-50	-60	-70	-80	-90	$\sum$
Frequencies	1	4	7	14	11	8	5	50

والمطلوب :-

1- رسم المنحنى التكرارى المتجمع الصاعد .

2- رسم المنحنى التكرارى المتجمع الهابط .

3- بإستخدام المنحنى التكرارى المتجمع الملائم حدد مايلى:-

أ- ماهو قيمة الاجر الاسبوعى الذى يتقاضى اقل منه 35 عاملا.

ب- اوجد قيمة الاجر الاسبوعى الذى يحصل عليه – كحـــد أدنــــى – ثلاثـــون عاملا .

جـــ– ماهو عدد العمال(وكذا ماهى نسبة العمال) الذين يتقاضون أجورا اسبوعية تقل عن64 جنيها.

د– ماهو قيمة الاجر الاسبوعى الذى يتقاضى أقل منه :

– 80 % من إجمالى عدد العمال .

– 52 % من إجمالى عدد العمال.

هـــ– حدد ماهو عدد العمال(وكذا ماهى نسبة العمال) الذين يتقاضون أجورا أسبوعية تبلغ 45 جنيها على الاقل .

و– حدد قيمة الاجر الاسبوعى الذى يحصل عليه كحد أدنى 50 % من إجمالى عدد العمال .

<u>الحل :</u>–

(1):لرسم منحنى التوزيع التكرارى المتجمع الصاعد يتم أولا ايجاد جدول التوزيع التكرارى المتجمع الصاعد وذلك على النحو المبين فى الجدول التالى :–

جدول (32–2)

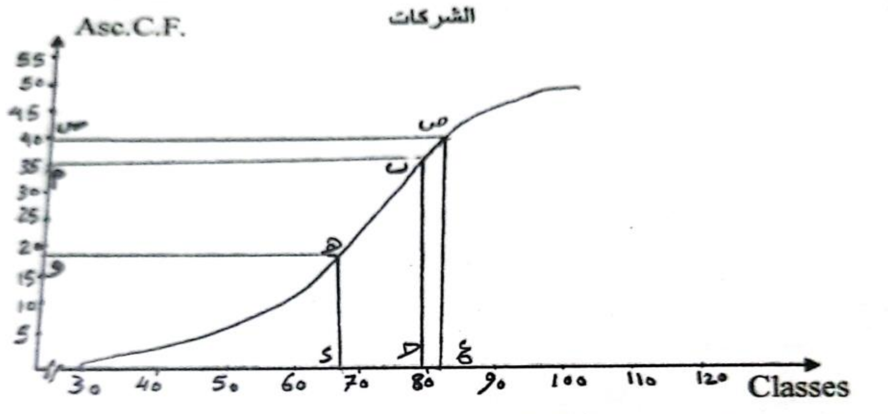
التوزيع التكرارى والتوزيع التكرارى المتجمع الصاعد لأجور العمال الاســبوعية فى إحدى الشركات:

Classes	F_i	Less than the Class Upper Limit	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 (2−16) يوضح المنحنى التكرارى المتجمع الصاعد للاجور الاسبوعية لمجموعة العاملين المعطاه .

شكل(2−16)

المنحنى التكرارى المتجمع الصاعد لأجور العمال الاسبوعية فى إحدى الشركات



(2): ولرسم المنحنى التكرارى المتجمع الهابط يجب أولا ايجاد جدول التوزيع التكرارى المتجمع الهابط وذلك على النحو المبين فى الجدول التالى:

جدول(2−33)

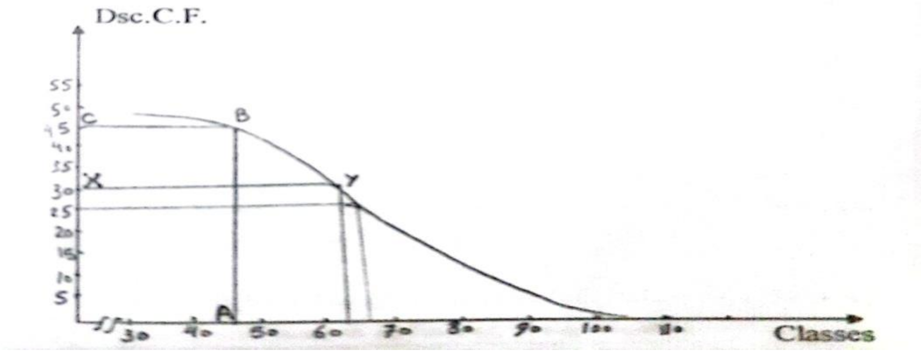
التوزيع التكرارى والتوزيع التكرارى المتجمع الهابط للأجور الاسبوعية لمجموعة من العاملين بإحدى الشركات

Classes	F_i	More than the Class Lower Limit	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
$\sum$	50	More than100	0

ومن هذا الجدول يتم رسم العمودين الأخيرين لتمثيل المنحنى التكرارى المتجمع الهابط. والشكل(2−17) يوضح المنحنى التكرارى المتجمع الهابط للأجور الاسبوعية لمجموعة العاملين المعطاه .

شكل(2−17)

المنحنى التكرارى المتجمع الهابط لاجور العمال الاسبوعية فى إحدى الشركات



(3): بعد رسم المنحنى التكرارى المتجمع الصاعد(شكل (2−16)) والمنحنى التكرارى المتجمع الهابط (شكل (2−17)) ومن خلال المرادفات التى سبق الاشارة إليها عند دراستنا للجداول التكرارية الصاعدة والهابطة فإنه:

أ− لتحديد قيمة الاجر الاسبوعى الذى يتقاضى أقل منه 35 عاملا فيكون ذلك من خلال المنحنى التكرارى المتجمع الصاعد. حيث يتم تعيين موضع تلك النقطة على المحور الرأسى فى الشكل(2−16) ولتكن النقطة(أ) والتى تمثل عدد العمال فى صورة تكرارات متجمعة صاعدة . ومن هذه النقطة يتم رسم خطا عموديا على المحور الرأسى وموازيا للمحور الافقى الى أن يقطع هذا الخط المنحنى التكرارى المتجمع الصاعد فى نقطة ولتكن النقطة (ب) فنسقط عمودا من نقطة إلتقاء هذا الخط الموازى للمحور الافقى بالمنحنى المتجمع الصاعد ليقطع هذا العمود المحور الافقى(فئات الاجور) فى نقطة ولتكن النقطة (جـ) فتكون تلك النقطة هى التى توضح قيمة الاجر الذى يتقاضى أقل منه 35 عاملا. ومن النقطة(جـ) على الرسم

يتضح لنا أن قيمة الاجر الذى يتقاضى أقل منه 35 عاملا هو الاجـــر 87 جنيهـــا تقريبا.

ب- لإيجاد قيمة الأجر الذى يتقاضاه <u>كحد أدنى ثلاثون عاملا</u> فيكون ذلـــك مـــن خلال المنحنى التكرارى المتجمع الهابط لانه كما سبق وأن أوضحنا أن لفظ كحـــد أدنى يكافئ فى المعنى كلمة فأكثر . وهنا يتم تعيين موضع النقطة 30 عاملا على المحور الرأسى للمنحنى التكرارى المتجمع الهابط كما هو موضح فى شـــكل(2-17) ولتكن النقطة (X)والتى تمثل عدد العمال فى صورة تكرارات متجمعة هابطة. ومن هذه النقطة يتم رسم خطا عموديا على المحور الرأسى وموازيـــا للمحـــور الافقى الى أن يقطع هذا الخط المنحنى التكرارى المتجمع الهابط فى نقطة ولتكن النقطة(Y) فنسقط عمودا من النقطة (Y) على المحور الافقى والذى يمثل الحدود الدنيا لفئات التوزيع التكرارى المتجمع الهابط الى ان يقطع هذا العمـــود المحـــور الافقى فى نقطة ولتكن (Z). فتكون النقطة (Z) هى عبارة عن قيمة الاجر الـــذى يتقاضاه كحد أدنى ثلاثون عاملا . ومن النقطة (Z) على الرسم الموضح فى شكل (2-17) يتضح لنا أن قيمة الاجر الذى يتقاضاه كحد أدنى ثلاثون عاملا هو الاجر 66 جنيه تقريبا .

ج- أما لتحديد عدد العمال (وكذا نسبة العمال) الذين يتقاضون أجرا أسبوعيا <u>يقل عن 64 جنيه</u> . فيتم استنتاج المطلوب بإستخدام المنحنـــى التكـــرارى المتجمـــع الصاعد ، حيث يتم تحديد موضع الاجر 64 جنيه بنقطة على المحور الافقى لفئات التوزيع التكرارى المتجمع الصاعد الموضحة فى الشكل (2-16) ولتكن النقطـــة (د) ومن النقطة (د) يتم إنشاء عمودا على المحور الافقى وموازيا للمحور الرأسى الى أن يقطع هذا العمود المنحنى التكرارى المتجمع الصاعد فـــى نقطـــة ولـــتكن النقطة (هـــ) . ومن النقطة (هـــ) يتم رسم خطا عموديا موازيا للمحور الافقـــى ويمتد على استقامته الى أن يقطع المحور الرأسى والذى يعبر عـــن التكـــرارات المتجمعة الصاعدة فى نقطة ولتكن النقطة (و) وهى تعتبر بمثابـــة عـــدد العمـــال المطلوب . هذا ومن الشكل (2-16) وبالتحديد عند النقطة (و) يتضح لنا أن عدد

العمال الذين يتقاضون أجرا اسبوعيا يقل عن 64 جنيه هو 18 عاملا . ومن ثـــم فإن :-

نسبة العمال الذين تقل أجورهم الاسبوعية عن 64 جنيه هى عبارة عن :

$$= (18 / 50) \times 100 = 36 \%$$

د- ولتحديد قيمة الاجر الذى يتقاضى <u>أقل منه 80%</u> من إجمالى عـــدد العمـــال فحيث أن النسبة 80% من إجمالى العمال هى:

50 × (80 / 100) = 40 عاملا. لذا فإنه يمكن استنتاج قيمة الاجر

الاسبوعى الذى يتقاضى <u>أقل منه 40 عاملا</u> باستخدام المنحنى التكرارى المتجمع الصاعد بنفس آلية المطلوب الاول رقم (أ) من هذا المثال وذلك من خلال الشـــكل رقم (2-16). حيث يتم تحديد موضع عدد العمال 40 عاملا على المحور الرأسى وليكن عند النقطة (س) ورسم خط عمود على المحور الرأسى وموازيا للمحـــور الأفقى الى أن يقطع المنحنى التكرارى المتجمع الصاعد فى نقطة ولتكن النقطـــة (ص) والتى يتم منها إسقاط عمودا على المحور الافقى وموازيا للمحور الرأســـى الى أن يقطع المحور الافقى فى نقطة ولتكن النقطة (ع) . ومنها فإن قيمة الاجـــر الذى يتقاضاها اسبوعيا أقل من 40عاملا (80% من إجمالى العمال) هو الاجـــر 84 جنيه تقريبا.

- أما لتحديد قيمة الأجر الأسبوعى الذى يتقاضى <u>أقل منه 52 %</u> من اجمالى عدد العمال . فحيث ان النسبة 52 % من إجمالى عدد العمال هى عبارة عن :

= 50 × (52 / 100) = 26 عاملا .

لذا فإنه يمكن استنتاج قيمة الاجر الذى يتقاضى <u>أقل منه 26 عاملا</u> بنفس الاسلوب السابق بإستخدام المنحنى التكرارى المتجمع الصاعد كما يوضحه الشكل(2-16)

حيث ينتج ان الاجر الذى يتقاضى أقل منه 26 عاملا (اى 52% من إجمالى عـــدد العمال) هو 70 جنيه تقريبا .

ملحوظة:-

لاحظ فى الجدول التكرارى المتجمع الصـــاعد ان عـــدد العمـــال26 يشـــكل أحـــد التكرارات المتجمعة الصاعدة يقابله الفئة 70جنيها كحـــد أقصـــى للأجـــر الـــذى يتقاضاه عدد26 عاملا.

ه- لتحديد عدد العمال (وكذا النسبة) الذين يتقاضون أجورا اسبوعية تبلـــغ 45 جنيها على الاقل (وهذه العبارة تعادل كلمة فأكثر) فأن ذلـــك يكـــون مـــن خـــلال المنحنى التكرارى المتجمع الهابط . حيث يتم تحديد هذا المطلوب بنفس الاسلوب الذى تم به إجابة المطلوب (ب) من نفس المثال . حيث يتم تحديد موضع الاجـــر 45 جنيها بنقطة على المحور الافقى للمنحنى التكرارى الهابط الموضح فى الشكل (2-17) ولتكن النقطة (A) ثم يتم إنشاء عمودا من النقطة (A) علـــى المحـــور الافقى وموازيا للمحور الرأسى الى أن يقطع هـــذا العمـــود المنحنـــى التكـــرارى المتجمع الهابط فى نقطة ولتكن (B) ومن النقطة (B) يتم رســـم خطـــا موازيـــا للمحور الافقى فى اتجاه المحور الرأسى الى ان يقطع المحور الرأسى فى نقطـــة ولتكن النقطة(C) . فتكون النقطة (C) هى بمثابة عدد العمال الـــذين يتقاضـــون على الاقل أجرا اسبوعيا يعادل 45 جنيها . ومن الرسم الموضح فى الشـــكل (2-17) يتضح لنا أن عدد العمال المطلوب هو 47 عاملا تقريبا ومن ثم فإن :-

النسبة المطلوبة = (47 / 50) × 100 = 94 % .

و- ولتحديد قيمة الاجر الاسبوعى الذى يتقاضاه كحد أدنى 50% من اجمالى عدد العمال . يتم استنتاج المطلوب من المنحنى التكرارى المتجمع الهابط . وذلك بعـــد تحويل النسبة 50% الى عدد تكرارات مطلقة فحيث أن:

50% من اجمالى عدد العمال = 50 × (50 / 100) = 25 عاملا .

ومن ثم فأن المطلوب الان هو تحديد قيمة الاجر الذى يتقاضاه كحد أدنى

(أى القيمة فأكثر) 25عاملا . وبنفس الاسلوب الذى تم به اجابة المطلـــوب (ب) من نفس المثال فإنه من خلال الشكل(2−17) يتضح لنا أن قيمـــة الاجـــر الـــذى يتقاضاه كحد أدنى 50 % من عدد العمال (أى 25 عاملا) هو الاجر 69 جنيـــه تقريبا .

4- مخطط الجذع والورقة :- Stem&Leaf

يعتبر مخطط الجذع والورقة من المخططات البيانية المستحدثة فى مجال العـــرض البيانى للظواهر الكمية المتصلة . وفى هذا الشكل أو هذا المخطط البيـــانى يـــتم تقسيم كل مفردة من مفردات الظاهرة أو المتغير محل الدراسة الى جزئين أساسين الاول يسمى بالجذع (Stem) والثانى يسمى بالورقه
(Leaf) حيث يمثل الجذع الجزء الايسر من البيانات أما الورقة فتمثل بالجزء الايمن . وهذا المخطط– الجذع والورقه – يشبه بدرجة كبيرة المدرج والمنحنـــى والمضلع التكرارى . والفارق الوحيد بين هذا المخطط البيانى والرسوم الســـابقة (مدرج– مضلع– منحنى تكرارى) فى أن التكرارات فى حالة المـــدرج والمضـــلع والمنحنى التكرارى تمثل بأعمدة بيانية وينتهى دور المفردات الحقيقية فى هـــذه الحالة ، أما فى حالة مخطط الجذع أو الساق والورقة فيتم تمثيل القيم الحقيقيـــة للمفردات . ولذلك فأن مخطط الجذع والورقة يعكس معلومات عن طبيعـــة القـــيم الموجودة أو الفعلية للظاهرة محل الدراسة ذاتها وذلـــك علـــى عكـــس المضـــلع والمنحنى التكراى والذى يفترض ان تكرارات كل فئة تأخذ قيمة مركـــز الفئـــة . والامثلة التالية توضح كيفية إنشاء هذا المخطط البيانى .

مثال(19):−

إذا كانت لديك مجموعة المفردات التالية والتى تعبر عـــن الـــدخل فـــى الســـاعة لمجموعة مكونة من 10 أشخاص :−

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

والمطلوب تمثيل مخطط الجذع (أو الساق) والورقة .

<u>الحل</u> :-

لتمثيل مخطط الجذع أو الساق والورقة فإننا نقسم المفردات التى تعبر عن الدخل أو الاجر فى الساعة لمجموعة المفردات 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 10جنيه فيعطى الشكل البيانى التالى (شكل 2-18).

شكل (2-18)

مخطط الجذع(الساق) والاوراق للدخل فى الساعة لمجموعة من العاملين

Stem- and- leaf plot

INCOME Stem & 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 (2-18) يتضح لنا انه تم رصد تكرارات كل فئة فى الجانب الايسر . حيث ان هناك مفردتين تبلغ دخولهم فى الساعة 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 الاعلى فيعطى صورة تفصيلية عن قيم المفردات ذاتها . حيث يتضح الاتى :-

- بالنسبة للصف الاول الذى يقع أسفل الجذع أو الساق والاوراق فإن هناك مفردتين فى خانة العشرات (الجذع) تحوى صفرا أما الورق فيحوى (5)،(7) وهما نتاج العملية الرياضية التالية:-

$5 = 0 \times 10 + 5 \quad , \quad 7 = 0 \times 10 + 7$

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 :

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

وهى المفردات الثلاثة التى تنحصرمابين 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 $20=10\times2$) هذا بالاضافة لأرقام الآحاد التى يوضحها الورق . وهذه الارقام السبعة تفصيلهم على النحو التالى :-

$20 = 2 \times 10 +0 \ , \ 21 = 2 \times 10 + 1 \ , 23 = 2 \times 10 + 3 \ ,$
$25 = 2 \times 10 + 5 \ , 27 = 2 \times 10 + 7 \ , \ 28 = 2 \times 10 + 8 \ , 29 =$
$2 \times 10 + 9 \ .$

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

- 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 وعليه يمكن تحديد مدى المتغير أو الظاهرة محل الدراسة وهو الفرق مابين أكبر قيمة وأصغر قيمة فى المفردات .

مثال(20):-

البيانات التالية تمثل أطوال مجموعة من نباتات القطن مقدرة بالسنتيمترات (56 مفردة).

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

والمطلوب :-

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 32 سم وأكبر قيمة فى المفردات هى99 سم . فإذا أعتبرنا طول الفئة 10سم فيكون المخطط الخاص بالجذع (أو الساق) والورقة يبدا أول رقم فى الجذع أو الساق بالرقم(3) ليعنى أن خانة العشرات الحد الادنى هو ثلاثون . وحيث ان المفردات تحتوى على العشرات المتتالية أى 40 ، 50 ، 60 ، 70 ، 80 ، 90 لذا فإن عمود الساق سينتهى بخانة العشرات بالرقم(9) اما عن الورق فيوضحه المخطط الموضح فى الشكل البيانى التالى:-

شكل(2-19)
مخطط الجذع (أو الساق) والورقة لأطوال عينة من نباتات القطن(بالسنتيمتر)

TALL Stem-and- Leaf

Frequency	Stem& Leaf
4.00	3 . 235
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)

ومن خلال شكل (2−19) يتضح لنا أن :−

− أصغر قيمة فى البيانات هى(32) وأكبر قيمة فى البيانات هى(99) .

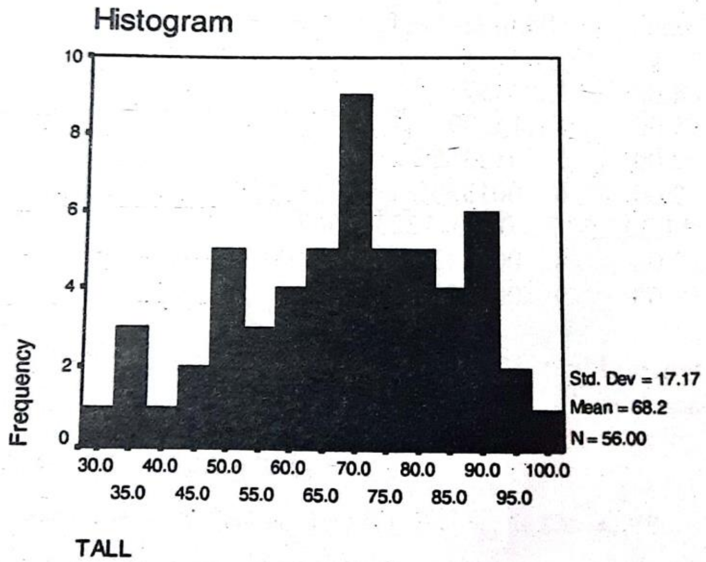
− طول الفئة فى هذا المخطط هو 10 .

− عدد المفردات والتى يمثلها مجموع التكرارات هو 56 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(2−20) يوضح رسم المدرج التكرارى لأطوال عينة نباتات القطن لهذا المثال :

شكل (2-20)

المدرج التكرارى لأطوال عينة من نباتات القطن



ثالثا :- الاشكال أو المخططات البيانية الخاصة بالسلاسل الزمنية :

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:-

1- 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 الراسى 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 من جانب وقيمة الظاهرة عند تلك النقطة من جانب آخر . وبعد الانتهاء من توقيع كافة النقاط التى تمثل السلسلة

الزمنية للظاهرة يتم توصيل كل نقطتين متتاليتين بخط مستقيم أو شبه مضــلع . فيكون الناتج هو عبارة عن الخط البيانى للظاهرة محل الدراسة . والمثال التـــالى يوضح ذلك .

مثال (21):-

فيما يلى بيان بتطور عدد العاملين(الذكور) بالوظائف العليا بالدولة وذلــك خــلال الفترة (1980-1984) :

جدول(34-2)

السنة	1980	1981	1982	1983	1984
عـــــــــدد العاملين	1407	1412	2003	2138	2187

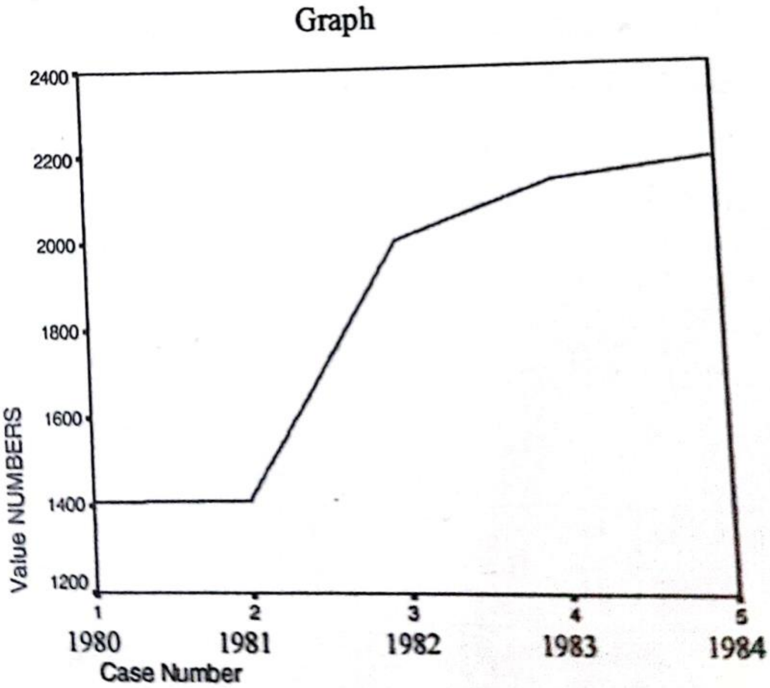
والمطلوب عرض بيانات تلك السلسلة بإستخدام خريطة الخط البيانى.

الحل :-

يتم تمثيل السنوات المختلفة للسلسلة على المحور الافقى بينما يتم تمثيــل عــدد العاملين على المحور الرأسى . وحيث أن أصغر قيمة لعدد العاملين هـــى 1407 عاملا فإنه يكون من الملائم كسر(قطع) المحور الرأسى من أسفل وذلك حتـــى يمكن أختيار مقياس رسم مناسب يوضح تطور أعداد العاملين دون ترك مســاحة كبيرة من ورقة الرسم البيانى غير مستخدمة . لذا فإننا ســوف نختــار مقيــاس الرسم للمحور الرأسى هو 1سم/ 100عاملا مع البدء بالقيمة1400 كمــا يتضــح ذلك فى الشكل (2-21) .

شكل(2−21)

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



ومن خلال هذا الشكل يمكن استنتاج أن أعداد العاملين بالوظائف العليا فى الدولة فى ذيادة مستمرة مع الزمن وذلك بمجرد النظر الى خريطة الخط البيانى الموضح فى الشكل (2−21) .

2− خريطة الشريط :− Band Chart

وتعتبر خريطة الشريط بمثابة امتدادا لشكل خريطة الخط البيانى السابق الاشارة اليها حيث أنها تشمل خطين بيانين أو أكثر حسب أقسام الظاهرة . على أن يمثل كل خط إحدى الظواهر المراد تمثيلها وبحيث يكون الفرق بين القيم المتناظرة لهذه الظواهر ذا معنى أو مدلول .

ولرسم خريطة الشريط يتم تمثيل كل ظاهرة من الظاهرتين بخط بيانى مع تلوين المساحة بين الخطين بلون أو تظليل معين . هذا وقد تأخذ الفروق فيما بين القيم

المتناظرة اتجاها واحدا (موجبا أو سالبا) أو قد تكون الفروق فيمــا بــين القــيم المتناظرة للظاهرتين موجبة أحيانا وسالبة احيانا أخرى وهنا يلزم استخدام لونين مختلفين أو طريقتين مختلفتين فى التظليل وذلك للتفرقة فيما بين الفروق الموجبة والاخرى السالبة بهدف توضيح مدلول كل منهما .

<u>مثال</u>(22):

فيما يلى لديك بيانات سلسلة زمنية للفترة(1960–1965) لأعداد الطلاب (ذكورا واناثا) فى إحدى الكليات.

جدول (35 –2)

السنة	1960	1961	1962	1963	1964	1965
عـــــدد الذكور	120	150	175	210	280	320
عـــــدد الاناث	90	100	120	150	160	175

والمطلوب عرض بيانات هذه السلسلة بإستخدام خريطة الشريط .

<u>الحل</u> :

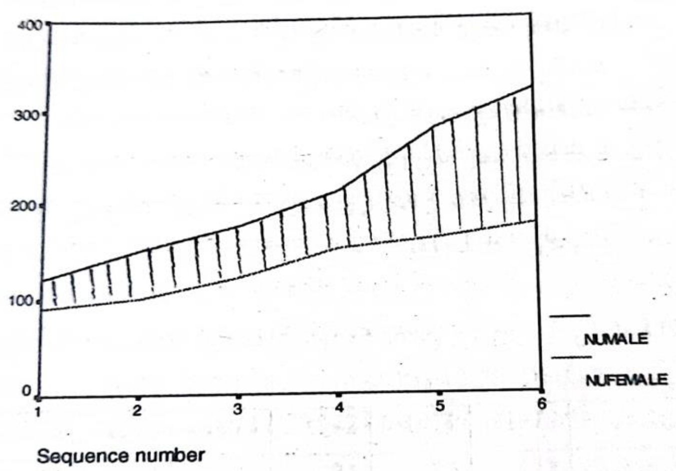
الشكل البيانى التالى (شكل(2–22)) يوضح خريطة الشــريط لبيانــات السلســلة الزمنية المعطاه.

شكل(2–22)

خريطة الشريط لأعداد الطلاب (ذكور وإناث)

فى إحدى الكليات عن الفترة 1960–1965

1965–1960



ويلاحظ من خريطة الشريط الموضحة فى الشكل (2–22) ان عدد الطلاب الذكور يكون دائما أكبر من عدد الطلاب الاناث . كما يتضح من الشكل ايضا ان الفرق بينهما فى ذيادة مستمرة من سنة لأخرى .

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

" تمارين الفصل الثانى "

(1): أخذت عينة من 40 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.

(2): أخذت تقديرات 50 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.

(3): فيما يلى تقديرات نجاح 20طالبا 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 البانات فى جدول تكرارى .

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

(4): اعتبر الارقام التالية تمثل الدرجات التى حصل عليها 54 طالبا فى احد الاختبارت (علما بأن الدرجة العظمى = 10 درجات) .

4	3	5	4	3	5	5	4	5
6	5	4	5	4	5	3	4	6
5	5	7	4	3	4	3	2	7
4	3	6	1	3	6	3	1	8
6	3	5	2	7	5	7	3	4
6	5	8	6	4	3	4	4	8

والمطلوب وضع هذه البيانات فى جدول تكرارى بسيط .

(5): 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	78	203	311	296	78	24	990
الكلية ب	274	236	229	291	41	16	1987

والمطلوب: عرض هذه البيانات بالأشكال البيانية المناسبة .

(6): الجدول التكرارى التالى يوضح التوزيع التكرارى لعدد أيام الاجازات التى قام بها 80 موظفا فى إحدى الشركات خلال عام كامل :

عدد أيام الاجازة	1	2	3	4	5	6	7
عدد الموظفين	4	7	11	18	22	10	8

والمطلوب عرض هذه البيانات بإستخدام الشكل الإحصائى المناسب .

(7): البيانات التالية تمثل الدرجات التى حصل عليها 100طالب فى إحدى المواد الدراسية:

45	52	26	50	56	67	57	55	36	45
93	93	40	54	27	57	34	31	46	78
27	35	54	80	30	38	50	44	47	20
71	25	51	61	12	42	45	57	33	54
66	68	12	52	62	65	43	96	25	56
42	51	28	52	67	55	50	40	48	74
93	49	66	63	59	52	39	51	83	15
84	40	46	50	56	53	50	61	31	45
79	56	55	91	86	52	75	75	54	67
63	77	58	62	95	86	51	55	36	56

والمطلوب :

1- وضع هذه البيانات فى جدول تكرارى .

2- رسم المدرج التكرارى والمضلع التكرارى لهذا التوزيع .

3- تكوين الجدول التكرارى المتجمع الصاعد والنازل .

(8): البيانات التالية تمثل أعمار 60 شخصا بالسنوات :

21	44	9	30	13	25	32	12	35	15
33	24	23	22	17	36	26	20	14	31
23	12	31	16	27	5	18	3	28	38
16	25	14	25	21	22	15	27	21	11
42	7	23	37	24	32	26	29	23	15
34	29	39	22	16	10	28	17	8	29

والمطلوب :

أ- ناقش بإيجاز الخطوات التى تستخدم فى تبويب هذه البيانـات ، ثم استخدم هذه الخطوات فى تكوين جدول تكرارى ذى فئات متساوية.

ب- بإسـتخدام الجـدول التكـرارى النــاتج فـى(أ) كـون الجـدول التكـرارى المتجمـع الصاعد مرة بإستخدام التكرارات الاصلية ومرة أخرى بإستخدام التكرارت النسبية.

ج- أوجد قيمة العمر الذى يقل عنه 33شخصا.

د- استخدم شكل المدرج التكرارى للتعبير عن التوزيـع النـاتج فـى (أ) ثـم علـق علـى شكل التوزيع .

(9): فيما يلى قيمـة الاربـاح (بـالآف الجنيـهات) لشـركتين متنـافسـتين (س،ص) فـى مجال الصناعات القطنية وذلك خلال الفترة (1970-1975) .

السنة	1970	1971	1972	1973	1974	1975
الشركة(س)	**12**	**16**	**21**	**23**	**24**	**32**
الشركة(ص)	**10**	**11**	**13**	**18**	**25**	**26**

<u>والمطلوب</u> :

أ- قارن بين أرباح الشركتين بإستخدام الاعمدة المجزأة والاعمدة المتلاصقة.
ب- إستخدام خريطة الخط البيانى لعرض البيانات الخاصة بأرباح الشركة(س).
ت- بين ماهى الشركة الاقدر على المنافسة بإستخدام خريطة الشريط .

(10): فيما يلى درجات 30 طالبا فى كل من مادتى الرياضـة والإحصـاء والمطلـوب وضع هذه البيانات فى صورة جدول تكرارى مزدوج.

الرياضة	الاحصاء	الرياضة	الاحصاء	الرياضة	الاحصاء
83	77	84	76	86	83
89	72	81	89	79	84
66	64	92	87	94	91
66	72	72	69	80	67
70	86	71	77	85	93
97	94	82	81	88	85
69	78	83	92	96	90
74	75	75	80	78	88
65	63	79	74	50	58
68	67	76	82	52	57

(11): البيانات التالية تمثل درجات 100 طالبا فى كل من مادتى الرياضيات (س) والإحصاء (ص):

س	ص	س	ص	س	ص	س	ص	س	ص
41	20	89	78	56	41	52	47	62	64
80	70	50	34	82	73	40	42	93	75
67	55	65	53	66	63	67	65	60	49
96	85	69	70	93	82	73	64	47	36
30	25	82	80	80	60	76	67	90	80
61	62	49	46	35	37	58	53	50	43
98	78	70	69	52	59	38	39	60	60
50	67	75	71	76	53	79	59	72	62
85	64	50	50	72	70	55	36	51	45
75	66	61	58	63	68	63	54	68	57

والمطلوب :

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

- طول الفئة فى كل من س ، ص = 10 درجة .

- الحد الادنى للفئة الاولى من فئات(س) = 30 درجة .

- الحد الادنى للفئة الاولى من فئات(ص) = 20 درجة .

ب - إستنتج التوزيع التكرارى 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 - بإستخدام المدرج التكرارى ، قارن بين شكلى التوزيعين من حيث التماثل والالتواء .

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

(ملحوظة : يعتبر الطالب ناجحا إذا حصل على 50 درجة فأكثر).

و - بإستخدام التوزيع المتجمع المناسب ، حدد قيمة الدرجة التى يقل عنها 14% من إجمالى عدد الطلاب فى كل من مادتى الرياضيات والاحصاء ثم علق على إجابتك .

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 بإستخدام الاعمدة المتلاصقة والاعمدة المجزأة .

ملحوظة : تقديرات النجاح والرسوب تحدد كما يلى :

التقدير	ضعيف جدا	ضعيف	مقبول	جيد	جيد جدا	ممتاز
فئة الدرجات	صفر-	35-	50-	65-	80	90-100

(12): الجدول التالى يمثل عدد 100شخص موزعين حسب فئات السن المختلفة :

فئات السن	20-	25-	35-	45-	50-	55-60	الجملة
عدد الاشخاص	5	22	30	21	19	3	100

والمطلوب :

أ- رسم المدرج التكرارى والمضلع التكرارى لهذا التوزيع .

ب- إنشاء الجدول التكرارى المتجمع الصاعد والنازل ثم أوجد عدد الاشخاص الذين يقل عمرهم عن 45 سنة بالرسم.

ت- ماهوالحد الادنى لأعمار 40 شخصا بالرسم .

(13): فيما يلى التوزيع التكرارى لأعمار 100 شخص :

فئات الاعمار	5-	10-	15-	20-	25-	35-	45-55	المجموع
عدد الاشخاص	5	12	13	22	18	16	14	100

<u>والمطلوب</u> :

أ- إرسم كلا من المدرج التكرارى والمضلع التكرارى والمنحنى التكرارى لهذا التوزيع .

ب- احسب قيمة العمر الذى يقل عنه 30% من إجمالى عدد الموظفين وكذا ما هو عدد الموظفين الذين تبلغ أعمارهم 35 سنة على الاقل .

جـ - بإستخدام المنحنى التكرارى المتجمع المناسب أحسب نسبة الاشخاص الذين تبلغ أعمارهم 18 سنة على الاقل.

(14): بإستخدام التوزيع التكرارى الوارد فى الجدول (2-14) :

<u>المطلوب</u> :

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 84% من إجمالى عدد الاسر .

ت- على ضوء نتائجك فى (ب) ، إستنتج قيمة الانفاق الاسبوعى الذى تنفقه كحد أدنى 36% من إجمالى عدد الاسر .

(15): فى إحدى الشركات الصناعية يوجد 200 عامل كان توزيع أجورهم فى الساعة بالجنيه كالآتى:

الأجر	20-	25-	30-	35-	40-	45-	50-	55-
العدد	10	17	24	43	34	30	23	19

والمطلوب :

1- عرض هذه البيانات بإستخدام المدرج التكرارى والمضلع التكرارى.
2- رسم المنحنى المتجمع الصاعد ومنه أوجد عدد العمال الذين يتقاضون أجرا أقل من 40 جنيه .
3- ماهو الحد الاعلى للأجر الذى يبلغه 120 عاملا .
4- رسم المنحنى التكرارى المتجمع النازل ومنه أوجد عدد العمال الذين يتقاضون أجرا قدره 37.5 جنيه فأكثر .
5- ماهو الحد الادنى الذى يتقاضاه 120 عاملا .

(16): سحبت عينة حجمها 130 عامل من شركة السكر وكانت كمية إنتاج العامل مبوبة فى الجدول التكرارى التالى :

كمية الانتاج	20-	30-	40-	50-	60-	70-80	المجموع
عدد العمال	14	20	28	36	22	10	130

والمطلوب :

أ- رسم المدرج التكرارى والمضلع التكرارى والمنحنى التكرارى للتوزيع السابق.
ب- إيجاد الجدول التكرارى المتجمع الصاعد والهابط .

(17): تم اختيار 40 طالب من مدرسة ما لدراسة مستوى الذكاء للطلبة وكانت النتائج كالتالى :

55	50	40	60	70	55	40	50	60	20
45	40	35	50	55	50	35	40	60	30
80	49	60	28	42	51	62	21	36	52
76	56	93	20	39	32	38	42	82	85

<u>والمطلوب</u> :

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

(18): فيما يلى كمية الانتاج من بعض الصناعات بالألف طن :

المصنع	السكر	الورق	الاسمنت	الحديد	غزل ونسيج	زيت البترول
الانتاج	230	800	1900	700	200	7

والمطلوب تمثيل هذه البيانات مستحدما فى ذلك :

أ- **الاعمدة البسيطة .** ب- **الدائرة .**

(19): البيانات التالية تمثل التحويلات السياحية (بالمليون جنيه) وعدد الليالى السياحية (بالليلة) خلال 25 شهرا ، والمطلوب وضع البيانات فى صورة جدول تكرارى مزدوج :

الليالى السياحية	التحويلات السياحية	الليالى السياحية	الليالى السياحية
46	82	36	89
40	74	32	66
42	85	36	65
42	88	43	70
43	91	47	96
30	83	37	68
45	78	36	74
27	50	35	84
46	96	44	72
46	85	34	73
46	80	38	72
32	94	41	81
45	79		

(20): أخذت عينة من 36 سائح وتم جمع البيانات الخاصة بالغرض من القدوم فكاتنت كالآتى:

متعة وترفيه	متعة وترفيه	عقيدة	عقيدة
عقيدة	عقيدة	متعة وترفيه	متعة وترفيه
متعة وترفيه	رياضة	عقيدة	عقيدة
عقيدة	عقيدة	متعة وترفيه	رياضة
رياضة	متعة وترفيه	عقيدة	عقيدة
زيارة للأسرة	رياضة	عقيدة	رياضة
عقيدة	متعة وترفيه	رياضة	زيارة للأسرة
رياضة	عقيدة	زيارة للأسرة	عقيدة
متعة وترفية	عقيدة عقيدة	رياضة	متعة وترفية

والمطلوب :

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

(21): البيانات التالية تمثل توزيع الانفاق اليومى ل 250 سائح فى إحدى محافظات ج.م.ع :

الإنفـــاق اليومى	20-	25-	30-	35-	40-	45-	50-	55-
عـــــــدد السائحين	18	27	32	51	43	35	24	20

المطلوب :

1- عرض هذه البيانات بإستخدام المدرج التكرارى والمضلع التكرارى .

2 - رسم المنحنى التكرارى المتجمع الصاعد ومنه أوجد :

أ- عدد السائحين الذين ينفقون أقل من 40 جنيه يوميا .

ب- الحد الادنى للانفاق اليومى لـ 120 سائحا .

3- رسم المنحنى المتجمع النازل ومنه أوجد :

a. عدد السائحين الذين ينفقون 37.5 جنيه فأكثر يوميا .

b. الحد الادنى للإنفاق اليومى لـ120سائحا.

(22): الجـدول التـالى يوضـح التوزيـع التكـرارى لأجـور العمـال الاسـبوعية (بالجنيهات) فى ثلاثة مصانع مختلفة (س،ص،ع) :

فئات الاجور		50-	60-	70-	80-	90-	100-	المجموع
عدد العمال فى....	المصنع(س)	1	3	4	12	12	8	40
	المصنع(ص)	8	12	12	4	3	1	40
	المصنع(ع)	15	17	18	18	17	15	100

<u>والمطلوب</u> :

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

*** مابين المصنعان : س،ص   * مابين المصنعان : س،ع**

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

3- بإستخدام التوزيع التكرارى المتجمع المناسب . قارن بين قيمة الاجر الذى يحصل عليه (كحد أدنى) 50% من اجمالى عدد العمال وذلك فى كلا من المصنعين س ، ع .

4- قارن بين قيمة الاجر الذى يحصل على أقل منه 20% من اجمالى عدد العمال فى كل من المصنعين س ، ص وذلك بإستخدام التوزيع التكرارى المناسب .

5- إحسب النسبة المئوية للعمال الذين يتقاضون أجورأقل من 80 جنيها فى كل من المصانع الثلاثة . علق على نتائجك على ضوء ماحصلت عليه فى المطلوب (1).

6- احسب قيمة الاجر الذى يقل عنه 40 عاملا فى المصنع (ع).

7- أدمج توزيعى أجور العمال فى المصنعين س ، ص فى كل فئة من الفئات المختلفة ، ثم قارن بين شكل التوزيع الجديد وبين شكل التوزيع فى كل من هذين المصنعين .

(23): فيما يلى لديك التوزيع التالى لأعمار عينة من الاشخاص حجمها 100 مفردة :

الفئة	اقل من 30	اقل من 35	اقل من 45	اقل من 60	اقل من 80
التكرار	19	27	48	73	100

والمطلوب :

1- أرسم المدرج التكرارى .

2- ماهى نسبة الاشخاص الذين تتراوح أعمارهم مابين 40 ، 67 سنة.

3- أذا كانت نسبة الاشخاص الذين تتراوح أعمارهم مابين 37سنةوالعمر(y) تبلغ 15% من اجمالى عدد الاشخاص . حدد ماهى قيمة العمر (y) .

4- ماهو الحد الادنى للعمر الذى يبلغه 75% من العمال على الاقل .

الفصل 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 لنفس الظاهرة وهذا يتنافى مع الخصائص الجيدة لطرق العرض. ويزداد الامر صعوبة إذا ما أردنا مقارنة مجموعتين أو أكثر من البيانات. وثمة امر آخر على جانب كبير من الاهمية يضع حدا لاستخدام الطرق البيانية والجدولية وهو صعوبة الاستفادة منها فى مجال الاسقراء الاحصائى . وربما اقتصرت فوائدها الإستقرائية على أن يقدم المضلع التكرارى مثلا لعينة من البيانات نقوم بتلخيصها تصورا عن شكل المضلع التكرارى للمجتمع الذى سحبت منه هذه العينة. لذا فإنه فى هذا الباب سنلجأ لتلخيص البيانات فى صورة رقمية وذلك عن طريق حساب بعض المقاييس الاحصائية والتى تعرف بإسم المتوسطات وهى التى تصف لنا كيفية توزيع الظاهرة محل الدراسة فى العينة أو فى المجتمع محل الدراسة بطريقة دقيقة ومختصرة وتصلح أيضا للمقارنة بين ظاهرتين أوأكثر. 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 يتم ترتيب مجموعة البيانات بحسب قيمتها العددية . فمن الملاحظ انه فى كثير من الظواهر السلوكية والاجتماعية تنزع معظم البيانات الى التمركز حول تلك القيمة الوسطية . فأولئك الذين يتصفون بالنحافة الشديدة 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 وهكذا ...". .... 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 :

(1):ان تكون طريقة حسابه سهلة وأن يكون أى تغير فى قيمة الظاهرة محل الدراسة لا يؤدى إلى تغير فى طريقة حسابه .

(2):يراعى أن تكون البيانات متجانسة ، فلا يجوز مثلا حساب دخل الوزير وبعض العمال فى وزارته ، بينما يجوز حساب متوسط السن مثلا للوزير والعمال معا .

(3):يفضل أن يأخذ المقياس عند حسابه – فى الاعتبار – جميع القيم ولايتأثر بالقيم المتطرفة او الشاذة .

(4):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 مبينين كيفية حسابها سواء فى حالة البيانات غير المبوبة أو فى حالة البيانات المبوبة تكراريا ، وتتلخص هذه المقاييس فيما يلى :

1. **الوسط الحسابى .**
2. **الوسيط ، وسندرس معه مقاييس شبيهه فى طريقة حسابة مثل الربيعين الادنى والأعلى وخلافه.**
3. **المنوال .**
4. **الوسط الهندسى .**
5. **الوسط التوافقى .**

<u>أولا:الوسط الحسابى</u>:- Arithmetic Mean (or Avarage)

يعتبر الوسط الحسابى من أهم مقاييس النزعة المركزية وأكثرها إستخداما فى الحياة العملية وذلك لبساطته وسهولة فهمه ووضوح معناه ، وهو الذى يشار له بإسم المتوسط فى لغتنا الدارجة .

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 .

(3−1−1) :حساب الوسط الحسابى من المفردات (بيانات غير مبوبة):

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:

أولا : الطريقة الحسابية المطولة (المباشرة) :−

بصفة عامة إذا كان لدينا مجموعة من المفردات ولتكن:

$x_1 ,\ x_2 ,\ \ldots,\ x_n$

حيث (n) تعبر عن عدد المفردات . فأن الوسط الحسابى لهذه المجموعة من هذه المفردات سنرمز له بالرمز $\overline{x}$ وهوعبارة عن النسبة مابين مجموع القيم وعددها أى أن :

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum x}{n} \qquad (1)$$

حيث يمثل البسط ($\sum x$) مجموع القيم التى يأخذها المتغير محل الدراسة والتى يبلغ عددها (n) مفردة .

مثال(1) :

أحسب الوسط الحسابى لمجموعة القيم : 500، 600 ، 700، 750، 800

الحل :حيث أن :

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

لذا فإن :

$$\overline{X} = \frac{500+600+700+750+800}{5} = 670$$

<u>مثال (2):</u>

احسب الوسط الحسابى لمجموعة القيم : 1 , 3- , 0 , 2 , 5 , 9

<u>الحل</u> : حيث أن :

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

لذا فإن :

$$\bar{X} = \frac{9+5+2+0+(-3)+(-1)}{6} = \frac{12}{6} = 2$$

<u>ثانيا :- طريقة الانحرافات البسيطة:</u>

وتهدف هذه الطريقة الى إختصار الأرقام وكذا العمليات الحسابية بعض الشئ وبصفة خاصة حينما يكون عدد المفردات وقيم المفردات كبير نسبيا. وهذه الطريقة تقلل الجهد الحسابى المبذول وكذا تقلل من احتمالات الخطأ خاصة حينما لاتتوافر الالات الحاسبة لإستخدامها فى هذا المجال . وتتلخص هذه اطريقة فى إختيار أى قيمة من قيم 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 معها 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 :

(1): بأفتراض أن لديك مجموعة القيم X_1 , X_2 , ... , X_n .

(2): يتم أعتبار إحدى 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 .

(3): يتم طرح قيمة الوسط الفرضى (A) من جميع المفردات لنحصل على انحرافات القيم عن هذا الوسط الفرضى ولتكن(di) حيث: i= 2,1,....,n أى أن:

$$d_1 = x_1 - A \quad , d_2 = x_2 - A , d_3 = x_3 - A , \ldots\ldots \quad , d_n = x_n - A$$

او بلغة أخرى فأن :

$$d_i = x_i - A \quad \text{for} \quad i= 1,2,\ldots..,n$$

وتسمى قيم (di) بالانحرافات البسيطة . فيتم إيجاد الوسط الحسابى لتلك الانحرافات البسيطة وذلك من خلال قسمة مجموع تلك الانحرافات على عددها أى ان :

$$\overline{d} = (\sum d_i) / n =$$

فيكون الوسط الحسابى لقيم المتغير (X) هو :

$$\overline{x} = \overline{d} + A = (\sum d_i) / n + A \qquad (2)$$

<u>مثال(3):</u>

أحسب الوسط الحسابى للمفردات الواردة بالمثال (1) 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 والذى سيتم طرحه من كافة المفردات هو القيمة 700 اى ان A=700 فيكون لدينا الجدول التالى :

جدول (3−1)

X_i	$d_i = x_i - 700$
500	500-700 = -200
600	600-700 = -100
700	700-700 = 0
750	750-700 = 50
800	800-700 = 100
$\sum$	-150

وحيث ان :

$$\bar{X} = A + \frac{\sum di}{n}$$

$$= 700 + \frac{(-150)}{5} = 700 - 30 = 670$$

وهو نفس الناتج الذى حصلنا عليه عند إتباع الطريقة الحسابية المطولة .

<u>ثالثا : طريقة الانحرافات المختصرة (المعدلة)</u> :-

إمتداد للطريقة السابقة (الانحرافات البسيطة) فإنه إذا كانت قيم الانحرافات البسيطة (di) تقبل القسمة على مقدار ثابت أو وسط فرضى آخر وليكن (B) بدون بواقى . فإنه يمكن إختصار العمليات الحسابية أكثر مما سبق عند حساب الوسط الحسابى وذلك على النحو التالى :

(1): بعد الحصول على الانحرافات البسيطة للمفردات كما سبق فى الطريقة السابقة ولتكن هذه الانحرافات هى (di) حيث أن :

$$d_i = x_i - A \qquad \text{for} \quad i = 1 , 2 , \ldots , n$$

فإنه يمكن أعتبار وسط فرضى آخر أو عامل مشترك فيما بين قيم الإنحرافات البسيطة وليكن (B) وبإعتبار هذا العامل المشترك وقسمة الإنحرافات البسيطة على هذا العامل نحصل على الانحرافات الأكثرإختصارا على النحو المبين التالى :

$$D_1 = \frac{d_1}{B} \;, \quad D_2 = \frac{d_2}{B} \;, \quad D_3 = \frac{d_3}{B} \;, \quad \ldots\ldots, \quad D_n = \frac{d_n}{B}$$

(2): لإيجاد الوسط الحسابى للقيم الأصلية(X_i) أى $\overline{X}$ مستخدمين فى ذلك الحساب الانحرافات الاكثر اختصارا (المعدلة) اى (D_i) فإنه يجب رد الحسابات بطريقة عكسية. فبعد حساب متوسط الانحرافات الاكثر إختصارا أى بعد حساب $\overline{D}$ ، فإنه يجب ضرب متوسط الإنحرافات المعدلة فى المقدار الثابت (B) الذى تم قسمة الانحرافات البسيطة عليه ثم إضافة المقدار الثابت الذى تم طرحه من المفردات أى (A) . وعليه فإن :

$$\overline{X} = A + B\,\overline{D} = A + B\,\frac{\sum Di}{n} \qquad (3)$$

<u>مثال (4):</u>

احسب الوسط الحسابى للمفردات الواردة بالمثال(1) بإستخدام الطريقة الاكثر إختصارا (الانحرافات المعدلة) .

<u>الحل</u> : لحساب الوسط الحسابى بالطريقة الاكثر إختصارا لاحظ انه فى المثال السابق فأن الانحرافات البسيطة (di) تقبل القسمة على الثابت 50 . لذا يمكن قسمة الإنحرافات البسيطة على 50 لتبسيط النواتج ثم حساب الوسط الحسابى بالطريقة المعدلة (أو الأكثر إختصارا)وهو ما يوضحه الجدول التالى :

جدول (3−2)

X_i	$D_i = X_i - 700$	$D_i = d_i /50$
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 :

$$\bar{X} = 700 + 50\left(\frac{-3}{5}\right) = 700 - 30 = 670$$

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) .

<u>(3−1−2) خصائص الوسط الحسابى :−</u>

بصفة عامة هناك مجموعة من الخصائص التى تميز الوسط الحسابى عن غيره من مقاييس المتوسطات نذكر منها مايلى :

<u>أولا</u>: 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 $X_n, \ldots, X_2, X_1$ ومتوسطه الحسابى هو($\bar{X}$) فتكون إنحرافات القيم(X_i) عن متوسطها هى عبارة عن :

$$(x_1 - \bar{x}),\ (x_2 - \bar{x}),\ (x_3 - \bar{x}),\ \ldots\ldots\ldots,\ (x_n - \bar{x})$$

وبذلك فإن مجموع هذه الانحرافات هو :

$$\sum_{i=1}^{n} (x_i - \bar{x}) = (x_1 - \bar{x}) + (x_2 - \bar{x}) + \ldots\ldots + (x_n - \bar{x})$$
$$= (x_1 + x_2 + x_3 + \ldots\ldots + x_n) - (\bar{x} + \bar{x} + \ldots\ldots + \bar{x})$$

لاحظ ان الحد الثانى عبارة عن مجموع المقدار الثابت ($\bar{x}$) عددا (n) من المرات فيكون مجموعه هو : ($n\bar{x}$)فيكون :

$$\sum_{i=1}^{n} (x_i - \bar{x}) = \sum_{i=1}^{n} x_i - n\bar{x} \qquad (4)$$

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 فبالتعويض فى العلاقة (4) فيكون :

$$\sum_{i=1}^{n} (x_i - \bar{x}) = n\bar{x} - n\bar{x} = 0$$

أى أن مجموع إنحرافات القيم عن متوسطها الحسابى يساوى دائما أبدا صفرا .

وهذه الخاصية توضح الدور المركزى الذى يلعبه قيمة الوسط الحسابى للمفردات.

<u>ثانيا</u> :مجموع مربعات إنحرافات القيم عن وسطها الحسابى أصغر مايمكن 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) عبارة عن أى مقدار ثابت أختيارى تختلف قيمته عن قيمة الوسط الحسابى $\bar{X}$ ، ولاثبات صحة ذلك فحيث ان :

$$\sum_{i=1}^{n} (x_i - A) = \sum_{i=1}^{n} \{ (x_i - \bar{x}) + (\bar{x}- A)\}^2$$

$$= \sum_{i=1}^{n} (x_i - \bar{x})^2 + 2 \sum_{i=1}^{n} (x_i - \bar{x}) (\bar{x} – A)+ \sum_{i=}^{n} (\bar{x} - A)^2$$

وذلك لأنA ، $\bar{X}$ مقادير ثابتة .

وحيث ان مجموع إنحرافات القيم عن وسطها الحسابى صفرا أى ان:

$\sum (x_i - \bar{x})=0$ لذا فإن الحد الاوسط فى العلاقة الأخيرة تتلاشى قيمته ويصبح لدينا :

$$\sum_{i=1}^{n} (x_i - A)^2 = \sum_{i=1}^{n} (x_i – \bar{x})^2 + n (\bar{x} - A)^2$$

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 (5):-</u>

إذا كان لديك مجموعة المفردات التالية : 7 ، 8 ، 12 ، 15 ، 3

فالمطلوب :

1- أوجد مجموع إنحرافات القيم عن وسطها الحسابى ، وعلق على نتيجتك

2-اوجد :

$$\sum_{i=1}^{5} (x_i - 10)^2 \quad , \sum_{i=1}^{5} (x_i - 7)^2 \quad , \sum_{i=1}^{5} (x_i - \overline{x})^2$$

وعلق على نتائجك .

الحل :

1- 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:

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{7 + 8 + 12 + 15 + 3}{5} = \frac{45}{5} = 9$$

والجدول التالى يبين الحسابات اللازمة :

جدول (3−3)

X_i	$x_i - \overline{x}$	$(x_i-\overline{x})^2$	(x_i-7)	$(x_i-7)^2$	(x_i-10)	$(x_i-10)^2$
7	-2	4	0	0	-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

فمن خلال العمود الثانى فى الجدول (3−3) يتضح لنا ان مجموع انحرافات القيم عن وسطها الحسابى مساويا للصفر أى ان :

$$\sum_{i=1}^{n} (x_i - \overline{x}) = \sum_{i=1} (x_i - 9) = 0$$

وهو ما يؤكد صحة الخاصية الاولى من خصائص الوسط الحسابى .

2- من خلال الاعمدة الموضحة فى جدول (3−3) يتضح لنا ان :

$$\sum_{i=1}^{5} (x_i - \overline{x})^2 = 86 \ , \ \sum_{i=1}^{5} (x_i - 7)^2 = 106 \ , \ \sum_{i=1}^{5} (x_i - 10)^2 = 91$$

وهو مايفيد أن مجموع مربعات إنحرافات القيم عن وسطها الحسابى أقل ما يمكن وهو مايتفق مع الخاصية الثانية من خصائص اى وسط فرضى اخر .

<u>(3−1−2) حساب الوسط الحسابى للبيانات المبوبة تكراريا :−</u>

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 مع طبيعة عرض البيانات المبوبة فى جداول تكرارية سواء كانت هذه البيانات تمثل متغيرات كمية منفصلة أم متصلة .

<u>أولا : حساب الوسط الحسابى لجدول تكرارى لمتغير كمى منفصل(متقطع):</u>

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 لحساب الوسط الحسابى فالأمر يتطلب حساب مجموع القيم منسوبا الى عددها . لذا فإنه يلزم ضرب كل قيمة من قيم المتغيرفى عدد مرات تكرارها ثم إيجاد مجموع حواصل الضرب ، وعليه فيمكن 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 , ... , X_n ذات تكرارات مقابلة لها فى هذا الجدول ولتكن F_1 , F_2 ,...., F_n على الترتيب . فأن الوسط الحسابى سيكون عبارة عن :

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i F_i}{\sum_{i=1}^{n} F_i} \qquad (6)$$

وهنا يمكن ملاحظة أن:

القانون الوارد بالعلاقة (6) وهو الخاص بحساب الوسط الحسابى لبيانات متغير منفصل مبوبا فى جدول تكرارى هو نفسه القانون الوارد بالعلاقة(1) 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 (1) بـــقيمة($X_iF\ i$) فى العلاقة (6). وكذلك استبدال (n) فى العلاقة (1) بـــ ($\sum Fi$) فى العلاقة (6) .أى ان التكرارت تؤخذ فى الاعتبار وذلك من خلال ضرب كل قيمة من قيم المتغير المنفصل (Xi) فى عدد مرات تكرارها أى (Fi) .

هذا مايتعلق بطريقة حساب الوسط الحسابى لبيانات متغير منفصل فى حالة البيانات المبوبة تكراريا . وأما إذا تم إستخدام كل من طريقة الانحرافات البسيطة أ و الانحرافات المعدلة فأنه يجرى أيضا نفس طبيعة التعديل على القوانين الواردة بالعلاقتين (2)،(3) لتكون صيغة حساب الوسط الحسابى للمتغير المنفصل فى تلك الحالة على الصورة التالية :

$$\bar{X} = A + \frac{\sum_{i=1}^{r} d_i F_i}{\sum_{i=1}^{r} F_i} \qquad (7)$$

وذلك فى حالة إستحدام طريقة الإنحرافات البسيطة . أما فى حالة إستخدام طريقة الإنحرافات المعدلة فيصبح القانون على الصورة التالية :

$$\bar{X} = A + B \frac{\sum_{i=1}^{r} D_i F_i}{\sum_{i=1}^{r} F_i} \qquad (8)$$

حيث تمثل (r) عدد اقسام أو فئات المتغير المنفصل فى التوزيع التكرارى.
والمثال التالى يوضح كيفية حساب الوسط الحسابى فى حالة الجدول التكرارى لمتغير منفصل باستخدام طريقة الإنحرافات المعدلة .

<u>مثال (6):</u>

فيما يلى التوزيع التكرارى التالى يبين عدد العمال الغائبين عن العمل فى إحدى الشركات خلال شهر نوفمبر من عام 2000 .

جدول (3-4)

No.of Absence	0	1	2	3	4	5	6	7	8	∑
Frequency	2	1	4	7	3	6	3	2	2	30

<u>والمطلوب</u>: 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(البسطة والإنحرافات البسيطة والإنحرافات المعدلة .**

جدول(3−5)

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
∑	30	120		0		0

***حساب الوسط الحسابى بالطريقة المباشرة (المطولة):**

فحيث ان :

$$\bar{X} = A + \frac{\sum_{i=1}^{r} X_i F_i}{\sum_{i=1}^{r} F_i}$$

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 (3−5) 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_iFi) مساويا للصفر لأن الوسط الفرضى المستخدم فى تلك الطريقة وهو قيمة (A) مساوية للوسط الحسابى . لذا فلابد أن يكون مجموع هذا العمود (diFi) مساويا للصفر وذلك لأن مجموع انحرافات القيم عن وسطها الحسابى هو صفرا (خاصية (1)) .

***<u>حساب الوسط الحسابى بإستخدام طريقة الانحرافات المعدلة</u> : 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

$$\bar{X} = 4 + 2\left(\frac{0}{30}\right) = 4 \quad \text{persons}$$

لاحظ ان النتيجة واحدة بأى من الطرق الثلاث .

<u>ثانيا : حساب الوسط الحسابى لبيانات متغير متصل (مبوبة 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 حتى 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 تكرارات 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 فيتمثل 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 وهذه المراكز 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) أم الغير منتظمة .

<u>مثال (7):</u>

فيما يلى الجدول التالى يبين توزيع فئات الاجر اليومى (بالجنيه) للعامل فى إحدى مصانع مدينة العاشر من رمضان .

جدول (3−6)

Classes	120-	140-	160-	180-	200-	220-	240-260	∑
Frequency	2	4	11	38	30	10	5	100

<u>والمطلوب</u>: حساب متوسط الاجر اليومى للعامل فى هذا المصنع بإستخدام الطرق الثلاثة .

<u>الحل</u>:

1- <u>بإستخدام الطريقة المباشرة (المطولة) :</u>

الجدول التالى يوضح الحسابات اللازمة لحساب متوسط الاجر اليومى :

جدول (3−7)

Classes	F_i	Mid Point (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 :

$$\bar{X} = \frac{19800}{100} = 198 \qquad (L.E)$$

2- <u>بإستخدام طريقة الانحرافات البسيطة :</u>

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 (Xi) للحصول على الانحرافات (di) . ثم ضرب هذه الانحرافات البسيطة فى التكرارت المقابلة وايجاد المجموع . والجدول التالى يلخص الخطوات اللازمة لحساب المتوسط الحسابى بطريقة الإنحرافات البسيطة :

جدول (3−8)

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
∑	100			800

وحيث أن :

$$\bar{X} = A + \frac{\sum_{i=1}^{r} d_i F_i}{\sum_{i=1}^{r} F_i}$$

وعليه فإن :

$$\bar{X} = 190 + (800 / 100) = 198 \qquad (L.E)$$

وهى نفس الناتج بإستخدام الطريقة المباشرة .

3- بإستخدام طريقة الانحرافات المختصرة (المعدلة) :-

لاحظ فى الجدول السابق ان عمود الانحرافات البسيطة (di) يقبل القسمة على (20) وهى بمثابة طول الفئة فى هذا الجدول المنتظم . لذا يمكن إعتبار طول الفئة بمثابة وسطا فرضيا آخر وحساب الانحرافات المعدلة (Di) وذلك من خلال قسمة عمود الانحرافات البسيطة(di) على (20)، ثم أيجاد حاصل ضرب هذه الإنحرافات المختصرة (Di) فى التكرارات المقابلة (Fi)، ثم إيجاد هذا المجموع (Σ Di Fi)والجدول التالى يوضح الحسابات اللازمة لايجاد متوسط الاجر اليومى بطريقة الإنحرافات المعدلة أو الأكثر إختصارا .

جدول (3-9)

Classes	F_i	X_i	$d_i=X_i-190$	$D_i = d_i/20$	$D_i F_i$
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 :

$$\bar{X} = A + B \left(\frac{\sum_{i=1}^{r} D_i F_i}{\sum_{i=1}^{r} F_i} \right)$$

وحيث أن : $A = 190$, $B = 20$. لذا فإن :

$$\bar{X} = 190 + 20\,(40 / 100) = 190 + 8 = 198 \quad \text{(L.E)}$$

وهى نفس النتائج المستخلصة من الطريقتين السابقتين لكن أهم مايميز هذه الطريقة هى تبسيط أرقام الجدول لأقصى درجة ممكنة لدرجة أنها تؤدى إلى عدم الحاجة الى إستخدام الآلآت الحاسبة فى حساب مجاميع الجدول اللازمة .

(3-1-3) :- 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 :-

فبالنسبة للقانون المستخدم لحساب الوسط الحسابى بإستخدام الطريقة المباشرة (المطولة) فمن العلاقة (6) السابق الاشارة اليها : فحيث ان :

$$\overline{X} = \frac{\sum_{i=1}^{r} X_i F_i}{\sum_{i=1}^{r} F_i}$$

$$= \sum_{i=1}^{r} X_i \frac{F_i}{\sum_{i=1}^{r} F_i}$$

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 فتكون قيمة الوسط الحسابى ماهى إلا :

$$\overline{x} = \sum_{i=1}^{r} X_i R_i \qquad (9)$$

اى أن الصيغة الرياضية لايجاد الوسط الحسابى بإستخدام الطريقة المباشرة فى حالة جداول التكرارات النسبية ماهى إلا عبارة عن إيجاد مجموع حواصل ضرب مراكز الفئات فى التكرارت النسبية المقابلة لتلك الفئات .

***كذلك وبنفس الاسلوب تصبح صيغة ايجاد الوسط الحسابى بإستخدام طريقة الانحرافات البسيطة فى حالة توافر تكرارت نسبية على الصورة :**

$$\bar{X} = A + \sum_{i=1}^{r} d_i R_i \qquad (10)$$

حيث (di) 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 نسبى فى الجدول التكرارى النسبى.

***وأخيرا فأن الصيغة الرياضية لحساب الوسط الحسابى بإستخدام طريقة الانحرافات المعدلة (الطريقة الاكثر إختصارا) فى حالة توافر جدول تكرارى نسبي تصبح على الصورة التالية :-**

$$\overline{X} = A + B \sum_{i=1}^{r} D_i R_i \qquad (11)$$

حيث (Di) تمثل الانحراف المختصر للفئة (i) كما سبق تعريفه .

والمثال التالى يوضح طريقة حساب الوسط الحسابى بإستخدام الطرق الثلاث فى حالة التوزيعات التكرارية النسبية .

<u>مثال(8):-</u>

بإستخدام الطرق الثلاث لحساب الوسط الحسابى ماهو متوسط قيمة الاجر للتوزيع التكرارى التالى لمجموعة من العاملين فى إحدى الشركات (بالجنيه) :

جدول (3−10)

التوزيع التكرارى النسبى للأجر فى إحدى الشركات

Classes	15-	25-	35-	45-	55-65
Relative Fre.	0.2	0.32	0.24	0.16	0.08

الحل :

1- ايجاد الوسط الحسابى للدخل اليومى بإستخدام الطريقة المباشرة (المطولة) . والجدول التالى يبين الحسابات اللازمة لايجاد متوسط الاجر اليومى للعامل فى تلك الشركة :

جدول (3−11)

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
∑	1		36

وحيث أن :

$$\overline{X} = \sum_{i=1}^{r} X_i R_i$$

لذا فإن :

$$\overline{X} = 36 \qquad (L.E)$$

2- ايجاد الوسط الحسابى بطريقة الانحرافات البسيطة :−

والجدول (3−12) يوضح الحسابات اللازمة لايجاد متوسط الاجر اليومى فى تلك الشركة بطريقة الإنحرافات البسيطة :

جدول(3−12)

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
∑	1			-4

وحيث أن :

$$\bar{x} = A + \sum_{i=1}^{r} d_i R_i$$

وحيث ان قيمة $A = 40$ (من الجدول(3−12)) . لذا فإن :

$$\bar{X} = 40 + (-4) = 36 \qquad \text{(L.E)}$$

وهى نفس النتيجة بإستخدام الطريقة المباشرة .

3- <u>ايجاد الوسط الحسابى بطريقة الانحرافات المختصرة (المعدلة)</u> :

والجدول (3−13) يوضح الحسابات اللازمة لايجاد متوسط الاجر اليومى فى تلك الشركة بإستخدام طريقة الإنحرافات المختصرة :

جدول(3−13)

Classes	R_i	X_i	$d_i=X_i-40$	$D_i = d_i/10$	$D_i R_i$
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
∑	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). لذا 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 أوطريقة الانحرافات البسيطة .

(3–1–4) : 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 هى :

$X_1 , X_2 , X_3 ,.....,X_m$ وكانت أوزانها الترجيحية المقابلة هى: $n_m ,.....,$

n_1 , n_2 , n_3 فأن الوسط الحسابى المرجح يكون عبارة عن :

$$\bar{X} = \frac{n_1\overline{x_1}+n_2\overline{x_2}+\dots\dots\dots\dots+n_m\overline{x_m}}{n_1+n_2+\dots\dots\dots\dots\dots\dots+n_m} = \frac{\sum_{i=1}^{m} n_i \overline{x_i}}{\sum_{i=1}^{m} n_i}$$

كذلك إذا كان لدينا مجتمعا محل دراسة وليكن حجمه (N) سحبت منه مجموعة من العينات وليكن عددها (m) عينة ذات الحجم $n_m,\dots,n_2,n_1$ وكانت اوساطها الحسابية هى:

$\overline{x_m},\dots,\overline{x_2},\overline{x_1}$ على الترتيب فإن: المتوسط العام (او المرجح) ماهو إلا :

$$\bar{X} = \frac{n_1\bar{x}_1+n_2\bar{x}_2+\dots\dots\dots\dots+n_m\bar{x}_m}{n_1+n_2+\dots\dots\dots\dots\dots\dots+n_m} = \frac{\sum_{i=1}^{m} n_i \overline{x_i}}{\sum_{i=1}^{m} n_i}$$

لاحظ أن كل حد من حدود البسط فى العلاقة الأخيرة 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 (ni) مقدار ثابت أى يكون :

$$n_1 = n_2 = n_3 = \dots\dots\dots\dots = n_m = n$$

فتصبح صيغة العلاقة (12) تأخذ الصورة التالية :

$$\bar{X} = \frac{n(\bar{x}_1+\bar{x}_2+\dots\dots\dots+\bar{x}_m)}{(n)(m)} = \frac{\bar{x}_1+\bar{x}_2+\dots\dots+\bar{x}_m}{m}$$

$$= \frac{\sum_{i=1}^{m} \overline{x_i}}{m} \qquad (12)$$

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 .

مثال (9):

مصنع به ثلاثة الآلآت انتاجية :-

-تنتج الآله الاولى 50 وحدة يوميا وتباع الوحدة منها بسعر 20 جنيه .

-وتنتج الآله الثانية 40 وحدة يوميا وتباع الوحدة منها بسعر 30 جنيه.

-بينما تنتج الآله الثالثة10وحدات يوميا وتباع منها الوحدة بسعر 40 جنيها .

والمطلوب حساب متوسط سعر بيع الوحدة من إنتاج هذا المصنع .

الحل:-

لاحظ أنه إذا استخدمنا صيغة الوسط الحسابى العادى للسعر دون أى أعتبار للكمية التى تنتجها الآله فأن متوسط سعر الوحدة يكون عبارة عن مجموع الاسعار الثلاثة على عددها (3) أى ان :

$$\bar{X} = \frac{20 + 30 + 40}{3} = 30 \qquad \text{(L.E)}$$

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 :

$$\bar{X} = \frac{\sum_{i=1}^{m} n_i \overline{x_i}}{\sum_{i=1}^{m} n_i} = \frac{20(50) + 30(40) + 40(10)}{50 + 40 + 10}$$

$$= \frac{2600}{100} = 26 \qquad (L.E)$$

<u>ملحوظة</u>: من عيوب الوسط الحسابى المرجح عملية تحديد الاوزان الترجيحية. ففى غالبية الاحيان تخضع الاوزان الترجيحية لعملية التقدير الشخصى مما يؤدى الى اختلاف النتائـج من شخص الى أخر .

<u>مثال (10):</u>

يتم تدريس مادة مبادى الاحصاء الوصفى لثلاث شعب فى كلية ما . وأجرى أختبار لطلاب تلك الشعب الثلاث فى هذا المقرر فكانت النتائج على النحو التالى :

Section	Account	Management	Economic
No.of Students	32	20	40
Mean of Test Degree	70	85	68

والمطلوب حساب درجة متوسط درجة الطالب فى هذا المقرر فى الشعب الثلاث:

<u>الحل:</u>– معطى لدينا المعلومات التالية :

$$n_1 = 32 , \quad n_2 = 20 \quad , \quad n_3 = 40$$
$$\bar{x}_1 = 70 , \quad \bar{x}_2 = 85 \quad , \quad \bar{x}_3 = 68$$

فيكون المتوسط العام لدرجة الطالب لكافة الشعب عبارة عن :

$$\bar{X} = \frac{\sum_{i=1}^{3} n_i \ \bar{x}_i}{\sum_{i=1}^{3} n_i} = \frac{30(70)+20(85)+40(68)}{32 + 20 + 40} = 72.39 \ \text{(degree)}$$

<u>ملحوظة</u> :

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 .

ملاحظات حول الوسط الحسابى (المزيا والعيوب):-

(1): قيمة الوسط الحسابى لاتتغير مهما اختلفت طرق 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 . كما أن طريقة الانحرافات البسيطة يفضل استخدامها فى حالة الجداول غير المنتظمة أما الطريقة المباشرة (المطولة) فغالبا ما تستخدم حينما يكون حجم المفردات صغيرا لأنها تحتاج كما رأينا الى عمليات حسابية مطولة واستخدام الآلآت الحاسبة نظرا لكبر حجم الارقام بإستخدام تلك الطريقة.

(2): 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 مايجعله أهم مقاييس النزعة المركزية ولذلك فإنه الاكثر استخداما فى التطبيقات العملية لأساليب التحليل الاحصائى .

(3): 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 وذلك إذا كان تكرارها صغير جدا نسبيا بالنسبة لمجموع تكرارت التوزيع التكرارى وهو مايترتب عليه إهمال جزء من البيانات ومن ثم عدم دقة قيمة الوسط الحسابى فى هذه الحالة . وربما من الافضل عدم استخدام الوسط الحسابى فى مثل هذه الحالات ويتم إستخدام 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 بأحداها . وهذه البدائل هى :

(أ): إستبعاد الفئات المفتوحة بتكرارها وحساب الوسط الحسابى لبقية فئات التوزيع. وإن كان هذا البديل يؤدى الى إهمال جزء من المعلومات الخاصة بالظاهرة محل الدراسة وبصفة 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 وذلك 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 .

(جـ): فى حالة التوزيعات التكرارية القريبة من التماثل فإنه يمكن ايجاد قيمة الوسط الحسابى بدلالة الوسيط والمنوال وذلك بموجب العلاقة التالية (لبيرسون) والتى تربط بين مقاييس النزعة المركزية الثلاث :

$$\text{الوسط الحسابى} = \frac{3\times \text{الوسيط} - \text{المنوال}}{2}$$

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 .

أما عن عيوب الوسط الحسابى: فإن أهم ما 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 مبتعدا 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:
185 ، 205 ، 195 ، 180 ،80 جنيها فإن الوسط الحسابى لمرتباتهم هو 169 جنيها وهو يقل عن مرتب أربعة موظفين من بين الخمسة (أى 80 % منهم) وهذا يعنى أن الوسط الحسابى لايمثل قيم الظاهرة تمثيلا صحيحا كقيمة متوسطه (لاحظ أن القيمة المتطرفة فى هذا المثال هى القيمة 80 جنيه).

(2-3) الوسيط (Q_2) : 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 والذى يعتبر مناسبا فى مثل هذه الحالة 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 ترتيبا تصاعديا او تنازليا . 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 50% من المفردات ويزيد عنها أو يساويها 50% من المفردات وسوف نستعرض فيما يلى طرق حساب الوسيط فى حالة كل من البيانات الغير مبوبة وكذا فى حالة البيانات المبوبة تكراريا .

(3−2−1) : <u>إيجاد الوسيط فى حالة 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 :

(1): إذا كان عدد المفردات (n) فرديا . فإن قيمة الوسيط هى عبارة عن قيمة المفردات ذات الرتبة : $\left(\frac{n+1}{2}\right) th$

(2): أما اذا كان عدد المفردات (n) زوجيا فيوجد مفردتان تتوسطان مفردات الظاهرة محل الدراسة ، الاولى رتبتها (n/2) والثانية التى تليها ورتبتها هى ((n/2)+ 1) ويكون الوسيط هوعبارةعن 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 (11):</u> احسب الوسيط فى الحلات الثلاثة التالية :

(أ):لمجموعة المفردات : 5 ، 18 ، 9 ، 10 ، 15 ، 7 ، 20 .

(ب):لمجموعة المفردات : 25 ، 18 ، 4 ، 30 ، 12 ، 8 ، 27 ، 22 .

(ج):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 ترتيبا تصاعديا أو تنازليا،وليكن الترتيب تصاعدى فتصبح لدينا المفردات على الصورة التالية :

5 ، 7 ، 9 ، 10 ، 15 ، 18 ، 20 .

وحيث أن عدد المفردلات (n=7) وهو عددا فرديا لذا فإن ترتيب الوسيط هو :

$\left(\frac{n+1}{2}\right)th$ = (7 + 1) /2) (= 4 أى المفردة الرابعة . وعليه فإن قيمة الوسيط = Q_2 = 10

(ب): بعد ترتيب المفردات ترتيبا تصاعديا تصبح المفردات 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 :

$(n/2) = 8/2 = 4$ ، $(n/2) + 1 = (8/2)+1 = 5$

فيكون الوسيط هو عبارة عن الوسط الحسابى للمفردتين الرابعة والخامسة أى أن:

$$Q_2 = \frac{18+22}{2} = 20$$

(جـــ): لتحديد وسيط مجموعة التقديرات المعطاه يتم بداية ترتيب هذه المجموعة من التقديرات تصاعديا أو تنازليا . فيكون ترتيب التقديرات ترتيبا تصاعديا على الصورة التالية :-

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 ($5=(9+1)/2$) أى المفردة الخامسة . ومن ثم فيكون وسيط مجموعة المفردات (التقديرات) هو التقدير "جيد".

ويتضح 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 فى اتجاهها . 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 أسر عدد افرادها هو 10 ، 8 ، 7 ، 6 ، 4 ، 3 فلما كان عدد المفردات زوجيا فأن الوسيط حسب التعريف السابق وهو الوسط الحسابى للمفردتين الثالثة والرابعة أى:

($6.5= 2 / (7+6)$) أى أن : $Q_2= 6.5$

وهذه القيمة لاوجود لها ولامعنى لها بأن تكون قيمة الوسيط لعدد أفراد الاسرة هو 6.5 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 70 ولاتوجد أى قيمة أصغر من قيمة الوسيط ويوجد 40% من المفردات اكبر من قيمة الوسيط .

وجدير بالذكر أنه 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 .

(3-2-2) : 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 ترتيبا 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 لإيجاد قيمة الوسيط فى حالة التوزيعات التكرارية يجب بداية الحصول على التوزيع التكرارى المتجمع الصاعد أو الهابط . ثم يتم تحديد رتبة الوسيط والتى تعادل قسمة مجموع التكرارت على(2) بغض النظر عن إجمالى التكرارات من حيث كونه عددا فرديا أم زوجيا ليتحدد بذلك رتبة الوسيط ومن ثم يتحدد الفئة التى سيقع فيها قيمة الوسيط اما عن تحديد قيمة الوسيط فيتم ذلك من خلا إستخدام 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 فى حالة الجداول التكرارية فيما يلى :-

(1):تكوين الجدول التكرارى المتجمع الصاعد أو الهابط ، وليكن التوزيع التكرارى المتجمع الصاعد .

(2): تحديد رتبة الوسيط وهى عبارة عن

$$\text{RanK of } (Q_2) = \frac{\sum_{i=1}^{r} Fi}{2}$$

حيث (r) عبارة عن عدد فئات التوزيع التكرارى المعطى .

(3): يتم تحديد الفئة الوسيطية وهى الفئة التى تقع فيها رتبة الوسيط والتى تم تحديدها فى الخطوة السابقة من الجدول المتجمع ومن ثم يتم تحديد الحد الادنى للفئة الوسيطية (Lower Bound of the Q_2 Class)

وكذا التكرار المتجمع الصاعد السابق وليكن $F_{(Pre.Asc)}$ واللاحق وليكن $F_{(Suc.Asc)}$

(4): يتم ايجاد قيمة الوسيط إما بقانون النسبة والتناسب أوبإستخدام قانون الوسيط حيث أن الوسيط (Q_2) يمكن تحديدها من خلال القانون التالى :

$$Q_2 = L.B.of\,(Q_2)Class + \frac{R.of\,(Q_2) - F_{(Pre.Asc)}}{F_{(Suc.Asc)} - F_{(Pre.Asc)}} \times length\ of(Q_2)\ Class \qquad (14)$$

وفى العلاقة(14) الأخيرة لاحظ أن :

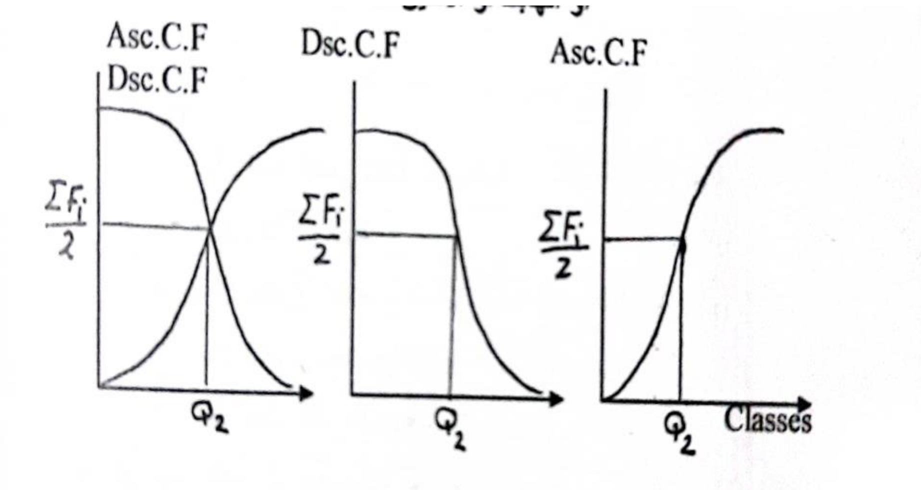
* $L.B.of\,(Q_2)Class$ تمثل الحد الادنى للفئة الوسيطية ،
* $R.\ of\,(Q_2)$ تمثل رتبة الوسيط ،
* $F_{(pre.Asc)}$ تمثل التكرار المتجمع السابق لرتبة الوسيط ،
* $F_{(suc.Asc)}$ تمثل التكرار المتجمع اللاحق لرتبة الوسيط ،
* $Length\ of\,(Q_2)class$ تمثل طول الفئة الوسيطية .

هذا ويمكن ايجاد قيمة الوسيط بيانيا من خلال رسم المنحنى التكرارى المتجمع الصاعد أو الهابط . ثم يتم تحديد رتبة الوسيط(($R.of(Q_2$) وهى عبارة عن قسمة مجموع التكرارات($\sum Fi$)على (2) . ويتم تحديد موقع هذه الرتبة على المحور الرأسى للشكل البيانى الخاص بالمنحنى التكرارى المتجمع الصاعد أو الهابط . ومن هذا الموقع يتم رسم خط عمودى على المحور الرأسى وموازيا للمحور الأفقى إلى ان يقطع هذا العمود المنحنى التكرارى المتجمع فيتم منها رسم عمود على المحور الأفقى وموازيا للمحور الرأسى الى أن يقطع المحور الأفقى

فى نقطه والتى تحدد قيمة الوسيط . أما أذا تم رسم المنحنيان التكراريان المتجمعان الصاعد والهابط معا فى شكلا بيانيا واحدا. فإن إحداثيات نقطة تقاطع المنحنيان تعطى قيمة الوسيط على المحور الأفقى وتعطى رتبة الوسيط على المحور الرأسى . أى أنه من نقطة تقاطع المنحنيين المتجمعين إذا تم اسقاط عمودا على المحور الافقى تحدد قيمة الوسيط ، أما اذا تم منها اسقاط عمودا على المحور الرأسى فهى تحدد رتبة الوسيط. والشكل البيانى التالى يوضح كيفية تحديد قيمة الوسيط بيانيا .

شكل (3−1)

ايجاد قيمة الوسيط بيانيا من خلال رسم المنحنى التكرارى المتجمع الصاعد أو الهابط أو الاثنين معا



<u>مثال (12)</u> : فيما يلى التوزيع التكرارى التالى يوضح الاجر اليومى (بالجنيه) لمجموعة من العاملين بإحدى الشركات :

جدول (3−14)

Classes	8-	18-	28-	38-	48-	58-68	$\sum$
Frequency	8	22	25	20	13	12	100

<u>والمطلوب</u> : أوجد قيمة الوسيط حسابيا وبيانا

الحل :

لايجاد قيمة الوسيط يجب بداية ايجاد الجدول التكرارى المتجمع الصاعد أو الهابط . دعنا نقوم بتكوين التوزيع التكرارى المتجمع الصاعد على النحو المبين بالجدول (3−15) التالى :−

جدول(3−15)

التوزيع التكرارى المتجمع الصاعد للأجر اليومى للعاملين بالشركة

Classes	F_i	Less than the Class U.L	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
∑	100	Less than68	100

ثم يتم تحديد رتبة الوسيط حيث ان :

$$\text{Rank of } (Q_2) = \frac{\sum^{r} F_i}{2} = (100/2) = 50$$

لاحظ أن الرتبة 50 تفيد أن الفئة الوسيطية هى الفئة من 28 حت أقل من 38 ومن ثم فإن الحد الادنى للفئة الوسيطية هو L.B = 28 والتكرار المتجمع السابق لرتبة الوسيط هو (30) وكذلك فإن التكرار المتجمع اللاحق لرتبة الوسيط هو (55) وطول الفئة الوسيطية هو (10) جنيه ومن ثم فبالتعويض فى العلاقة (14) السابقة فتكون قيمة الوسيط على النحو التالى :

$$Q_2 = 28 + \left(\frac{50 - 30}{55 - 30}\right) \times 10$$

$$= 28 + (20/25) \times 10 = 36 \quad (L.E)$$

أى أن وسيط الأجر اليومى للعامل فى هذا التوزيع هو 36 جنيه .

ملحوظة : يمكن ايجاد قيمة الوسيط حسابيا بإستخدام قاعة النسبة والتناسب السابق الاشارة اليها وذلك بعد تحديد رتبة الوسيط فيكون لدينا الرسم الكروكى التالى :-

Less than 28 (L.E) $\longrightarrow$ 30 persons
Less than (Q2) (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 :

$$5 = \frac{10}{38 - Q_2} \quad , \text{ i.e., } 5(38 - Q_2) = 10 \;,\; 190 - 10 = 5Q_2$$

فتكون قيمة الوسيط هى :

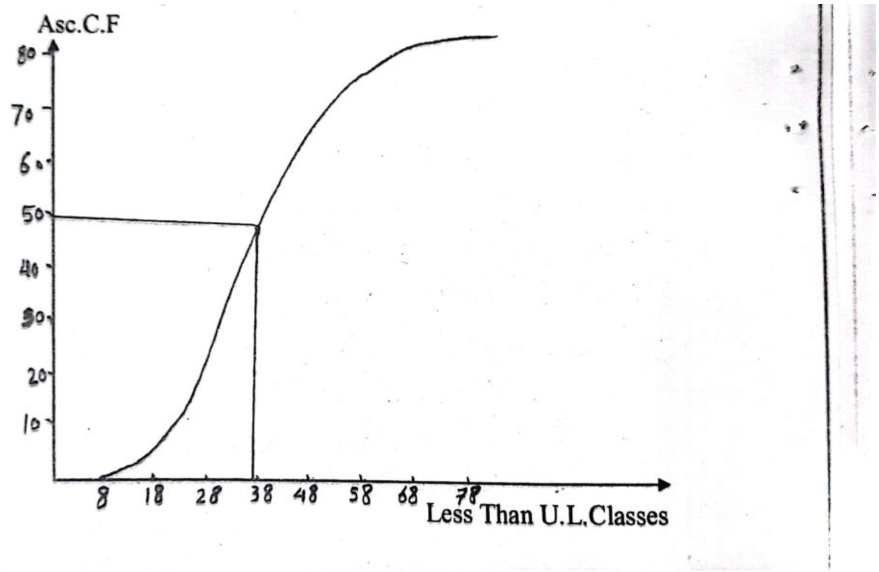
$$Q_2 = \frac{180}{5} = 36 \quad (L.E)$$

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 كما هو موضح بالشكل (3 – 2) تاركين للقارئ تحديد الوسيط سواء بإستخدام المنحنى التكرارى المتجمع الهابط أو تحديد الوسيط بإستخدام رسم المنحنيين التكراريين المتجمعين الصاعد والهابط معا .

شكل (3−2)

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 ما يلى :

(1):يتميز الوسيط بأنه لايتأثر بالقيم الشاذة (Outliers) أو المتطرفة (Extreme Points) لأنه 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 التكرارية الملتوية أوشديدة الالتواء .

(2):يمكن ايجاد الوسيط فى حالة التوزيعات التكرارية المفتوحة مالم يكن التوزيع شديد الالتواء بحيث لا تقع رتبة الوسيط فى الفئة الاولى فى حالة الجداول التكرارية المفتوحة من أعلى ، اوالاخيرة من التوزيع فى الفئة الأخيرة فى حالة الجداول التكرارية المفتوحة من أسفل .

(3):يمكن ايجاد الوسيط بيانيا من خلال رسم المنحنى التكرارى المتجمع الصاعد او المتجمع الهابط أو الاثنين معا .

(4):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:

$$\sum_{i=1}^{n} | x_i - Q2 | < \sum_{i=1}^{n} (x_i - A)$$

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 :

(1): من أهم عيوب الوسيط أنه يعتمد على قيمة واحدة تقع فى منتصف المفردات وذلك بعد ترتيبها ولايدخل فى حسابه باقى المفردات كما هو الحال فى الوسط الحسابى .

(2): كذلك يعاب على الوسيط أنه لايخضع للعمليات الجبرية خضوع الوسط الحسابى . لذلك ترانا تجاوزنا الدقة التامة عند تحديد رتبة الوسيط فى حالة التوزيعات التكرارية فأعتبرنا أن رتبة الوسيط هى ناتج خارج قسمة مجموع التكرارت على (2) سواء كان هذا المجموع ($\sum Fi$) عددا زوجيا أو فرديا .

(3): 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 لكل عينة.

(3-2-3) شبيهات الوسيط :

هناك بعض المقاييس الاحصائية تتشابه مع الوسيط فى تعريفها وفى طريقة حسابها وتتلحص هذه المقاييس فيما يلى :

- الربيعين **Quartiles**
- العشيرات **Deciles**
- المئينيات **Percentiles**

وسوف نتناول دراسة تلك المقاييس بشئ من التفصيل على النحو التالى :

أولا : الربيعين (الربيع الادنى والربيع الاعلى):-

والمقصود بالربيعين هما الربيع الادنى أو مايسمى بأسم الربيع الاول (First(1st Quartile والربيع الأعلى أو الربيع الثالث Third(3rd) Quartile.

هذا ويمكن تعريف الربيع الادنى (او الاول أو (Q_1)) بانه القيمة التى يقل عنها ربع المفردات وكذلك القيمة التى يزيد عنها ثلاثة أرباع المفردات . أما الربيع الاعلى (أو الثالث أو (Q_3)) فيمكن تعريفه بأنه القيمة التى يقل عنها ثلاثة أرباع القيم ويزيد عنها ربع القيم . ومن خلال هذا المنطلق فأنه يتم تعريف الوسيط بأنه بمثابة الربيع الثانى أى (Q_2) حيث ان الوسيط هو القيمة التى يقل عنها نصف (أو ربعى) القيم ويزيد عنها النصف الاخر (أو الربعين الآخرين) من تلك القيم .

هذا ويتشابه الربيعيين الادنى والاعلى مع الوسيط فى طريقة ايجادهما سواء كان ذلك بالطريقة الحسابية (سواء من خلال قانون الوسيط او قاعدة النسبة والتناسب) أو بالطريقة البيانية وذلك من خلال رسم أحد المنحنيين التكراريين المتجمعين الصاعد أو الهابط أو المنحنيين معا كما سبق وأن أوضحنا عند تعرضنا للوسيط. ويمكن توضيح ذلك كما يلى :

(أ):إيجاد الربيعيين (Q_1 أو Q_3) حسابيا :

لايجاد قيمة الربيعيين الأدنى والأعلى حسابيا فأنه يمكن استخدام أما القانون الوارد بالعلاقة رقم (14)السابق تناولها والمستخدمة فى حساب الوسيط مع

اختلاف مسمى المقياس الذى يتم حسابه وذلك على النحو التالى : فمثلا يتم حساب قيمة الربيع الأول من خلال المعادلة التالية :

$$Q_1 = L.B\ of\ (Q_1)class + \left(\frac{R.of(Q_1)-F_{(pre.Asc)}}{F(suc.Asc)-F(pre.Asc)} \right) \times length\ of\ (Q_1)class \qquad (15)$$

أما الربيع الثالث فيتم حسابة من المعادلة التالية :

$$Q_3 = L.B\ of\ (Q_3)class + \left(\frac{R.of(Q_3)-F(pre.Asc)}{F(suc.Asc)-F(pre.Asc)} \right) \times length\ of\ (Q_3)class \qquad (16)$$

هذا ويمكن استخدام قاعدة النسبة والتناسب السابق الاشارة اليها فيما سبق 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 أوالاعلى على المحور الرأسى للشكل البيانى للمنحنى التكرارى المتجمع الصاعد مثلا . ثم يتم رسم خطين موازيين للمحور الافقى إلى أن 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) .

<u>ملحوظه</u>:- فى حالة استخدام المنحنى التكرارى المتجمع الهابط لاحظ أنه من خلال دراستنا لمفهوم الربيعيين الادنى والاعلى فإن :

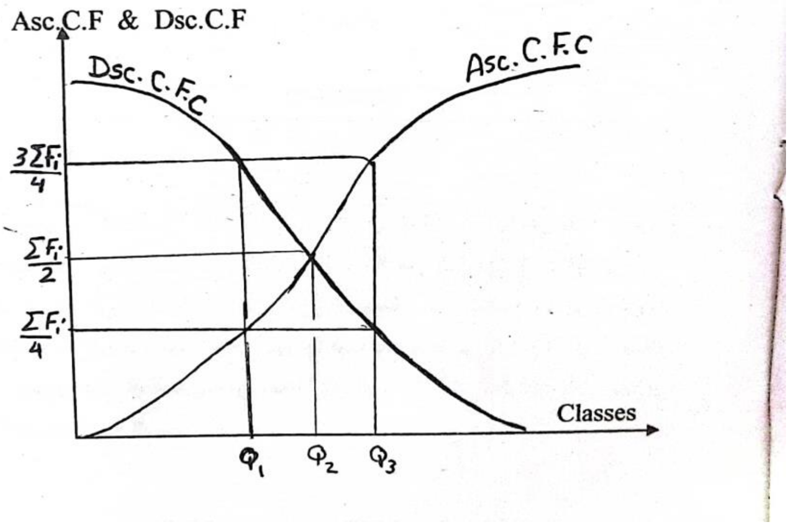
الربيع الادنى هو القيمة التى يقل عنها 25% من المفردات ويزيد عنها 75% من المفردات . أى انه إذ تم رسم المنحنيين التكراريين المتجمعين الصاعد والهابط

معا وبتحديد موقع النقطة التى تعبر عن ترتيب (Q_1). فاذا تم منها رسم خطا مسقيما موازيا للمحورالافقى الى ان يقطع المنحنى التكرارى المتجمع الصاعد والهابط فى نقطة . 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) كما هو موضح فى الشكل البيانى (3−3) التالى :

شكل (3−3)

ايجاد الربيعين Q_3,Q_1

من خلال رسم المنحنيين المتجمعيين الصاعد والهابط معا



كذلك فإنه من نقطة ترتيب الربيع الاعلى ($3\sum Fi / 4$) 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 قيمتى 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 (3−4) فإذا تصورنا ترتيب المفردات تصاعديا مثلا فإن الشكل (3−4) يوضح التعريف الذى تناولناه سابقا لمعنى الربيعيين .

شكل (3−4)

نسبة المشاهدات أو المفردات التى تقع مابين الربيعيين

Q_3, Q_1

Ascending

$X_{(1)}$ Q_1 Q_2 Q_3 $x_{(n)}$

حيث أن هناك 25% من المفردات تقل قيمتها عن قيمة الربيع الادنى Q_1 وان هناك نفس النسبة اى 25% من المفردات تزيد قيمتها عن الربيع الاعلى Q_3 . وكذلك 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 قيمتى الربيعيين تمكننا من الحصول على معلومات أكثر مما لو استخدمنا قيمة الوسيط بمفرده .

<u>مثال (13):</u>

أوجد قيمة الربيعيين الادنى والاعلى حسابيا بالطرق المختلفة وبيانيا للتوزيع التكرارى المعطى فى المثال السابق .

<u>الحل :</u>

<u>أولا:ايجادة قيم الربيعيين Q_3, Q_1 حسابيا :</u>

(1):باستخدام قانون الوسيط مع إجراء التعديل اللازم عليه . فبعد الحصول على الجدول التكرارى المتجمع الصاعد كما هو مبين فى الجدول(3−15) . يتم تحديد رتبة كل ربيع على حده على النحو التالى :

$$\text{Rank of } (Q_1) = \frac{\sum Fi}{4} \times 1 = \frac{100}{4} = 25 \quad ,$$

$$\text{Rank of } (Q_3) = \frac{\sum Fi}{4} \times 3 = \frac{3 \times 100}{4} = 75 \quad ,$$

ومن ثم فيتطبق كل من القانونين(15)،(16) السابق الاشارة اليهما فيكون:

$$Q_1 = \text{L.B of } (Q_1)\text{class} + \left(\frac{\text{R.of}(Q_1) - F_{(Pre.Asc)}}{F_{(Suc.Asc)} - F_{(Pre.Asc)}} \right) \times \text{length of } (Q_1)\text{class}$$

$$= 18 + \left(\frac{25 - 8}{30 - 8} \right) \times 10 = 18 + (17/22) \times 10 = 25.727 \text{ L.E}$$

أما عن قيمة Q_3 فبالنظر لعمود التكرارات المتجمعة الصاعدة فى الجدول (3-15) يتضح لنا أن رتبة الربيع الاعلى (75)والتى تم تحديدها فى الخطوة السابقة ماهى إلا أحد التكرارات المتجمعة الصاعدة وبالتحديد امام الفئة (48) 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 :

$$Q_3 = 48 \quad \text{(L.E)}$$

(2):<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 أن ترتيب الربيع الأعلى يقع أمام أحد الفئات الموجودة فى جدول التوزيع التكرارى المتجمع .

وعلية فلإيجاد قيمة Q_1 لاحظ ان الرتبة (25) الخاصة بترتيب Q_1 تقع مابين التكرارين المتجمعيين الصاعديين 8 ، 30 ومن ثم فإن فئة الربيع الادنى المقابلة لتلك التكرارت المتجمعة هى من (18) جنيها الى أقل من (28) جنيها. أى أن :

Less than 18 (L.E) $\longrightarrow$ 8 persons

Less than (Q_1) (L.E) $\longrightarrow$ 25 persons

Less than 28 (L.E) $\longrightarrow$ 30 persons

ومن ثم فإن :

$$(28 - 18) / (28 - Q1) = (30 - 8) / (30 - 25)$$

وعليه فأن :

$$\frac{10}{28 - Q_1} = \frac{22}{5}$$

أى أن :

$$50 = 22 (28 - Q_1) \quad , \quad 22Q1 = 616 - 50$$

فيكون :

$$Q_1 = \frac{566}{22} = 25.272 \qquad (L.E)$$

وهى نفس النتيجة السابقة .

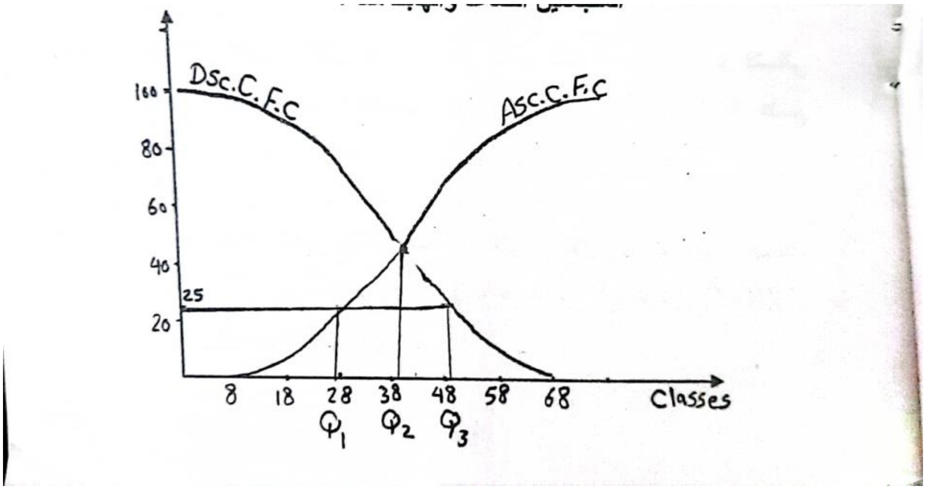
<u>ثانيا : إيجاد قيمة الربيعيين Q_3,Q_1 بيانيا</u> :

يمكن تحديد قيمة الربيعيين الأدنى والأعلى 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 الاول

والثانية ونقوم بإيجاد قيمة الربيعيين من خلال رسم المنحنيين معا كما هو موضح فى الشكل(3−5)

شكل (3−5)

ايجاد قيمة الربيعيين Q_3,Q_1 من خلال رسم المنحنيين المتجمعين الصاعد والهابط معا.



ثانيا : <u>العشيرات</u> :− Deciles

يمكننا تعريف العشير الاول First(1st) 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 حساب الوسيط والربيعيين الادنى والاعلى مع مراعاة انه عند تحديد رتبة العشير فإنه يتم قسمة مجموع التكررات على (10) ثم ضرب ناتج القسمة فى رتبة العشير المطلوب ايجاد قيمته . وعلى ذلك فإنه لإيجاد رتبة العشير الثالث مثلا يتم قسمة مجموع التكرارت ($\sum Fi$) على (10) ثم ضرب ناتج القسمة فى (3) والتى

تمثل رتبة العشير الثالث. ويتم استخدام القانون الوارد بالعلاقة (14) او (15)او (16) 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 .

مثال(14):

بإستخدام التوزيع التكرارى الوارد فى المثال (12) احسب العشير الثانى للأجر اليومى لمجموعة العاملين . ثم احسب العشير الخامس وعلق على نتيجتك التى حصلت عليها .

الحل:

لايجاد العشير الثانى والعشير الخامس للتوزيع التكرارى يتم بداية ايجاد الجدول التكرارى المتجمع الصاعد (كما هو موضح فى الجدول (3−15)) ثم يتم تحديد رتبة العشير المطلوب حسابه . فحيث أن :

رتبة العشير الثانى هى :

$$\text{Rank of } (D_2) = \frac{\sum F_i}{10} \times 2 = \frac{100}{10} \times 2 = 20 \,,$$

كذلك فأن رتبة العشير الخامس هى :

$$\text{Rank of } (D_5) = \frac{\sum F_i}{10} \times 5 = \frac{100}{10} \times 5 = 50 \,,$$

لاحظ أن رتبة العشير الخامس هى نفس رتبة الوسيط . ومن ثم فنتوقع 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 (20)تقع مابين التكرارين المتجمعين الصاعدين 8 ، 30 فى جدول (3–15) وعليه فإن فئة العشير الثانى هى من 18 الى أقل من 28 جنيه . وعليه يكون لدينا الشكل الكروكى التالى :

Less than 18 (L.E) $\longrightarrow$ **8 persons**
Less than (D2) (L.E) $\longrightarrow$ **20 persons**
Less than 28 (L.E) $\longrightarrow$ **30 persons**

وعليه فأن :

$$\frac{30-8}{30-20} = \frac{28-18}{28-D_2}$$

فيكون :

$$22(28 - D_2) = 100$$

أى ان :

$$D_2 = \frac{516}{22} = 23.45 \qquad (L.E)$$

لاحظ أن قيمة العشير الثانى D_2 تقع مابين حدى الفئات المتجمعة الصاعدة أى مابين 18 ، 28 جنيه أى داخل فئة العشير السابق تحديدها . وهذا يعنى أن 20% من العمال تقل أجورهم عن 23.45 جنيه وأن 80% من العمال تزيد اجورهم اليومية عن 23.45 جنيه . اما العشير الخامس فحيث ان رتبته هى 50 وهى مساوية لرتبة الوسيط فستكون النتيجة مساوية لقيمة الوسيط كما سبق أن أوجدناها وهى قيمة 36 جنيه . أى أن :

$$D_5 = Q_2 = 36 \quad (L.E)$$

وهو مايفيد أن 50% من العمال تقل أجورهم عن 36 جنيه وأن 50% من العمال تزيد أجورهم عن 36 جنيه.

<u>ثالثا : المئينيات</u> :- Percentiles

بنفس الدلالة السابقة مع اختلاف المفهوم فيمكن تعريف المئين الاول P_1 بان القيمة التى يقل عنها 1% من عدد القيم ويزيد عنها 99% من باقى المفردات كما إن المئين السابع والعشرون مثلا P_{27} وهو عبارة عن القيمة التى يقل عنها 27% من المفردات أو القيمة التى يزيد عنها 73% من باقى المفردات . وهكذا يمكن تعريف العديد من المئينيات بنفس الاسلوب . ولايجاد قيم المئينيات يمكن ايضا استخدام قانون الوسيط الوارد بالعلاقة رقم (14) وذلك بعد اجراء التعديلات اللازمة عليه لكى يتفق ومفهوم المئين المطلوب . كما يمكنناحساب المئينيات بإستخدام قاعدة النسبة والتناسب وذلك بعد تحديد رتبة المئين فى تلك الحالة . كما يمكن حساب المئينيات بيانيا من خلال رسم المنحنى التكرارى المتجمع الصاعد او الهابط أو من خلال رسم المنحنيين معا .

<u>مثال (15)</u>: باستخدام التوزيع التكراري التالي للأجر اليومى ل50 عاملاً (بالجنيه) بإحدى الشركات المطلوب حساب المئين الثامن والستون ؟

جدول (3 – 16)

Classes	30-	40-	50-	60-	70-	80-	90-	∑
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 القيمة سواء من خلال قانون الوسيط

وشبيهاته أو قاعدة النسبة والتناسب أو بيانياً . وسوف نقوم بحسابه من خـــلال القانون تاركين للطالب التأكد من صحة النتائج باستخدام قاعدة النسبة والتناســب وكذا الطريقة البيانية .

جدول (3 – 17)

Classes	F_i	Less than the Class U.L	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
	50		

وحيث أن رتبة المئتين الثامن والستون هي :

Rank of $P_{(68)} = (\sum F_i / 100) \times 68 = 34$

والرتبة (34) تنحصر ما بين التكرارين المتجمعين الصاعدين 26 ، 37 . لذا فإن فئة المئتين الثامن والستون هي من (70) حتى أقل من 80 جنيه وعليه فإن :-

$$P_{68} = \text{L.B of } (P_{68})\text{Class} + \left(\frac{\text{R.of}(P_{68}) - F_{(\text{Pre.Asc})}}{F_{(\text{Suc.Asc})} - F_{(\text{Pre.Asc})}} \right) \times \text{length of Class}$$

$$= 70 + \frac{(34 - 26)}{(37 - 26)} \times 10$$

$$= 70 + (8 / 11) \times 10 = 77.27 \qquad (\text{L.E})$$

وهو ما يعنى أن 68% من إجمالي عدد العمال تقل أجــورهم الاســبوعية عــن 77.27 جنيه .

<u>إستخدام المئنيات</u> :-

لاشك أن مفهوم المئنيات يعتبر مفيد جداً في بعض النواحي التطبيقية ولنضرب على ذلك أمثلة من واقع حياتنا العملية فمثلا :-

(1):في حالة الطلاب المتقدمين لشغل أماكن المدن الجامعية (الإسكان الجامعي) في إحدى الجامعات . لو افترضنا أنه وفقاً لعدد الأماكن المتاحة فإن سيتم قبول 73% فقط من إجمالي عدد الطلاب المتقدمين وذلك باستخدام النسبة المئوية للنجاح كمعيار لقبول الطالب من عدمه .

وفي هذه الحالة فإن قيمة المئين السابع والعشرون أي $P_{(27)}$ هي الدرجة التي لو حصل عليهاالطالب أو أكبر منها فإنه يكون من المقبولين في المدينة الجامعية وفيما عدا ذلك يتم عدم قبول الطالب .

(2):في حالة الطلاب المتقدمين إلى مكتب التنسيق وكانت رغباتهم الأولى هي الالتحاق بإحدى كليات التجارة مثلاً فإذا كانت أعداد هؤلاء الطلاب المتقدمون هو 2000 طالباً على سبيل المثال . وكان إجمالي الأماكن المتاحة في هذه الكلية هو 1640 طالباً (أي 82 % من إجمالي المتقدمين) . فإنه في هذه الحالة يمكن تحديد درجة النجاح التي لو حصل عليها الطالب أو أكبر منها فإنه يتم قبوله وفيما عدا ذلك فإنه يتم رفض طلبه . وتحدد هذه النسبة بقيمة المئين الثامنة عشر أي $P_{(18)}$.

(3):إذا كان الهدف من إجراء أحد أبحاث ميزانية الأسرة هو تحديد قيمة الدخل الشهري الذي يقل عنه 67% (على سبيل المثال) من إجمالى عدد الأسر التي شملها البحث . فإن قيمة هذا الدخل الشهري تتحدد بإيجاد المئين السابع والستون أي $P_{(67)}$.

(4):المئين الثالث والخمسون $P_{(53)}$ هو القيمة المستهدفة إذا كان الهدف من إجراء إحدى الدراسات السكانية هو تحديد قيمة العمر الذي يقل عنه 53% (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 :

(1): 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 تصح العلاقة التالية :

$$\text{الوسيط } (Q_2) = \frac{\text{الربيع الأدنى } (Q_1) + \text{الربيع الأعلى } (Q_3)}{2}$$

ومدى الإنحراف (أي القرب أو البعد) عن قيمة الوسيط الحقيقية والقيمة الناتجة من هذه العلاقة يعطي مؤشرًا عن مدى تماثل او قرب أو بعد التوزيع التكراري عن التماثل . وهو ما سنتعرض له بالدراسة عند دراستنا لمقاييس التشتت والالتواء فيما بعد .

(2): 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 عبارة عن متوسط قيمتي الأزواج التالية من المقاييس :

– العشير الثاني والعشير الثامن – وكذلك العشير الرابع والعشير السادس .

– المئين الرابع والخمسون والمئين السادس والأربعون . وكذلك المئين الثامن والستون والمئين الثاني والثلاثون ....وهكذا .

فباعتبار الزوج الأول من هذه المقاييس فإنه يمكن التحقق من كلاً مـــن العشــير 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 تلــك المقــاييس للتوزيع التكراري التالي :

جدول (3-18)

التوزيع التكراري لدرجات عينة من الطلاب في

إحدى المواد بإحدى الكليات

Classes	10-	22-	34-	46-	58-	70-	82-	$\sum$
Frequency	4	8	13	15	13	8	4	65

(3):في حين أن الوسيط لا يساوي الوسط الحسابي للربيعين الأدنى والأعلــى أو الوسط الحسابي لأي زوج من أزواج العشيرات أو المئنيات إلا في حالة التوزيعات المماثلة فقط إلا أن العلاقات التالية صحيحة دائماً لكافة التوزيعات (المتماثلــة أو الملتوية) فيما يتعلق برتبة تلك المقاييس . فمثلاً : في أي توزيع تكراري نجــد 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 كمــا يلــي : فحيث أن :

$$\frac{\text{Rank of } (Q_1) + \text{Rank of } (Q_3)}{2} = \frac{\sum F_i + 3\sum F_i}{4 \times 2}$$

$$= \frac{4 \times \sum F_i}{8}$$

$$= \frac{\sum F_i}{2} = \text{Rank of } (Q_2)$$

(4):في حالة التوزيعات التكرارية النسبية فإن رتبة الوسيط هي (0.5) أو 50 % سواء باستخدام التوزيع التكراري المتجمع الصاعد أو 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 أو 30 % والمئين الستون ترتيبه هو 0.6 أو 60 % وهكذا .

(3-3) المنوال : Mode

بصفة عامة يمكن تعريف المنوال بأنه القيمة الأكثر شيوعاً أو تكراراً أي أنه هو القيمة التي تتكرر أكثر من غيرها . فمثلاً إذا كانت لدينا مجموعة من المفردات أى بيانات الغير مبوبة ولتكن درجات مجموعة من الطلاب هي 15 ، 12 ، 15 ، 20 ، 14 ، 15 ، 6 ، 8 ، 15 فإن القيمة المنوالية لدرجة هؤلاء الطلاب هي الدرجة 15 درجة وذلك لأنها تكررت أكثر من غيرها . في حين أن مجموعة الدرجات التالية : 15 ، 18 ، 10 ، 7 ، 5 ، 4 ، 3 لا توجد لها قيمة منوالية حيث لا توجد درجة متكررة أكثر من غيرها . أما مجموعة الدرجات 15 ، 8 ، 5 ، 5 ، 4 ، 3، 8 ، 5 فلها قيمتين متواليتين هما 5 ، 8 درجة حيث تكررت كل منهما عدداً متساوياً من مرات التكرار وهو ثلاث مرات لكل واحدة منهما .

وكمثال آخر إذا كان لدينا 500 شخصاً موزعين فى جدول تكرارى مزدوج حسب حالتهم التعليمية وممارستهم للأنشطة الرياضية كما يلي :-

جدول (3 – 19)

الحالة التعليمية / ممارسة الرياضة	متعلم	أمي
يمارس الأنشطة الرياضية	110	95
لا يمارس الأنشطة الرياضية	195	100

فإن منوال هذا التوزيع التكراري المزدوج هو "الأشخاص المتعلمون ولا يمارسون الأنشطة الرياضية فهي الصفة السائدة وذلك لأن تكرارها 195 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 لأى مجموعة من البيانات قدلا 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 حساب قيمة المنوال من التوزيعات التكرارية (البيانات المبوبة) بشكل تقريبي بطريقة حسابية أو بالرسم شأنه في ذلك شأن الوسيط .

(3–3–1) إيجاد قيمة المنوال في حالة البيانات المبوبة تكرارياً :

إن عملية إيجاد المنوال لمجموعة من البيانات غير المبوبة (المفردات) تتمثـــل كما ذكرنا حالاً في مجرد البحث عن تلك المفردة التي تتكرر أكثر مـــن غيرهـــا . ومن ثم فهي تنطوي على قدر كبير من البساطة والوضوح . ولكن الأمـــر لـــيس 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 القيمـــة التـــي تتكرر أكثر من غيرها . وأفضل ما يمكن عمله في هذا الشأن لتحديد قيمة المنوال هو اختيار تلك الفئة التي تحتوي على أكبر تكرار أصلي (فـــي حالـــة الجـــداول التكرارية المنتظمة أي المتساوية من حيث أطول الفئات) أو أكبر تكرار معـــدل(وذلك في حالة الجداول التكرارية الغير منتظمة) وهي ما تسمى بالفئة المنواليـــة **Model Class** . وبعد ذلك يمكن تحديد قيمة المنوال داخل تلك الفئة المنوالية باستخدام عدة أساليب . وعليه 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 قيمـــة المنـــوال 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 والأعلـــى للفئـــة المنواليـــة وبالتالي يمكن تحديد مركز هذه الفئة فتكون هي بمثابة قيمـــة المنـــوال . 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 الأعلى للفئة المنوالية .

<u>مثال (16)</u> :-

الجدول التالي يبين توزيع أعمار 60 بطارية حسب مدة صلاحية كل منها بالشهور :

جدول (3−20)

Classes	6-	9-	12-	15-	18-21	$\sum$
Frequency	4	12	20	18	6	60

والمطلوب حدد قيمة المنوال مستخدماً طريقة الفئة المنوالية .

<u>الحل:</u>

بداية وبملاحظة طول الفئة في هذا التوزيع يتبـــين لنـــا أن الجـــدول التكـــرارى متساوي من حيث أطوال الفئات بمعنى أن التوزيع التكراري توزيعاً منتظماً لذا يتم حساب المنوال مباشرة . وبملاحظة قيم التكرارات نجد أن أكبر تكرار هو 20 لذا فإن الفئة المنوالية التي تقابل أكبر تكرار هي الفئة من (12) حتى أقل مـــن (15) شهر ومن ثم فإن المنوال هو عبارة عن مركز تلك الفئة.أى أن :

$$\text{Mode(M)} = \frac{(\text{L.B} + \text{U. B})\text{ for the Model Class}}{2}$$

$$= \frac{12 + 15}{2} = \frac{27}{2} = 13.5 \quad (\text{ month })$$

مستخدمين في ذلك طريقة مركز الفئة المنوالية كأحد طرق حساب المنوال .

ملحوظة :- إذا كان التوزيع التكراري غير منتظم فلابد من حســـاب التكـــرارات المعدلة أولاً . وتكون الفئة المنوالية هي الفئة المقابلة لأكبر تكرار معدل .

مثال (17) : الجدول التالي يبين التوزيع التكراري لدرجات مائة طالب في مـــادة الإحصاء :

جدول (3−21)

Classes	10-	-25	-30	-40	-50	-60	-80	90-100	∑
Frequency	5	7	13	21	19	16	10	9	100

والمطلوب : حساب المنوال مستخدماً طريقة مركز الفئة المنوالية .

الحل : حيث أن الجدول (3−21) توزيع تكراري غير متساوي من حيث طــول الفئة (توزيع تكرارى غير منتظم). لذا فإنه يجب حساب التكرارات المعدلة أولاً وهو ما يوضحه الجدول التالي حيث يتم حساب التكرارات المعدلة لهذا التوزيع .

جدول (3−22)

Classes	F_i	Length(L_i)	Modified Fre. (MF_i) = F_i/ L_i
10-	5	15	0.33
25-	7	5	1.4
30-	13	10	1.3
40-	21	10	2.1
50-	19	10	1.9
60-	16	20	0.8
80-	10	10	1.00
90-100	9	10	0.9
∑	100		

ومن خلال عمود التكرارات المعدلة يتضح لنا أن أكبر تكرار معدل هو (2.1) لــذا فالفئة المقابلة له (الفئة المنوالية) هي من (40) حتى أقل من (50) . لذا فــإن قيمة المنوال باستخدام طريقة مركز الفئة المنوالية هي :

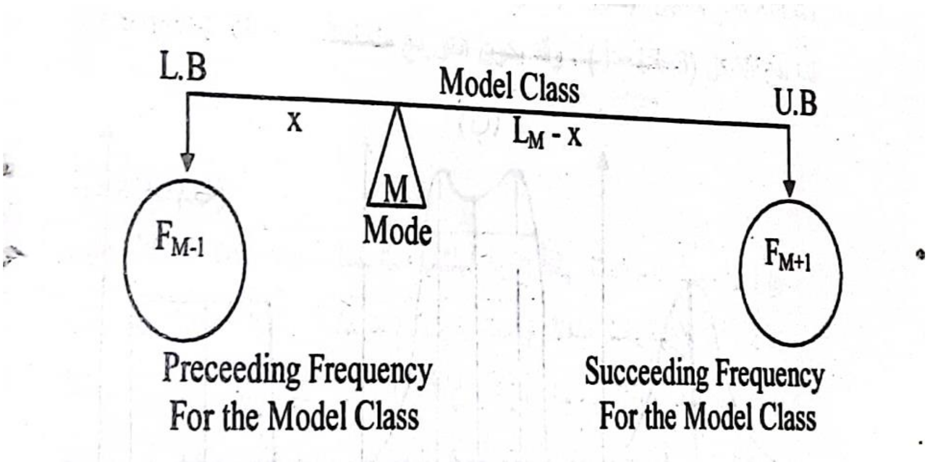
$$(M) = \frac{(L.B + u.B)}{2} = \frac{(40 + 50)}{2} = 45 \quad (degree)$$

<u>ثانياً: طريقة الرافعة لحساب المنوال</u> :

وهذه الطريقة تعتبر أن الفئة المنوالية بمثابة ذراع أفقي تتجاذبه قوتان(رافعة من الدرجة الأولى) أحدهما التكرار السابق للفئة المنوالية و الأخرى التكرار اللاحــق للفئة المنوالية . فإذا كان التكرار السابق للفئة المنوالية أكبر من التكرار اللاحــق للفئة المنوالية فإن قيمة المنوال تكون قريبة من الحد الأدنى للفئـــة المنواليـــة . 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 للفئـــة المنوالية . أما إذا تساوت القوتان أي تساوي التكرار الســـابق واللاحــق للفئـــة المنوالية فإن قيمة المنوال ستكون عند مركز الفئة المنوالية .

فإذا رمزنا لتكرار الفئة المنوالية بالرمز F_M ولتكـــرار الفئـــة الســـابقة للفئـــة المنوالية بالرمز $F_{(M-1)}$ ولتكرار الفئة اللاحقة للفئة المنواليـــة بـــالرمز $F_{(M+1)}$ 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 (3−6) . حيث يمثل هذا الشكل خطاً أفقياً يمثل رافعة طولها يساوي طول الفئة المنوالية أي (L_M) . فإذا اعتبرنا أن نقطة التوازن أو الارتكاز هي بمثابة نقطة المنوال(M). وبافتراض أن نقطة التوازن (أي قيمة المنوال) تبعد مسافة قـــدرها (X) عـــن بداية الفئة المنوالية فيكون بعد المنوال عن نهاية الفئة المنوالية ما هو إلا طـــول الفئة المنوالية أي L_M مطروحاً منها البعد (X) كمـــا هـــو موضـــح بالشـــكل الكروكى (3−6).

شكل (3−6)



هذا وبتطبيق قانون الرافعة والذي يحوي في مؤداه كما نعلم أن:−

القوة × ذراعها = المقاومة × ذراعها

فنستطيع الحصول على البعد (X) كالآتي :−

$$X \times (F_{M-1}) = (L_{M-x}) \times (F_{M+1}) \qquad \text{(I)}$$

أي أن

$$X \times (F(_{M-1}) + F(_{M+1})) = L_M \times (F_{M+1})$$

ومنها فإن :

$$X = \frac{F_{M+1}}{(F_{M-1} + F_{M+1})} \times (L_M) \qquad \text{(II)}$$

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 :

$$M = L.B \text{ for the Model Class} + X$$

أو باستخدام الرموز الواردة في العلاقة II فإن:

$$M = L.B \text{ (Model Class)} + \frac{F_{M+1}}{F_{M-1} + F_{M+1}} \times L_M \qquad \text{(III)}$$

هذا ويجب أن ننوه بالتأكيد على أنه قبل حساب المنوال بأي من طرق حسابه يجب التحقق من مدى انتظام الجدول التكرارى من عدمه . بمعنى إذا كان الجدول منتظماً يتم حساب المنوال مباشرة بأى من الطرق السابقة وإلا فيجب حساب التكرارات المعدلة (في حالة عدم الانتظام) ثم حساب المنوال بهذه الطرق .

وهذه الطريقة (طريقة الرافعة) وإن كانت تتميز بأنها تأخذ في الاعتبار التكرار السابق والتكرار اللاحق للفئة المنوالية إلا أن من أهم عيوبها أنها تتجاهل تكرار الفئة المنوالية ذاتها .

<u>مثال (18)</u> : احسب قيمة المنوال باستخدام طريقة الرافعة للتوزيع التكراري المعطى في مثال (16) .

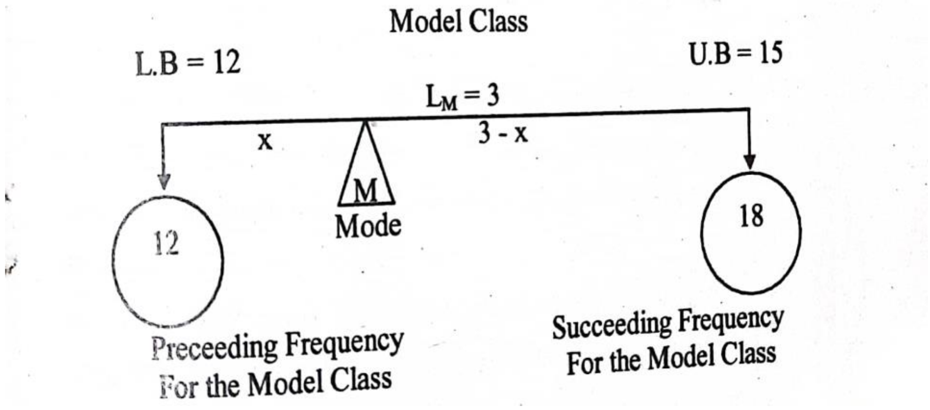
<u>الحل</u> :-

بداية حيث أن التوزيع التكراري الوارد في مثال (16) توزيعاً تكراريا منتظماً لذا يتم حساب المنوال مباشرة . وباستخدام طريقة الرافعة فإنه من خلال الجدول (3-20) يتضح لنا الآتي :

- **الفئة المنوالية** — **Model Class = (12-15)**
- **طول الفئة المنوالية** — $L_M = 3$
- **التكرار السابق للفئة المنوالية** — $F_{M-1} = 12$
- **التكرار اللاحق للفئة المنوالية** — $F_{M+1} = 18$
- **الحد الأدنى للفئة المنوالية** — **L . B for the Model Class = 12**

وبافتراض أن المنوال (M) هو نقطة التوازن لرافعة من الدرجة الأولى ويبعد مسافة قدرها (X) عن بداية الفئة المنوالية فنكون بصدد الشكل التالي التوضيحى التالى شكل (3 – 7) :

شكل (3 – 7)



فيكون :

$$12X = 18(3 - x)$$
$$30X = 54$$
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 (19)</u> :- احسب قيمة المنوال باستخدام طريقة الرافعة للتوزيع التكـــراري المعطى في مثال (17) .

<u>الحل</u> :-

بداية حيث أن التوزيع التكراري المعطى في مثال (17) توزيعاً تكرارياً غير منتظم لذا يجب حساب التكرارات المعدلة أولاً قبل حساب قيمة المنوال .

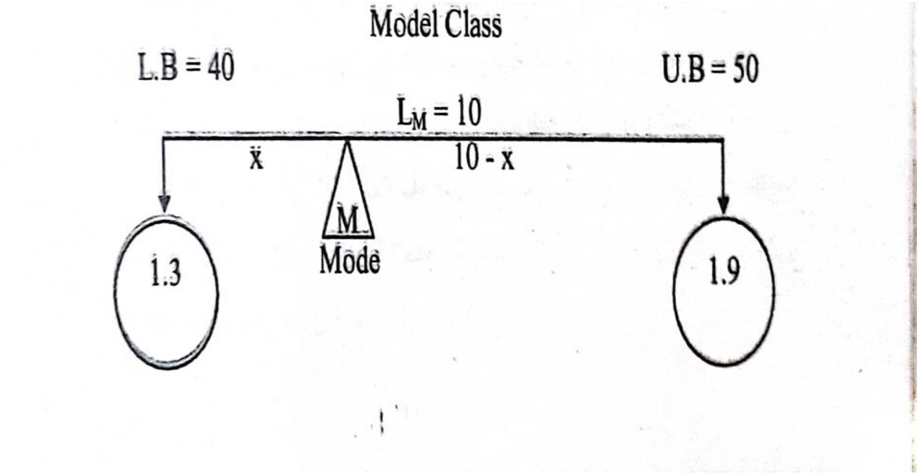
وعليه يجب استخدام بيانات الجدول (3 – 22) في تحديد البيانـــات اللازمـــة لتحديد قيمة المنوال . ومن هذا الجدول فإن :

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 صورة الرافعة في الشكل (3−8) . فبافتراض أن المنوال (M) هو نقطة التوازن لرافعة من الدرجة الأولى ويبعد مسافة قــدرها (x) عــن بداية الفئة المنوالية فنكون بصدد الشكل التالى:

شكل (3 − 8)



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 (3−22) فيمكن استنتاج أن قيمة المنوال لابد وأن تزيد عــن 45 درجة (وهو بمثابة مركز الفئة المنوالية) وذلك لأن التكرار المعدل اللاحق للفئة المنوالية أكبر من التكرار المعدل السابق للفئة المنوالية .

<u>ثالثاً : طريقة الفروق لبيرسون لحساب المنوال</u> :-

قلنا سابقاً أن من أهم ما يؤخذ على طريقة الرافعة بأنها تأخذ في الاعتبار كلاً من بداية الفئة 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 المنوال($F\ (_{M-1})$)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 الفرق الأول D_1 هو عبارة عن القوة أي أن :

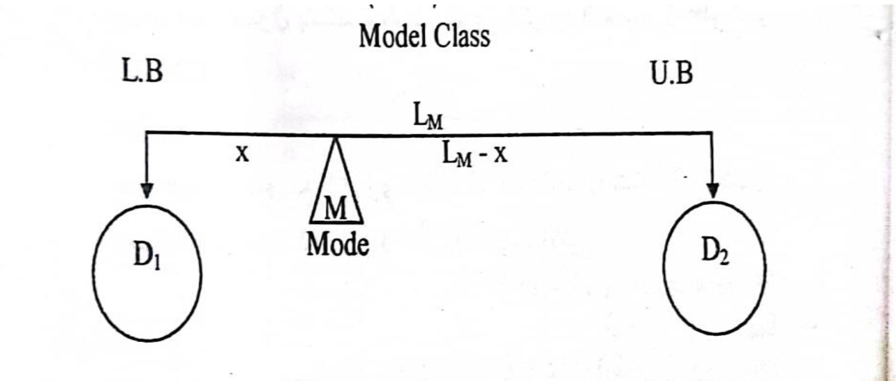
$$D_1 = F_M - F_{M-1}$$

وأن الفرق الثاني D_2 هو عبارة عن المقاومة أي أن : -

$$D_2 = F_M - F_{M+1}$$

فتأخذ الرافعة الشكل البياني التالي :-

شكل (3-9)



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 أن :

$$M = L . B \text{ for } (\text{Model Class}) + x$$

$$= L . B (M) + L_M \times \left(\frac{D_1}{D_1 + D_2} \right) \qquad (Iv)$$

والأمثلة التالية توضح كيفية حساب المنوال باستخدام طريقة الفروق لبيرسون.

<u>مثال (20)</u> : احسب قيمة المنوال باستخدام طريقة الفروق للتوزيـــع التكـــراري المعطى في مثال (16) .

<u>الحل</u> :-

بداية حيث أن التوزيع التكراري منتظماً لذا يمكن حساب المنوال مباشرة. فمـــن خلال الجدول (3−20) 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 السابقة (IV) فإن قيمة المنوال تكون عبارة عن :

$$M = 12 + 3 \left(\frac{8}{(8 + 2)} \right)$$

$$= 12 + 2.4 = 14.4 \qquad (\text{ month })$$

<u>مثال (21)</u> : احسب قيمة المنوال باستخدام طريقة الفروق للتوزيع التكراري المعطى في مثال(17) .

<u>الحل</u> :-

بملاحظة التوزيع التكراري الوارد في الجدول (3−21) نجد أنه توزيعاً تكرارياً غير منتظم لذا يجب الحصول على التكرارات المعدلة قبل حساب قيمة المنوال . لذا فمن بيانات التكرارات المعدلة الموضحة في جدول (3−22) يتضح لنا أن أكبر تكرار معدل هو (2.1) وعليه فإن :

التالية :

Model Class = (40 - 50)

$L_M = 10$

$D_1 = F_M - F_{M-1} = 2.1 - 1.3 = 0.8$

$D_2 = F_M - F_{M+1} = 2.1 - 1.9 = 0.2$

وعليه يتم استنتاج المنوال من الصيغة التالية باستخدام طريقة الفروق لبيرسون والتي تأخذ الصورة :

$$M = L\,.\,B \text{ for (Model Class)} + L_M \times \left(\frac{D_1}{(D_1 + D_2)} \right)$$

$$= 40 + 10 \left(\frac{0.8}{(0.8 + 0.2)} \right)$$

$$= 40 + 8 = 48 \qquad \text{(degee)}$$

<u>ملاحظات</u> :-

(1): 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 التكرارية المتماثلة أو التوزيعات التي تتساوى فيها التكرار الأصليةالسابقة (في حالة الجداول المنتظمة) واللاحقة للفئة المنوالية أو فى حالة تساوى التكرارات المعدلة السابقة واللاحقة للفئة المنوالية (في حالة الجداول الغير منتظمة) .

(2): يمكن القول أن طريقة مركز الفئة المنوالية هي أقل الطرق الثلاثة التي تحدد قيمة المنوال من حيث الدقة .وأن طريقة الفروق لبيرسون تفضل علــى طريقـــة الرافعة من حيث تقديرها لقيمة المنوال وذلك لأن طريقة الفروق تستخدم معلومات أكثرمن تلك التي تستخدمها طريقة الرافعة في تحديدها لقيمة المنوال .

(3): إذا تساوى التكرار السابق للفئة المنوالية أي (F_{M-1}) مع التكــرار اللاحــق للفئة المنوالية أي (F_{M+1}) فإن بعد المنوال عن بداية الفئة المنوالية (أي قيمـــة (x) سوف يأخذ نفس القيمة في الطرق الثلاث . وبالتالي فإن كل من هذه الطرق سيعطي نفس القيمة للمنوال وتكون قيمة x في هذه الحالة عبـــارة عــن خـــارج قسمة طول الفئة المنوالية أي L_M على (2) أى مركز الفئة المنوالية . وبمعنى أكثر بساطة إذا تساوى التكرار السابق واللاحق للفئة المنوالية فإن قيمة المنـــوال للطرق الثلاث تكون متساوية وتقدر قيمة المنوال في هذه الحالة بمركـــز الفئـــة المنوالية . أي أنه إذا كان :

$$F_{M-1} = F_{M+1}$$

فإن:

$$\text{Mode} = \left(\frac{(L.B + U.B) \text{ for the Model Class}}{2} \right)$$

<u>(2-3-3) إيجاد قيمة المنوال بيانياً من خلال المدرج التكراري</u> :

رأينا فيما سبق أن حساب المنوال يعتمد على محددات رئيسية وهي الفئة المنوالية وتكرار الفئة المنوالية وتكرار الفئة السابقة للفئة المنوالية وتكرار الفئة اللاحقة للفئة المنوالية . لذلك فإن لإيجاد المنوال يراعى أولاً تحديد طبيعة المتغير أو الظاهرة محل الدراسة . فإذا كان المتغير الإحصائي متغيراً منفصلاً وموزعاً توزيعاً تكرارياً فإن المنوال يكون عند تلاقي العمود الذي له أعلى تكرار في فئات أو أقسام الظاهرة مع المحور الأفقي ليبين قيمة الظاهرة التي تقابل أكبر تكرار(أي المنوال) .

أما في حالة المتغيرات الكمية المتصلة فإنه لتحديد قيمة المنوال بيانياً يتم رسم المدرج التكراري للتوزيع المعطى أو بالتحديد يكتفي رسم الفئات الثلاث التي تشكل أهمية عند تحديد المنوال وهي رسم الفئة المنوالية والفئة السابقة للفئة المنوالية و الفئة اللاحقة لها . وهنا يكون لدينا ثلاث مستطيلات متلاصقة أوسطها يعبر عن الفئة المنوالية والسابق لها يمثل الفئة السابقة للفئة المنوالية واللاحق لها يمثل الفئة اللاحقة للفئة المنوالية فيكون :

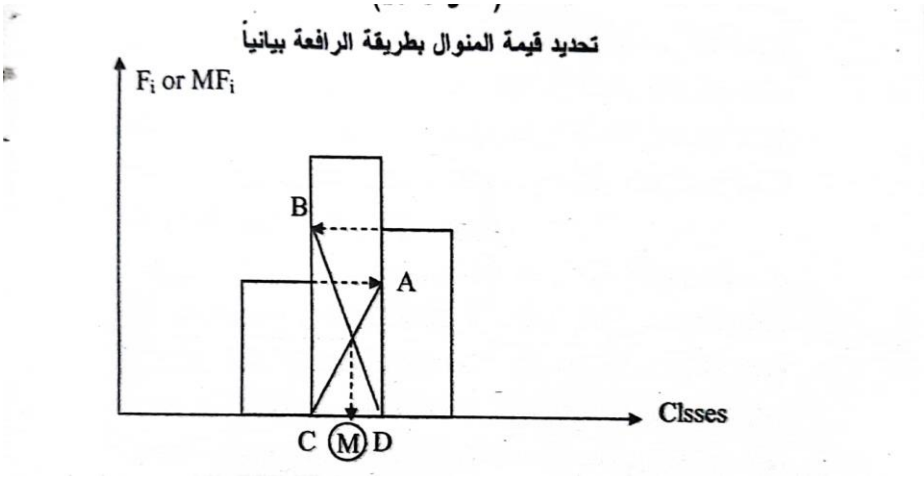
<u>(ا):وفقاً لطريقة الرافعة</u> :-

بعد رسم الفئات الثلاث محل الاهتمام في المدرج التكراري (الفئةالمنوالية والفئة السابقة والفئة اللاحقة للفئة المنوالية) حسب مقياس الرسم المناسب يتم استكمال قمم 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 (3−10) .

(شكل 3−10)

تحديد قيمة المنوال بطريقة الرافعة بيانياً

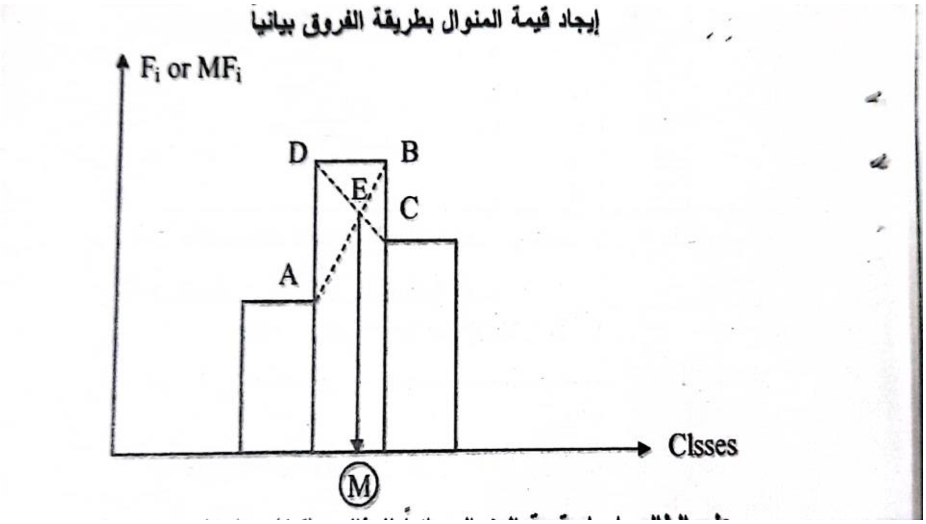


(ب):<u>وفقاً لطريقة الفروق</u> :−

يتم توصيل القمم العليا للمستطيلات الثلاثة المتلاصقة بخطين مستقيمين بطريقة عكسية . حيث يتم توصيل نهاية قمة المستطيل الممثل للفئة المنوالية بخط مستقيم مع نهاية قمة المستطيل الممثل لنهاية الفئة السابقة للفئة المنوالية . 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 (3−11) .

شكل (3−11)

إيجاد قيمة المنوال بطريقة الفروق بيانياً



وعلى الطالب إيجاد قيمة المنوال بيانياً للمثالين (16) ، (17) باستخدام كل من طريقتي الرافعة والفروق .

<u>مثال (22)</u> :

فيما يلي لديك الجدول التالي يعبر عن توزيع الدخل اليومي (بالجنيه) لمجموعة من العاملين بإحدى الشركات :

جدول (3− 24)

Daily Income Classes	5 -	10 -	20 -	40-	50-60
Frequency	30	150	240	50	20

والمطلوب إيجاد قيمة المنوال حسابياً بكافة الطرق المستخدمة لحسابه.

<u>الحل</u> :− بملاحظة الجدول التكراري المعطى نجد أنه توزيعاً تكرارياً غير منتظم لذا يجب الحصول على التكرارات المعدلة قبل حساب المنوال . والجدول (3−25) يعطي الحسابات اللازمة لحساب التكرارات المعدلة :

جدول (3−25)

Classes	F_i	L_i	Modified Frequency (MF_i)	
5 -	30	5	$30 \div 5 = 6$	F_{M-1}
10 -	150	10	15	F_M
20 -	240	20	12	F_M
40 -	50	10	5	
50-60	20	10	2	

ومن خلال الجدول (3−25) فإن حساب المنوال باستخدام :−

<u>أولاً : طريقة مركز الفئة المنوالية فإن</u> :

$$\textbf{Mode(M)} = \left(\frac{(\text{L . B} + \text{U.B}) \text{ for the Model Class}}{2} \right)$$

وحيث أن أكبر تكرار معدل هو (15) لذا فإن الفئة المنوالية هي الفئة :−

Model Class = (10- 20)

وعليه فإن قيمة المنوال باستخدام طريقة مركز الفئة المنوالية هى :

$$M = \frac{(10 + 20)}{2} = 15$$

<u>ثانياً : طريقة الرافعة فإن</u> :−

من خلال الجدول السابق يتضح لنا أن :

Model Class = (10- 20)

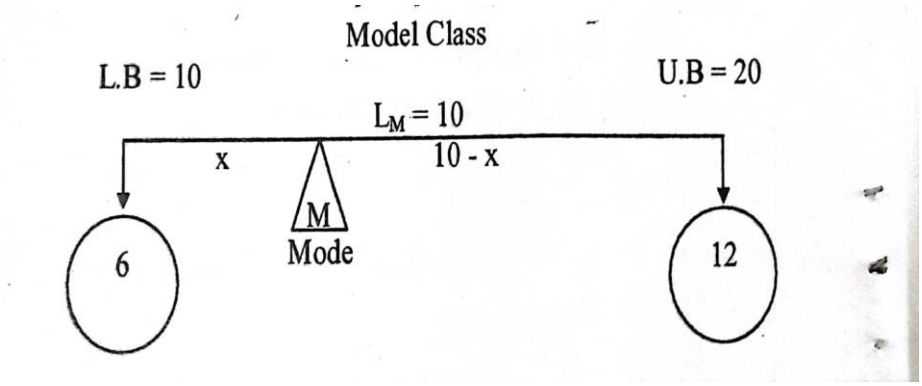
L . B for the Model Class = 10

$F_{M-1} = 6$

$F_M = 15$

$F_{M+1} = 12$

وعليه فإنه يمكن تصور الرسم الكروكي التالي :



وطبقاً لقاعدة الرافعة فإن :

6 x = 12 (10-x)

i.e., 18x = 120

وعليه فإن قيمة :

X = (120 / 18) = 6.667 (L.E)

فتكون قيمة المنوال عبارة عن :

M = L . B for the Model Class + x

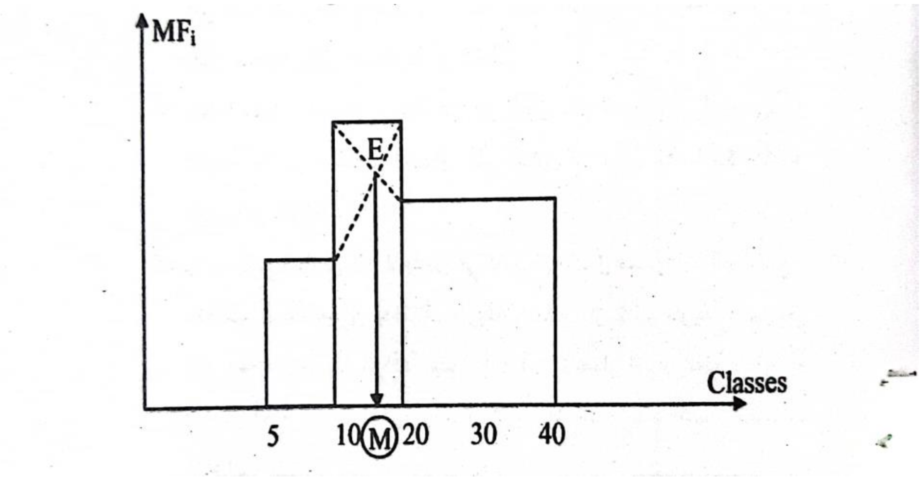
= 10 + 6.667

= 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 (3−12)

تحديد قيمة المنوال باستخدام طريقة الرافعة بيانياً



ومن الرسم البياني فإن قيمة المنوال هي تقريبا $M = 16.667$

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 :−

$$M = L.B \text{ for the Model Class} + L_M \left(\frac{D_1}{D_1 + D_2} \right)$$

$$= 10 + 10 \left(\frac{9}{9 + 3} \right)$$

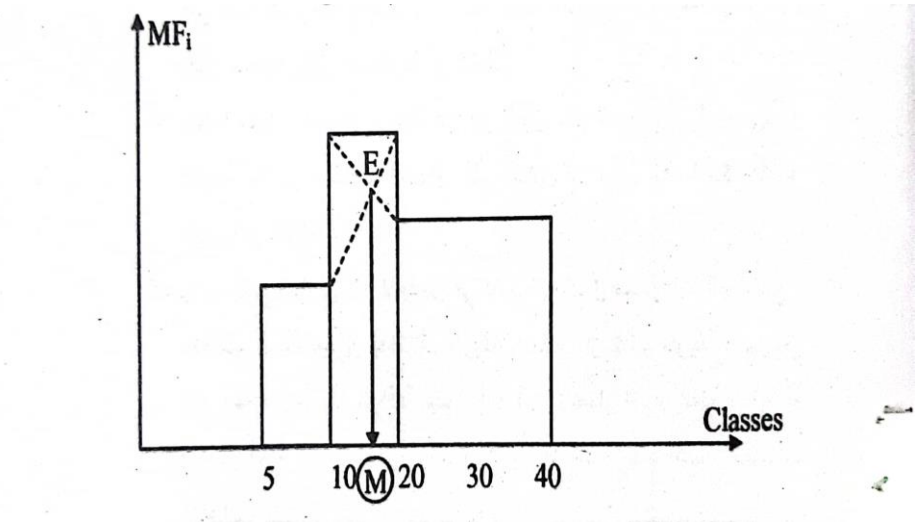
وعليه فإن :

$$= 10 + \frac{90}{12} = 17.5 \qquad (L.E)$$

أما عن تقدير المنوال بيانياً باستخدام طريقة الفروق فالشكل (3−13) يوضح إيجاد قيمة المنوال :

شكل (3−13)

تحديد قيمة المنوال باستخدام الفروق بيانياً



ومن الرسم يتضح أن المنوال تقريبا هو :

M = 17.5 L.E

<u>ملاحظات حول المنوال :</u>

(1):المنوال هو الصفة أو الظاهرة الغالبة في بيانات وصفية أو رقمية ، والصفة الغالبة تعني أنها الصفة التي تتحقق في عدد من المفردات والتي تصنيفها أو تكرارها يفوق عدد من المفردات المحققة لأية صفة أخرى . وهذا لا تعني بالضرورة أنها الصفة التى تتحقق في أغلبية العناصر أي في أكثر من 50 % منها .

(2): المنوال هو الصفة أو الظاهرة الأكثر تكراراً وليس تكرار تلك الظاهرة .

(3): يتميز المنوال بسهولة حسابه خاصة فى التوزيعات التكرارية، إذ أننا نحتاج إلى الفئة المنوالية والفئتين السابقة للفئة المنوالية واللاحقة لها ، ولا يتأثر

حسابه بالقيم المتطرفة أو الشاذة .

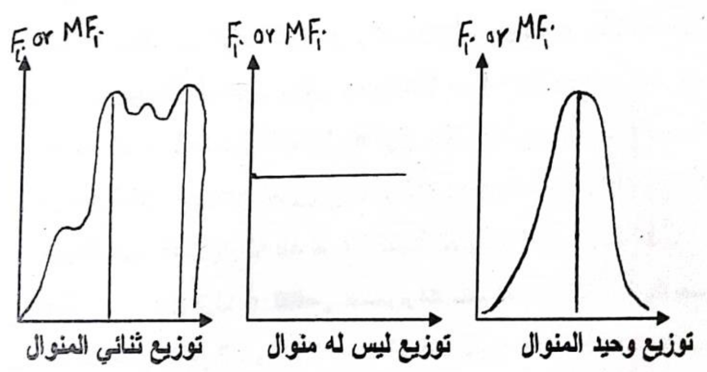
(4): حينما تكون القياسات نوعية كما هو الحال فى ظاهرة النوع أو الجنسية أو الديانة فيفضل استخدام المنوال على أساس أنه يمثل الظاهرة الأكثر ظهوراً أو تكراراً .

(5): 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 .

(6): يمكن إيجاد المنوال من المنحنى التكراري (أو المضلع التكراري) حيث يمثل المنوال نقطة النهاية العظمى للمنحنى (أو المضلع التكراري) ، فإذا كان لمنحنى (أو للمضلع) التكراري الذي يمثل توزيع ظاهرة ما نقطة نهاية عظمى وحيدة فيوجد للظاهرة منوال واحد كما في شكل (3–14 – 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 (3–14 – 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 (3 –14 – جـ) فهذا يعني وجود منوالين أو أكثر لهذا التوزيع كما هو في الحالتين الأخيرتين .

شكل (3−14)

(يوجد منوال وحيد أو لايوجد منوال أو يوجد أكثر من منوال وهنا لا يصلح المنوال لأن يكون مقياساً للنزعة المركزية).



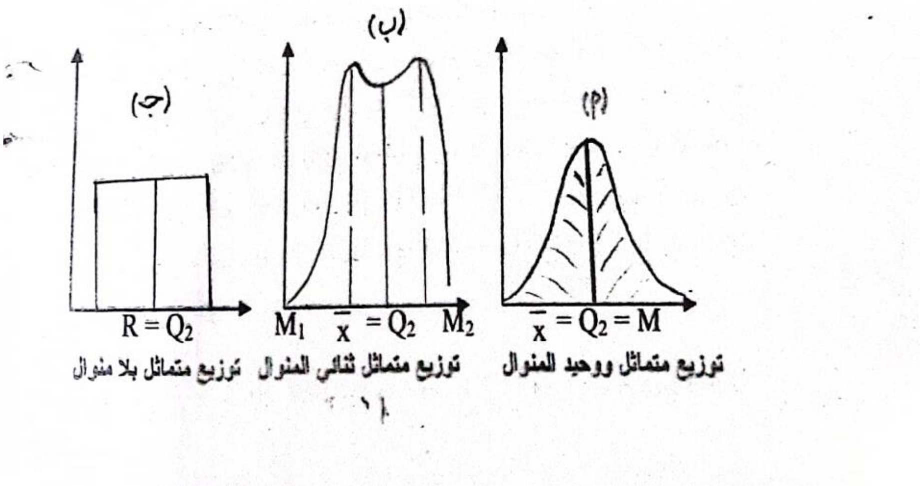
<u>(3−3−3) العلاقة بين الوسط الحسابي والوسيط والمنوال</u> :

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 مجموعة 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 كانت قيمة تلك المفردة ولكن الزيادة في الوسط الحسابي ستصبح أكثر كلما زادت فيه قيمة 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 .

هذا 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 المنـــوال ويسمى التوزيع فى هذه الحالة بالتوزيع المتماثـــل **Symetric Distribution** وهنا يكون المنحنى التكراري للتوزيع ينقسم عند القيمة المتوسطة له إلى قسمين متطابقين ويكون تزايد أو تناقص التكرارات متشابهاً ومنتظماً بطريقـــة متماثلـــة على جانبي المحور الرأسي المقام عند القيمة المتوسطة كما يتضح لنا من الشكل (3−15−أ) و يكون التوزيع متماثلاً فى هذه الحالة .

شكل (3−15)



أما التوزيعات التكرارية ثنائية المنوال فإنه في هذه الحالـــة يتســـاوى الوســـط الحسابي والوسيط بينما يوجد منوالين يختلفان عـــن قيمـــة الوســـط الحســـابي والوسيط كما يتضح من الشكل (3−1ب) . هذا وقد يكون التوزيع متمـــاثلاً بـــلا منوال فيظل الوسط الحسابي مساوياً للوسيط كما في الشكل (3−15−جـــ) .

وفي حالة التوزيعات الملتوية حيث تميل التكرارات الكبيرة إلى التركز في ناحية المتوسط وتنتشر التكرارات الصغيرة بعيداً عن شكل الذيل في الجانب الآخر منـــه ويكون اتجاه الذيل هو اتجاه الالتواء ، فإذا كان الذيل على اليسار قلنا أن التوزيع ملتو إلى اليسار أي سالب الالتواء **Negative Skewness** أما إذا كـــان الـــذيل على اليمين قلنا أن التوزيع ملتو إلى اليمـــين أي موجـــب الالتـــواء **Positive Skewness** وفي التوزيعات الملتوية يقع الوسيط دائماً بين الوســـط الحســـابي

والمنوال ، والوسط الحسابي كما رأينا شديد الحساسية للقيم الشاذة لـــذلك نـــراه مائلاً إلى اتجاه الذيل بينما يرشح المنوال نفسه عند قمة التوزيع .

فإذا كان التوزيع ملتوياً جهة اليسار (أي سالب الالتواء) حيث تميل التكـــرارات الكبيرة إلى التركز عند فئات التوزيع العليا ويكون اتجاه الذيل على اليســـار كمـــا يتضح من الشكل (3–16–أ) وهنا تكون العلاقة فيمـــا بـــين مقـــاييس النزعـــة المركزية الثلاثة كالآتي :

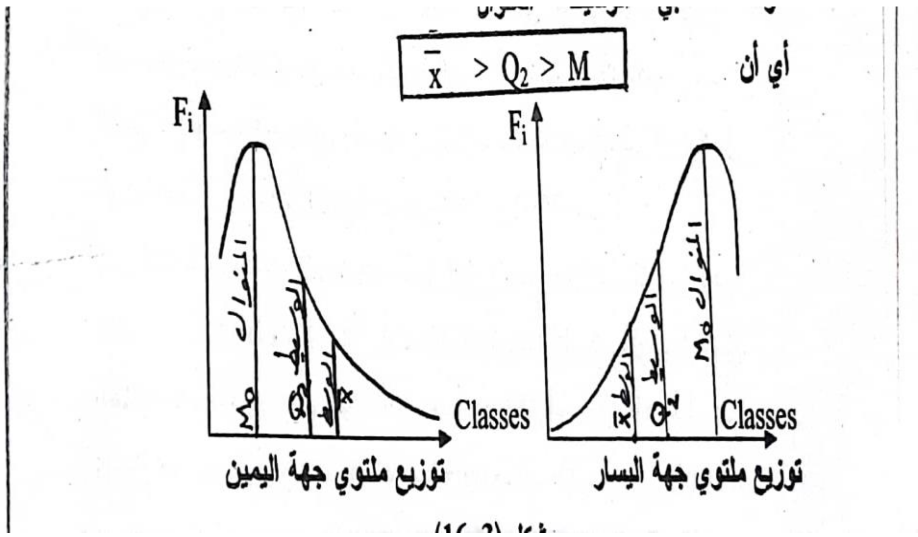
الوسط الحسابي < الوسيط < المنوال . أى أنه إذا كانت :

$M > Q_2 > \overline{X}$ فيكون التوزيع سالب الإلتواء.

أما إذا كان التوزيع ملتوياً جهة اليمين أي موجب الالتواء حيث تميل التكـــرارات الكبيرة إلى التركز عند فئات التوزيع الدنيا ويكون اتجاه الذيل جهة اليمـــين كمـــا يتضح من الشكل (3–7 – ب) فتكون العلاقة بين مقاييس النزعـــة المركزيـــة الثلاثة هي :

الوسط الحسابي > الوسيط > المنوال . أى أنه إذا كانت :

$M < Q_2 < \overline{X}$ فيكون التوزيع موجب الإلتواء.



وعلى ذلك فإن هناك علاقة بين قيمة مقاييس النزعة المركزيـــة الثلاثـــة وشـــكل التوزيع كالآتي : فإذا كان :

الوسط الحسابي – الوسيط   <   صفر   يكون التوزيع سالب الالتواء .

=   صفر   يكون التوزيع متماثلاً .

الوسط الحسابي – المنوال   >   صفر   يكون التوزيع موجب الالتواء.

وفي حالة التوزيعات القريبة جداً من التماثل تتحقق العلاقة الآتية بين المقـــاييس الثلاثة السابقة (الوسط الحسابى ، الوسيط ، المنوال) وهي :

الوسط الحسابي – المنوال = 3 (الوسط الحسابي – الوسيط)

وإذا استخدمنا الرموز الدالة على تلك المقاييس فإن :

$$\overline{X} - M = 3\,(\,\overline{x} - Q_2\,)$$

وتفيد هذه العلاقة في تقدير قيمة المنوال بطريقة ثالثة بالإضـــافة إلـــى طريقتـــي الرافعة والفروق وبصفة خاصة إذا كانت الفئة المنوالية هـــى الفئـــة الأولـــى أو الأخيرة فى التوزيع التكرارى . كما تمكن من تقدير قيمة مقربة للوسط الحســـابي في حالة التوزيعات التكرارية المفتوحة بعد معرفة قيمة الوسيط والمنـــوال لهـــذه التوزيعات بشرط أن تكون هذه التوزيعات قريبة من التماثل .

<u>مثال (2–20)</u>:

إذا علمت أن الجدول الآتي يمثل توزيع تكراري قريب من التماثل :

Classes	Less than 6	6 -	12-	18-	24and more	$\sum$
F_i	8	20	40	25	7	100

والمطلوب تقدير قيمة الوسط الحسابي .

<u>الحل</u> :

حيث أن الجدول التكراري مفتوح من أعلى ومن أسفل لذلك لا يمكن إيجاد الوسط الحسابي مباشرة ، وحيث أن التوزيع قريب من التماثل وهو مايتضح من توزيـــع التكرارات عبر فئات التوزيع حيث تتزايد تقريبا تكرارات التوزيع إلـــى أن تصـــل التكرارات لتصل لأكبر قيمة لها أمام الفئة التى تتوسط الجدول ثم تبدأ التكـــرارات

فى التناقص تكاد تكون بنفس معدل زيادة التكرارات فيما قبل الفئة التى تتوسط التوزيع . وعليه فالتوزيع يقترب من التماثل ، لذا فيمكن إيجاد قيمة كل من الوسيط والمنوال مستخدمين فى ذلك العلاقة السابقة (علاقة كارل بيرسون) لايجاد قيمة الوسط الحسابي وذلك على النحو المبين التالى :

فيتم إيجاد التوزيع التكرارى المتجمع الصاعد لحساب الوسيط :

Classes	F_i	Less than the Class U.L	Asc.C.F
Less than	8	Less than 6	8
6 -	20	Less than 12	28
12 -	40	Less than 18	78
18 -	25	Less than 24	93
24and more	7	Less than Latest U.L	100
$\sum$	100		

وعليه فلايجاد الوسيط فإن :

ترتيب الوسيط

$$\text{Ran of } (Q_2) = \frac{\sum F}{2} = \frac{100}{2} = 50$$

ومن ثم تكون قيمة الوسيط هي :

$$Q_2 = 12 + \left(\frac{50-28}{68-28}\right) = 15.3$$

أما لإيجاد المنوال دعن نستخدم طريقة الفروق ، فبالتطبيق على المثال يكون :

$F_M = 40$

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 :-

$$M = L.B \text{ for the Model Class } + L_M \left(\frac{D_1}{D_1 + D_2} \right)$$

$$= 12 + 6 \left(\frac{20}{20+15} \right) = 15.43$$

لاحظ قرب قيمتى الوسيط (15.3) والمنوال (15.43) وهو ما يؤكد أن التوزيع قريب جدا من التماثل لذا يصح تطبيق قاعدة بيرسون التالية:

$$\bar{X} - M = 3 \ (\bar{x} - Q_2) \quad \text{i.e.,} \quad \bar{x} - 15.43 = (3(\bar{x} - 15.3))$$

وبحل المعادلة ينتج أن قيمة الوسط الحسابى هى: $\bar{x} = 15.235$

هذا 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 :

1400 , 400 , 5000 , 7200 , 400 , 400 , 5600 , 400

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 . فلو أن مراقبًا للحسابات أراد أن يظهر الشركة بمظهر أنها تدفع رواتب متدنية جدًا في المتوسط لأختار المنوال وهو 400 جنيهًا كمقياسًا للنزعة المركزية ، في حين أن مدير الشركة سيختار الوسط الحسابي وهو 2600 جنيهًا كمقياسًا للنزعة المركزية ليثبت أن رواتب الشركة مرتفعة ، ولو أن باحثًا محايدًا أراد أن يعبر بموضوعية عن حقيقة

الرواتب بالشركة فسيختار الوسيط وهو 900 جنيهًا كمقياسًا للنزعة المركزيـــة . هذا ويظل الوسط الحسابي في أغلب الأحوال مقياسًا هامًا للنزعة المركزية يتمتع بخصائص جيدة تؤهله لأن يستخدم على نطاق واسع فـــي علـــم الإحصـــاء كمـــا سيتضح لنا فيما بعد . ولو أخذنا عينات مختلفة ذات أحجام متساوية من مجتمـــع الدراسة وحسبنا لكل عينة الوسط الحسابي والوسيط والمنوال لوجدنا أن التغييـــر من عينة إلى أخرى هو أقل في قيم الوسط الحسابي منه عن قيم كل من الوســـيط والمنوال ، مما يعني أن الوسط الحسابي أكثر استقرارًا من كـــل مـــن الوســـيط والمنوال عبر عينات نسحبها من مجتمع الدراسة .

<u>(3−4): الوسط الهندسي</u> Geometric Mean (G.M)

يعتبر الوسط الهندسي أحد مقاييس النزعة المركزية والـــذى يفضـــل اســـتخدامه عندما تكون قيم الظاهرة فى صورة نسب أو معدلات وكذلك عنـــد تقـــدير عـــدد السكان بين سنوات التعداد .

هذا ويعرف الوسط الهندسي لمجموعة من القيم عددها وليكن (ن)من المفـــردات بأنه عبارة عن الجذر النوني لحاصل ضرب هذه القيم أو هذه المفردات .

<u>(3−4−1): حساب الوسط الهندسي في حالة البيانات الغير المبوبة</u> :

إذا كان لدينا المتغير (X_i) يأخذ القيم X_1 ، X_2 ، …….. ، X_n الموجبة والغيـــر صفرية فإن الوسط الهندسي لهذه القيم وسنرمز له بالرمز (G.M) يأخذ الصورة الرياضية التالية :

$$(G.M) = \sqrt[n]{X_1 \,.\, X_2 \,.\, X_3 \ldots\ldots X_n}$$

$$= \sqrt[n]{\pi^n X_i}$$

$$\text{Log}\,(G.M) = {}^1/_n \sum^n \text{Log}\, X_i$$

وعلى ذلك فإن لوغاريتم الوسط الهندسي لمجموعة من القيم هو عبارة عن الوسط الحسابي للوغاريتمات هذه المجموعة من القيم . وعليه يكون الوسط الهندسي (G.M) هو العدد المقابل للمقدار Log(G.M) الذي نحصل عليه من جدول الأعداد المقابلة للوغاريتمات أو من خلال الآلة الحاسبة (العملية العكسية للوغاريتم) . وبصفة عامة فإن الوسط الهندسي لمجموعة من القيم دائمًا أقل من الوسط الحسابي ويمكن إثبات ذلك رياضيًا للتبسيط من خلال مفردتين ولتكن X_1 ، X_2 وذلك على النحو التالي :

الوسط الحسابي = $\bar{X} = \dfrac{X_2 + X_1}{2}$

أما الوسط الهندسي (G.M) فهو: $G.M = \sqrt{X_1X_2}$

وحيث أن :

$$(\sqrt{X_1} - \sqrt{X_2})^2 > 0$$

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 (3−21)</u> :

احسب الوسط الهندسي للنسب التالية :

43% , 12% , 7% , 6%

الحل :

$$G.M = \sqrt[4]{6 \times 7 \times 12 \times 43}$$

$$\text{Log (G.M)} = \tfrac{1}{4}(\text{Log } 6 + \text{Log } 7 + \text{Log } 12 + \text{Log } 43)$$
$$= \tfrac{1}{4}(0.7782 + 0.8451 + 1.0792 + 1.6335)$$
$$= \tfrac{1}{4}(4.336) = 1.084$$

وبالكشف في جدول الأعداد المقابلة للوغاريتمات نجد أن : $G.M = 12.13$

أى أن : $(G.M) = 12.13\%$

في حين أن الوسط الحسابي ($\bar{x}$) لهذه النسب = 17 % ، وهذا يتفق مع ما هو معروف من أن الوسط الهندسي لمجموعة من القيم الموجبة أصغر من الوسط الحسابي لها .

مثال (3−22) :

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 :

$$G.M = \sqrt{1500000 \times 2000000} = 1732050 \quad \text{person}$$

ومن هذا المثال يتضح أن الوسط الهندسي يكون أكثر دقة من الوسط الحسابي عندما تكون البيانات في صورة معدلات أو نسب .

(3−4−2) الوسط الهندسي في حالة البيانات المبوبة:

إذا كان لدينا توزيع تكراري وكانت (xi) تمثل مراكز فئات التوزيع ، Fi هي التكرارات المناظرة حيث (i= 1,2,….. , r) ، r هي عدد فئات التوزيع ، فإن الوسط الهندسي لهذا التوزيع هو :

$$G.M = \sqrt[\sum F_i]{(x_1)^{F_1} \cdot (x_2)^{F_2} \ldots\ldots (x_r)^{F_r}}$$

$$\text{Log}\,(G.M) = {}^{1}/_{\sum F_i}\,(F_1 \text{ Log } x_1 + F_2 \text{ Log } x_2 + \ldots\ldots + F_r \text{ Log } x_r)$$

$$= {}^{1}/_{\sum F_i}\,(\sum F_i \text{ Log } x_i)$$

مثال (3−24) :

احسب الوسط الهندسي للتوزيع التكراري التالي :

Classes	55 -	65 -	75 -	85-95	$\sum$
F_i	10	30	40	20	100

الحل :

لحساب الوسط الهندسي نكون الجدول التالي :

Classes	F_i	x_i	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

$$\text{Log(G.M)} = \frac{\sum F_i \log x_i}{\sum_1^r F_i} = \frac{188.343}{100} = 1.88343$$

ومن جدول الأعداد المقابلةللوغاريتمات أو من خلال العملية العكسية للوغاريتم على الآلة الحاسبة نجد أن الوسط الهندسى للتوزيع هو:

G.M = 76.38

<u>(5−3): الوسط التوافقي</u> : Harmonic Mean (H . M)

الوسط التوافقي يمثل أحد مقاييس النزعة المركزية ويفضل استخدامه في حالة ما إذا كان المتغير أو الظاهرة محل الدراسة فى صورة معدلات زمنية أو فى صـــورة معدل تغيير أو معدل السرعة بالنسبة للزمن ، ويعرف الوسط التوافقي لمجموعة من القيم بأنه مقلوب الوسط الحسابي لمقلوبات القيم .

<u>(1−5−3): الوسط التوافقي في حالة البيانات غير المبوبة :</u>

إذا كان لدينا مجموعة القيم $x_n, \ldots, x_2, x_1$ الغير صفرية فإن الوسط التوافقي والذى نرمز له بالرمز (H . M) يتم حسابه على النحو التالى :

$$(H.M) = \frac{n}{{}^1/_{x1} + {}^1/_{x2} + \ldots + {}^1/_{xn}} = \frac{n}{\sum {}^1/_{xi}}$$

مثال (3−25) :

احسب الوسط التوافقي والوسط الهندسي لمجموعة القيم التالية :

1 , 5 , 25 , 5 وماذا تستنتج ؟

الحل :

الوسط التوافقي هو :

$$(H.M) = \frac{4}{{}^1/_1 + {}^1/_5 + {}^1/_{25} + {}^1/_5} = \frac{4}{36/25} = 2.78$$

$$(G.M) = \sqrt[4]{1 \times 5 \times 25 \times 5} = \sqrt[4]{625} = 5$$

ومن هذه النتائج نستنتج أن الوسط التوافقي أصغر من الوسط الهندسي .

مثال(3−26) :

إذا كانت المسابقة 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 ، ومن المدينة(جـ)إلى المدينة (د)هي 30 كم قطعها القطار بسرعة 120 كم/ساعة . احسب متوسط سرعة القطار خلال الرحلة .

<u>الحل</u> :

إذا أردنا تقدير متوسط سرعة القطار في الساعة باستخدام الوسط الحسابي والوسط التوافقي لمعرفة أيهما أكثر ملائمة فإن :

<u>أ – باستخدام الوسط الحسابي :</u>

متوسط سرعة القطار في الساعة بإستخدام مفهوم الوسط الحسابى هو :

$$\overline{X} = \frac{80+90+120}{3} = 96.667 \quad K.M/hour$$

<u>ب– باستخدام الوسط التوافقي :</u>

متوسط سرعة القطار فى الساعة بإستخدام الوسط التوافقى هو :

$$\overline{X} = \frac{3}{1/_{80} + 1/_{90} + 1/_{120}} = 93.913 \quad K.M/hour$$

ولمعرفة أي المقياسين أكثر ملائمة فنجد 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 (جــ) في زمن قدره : 60 / 90 = 0.67 ساعة ، وقطع المسافة من المدينة (جــ) إلى المدينة(د) في زمن قدره : 30 / 120 = 0.25 ساعة وبالتالى فإن :

الزمن الإجمالي للرحلة = 0.5 + 0.67 + 0.25 = 1.42 ساعة.

المسافة الإجمالية للرحلة = 40 + 60 + 30 = 130 كم /ساعة .

وعليه فإن متوسط سرعة القطار في الساعة عبارة عن :

= 130 /1.42 = 91.55 كم/ ساعة

ومن خلال مقارنة متوسط سرعة القطارالمتحصل عليها أخيرا (91.55) وقيمتى الوسط الحسابى (96.667) والوسط التوافقى (93.913) يتضح لنا أن الوسط التوافقي أكثر ملائمة لقياس متوسط سرعة القطار من الوسط الحسابي .

(3-5-2) الوسط التوافقي في حالة البيانات المبوبة:

إذا كان لدينا توزيع تكراري وكانت (xi) هي مراكز الفئات للتوزيع ، (Fi) هي التكرارات المناظرة لها فيتم حسب الوسط التوافقي للتوزيع التكراري كما يلي :

- نوجد مقلوب مراكز الفئات ، أي (1 / xi) .
- نضرب مقلوب مركز كل فئة في التكرار المقابل له ، أي نوجد $Fi \times (1 / xi)$.
- نقسم $\sum_1^r Fi$ على $\sum_1^r (Fi/xi)$ فنحصل على الوسط التوافقي
- أي أن :

$$H.M = \frac{\sum_1^r Fi}{\sum_1^r (Fi/xi)}$$

مثال (3-27) :

احسب الوسط التوافقي للتوزيع التكراري التالي :

Classes	5 -	15 -	25 -	35-45	∑
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 -15 -25 35-45	20 50 90 40	10 20 30 40	1.0 05.0 0.033 0.025	2 2.5 3 1
∑	200			8.5

فيكون الوسط التوافقي عبارة عن :

$$H.M = \frac{\sum_1^r Fi}{\sum_1^r \left(^{Fi}/_{xi}\right)} = \frac{200}{8.5} = 23.53$$

ملاحظات حول الوسط الهندسي والوسط التوافقي:

(1):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 :

$$\overline{X} \geq G.M \geq H.M$$

(2):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 .

(3):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 (3−28) :

فيما يلي توزيع تكراري يبين درجة تلوث الهواء (مقاسه بالميكروجرام في المتر المكعب) في 80 مدينة كبيرة في العالم :

فئات درجة التلوث	10 −	20 −	30 −	40 −	50 −	60 −	80−70
عدد المدن	9	11	16	20	12	8	4

والمطلوب :

(1):حساب الوسط الحسابي والوسط الهندسي والوسط التوافقي والتحقق من أن الوسط الحسابي > الوسط الهندسي > الوسط التوافقي .

(2):حساب الوسيط والمنوال والربيعين الأدنى والأعلى بالحساب وبالرسم .

(3):ما هي نسبة المدن (في عينة الدراسة) التي تزيد فيها درجة التلوث عن 45 ميكروجرام / متر مكعب .

(4):ما هي درجة التلوث التي تكون عندها 30% من المدن في عينة الدراسة أكثر نظافة ؟ .

الحل :

(1):لحساب الوسط الحسابي والوسط الهندسي والتوافقي يتم حساب المجاميع اللازمة لهذه المقاييس والتى يوضحها الجدول التالي:

Classes	F_i	X_i	$D_i = x_i - 45$	$D_i = d_i/10$	$D_i F_i$	$\log x_i$	$F_i \log x_i$	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
$\sum$	80				-25		126.6198	2.3361

$$\overline{X} = A + \frac{\sum D_i F_i}{\sum F_i}) = 45 + 10(\frac{-25}{80}) = 41.875$$

وبفرض أن الوسط الهندسي هو G.M فإن :

$$Log(G.M) = \frac{\sum F_i \log x_i}{\sum F_i} = \frac{126.6198}{80}$$

ومن جدول الأعداد المقابلة للوغاريتمات أو من خلال العملية العكسية للوغاريتم بالآلة الحاسبة نجد أن :

$G.M = 38.26$ ميكروجرام/م^3

أما عن الوسط التوافقي فإن :

$$H.M = \frac{\sum F_i}{\sum_1^r ({}^{Fi}/_{xi})} = \frac{80}{2.3361} = 34.2451$$ ميكروجرام/م^3

ويتضح لنا من النتائج الثلاث السابقة أن الوسط الحسابي أكبر من الوسط الهندسي والوسط الهندسي أكبر من الوسط التوافقي .

(2):لحساب الوسيط والربيعين الأدنى والأعلى يلزم تكوين الجـدول التكـراري 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
∑	80		

ولحساب الوسيط أي Q_2 فإن :

$$\text{Rank of } (Q_2) = \frac{\sum F_i}{2} = \frac{80}{2} = 40$$

وعليه فإن قيمة الوسيط هى :

$$Q_2 = 40 + 10\,(40 - 36) / (56 - 36) = 42$$ **ميكروجرام/م^3**

ولحساب الربيع الأدنى (Q_1) فإن :

$$\text{Rank of } (Q_1) = \frac{\sum Fi}{4} = \frac{80}{4} = 20$$

وحيث أن الترتيب 20 يوجد مباشرة في عمود التكرار المتجمع الصاعد فيكون الربيع الأدنى هو القيمة التي تناظر هذا الترتيب من الجدول،أي أن :

$Q_1 = 30$ ميكروجرام/م^3 .

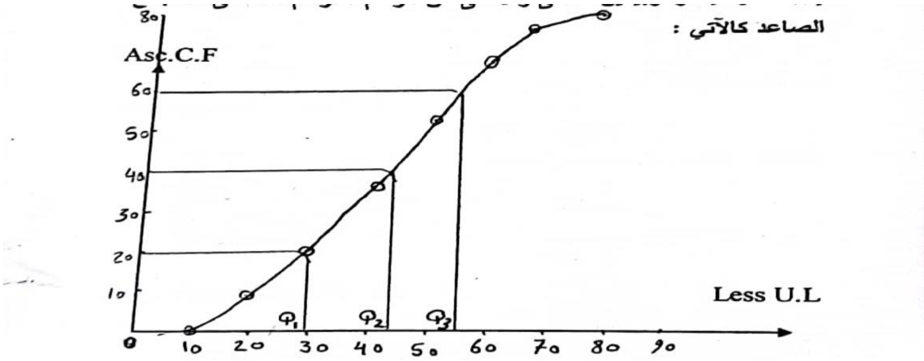
ولحساب الربيع الأعلى أي Q_3 فإن :

$$\text{Rank of } (Q_3) = \frac{3\sum F}{4} = \frac{80}{4} \times 3 = 60$$

وعليه فتكون قيمة Q_3 هي : -

$$Q_3 = 50 + 10\left(\frac{60-56}{68-56}\right) = 53.33$$ **ميكروجرام/م^3**

ولإيجاد الوسيط والربيعين الأدنى والأعلى من الرسم ، يتم رسم المنحنى التكرارى المتجمع الصاعد وذلك كما يوضحه الشكل البيانى التالى :



وعلى الرسم المبين أعلاه تتضح قيم الوسيط والربيعين الأدنى والأعلى .

أما المنوال فحيث أن الجدول التكراري منتظم فيتم حساب المنوال مباشرة ، ويمكن حساب المنوال بطريقة الرافعة أو بطريقة الفروق ، فيكون :

$F_M = 20$

$F_{M-1} = 16$

$F_{M+1} = 12$

وفئة المنوال هي من 40 إلى أقل من 50 أي أن :

Model Class = (40-50)

والفرق الأول هو :

$$D_1 = F_M - F_{(M-1)} = 20 - 16 = 4$$

والفرق الثاني هو

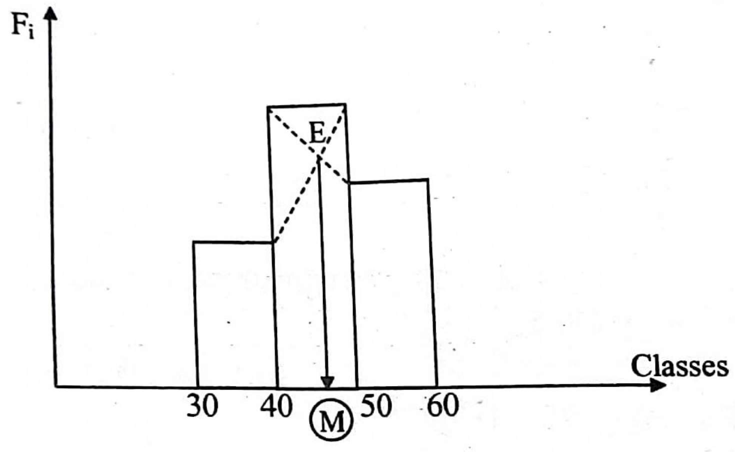
$$D_2 = F_M - F_{(M+1)} = 20 - 12 = 8$$

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 لها والفئة اللاحقة لها وذلك كما يوضحه الشكل التالى :



ومن الرسم نستنتج أن : المنوال = 43.33 ميكروجرام / م3

(3):لإيجاد نسبة المدن التي تزيد فيها درجة التلوث عن 45 ميكروجرام /م^3. فمن المنحنى التكرارى المتجمع الصاعد والسابق رسمه، ومن خلال رسم خط مستقم على المحور الأفقى وذلك عند النقطة 45 حيث يتم رسم خطاً رأسيا يوازى المحور الرأسى إلى أن يقابل المنحنى المتجمع الصاعد في نقطة نرسم منها خطاً افقياً إلى أن يقابل المحور الرأسى نجد أنه عند النقطة 47 مدينة ، وبالتالي فإن عدد المدن التي تقل فيها درجة التلوث عن 45 ميكروجرام /م3= 47 مدينة .

وعليه فإن :

عدد المدن التي تزيد فيها درجة التلوث عن 45 ميكروجرام /م3

= 80 − 47 = 33 مدينة .

وعليه فإن نسبة المدن التي تزيد فيها درجة التلوث عن 45 ميكروجرام /م3 هى

=(33 / 80) % = 41.25 %

(4):لإيجاد درجة التلوث التي تكون عندها 30 % من المدن التـــي فـــي عينـــة الدراسة أكثر نظافة ، فحيث أن نسبة ال 30 % من المدن عبارة عن:

= 80 × (30 / 100) = 24 مدينة .

لذا فإنه من المنحنى التكرارى المتجمع الصاعد نرسم على المحور الرأسي وبالتحديد عند النقطة 24 خطا أفقياً يوازى المحور الأفقى إلى أن يقابل المنحنى الصاعد في نقطة نرسم منها خطا رأسياً موازيا للمحور الرأسى إلى أن يقابل المحور الأفقي عند النقطة 32 وبالتالي فإن :

درجة التلوث التي عندها 30 % من المدن في عينة الدراسة تعد أكثر نظافة هي 32 ميكروجرام /م^3 .

تمارين الفصل الثالث

(1):-الجدول التالي يبين توزيع الطلاب حسب أيام الغياب خلال العام الماضي:

عددأيام الغياب	صفر-	5 -	10-	15-	20-	25-	30-	35-	40-	$\sum$
عدد الطلبة	29	195	241	117	52	10	6	3	2	655

والمطلوب حساب الوسط الحسابي والوسيط لهذا التوزيع .

(2):-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 .

(3):-الجدول التالي يوضح توزيع عينة 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 :

1-حساب قيمة الوسط الحسابي لإنفاق هذه الأسر .

2-حساب قيمة الوسيط وذلك من خلال المنحنى التكراري المتجمع الصاعد والنازل معا.

3- حساب قيمة المنوال وذلك من خلال :

- **استخدام طرق الحساب المختلفة .**
- **بالرسم من المدرج التكراري .**

4- حساب الوسط الهندسي .

5 – حساب الوسط التوافقي .

(4):-الجدول التالي يوضح توزيع عدد من المحلات التجارية حسب جملة المبيعات السنوية بآلاف الدولارات :

جملة المبيعات	110-	115-	120-	125-	130-	135-139
عدد المحلات	4	20	38	24	10	4

والمطلوب هو :

1- حساب الوسط الحسابي لجملة المبيعات .

2- حساب الوسيط.

(5):-احسب متوسط الطول(بالمليمتر) للبيانات التالية :

الطول بالمليمتر	198	199	200	201	202
التكرار	1	4	17	2	1

(6):- فيما يلى توزيع درجات الحرارة :

درجات الحرارة	16	17	18	19	20	21	22
التكرار	1	4	6	8	9	1	1

والمطلوب حساب الوسط الحسابي لدرجة الحرارة .

(7):–الجدول التالي يمثل توزيع مستويات أسعار 100 سلعة .

فئات أسعارالسلعة	115 –	125 –	–135	145 –	155 –	إجمالي
التكرار	10	20	40	20	10	100

والمطلوب حساب :

- **الوسط الحسابي .**
- **الوسيط .**
- **المنوال .**
- **الوسط الهندسي والوسط التوافقي .**

(8):الجدول التالي يبين أطوال عدد معين من السيدات يعملن في إحدى المصانع:

الطول بالسنتيمتر	–156	–159	–162	–165	170–168
العدد	2	9	18	7	3

والمطلوب حساب الوسط الحسابي والوسط الهندسي والوسط التوافقي للطول.

(9):–الجدول التكراري يبين حجم مبيعات المنتج أ حسب السعر:

سعر المنتج أ	75 –	77 –	79 –	81 –	83 –	85 –	المجموع
مبيعات المنتج أ	3	23	52	15	7	2	102

والمطلوب حساب :

أ – الوسط الحسابي للسعر.

ب – الوسط الحسابي بالطريقة المختصرة .

جـ– الوسيط .

د – القيمة الأكثر شيوعاً للسعر حسابياً وبيانياً .

(10):- الجدول التالي يبين توزيع العمال حسب الدخل :

فئات الدخل	أقل من20	20 -	30 -	40 -	50-	60 فأكثر	المجموع
عدد العمال	5	14	20	15	10	6	70

المطلوب إيجاد :

أ – مقياس مناسب للنزعة المركزية موضحاً سبب ذلك .

ب – مقياس مناسب للنزعة المركزية عن طريق الرسم وحسابياً .

(11):-من الجدول السابق في التمرين رقم (10) إذا علمت أن أقل قيمة لدخل العامل هى 10 جنيهاً وأكبر قيمة هى 70 جنيهاً ، فأوجد قيمة المقياس الآخر المناسب من مقاييس النزعة المركزية .

(12):-المطلوب حساب الوسط الحسابي والوسيط والمنوال للقيم التالية :

6 ، 37 ، 21 ، 58 ، 36 ، 33 ، 45

(13):- إذا كانت متوسطات درجات الطلبة في الإحصاء في ثلاث اختيارات: أ ، ب ، جـــ هي 75 ، 82 ، 84 درجة ، فإذا كان عدد الطلبة في كل من هذه الفصول هو 32 ، 25 ، 17 طالباً على التوالي .

المطلوب حساب الوسط الحسابي لجميع الفصول .

(14): – الآتي بيان بتوزيع أفراد عينتين من نزلاء أحد فنادق الدرجة الأولى بجمهورية مصر العربية حسب فئات تكاليف الإقامة(بالجنية) بأحد الليالي السياحية:

فئات التكاليف	100 -	150 -	250 -	300 -	350-400
عدد أفراد عينة الفندق الأول	8	20	12	10	5
عدد أفراد عينة الفندق الثانى	6	15	12	7	2

والمطلوب :

1- رسم المدرج التكراري لبيانات عينة الفندق الأول .

2- رسم المنحنى المتجمع الصاعد لبيانات عينة الفندق الثاني وأيضاً تحديد قيمة الوسيط .

3- إيجاد الوسط الحسابي والوسيط والمنوال لعينة الفندق الأول .

4- إيجاد المنوال بالرسم فقط لعينة الفندق الثاني .

(15):-فيما يلي التوزيع التكراري لـــ 300 أسرة حسب دخلها اليومى(بالجنيه):

Weekly Income Classes	25 -	30 -	35 -	40 -	45 -	50 -	55 -	∑
No. of Families	14	30	45	68	62	49	32	300

والمطلوب :-

(أ):حساب الوسط الحسابي والوسيط والمنوال .

(ب):حساب الربيعين الأدنى والأعلى .

(ج):ادرس شكل التوزيع مرة باستخدام العلاقة مابين الوسط الحسابي والوسيط والمنوال ومرة أخرى باستخدام الربيعين الأدنى والأعلى والوسيط ، قارن بين النتيجة في الحالتين .

(د):احسب قيمة الدخل الذى يقل عنه 60% من إجمالي عدد الأسر .

(ه):احسب النسبة المئوية لعدد الأسر التي يقل دخلها عن 43 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 .

(ى):احسب عدد الأسر الذي يتراوح دخلها الاسبوعي ما بين 37 ، 46 جنيهاً

(16):- البيانات التالية هي قيمة الإيجارات الشهرية (بالجنيه) لعينة من المساكن حجمها 50 شقة .

103	84	99	86	105	90	95	76	87	112
88	118	75	109	67	100	87	104	94	101
112	74	111	80	115	65	92	112	72	122
98	104	93	97	90	108	87	92	101	90
94	123	69	117	84	98	77	103	86	120

والمطلوب :

(أ):حساب الوسط الحسابي والوسيط لهذه البيانات .

(ب):عرض هذه البيانات في صورة جدول تكراري متساوي الفئات طول كل فئة 10 مبتدئا بالفئة 65 - . 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 95 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 بين70، 90 جنيهاً.

(17):- الجدول التالي يوضح التوزيع التكراري لدرجات الطلاب بإحدى كليات التجارة حسب نظام الدراسة (انتظام- انتساب) وذلك في دور مايو 1987 .

فئات الدرجات / نظام الدراسة	0 –	20 –	30 –	40 –	50 –	60 – 100	$\sum$
إنتظام	166	148	170	167	36	15	702
إنتساب	108	88	59	24	5	1	285

والمطلوب :

1–حساب نسبة النجاح لكل من الطلاب المنتظمين والطلاب المنتسبين .

2–قارن بين القيمة المتوسطة في كل من نظامي الدراسة مستخدماً الوسط الحسابي والوسيط والمنوال . استخدم الأشكال البيانية المناسبة للتعبير عن هذه البيانات .

3–احسب نسبة الطلاب اللذين تنحصر درجاتهم ما بين 35 ، 45 درجة لكل من الطلاب المنتظمين والطلاب المنتسبين ثم قارن بين النسبة في الحالتين .

(18): – 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-	$\sum$
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 74 جنيهاً على الأقل في كل من المصنعين ثم قارن بينهما .

(ج):احسب الوسط الحسابي لأجور العمال في المصنعين معاً وذلك باستخدام ما يلي:

1-خاصية الوسط الحسابي :
2-تكوين الجدول التكراري الذي يمثل توزيع العمال في المصنعين معاً
(وذلك بالحصول على مجموع تكراري المصنع أ والمصنع ب بالنسبة لكل فئة) .
3-قارن بين النتيجة في الحالتين .
د – احسب نسبة العمال الذين تتراوح أجورهم ما بين 60 ، 95 جنيهاً في كل من المصنعين . قارن بين النسبة في الحالتين .

(19):- الجدول التالي يوضح التوزيع التكراري لـــــ 40 أسرة حسب الحجم :

حجم الأسرة	1	2	3	4	5	6	المجموع
عدد الأسر	4	5	8	12	8	3	40

<u>والمطلوب</u> : هو إيجاد قيمة متوسطة لحجم الأسرة باستخدام المتوسط الملائم.

(20):- فيما يلي التوزيع التكراري لدرجات 100 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 Comm. Students	3	5	9	15	23	26	14	9	6
No. of Education Students	2	9	12	18	25	14	11	6	3

والمطلوب :
أ – قارن بين الوسط الحسابي والوسيط والمنوال في كل من الكليتين .
ب- احسب النسبة المئوية لعدد الطلاب الحاصلين على تقدير جيد فأكثر في كـــل من الكليتين وقارن بينهما (الحد الأدنى لتقدير جيد هو 65 درجة).
جـــ - قارن بين عدد الطلاب الذين تنحصر درجاتهم بين 42 ، 76 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 .

(21):– احسب الوسط الحسابي العام من البيانات الآتية

أ – ن$_1$ = 10 ، $\overline{x}_1$ = 25 ، ن$_2$ = 12 ، $\overline{x}_2$ = 15

ب– ن$_1$ = 30 ، $\overline{س}_1$ = 8 ، ن$_2$ = 8 ، $\overline{س}_2$ = 12

(22):– إذا كان :

ن$_1$ = 10 ، $\overline{س}_1$ = 10 ، ن$_2$ =؟ ، $\overline{س}_2$ = 20

فإذا كان الوسط الحسابي العام للعينتين هو 16،احسب حجم العينة الثانية .

(23):– الجدول التالي يوضح توزيع 80أسرة حسب انفاقها اليومي بالجنيه:

فئات الإنفاق	12 –	14 –	16 –	18 –	22 –	26 –	المجموع
عدد الأسر	7	11	17	22	15	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 15 ، 19 جنيها .

د– من النتائج أ ، ب إدرس تماثل التوزيع .

(24):- باستخدام التوزيع التكراري الوارد بالجدول (2-8) ، والمطلوب :

- **أ- تحقق من تماثل التوزيع باستخدام الطرق المختلفة .**
- **ب- أوجد قيم المئين الخامس عشر والمئين الخامس والثمانين .**
- **ت- قارن مابين ما تحصل عليه في (ب) بقيمة الوسيط . علق على ما تحصل عليه من نتائج .**

(25):-التوزيع التكراري التالي يوضح الدخول الاسبوعية (بالجنيهات) لعينة حجمها 50 أسرة :

فئات الدخول	40	50	60	70	80	90-100	المجموع
عدد الأسر	6	8	11	14	7	4	50

والمطلوب :

- **أ- حساب كل من الوسط الحسابي والوسيط .**
- **ب- احسب قيمة الدخل الشهري الذي يقل عنه 40% من إجمالي عدد الأسر.**
- **ت- استخدم شكل المدرج التكراري للتعبير عن مدى تماثل هذا التوزيع .**

(26):- فيما يلي التوزيع التكراري للأجور اليومية لمائة عامل في إحدى الشركات :

فئات الأجور	35-	45-	55-	65-	75-	85-95	المجموع
عدد الموظفين (التكرارات)	9	13	18	26	19	15	100

والمطلوب :

- **أ- حساب الوسط الحسابي .**

ب- تكوين الجدول التكراري المتجمع الصاعد وحساب قيمة الوسيط .

ت- باستخدام طريقة الفروق ، احسب المنوال .

(27):- فيما يلي التوزيع التكراري لدرجات 40 طالباً في مادة الإحصاء :

فئات الدرجات	25-	35-	45-	55-	65-	75-	85-95
عدد الطلاب	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 50 درجة) .

- إذا كان عدد الطلاب الذين تتراوح درجاتهم ما بين (الدرجة 45 ، الدرجة س) يمثل 65 % من إجمالي عدد الطلاب ، احسب قيمة الدرجة (س) .

(د): من نتائجك في أ ، ب ناقش بإيجاز شكل التوزيع من حيث التماثل والالتواء .

الفصل الــــــــرابع

مقاييس التشتت

Measures of Dispersion

فى الفصل السابق لاحظنا أن مقاييس النزعنة المركزية أو المتوسطات (الوسط الحسابى- الوسيط – المنوال – الوسط الهندسى – الوسط التوافقى) تمثل قيمة متوسطة تتركز 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 تباعدها 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 فى كل من المجتمعين على النحو المبين التالى:

الدخل الأسبوعى فى المجتمع (أ) هو 5 ، 50 ، 100 ، 150 ، 195 جنيه وأن الدخل الاسبوعى فى المجتمع (ب)هو98 ، 99 ، 100 ، 101 ، 102 جنيه . فنجد أن الوسط الحسابى للدخل الاسبوعى لكل من المجتمعين(أ ، ب) هو مساوياً 100 جنيه . كما أن وسيط الدخل الأسبوعى لكل من المجتمعين (أ ، ب) أيضاً مساوياً 100 جنيه .أى أنه وفقاً لمقاييس النزعة المركزية فإننا نجد أن البيانات فى المجتمعين (أ ، ب)لا يختلفان من حيث متوسط أو وسيط الدخل الاسبوعى . ولكن هذا مخالفاً لواقع البيانات الفعلية فى المجتمعين(أ ، ب) حيث يوجد هناك تفاوتاً واضحاً فى توزيع مفردات الدخل فى كل مجتمع عن الآخر. فالبيانات فى المجتمع الأول (أ) يتراوح فيها الدخل الأسبوعى فيما بين 5 جنيهات و 195 جنيه أى أن المدى هو :-

Range = Maximum value – Minimum value.

= 195 – 5 = 190 (L.E)

بينما نجد أن فى المجتمع الثانى (ب) يتروراح فيه قيمة الدخل الأسبوعى فيما بين 98 ، 102 جنيه. اى أن المدى هو :

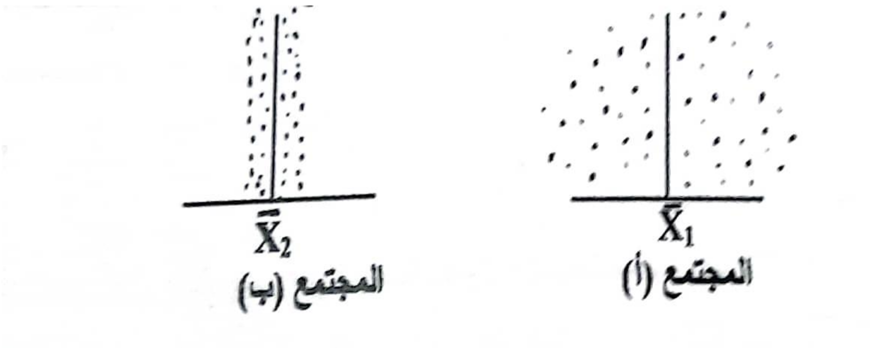
Range = 102 – 94 = 8 (L.E)

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 قريبة 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 عن 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 (4-1) نجد أن مفردات المجتمع (أ) أكثر تشتتاً (أو أقل تجانساً) من مفردات المجتمع (ب) :



شكل (4-1)

بيان 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 آخر 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:-

<u>النوع الأول</u> : 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 (فكل من مفردات المجموعتين وحدات قياسها هى الدرجة) أو فى دراسة مقارنة تشتت أعمار (بالسنوات) مجموعتين مختلفين من الطلاب أو الطالبات (فكل من مفردات المجموعتين وحدات قياسها هى السنة) .

وبالإضافة إلى ان مقاييس التشتت المطلقة تقيس تشتت البيانات أو بعثرتها فأهم ما يميزها أنها تأخذ فى النهاية قيمة وحدة قياسها 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:-

1-<u>المـدى</u> : Range: ويعرف المـدى لمجموعـة مـن المفـردات كمـا سبق أن ذكرنـا بأنـه عبارة عن الفرق ما بين أكبر القيم وأصـغرها أو مـا يعرف بـالقيمتين الصغرى والكبرى معاً . هذا ويستخدم المدى فى بيان مدى الإتساع الذى تنتشـر عليه المفردات. أى أن المدى :-

أ-<u>فى حالة المفردات (البيانات الغير مبوبه)</u> : هو :

Range = Maximum value – Minimum value.

فـإذا كـان لـدينا متغيـراً يأخـذ مجموعـة القيم X_1 ، X_2 ، ..... ، X_n وتـم ترتيبهـا تصـاعدياً أو تنـازليـاً ولـيكن $X_{(1)}$ ، $X_{(2)}$ ، ..... ، $X_{(n)}$ حيـث يـتم التعبيـر عـن أصغر القيم بالقيمة $X_{(1)}$ وأكبر القيم 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 $X_{(n)}$.

<u>فمثلاً</u>

إذا كـان لـدينا مجموعـة القـيم : 20 ، 15 ، 38 ، 12 ، 30 فـإن المـدى لتلـك المجموعة هو :

Range = 38 - 12 = 26

أو نقول بأن المـدى مـن 12 حتـى 38 وإن كـان هـذا التعبيـر الأخيـر هـو الأفضـل حيث أنـه يعطى معلومات أكثر من التعبير الأول (القانون) .

<u>ب- المدى فى حالة البيانات المبوبة تكرارياً :-</u>

إذا كـان لـدينا توزيعـاً تكراريـاً فـإن هنـاك مجموعـة مـن الصيغ التـى يمكـن مـن خلالها حساب المدى . الأولى والتـى تفيـد بـأن المـدى هـو عبـارة عـن الفـرق مـا

بين الحد الأعلى للفئـة الأخيرة والحد الأدنـى للفئـة الأولى للتوزيع التكـرارى المعطى أى أن :-

Rang = U.L for the Latest Class – L.L for the First Class.

حيث :

تمثل الحد الاعلى للفئة الاخيرة **U.L for the Latest Class**

تمثل الحد الادنى للفئة الاولى **L.L for the First Class**

أما الصيغة الثانية والتى يمكن من خلالها حساب المدى ايضـا وذلك من خلال ايجاد الفرق بين مركزى الفئة الاخيرة والفئة الاولى للتوزيع التكرارى .ووفقا لهذا التعريف فان المدى يمكن كتابته على الصورة :

Range=Midpoint for theLatest 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 هنا إلى 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 :-

(1):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.

(2): من خلال مفهومنا السابق عن المدى ونظراً لإعتماد المدى فى حسابه على المفردتين المتطرفتين فى القيمة فقط 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 قيمة المدى فيستدل منه على أن مفردات الظاهرة متشتتة جداً بينما قد تكون مفرداتها كلها تقريباً – أى فياً عدا القيمة الشاذه أو المتطرفة – متقاربة . فمثلاً إذا كان لدينا مجموعة القيم 15، 18 ، 20 ، 13 ، 19 فإن المدى لهذه القيم مساوياً :

Range = 20 – 13 = 7

فإذا ما أضفنا قيمة واحده ولتكن المفردة 100 مثلاً لمجموعة المفردات السابقة فإن المدى سيصبح :-

Range = 100 – 13 = 87

أى أن المدى تغيرت قيمته لتصبح 87 بدلاً من 7 نتيجة إضافة المفردة التى تأخذ القيمة الشاذه أو المتطرفة عن باقى المفردات وهى المفردة (100).

(3):يعتبر المدى حساس جداً بالنسبة لحجم العينة. فغالباً ما يزداد المدى نتيجة زيادة حجم العينة . ولذلك فلا يعتمد الإحصائيون على المدى لتقدير التغير فى قيم مفردات المجتمع إذ أن قيمته تتوقف على حجم العينة.

(4):المدى يتغير كثيراً من عينة لأخرى – وهذا الأمر يكون دائماً غير مرغوب فيه لدى الإحصائي – وهذا الإختلاف أكثر من الأختلاف الذى نحصل عليه من بعض مقاييس التشتت الأخرى.

(5):يصعب حساب المدى فى حالة الجداول التكرارية المفتوحة حيث لا يكون معلوماً الحد الأدنى للفئة الأولى أو الحد الأعلى للفئة الأخيرة هذا بالإضافة لعدم إمكانية حسابه بيانياً.

هذا وبالرغم من هذه العيوب المتعددة إلا أن المدى كما سبق وأن ذكرنا يستخدم على نطاق واسع فى مجال واسع فى مجال دراسات ضبط جودة الإنتاج Statistical Quality Control . والسبب فى ذلك يرجع إلى أنه كما ذكرنا أنه فى مثل هذه الدراسات (دراسات ضبط الجودة) فإننا نحتاج إلى حساب مقياس للتشتت بسرعة . والمدى بطبيعة الحال هو أسرع مقياس يمكن إيجاده ويعطى دلاله على التشتت. هذا وتجدر الإشارة إلى أنه فى مثل تلك الحالات فأن المدى يتم حسابه أكثر من مره وذلك بإستخدام عدد كبير من العينات 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.

2-شبيهات المدى :- Quasi or Semi 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:

$$\text{First } (1^{st}) \text{ Range} = X_{(n-1)} - X_{(2)}$$

Second (2^{nd}) Range = $X_{(n-2)} - X_{(3)}$

وهكذا بالنسبة لبقية مقاييس شبيهات المدى حيث يتم التعريف المدى الرائـى أو ذو الرتبة (r) بأنه على الصورة:

$r^{th} \sim$ Range = $X_{(n-r)} - X_{(r+1)}$

<u>فمثلاً</u>

إذا كان لدينا مجموعة البيانات التالية :

12 ، 11 ، 10 ، 13 ، 14 ، 17 ، 15 ، 18 ، 84 ، 16.

فأنه يمكن إيجاد المدى وشبيهاته وذلك على النحو التالى :-

فبعد ترتيب المفردات تنازليا أو تصاعدياً مثلاً فإن البيانات تأخذ الصورة التالية

84 ، 18 ، 17 ، 16 ، 15 ، 14 ، 13 ، 12 ، 11 ، 10

فيكون المدى هو :

Range = $X_{(n)} - X_{(1)} = X_{(10)} - X_{(1)} = 84 - 10 = 74$

أما شبيهات المدى فهى عبارة عن :

فالمدى الأول هو: **First (1^{st}) Range = $X_{(n-1)} - X_{(2)}$**

or 1st Range = $X_{(9)} - X_{(2)} = 18 - 11 = 7$

والمدى الثانى هو :

Second(2^{nd}) range = $X_{(n-2)} - X_{(3)}$ Or,

2^{nd} Range = $X_{(8)} - X_{(3)}$

$= 17 - 12 = 5$

~

وكذلك فإن المدى الثالث هو :

$$\text{Third Range} = X_{(n-3)} - X_4 \quad \text{Or,}$$

$$3^{rd}\ \text{Range} = X_{(7)} - X_4$$

$$= 16 - 13 = 3$$

لاحظ مدى الأثر الذى يقع على شبيهات المدى من حيث عدم تأثر مثل هذه المقاييس(شبيهات المدى) 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 وبين المئين الخامس أو بين المئين السادس والتسعون وبين المئين 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 .

1- 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 :

$$\textbf{Semi. Int. qu. Range} = \frac{Q_3 - Q_1}{2}$$

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 فإن نصف المدى الربيعى يصبح غير ملائماً لقياس تشتت الظاهرة محل الدراسة.

مثال :- فيما يلى لديك التوزيع التكرارى لدرجات مجموعة من الطلاب فى اختبار مادة الإحصاء.

جدول (4-1)

Classes	0-	10-	20-	30-	40-50	Σ
Frequency	5	10	24	12	9	60

والمطلوب :-

1- أحسب المدى والمدى الأول.

2- أحسب نصف المدى الربيعى لهذا التوزيع.

الحل :

حيث أن البيانات المعطاه لتوزيعاً تكرارياً لذا فأن :

1- المدى هو :

Range = Midpoint for the Latest Class - Midpoint for the1stClass

= 45 – 5 = 40 (degree)

or

Range = U.L. for the Latter Class – L.L for the 1st Class

= 50 – 0 = 50 (degree)

لاحظ أن إختلاف القيم ليس إلاّ 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 (4 – 2)

Classes	Frequency	Less than the U.L for 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 :

$$\textbf{Rank of } Q_1 = \frac{\sum F}{4} = \frac{60}{4} = 15$$

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 المتجمعة اى ان:

$$Q_1 = 20 \quad \text{degree}$$

اما عن الربيع الثالث فحيث أن 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 39 , 51 ومن ثم فان بداية فئة الربيع الثالث هى 30 درجة وبتطبيق الصيفية الحسابية للربيع الثالث نجد أن :

$$Q_3 = \text{L.L for } (Q_3) \text{ class} + \frac{\frac{3\sum F}{4} - \text{Pre. Asc. C.F. for the rank of } (Q_3)}{\text{Suc. Asc. C.F. for} - \text{Pres. Asc. C.F. for}} \times LQ_3$$

أى أن :

$$Q_3 = 30 + 10\left(\frac{45\text{-}39}{51\text{-}39}\right) = 30 + 10\left(\frac{6}{12}\right)$$

i.e.,

$$Q_3 = 35 \qquad \text{(degree)}$$

وعليه فان نصف المدى الربيعى يكون :

$$\text{Semi – Int. Qu. Range} = \frac{Q_3 - Q_1}{2}$$

$$= \frac{35\text{-}20}{2} = 7.5 \qquad \text{(degree)}$$

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 بينهما 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 $X_{(1)}$ ، $X_{(n)}$ كما هو فى حالة المدى .

وفيما يلى نتعرض بالدراسة الى مقياسين آخيرين 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 .

3- 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 المحسوب 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 ($\bar{X}$) . والقصود بالإنحراف المطلق للفرق (di) هو عبارة عن الفرق بين كل قيمة أو مفردة من مفردات الظاهرة وبين الوسط الحسابى مع اهمال اشارة الفرق اى ان :

$$d_i = | X_i - \bar{X} | \qquad \text{where} \quad i = 1,2,\ldots\ldots,n$$

وقد تم اعتبار الفرق المطلق اى تم أهمال 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 خلال 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 :

$$\textbf{Mean Deviation about Mean (MD)} = \frac{\sum_{i=1}^{n} \left| X_i - \bar{X} \right|}{n}$$

كما ان الانحراف المتوسط عن الوسيط يعرف بالصورة التالية :

$$\text{Mean Deviation about Median} = \frac{\sum_{i=1}^{n} \left| X_i - Q_2 \right|}{n}$$

حيث تعنى $\sum|X_i - \bar{X}|$ ، $\sum|X_i - Q_2|$ 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 أو التوافقى الا ان هذا نادر الحدوث . هذا ويعتبر المقياس الاكثر استخداما هو الانحراف المتوسط عن الوسط الحسابى او الوسيط . واذا ذكرنا الانحراف المتوسط بدون تخصيص بانه حول اى من قيم المتوسطات المختلفة فيكون المقصود به هو الانحراف المتوسط عن الوسط الحسابى .

مثال :

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 أولا وذلك على النحو التالى :

$$\bar{X} = \frac{\sum X}{n} = \frac{133 + 137 + \dots\dots + 131 + 150}{7}$$

$$= \frac{959}{7} = 137$$

والجدول التالى يعطى الفرق فيما بين محموعة المفردات والوسط الحسابى ومن ثم الانحرافات المطلقة لتلك الفروق :

جدول (4-3)

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 :

$$\text{Mean Deviation} = \frac{\sum |X_i - \bar{X}|}{n} = \frac{\sum |d_i|}{n} \quad i.e.,$$

$$MD_{\bar{X}} = \frac{30}{7} = 4.28$$

أمـا لايجـاد الانحـراف المتوسـط عـن الوسـيط فيجـب ترتيـب المفـردات ولـيكن 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 هى:

$$Rank\ of\ Q_2 = \frac{n+1}{2} = \frac{7+1}{2} = 4$$

اى ان الوسيط هو عبارة عن المفردة الرابعة للبيانـات بعد ترتيبهـا اى ان قيمـة الوسيط هـى 135 اى ان: $Q_2=135$. وعليـه فانـه لحسـاب قيمـة الانحـراف المتوسط عن الوسيط يتم تكوين الجدول التـالى لحساب الفروق المطلقـة للقيم عن الوسيط أى قيمة : $|X_i - 135|$.

جدول (4-4)

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 (4-4) فان :

الانحراف المتوسط من الوسيط هو :

$$\text{Mean Deviation} = \frac{\sum|X_i - Q_2|}{n} = \frac{\sum|d_i|}{n}$$

about Median

$$i.e., MDQ2 = \frac{28}{7} = 4$$

<u>* حساب الانحراف المتوسط فى حالة البيانات المبوبة :</u>

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 :

$$\text{Mean Deviation} = \frac{\sum|X_i - \bar{X}|\, F_i}{\sum F_i}$$

$$MD_{\bar{X}} = \frac{\sum|d_i|F_i}{\sum F_i}$$

حيث X_i , F_i تعبر عن مركز وتكرار الفئـة (i) علـى الترتيب حيث $i = 1,2,\ldots..r$ ، كما أن (r) 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 :

$$\text{Mean Deviation} = \frac{\sum|X_i - Q_2|F_i}{\sum F_i}$$

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 (4-5)

Classes	20-	28-	36-	44-	52-60	∑
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 (4-6) والذى يوضح الحسابات اللازمة لحساب الوسط الحسابى ومن ثم الانحراف المتوسط عن قيمة الوسط الحسابى

جدول (4-6)

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**
∑	**40**				**-10**		**256**

فيكون :

$$\bar{X} = A + B\,\frac{\sum D_i F_i}{\sum F_i}$$

$$= 40 + 8\left(\frac{-10}{40}\right) = 40 - 2 = 38 \qquad (year)$$

وعليه فان الانحراف المتوسط عن الوسط الحسابى هو:

$$MD_{\bar{X}} = \frac{\sum |X_i - \bar{X}| F_i}{\sum F_i} = \frac{256}{40} = 64 \qquad (year)$$

مثال: احسب الانحراف المتوسط عن الوسيط للتوزيع التكرارى التالى لأعمار مجموعة من الأشخاص (بالسنة) :

جدول (4-7)

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 (4-8) يوضح الحسابات اللازمة لذلك :

جدول (4-8)

Classes	F_i	X_i	Less than the Class 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

$$Rank\ of\ Q_2 = \frac{\sum F}{2} = \frac{100}{2} = 50$$

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:

$Q_2 = (48 + 54) / 2 = 51$ (year)

(ويمكنك عزيزى القارئ التأكد من صحة قيمة الوسيط باستخدام صيغة القانون المعروف بقانون الوسيط والمستخدم لحساب الوسيط أو باستخدام قاعدة النسبة والتناسب).

وعليه فان الانحراف المتوسط عن الوسيط هو :

$$\frac{\sum \lvert X_i - Q_2 \rvert F_i}{\sum F_i}$$

Mean Deviation =

About Median

ومن خلال الجدول السابق فان :

$$\text{M.D.about } Q_2 = \frac{454}{100} = 4.45 \qquad \text{(year)}$$

وأخيرا فان الانحراف المتوسط كمقياس للتشتت المطلق يعتبر غير شائع الاستخدام وذلك لعدة اسباب هى :

1- **عدم سهولة العمليات الحسابية اللازمة لايجاده .**
2- **صعوبة التعامل معه جبريا 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 بدلالة 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 الآن 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 .

4 - التباين والانحراف المعيارى :

Variance and Standard Deviation

سبق وأن راينا ان انحرافات القيم عن وسطها الحسابى تعطى صورة جيدة لدرجة تباين(تشتت) 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 (σ^2).

فالتبـاين يعرف بانـه عبـارة عن متوسط مربعـات انحرافـات القيم عن وسطها الحسابى اما الانحرافات المعيارى والـذى يرمـز بـالرمز (σ) والـذى يعتبـر مـن اهم مقاييس التشتت المطلقة فما هو الا الجزر التربيعى (الموجب) للتبـاين اى ان الانحراف المعيارى ما هو الا الجذر التربيعى (الموجب) للوسط الحسـابى لمربعـات انحرافـات القـيم عـن وسـطها الحسـابى وعـادة مـا يسـتخدم الانحـراف المعيارى (σ) كمقياس للتشتت عندما لا يكون قياس تشتت الظاهرة هو نـهايـة المطاف بل بداية لعملية التحليل الاحصـائى الاكثـر اهميـة والمتعلقـة بالاسـتنتاج الاحصائى لمقاييس المجتمع محل الدراسة. وترجع تلك الاهميـة الـى خضـوعه للطرق الرياضية وسـهولـة التعامـل معـه جبريـا فهـو يستخدم علـى نطـاق واسـع للغاية فى نظرية التقدير الاحصائى واختبارات الفروض الاحصائية كأحد المـهـام الرئيسية لفرع الاحصاء الاستدلالى – الاحصـاء التطبيقـى - وفيمـا يلـى نوضـح كيفية حساب الانحراف المعيارى ومـن ثـم التبـاين فـى حالـة المفـردات وكـذا فـى حالة البيانـات المبوبة تكراريا .

أولا:الانحـراف المعيـارى للبيانـات غيـر المبوبـة (المفـردات) :

اذا كـان لـدينـا متغيـرا ولـيكن X يأخـذ مجموعـة القيـم $X_1 , X_2 , \ldots\ldots , X_n$ **لظاهرة معينـة وكـان متوسطها الحسابى هو** $(\bar{X})$ **فـان انحرافـات هـذه القيم عن وسطها الحسابى هى :**

$$(X_1-\bar{X}) , (X_2-\bar{X}) , (X_3-\bar{X}) , \ldots\ldots\ldots\ldots , (X_n-\bar{X})$$

ومن ثم يكون التباين عبـارة عن متوسط مربعـات انحرافـات القيم عن وسطها الحسابى اى ان :

$$\sigma^2 = \frac{1}{n} [(X_1-\bar{X})^2 + (X_2-\bar{X})^2 + \ldots\ldots + (X_n-\bar{X})^2]$$

اما عن الانحراف المعيارى(σ) فما هو الا الجذر التربيعى للتباين والعكس صحيح اى ان التباين ما هو الا مربع الانحراف المعيارى. اى ان :

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n}} \quad , \text{ where } i=1 , 2 , 3 , \ldots , n$$

والسبب فى اخذ الجذر التربيعى للتباين σ^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 الاساسية السابقة .

<u>مثال :</u>

احسب الانحراف المعياري لمجموعه القيم التالية :

3 ، 12 ، 9 ، 8 ، 15 ، 6

<u>الحل : -</u>

لاستخدام الصيغه الرياضية السابقه في حساب الانحراف المعياري يتم حساب الوسط الحسابي لتلك المفردات أولا : 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 (4 – 10)

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 :

$$\sigma = \sqrt{\frac{\sum(X_i - \bar{X})^2}{n}} = \sqrt{\frac{50}{6}}$$

$$= \sqrt{8.333} \simeq 2.887$$

اي ان الانحراف المعياري لتلك القيم هو تقريبا 2.887 اما التباين فهو عبارة عن الناتج تحت الجذر التربعي . اي ان : $\sigma^2 = 8.333$

وجدير بالذكرانه يفضل 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 الصيغة التاليه عنها .

1- 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 أن التقريب المستخدم غالبا ما يؤدي الي عدم الدقة في النتائج لذا ينصح بالتغلب علي ذلك من خلال ما يلي :

فحيث ان :

$$\sigma = \sqrt{\frac{\sum(X-\bar{X})^2}{n}}$$

لاحظ أن:

$$\begin{aligned}\sum(X_i-\bar{X})^2 &= \sum_{i=1}^{n}(X_i^2 - 2\,\bar{X}\,X_i + \bar{X}^2)\\ &= \sum_{X_i^2} - 2\,\bar{X}\sum_{Xi} + n\,\bar{X}^2\\ &= \sum_{X_i^2} - 2\,n\,\overline{X^2} + n\,\bar{X}^2\\ &= \sum_{X_i^2} - n\,\bar{X}^2\end{aligned}$$

$$= \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

وعليه فإنه :-

$$\sigma = \sqrt{\frac{1}{n}\left[\sum x_i^2 - \frac{(\sum xi)^2}{n}\right]}$$

$$= \sqrt{\frac{\sum Xi^2}{n} - \left(\frac{\sum Xi}{n}\right)^2} \qquad (2)$$

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 قيمة الوسط الحسابي من كل قيمة من قيم المتغير، وما يترتب علي ذلك من صعوبة في اجراء العمليات الحسابية . وإنما يتطلب الأمر وفقا لتلك الصيغه مجرد حساب مجموع القيم اي $\sum Xi$ ومجموع مربعات هذه القيم اي ($\sum Xi^2$) . ففي مثالنا فأن :-

$$\sum X_i = 48 \quad , \quad \sum X_i^2 = 434$$

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 مربع الانحراف المعياري او الناتج تحت الجذر التربيعي اي ان :

$$\sigma^2 = 8.333$$

وهي نفس النتائج السابقة .

وفي هذا المثال قد لا تبرز اهمية صيغة القانون الاخير بشكل ملموس حيث سهولة استخدامه وذلك لان الوسط الحسابي لمجموعة القيم يأخذ صورة كسرية مما يجعل معه طريقة حساب الانحراف المعياري بأتباع الصيغه الاولي امرا اكثر تعقيدا . ولكن وكما سبق وأن ذكرنا فأن الصيغه الثانية تبسط كثيرا من العمليات الحسابية في الحالات التي يكون قيمة الوسط الحسابي لمجموعة المفردات تحتوي علي كسور مما يجعل معه ايجاد الانحرافات عنه وكذا مربعات هذه الانحرافات 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 عن الوسط الحسابي أوباستخدام مجموع القيم الصلية ومجموع مربعاتها دعنا اولا نقوم بحساب الوسط الحسابي للمفردات . فحيث ان :

$$\bar{X} = \frac{\sum x}{n} = \frac{3 + 1 + 9 + \dots\dots + 6}{7} = \frac{33}{7} = 4.71$$

ثم دعنا نجري الحسابات اللازمة للطريقتين من خلال الجدول التالي :

X_i	X_i^2	$(x_i - \bar{X})$	$(x_i - \bar{X})^2$
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 – <u>طريقة حساب الانحراف المعيارى باستخدام الأنحرافات عن الوسط فإن</u> :

$$\sigma = \sqrt{\frac{\sum(X - \bar{X})^2}{n}} = \sqrt{\frac{55.4287}{7}} = \sqrt{7.918385714} = 2.813962636$$

إما بإستخدام مجموع القيم الأصلية ومجموع مربعاتها فإن :

$$\sigma = \sqrt{\frac{\sum X^2}{n} - \left(\frac{\sum x}{n}\right)^2} = \sqrt{\frac{211}{7} - \left(\frac{33}{7}\right)^2}$$

$$= \sqrt{7.918367347} \simeq 2.813959372$$

لاحظ ان الاختلاف فيما بين قيمة الانحراف المعياري يأتي 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 مساويا 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 أردنا ذلك وكما يتضح لنا من المثالين السابقين فانه اذا كان الوسط الحسابي قيمته عددا صحيحا (اي لا يحتوي علي فيمة كسرية) فسوف يتساوي قيمة الانحراف المعياري باستخدام ايا من الاسلوبين سواء باستخدام طريقة الانحرافات حول الوسط او باستخدام مجموع القيم الاصلية ومجموع مرعاتها لانه لن يكون هناك اي عمليات تقريب وإن كان من المفضل استخدام الطريقة الخاصه بالقيم الاصلية لانها تحتوي علي عمليات حسابية اقل كما سبق ان أوضحنا . اما اذا كان الوسط الحسابي يحتوي علي قيمة كسرية فانه يفضل حساب الانحراف المعياري باستخدام القيم الاصلية . لانها تتجنب اخطاء التقريب في الحسابات وبالتالي تكون النتيجه اكثر دقة .

خصائص الانحراف المعياري :

1 – لا يختلف التشتت فى قيمة الظاهرة بعمليتى الجمع أو الطرح لمقدار ثابت على أو من القيم الأصلية حيث يظل التشتت فيما بين المشاهدات كما هو دون أدنى تغير نتيجة جمع أو طرح مقدار ثابت على أومن مجموعة المفردات . وعليه فلا يتاثر الانحراف المعياري بعمليتي الجمع او الطرح : بمعني أنه إذا كان لدينا قيم المتغير X_i هي مجموعه القيم $X_1 , X_2 , ..., X_n$ فان :-

الوسط الحسابي لتلك القيم هو :

$$\bar{X} = \frac{\sum x}{n}$$

كما ان الانحراف المعياري للمتغير X 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 :

$$d_1 = X_1 - a \; , \; d_2 = X_2 - a \; , \; ...... \; , \; d_n = X_n - a$$

ومن ثم فان :

الوسط الحسابي للقيم d_i حيث $i = 1 , 2 ,......,n$ 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 X_i عن الوسط الفرضي (a) هي نفسها الانحراف المعياري لقيم الاصلية . وهو ما يفيد بأن الانحراف المعياري لم يتأثر بعملية الطرح . والعكس صحيح تماما فيمكن 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 تقيس 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 .

2- <u>الانحراف المعياري يتأثر بعمليتى الضرب والقسمه</u> :

ويمكن اثبات ذلك رياضيا بنفس الالية السابقة المستخدمه في الخاصية الاولي وذلك علي النحو التالي :

فنفرض ان لدينا محموعة القيم $X_1 , X_2 , ... , X_n$ للمتغير (X_i) 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 أن قيمة** (a) **لا تساوي الصفر اي** ($a \neq 0$) **فان مجموع الصيغ الجديده ما هي الا :** $D_1 = ax_1$, $D_2 = ax_2$, ……. , $D_n = ax_n$

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 فيكون عبارة عن :-

$$\sigma_D = \sqrt{\frac{\sum(Di-\bar{D})^2}{n}} = \sqrt{\frac{\sum(ax_i-a\bar{x})^2}{n}}$$

أى أن :

$$\sigma_D = \sqrt{\frac{a^2 \ \Sigma(X_i - \bar{X})^2}{n}}$$

$$= a\sqrt{\frac{\Sigma(X_i - \bar{X})^2}{n}} = a\,\sigma_x$$

وهو ما يفيد أن الانحراف المعياري للقيم الجديدة (اي بعد ضرب قيم المتغير الاصلى في المقدارالثابت) ما هو الا حاصل ضرب الانحراف المعياري للقيم الاصلية (اي قبل عملية الضرب) فى المقدار الثابت والذى تم ضرب قيم المتغير الأصلى فيه . وعليه فأن :

$$\sigma_x = \frac{\sigma_D}{a}$$

اي ان الانحراف المعياري للقيم الاصلية (x_i) = الانحراف المعياري للقيم الناتجة بعد اجراء عملية ضرب المفردات الاصلية في المقدار القابت (a) مقسوما على هذا المقدار الثابت (a) الذى تم ضرب مفردات الظاهرة فيه .

كذلك يتم التعامل مع العملية العكسية لعملية الضرب (اي عملية القسمة) نفس معاملة عملية الضرب ولكن بطريقه عكسية. بمعني انه اذا تم قسمه قيم المتغير (x_i) على مقدار ثابت وليكن (a) حيث $(a \neq 0)$ فان قيمة الانحراف المعياري للقيم الاصلية اي σ_x سوف يكون مساويا للانحراف المعياري للقيم الجديدة أى (d_X) مضروبا في هذا المقدار الثابت (a) . هذا ويمكن بيان مدي صحة هاتين الخاصيتين السابقتين من خلال المثال البسيط التالى :

افرض ان لديك المفردات الخمس التاليه 100 , 200 , 300 , 400 , 500 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 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 عبارة عن :

$$d_i = 0, 100, 200, 300, 400$$

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 الموجودة فى هذا المثال علي المقدار الثابت (100) فتكون نواتج قسمة المفردات الاصلية علي 100هي :

$D_i = 1 , 2 , 3 , 4 , 5$

فيكون 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 :

$$\sigma_D = \sqrt{\frac{\sum(Di - \bar{D})^2}{n}}$$

$$= \sqrt{\frac{(1-3)^2 + 2-3^2 + \ldots\ldots + (5-3)^2}{5}} = 1.414213562$$

أى أن :-

$$\bar{X} = 300 \quad = 100\ (3) \quad = a.\bar{D}$$

وكذلك فأن :-

$$\sigma_x = 141.4213562 \quad = 100\,(1.414213562) \quad = a.\sigma_D$$

اي ان الوسط الحسابي وكذا الانحراف المعياري للقيم الاصلية هو عبارة عن حاصل ضرب المقدار (a) في الوسط الحسابي أوالانحراف 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 رأينا 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 $d_i = X_i - a$ ، حيث $i = 1, 2, \ldots, n$ تمثل الانحرافات البسيطة حول الوسط الفرض (a) فان :

$$\sigma_X = \sigma_d = \sqrt{\frac{\sum di^2}{n} - \left(\frac{\sum di}{n}\right)^2}$$

والوسط الفرضى (a) يتم اختياره بالشكل الذي يؤدي الي تبسيط العمليات الحسابية . ويفضل إستخدام قيمة هذا الوسط الفرضى بمثابة قيمة المفردة التى تتكرر أكثر من غيرها فى باقى المفردات وذلك لتبسيط الحسابات .

<u>* طريقة الانحرافات المعدلة (الاكثر اختصارا)</u> :- بعد الحصول علي الانحرافات البسيطة $d_i = X_i - a$ ، فاذا كانت جميع الانحرافات البسيطة d_i تقبل

القسمة علي وسط فرضى آخر و بدون بواقي . فأنه يتم قسمة قيم الانحرافات البسيطة (d_i) على هذا الوسط الفرضى وليكن (b) فنحصل علي الانحرافات المعدلة D_i حيث ان :

$$D_i = \frac{X_i - a}{b} = \frac{di}{b} \qquad where \;\; i = 1, 2, \ldots\ldots, n$$

والانحرافات المعدلة لا شك انها تحقق وفرا اكثر في الجهد الحسابي حيث تؤدي الي تبسيط قيم المفردات ويكون الانحراف المعياري للقيم الاصلية (X_i) ما هو الا عبارة عن الانحراف المعيارى للانحراف المعدلة(D_i) وذلك بعد ضربها في قيمة الوسط الفرضي (b) حيث ان الانحراف المعياري يتأثر فقط بعملية الضرب والقسمة كما سبق وان اسلفنا(الخاصية الثانية من خصائص الانحرافات المعيارية) اى ان :

$$\sigma_X = b\,\sigma_d = b\sqrt{\frac{\sum di^2}{n} - \left(\frac{\sum di}{n}\right)^2}$$

حيث أن : $d_i = (X_i - a) / b$

<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 .

<u>الحل</u> :-

<u>أ-حساب الإنحراف المعيارى باستخدام القيم الاصلية :</u>

الجدول التالى يوضح الحسابات اللازمة لحساب الإنحراف المعيارى من واقع استخدامنا للقيم الاصلية :

X_i	X_i^2
625	390625
700	490000
650	422500
725	525625
2700	1828750

وحيث أن :

فإنه :

$$\sigma = \sqrt{\frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2}$$

$$\sigma = \sqrt{\frac{1828750}{4} - \left(\frac{2700}{4}\right)^2} = 39.53$$

<u>ب- حساب الانحراف المعيارى باستخدام طريقة الانحرافات البسيطة :</u>

يتم اختيار اى من قيم المفردات الأصلية بمثابة وسط فرض ولتكن $a = 650$

والجدول التالى يوضح الحسابات اللازمة لحساب الانحراف المعيارى باستخدام طريقة الانحراف البسيطة :

وحيث أن :

$$\sigma = \sqrt{\frac{\sum di^2}{n} - \left(\frac{\sum di}{n}\right)^2}$$

فإنه :

$$\sigma = \sqrt{\frac{8750}{4} - \left(\frac{100}{4}\right)^2} = 39.53$$

X_i	$di = Xi - 650$	d_i^2
625	625-650= -25	625
700	50	2500
650	0	0
725	75	5625
	100	8750

وهى نفس النتيجة التى تم الحصول عليها في المطلوب السابق (أ) . لاحظ اثر تبسيط الحسابات فى هذه الطريقة عن السابقة .

ج – حساب الانحراف المعيارى باستخدام طريقة الانحرافات المعدلة :

بملاحظة عمود قيم المتغير (di) فى الطريقة السابقة يلاحظ انه يمكن قسمة قيم الانحرافات البسيطة عن الوسط الفرضى b=25 دون بواقى لذا يتم تكوين الجدول

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

التالى لايجاد المجاميع اللازمة لحساب الانحراف المعيارى باستخدام طريقة الانحرافات المعدلة وذلك على النحو التالى :

$$\sigma_X = b\,\sigma_{\triangleright} = b\sqrt{\frac{\sum Di^2}{n} - \left(\frac{\sum Di}{n}\right)^2}$$

وعليه فإن :

$$\sigma_X = 25\left(\left(14/4\right) - \left(4/4\right)^2\right)^{0.5} = 39.53$$

وهى نفس النتيجة التى حصلنا عليها فى (أ) ، (ب) مع تبسيط اكثر فى العمليات الحسابية وجهد التجميع .

ثانيا: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 وذلك من خلال إستخدام أيا من الطرق الثلاث السابقة (الطريقة المباشرة (او القيم الأصلية) – طريقة الانحرافات البسيطة – طريقة الانحرافات المعدلة أو المختصرة) وذلك مع اعتبار ان مراكز فئات التوزيع هى قيم (X_i) ويراعى ترجيحها بالتكرارات المناظرة لتلك المراكز او تلك الفئات اى (Fi) .

1 – الطريقة المباشرة لحساب الانحراف المعيارى فى حالة البيانات المبوبة :

اذا كانت لدينا مراكز فئات توزيع تكرارى هى $X_n,\ldots,X_2,X_1$ وما يناظرها من تكرارات $F_n,\ldots F_2,F_1$ 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 إنحرافات قيم مراكز فئات التوزيع (X_i) عن الوسط الحسابى للتوزيع ($\overline{X}$) اى نوجد $(X_i - \overline{X})$ لكل

فئـة مـن فئـات التوزيـع ثـم يـتم ايجـاد مربعـات انحرافـات المراكـز عـن الوسـط الحسـابى اى $(X_i - \overline{X})^2$ ثـم ضـرب نـواتج مربعـات انحرافـات المراكـز عـن وسطها الحسابى فى التكرارات المقابلة لها ثم ايجاد المجموع اى :

$$\sum(X_i - \overline{X})^2 \, F_i$$

فيكون الانحراف المعيارى للتوزيع التكرارى عبارة عن :

$$\sigma = \sqrt{\frac{\sum(X_i - \overline{X})^2 \, F_i}{\sum F_i}}$$

ويمكن تبسيط هذه الصورة الاخيرة كما سبق لتأخذ الصورة التالية :

$$\sigma = \sqrt{\frac{\sum {X_i}^2 \, F_i}{\sum F_i} - \frac{(\sum X_i \, F_i)^2}{\sum F_i}}$$

ويفضـل اسـتخدام هـذه الصـيغة عـن السـابقة لهـا لحسـاب الانحـراف المعيـارى خصوصا اذا كانت قيمة الوسط الحسابى تحتوى على قيما كسرية كما سبق وان بينا ذلك .

2- <u>طريقة الانحراف البسيطة</u> :

وهنا يتم اخذ احد مراكز الفئات بمثابة وسط فرضى وليكن (a) ويفضـل اختيـار قيمة الوسط الفرضى (a) على ان يكون مركز الفئة التى تتوسط فئات التوزيـع التكرارى او مركز الفئة الذى يقابل اكبر تكرار وذلك لتبسيط العمليات الحسـابية بشكل ملحوظ . فـاذا افترضـنا ان انحرافـات مراكـز فئـات التوزيـع التكـرارى عـن الوسط الفرض بمثابة القيم (d_i) حيث :

$$d_i = x_i - a \qquad \text{where} \quad i = 1 , 2 , \ldots\ldots , r$$

وتعبـر (r) عـن عـدد فئـات التوزيـع التكـرارى ثـم يـتم ايجـاد حاصـل ضـرب الانحرافات (di) فى التكرارات المقابلة ومن ثم المجموع اى $\sum$ di Fi ومـرة اخرى يتم ضرب $d_i F_i$ فى عمود الانحرافات (di) ثم الجمع لنحصل على المجموع $\sum d_i^2 F_i$ فيكون الانحراف المعيارى عبارة عن :

$$\sigma_X = \sigma_d = \sqrt{\frac{\sum di^2 F_i}{\sum F_i} - \left(\frac{\sum_{di} F_i}{\sum F_i}\right)^2}$$

<u>3 – طريقة الانحرافات المعدلة (الاكثر اختصارا) :</u>

اذا كانـت قـيم الانحـراف البسـيطة (d_i) المحسـوبة فـى الطريقـة السـابقة تقبـل القسمة علـى مقدار اوسـط فرضـى آخـر ولـيكن (b) بـدون بـواقى (وغالبـا مـا تكون قيمة (b) بمثابة طول الفئـة فى حالـة الجدول التكراريـة المنتظمـة) فانـه يفضل قسمة الانحرافات البسيطة (d_i) على الوسط الفرضى (b) لينـتج لنـا 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 التكـرار المقابل ثم ايجـاد المجموع اى ($\sum D_i F_i$) ومرة اخرى يتم ضرب عمـود ($D_i F_i$) فى عمـود الانحرافـات المختصـرة (D_i) ثـم الجمـع لينـتج $\sum D_i^2 F_i$ ثـم يـتم حسـاب 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 (4-11)

No. of defective units	0	1	2	3	4	5
Frequency	13	15	18	25	20	19

والمطلوب :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 (4 – 12)

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 الإنحراف المعيارى لعدد الوحدات المعيبة هو :

$$\sigma \sqrt{\frac{1107}{100} - \left(\frac{301}{100}\right)^2} = 1.42 \text{ unite}$$

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 حساب المجاميع اللازمة لحساب هذين المقياسين وذلك كما هو موضح بالجدول التالى :

جدول (15-4)

Classes	Fi	Mid point(Xi)	di =Xi-35	di fi	$d_i^2 F_i$
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 :

$$\sigma_X = \sigma_d = \sqrt{\frac{\sum di^2\, F_i}{\sum F_i} - \left(\frac{\sum di\, F_i}{\sum F_i}\right)^2}$$

$$= \sqrt{\frac{8900}{50} - \left(\frac{-124}{50}\right)^2} = 13.10914185$$

<u>مثال</u> :

فيما يلى التوزيع التكرارى التالى لاطوال عينة من طلاب كلية التجارة بالسنتيمتر .

جدول (4 -16)

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 ؟

<u>الحل</u> :

حيث ان التوزيع التكرارى المعطى توزيعا تكراريا منتظما(اى متساوى من حيث اطوال فئاته) لذا يفضل استخدام طريقة الانحرافات المعدلة والجدول

(17-4) يبين الحسابات اللازمة لايجاد كلا من الوسط الحسابى والانحراف المعيارى والتباين لطول الطالب فى العينة باستخدام طريقة الانحرافات المعدلة:

جدول (4 – 17)

Classes	F_i	X_i	$d_i=X_i-165$	$D_i = \frac{di}{10}$	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 هو :

$$\bar{X} = a + b\,\frac{\sum \text{Di Fi}}{\sum \text{Fi}}$$

لذا فان :

$$\bar{X} = 165 + 10\left(\frac{61}{50}\right) = 177.2 \qquad \text{(C.M)}$$

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 (σ^2) فهو عبارة عن مربع الانحراف المعيارى

اى ان : $\sigma^2 = (11)^2 = 121 \quad (C.M)$

<u>*الانحراف المعيارى المصحح لشبرد (تصحيح الانحراف المعيارى للبيانات المبوبة) :</u>

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 كما ذكرنا سابقا تحت افتراض ان قيم المفردات تتوزع بانتظام (اى متجانسة الانتشار) على مدار طول كل فئة من فئات الجدول وهو ما يدعونا الى استخدام مركز كل فئة كقيمة تأخذها كل مفردة من المفردات(تكرار) داخل هذه الفئة. وهذا الافتراض وإن كان لا يشوبة الدقة الا إنه 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 واقع القيم الاصلية قبل تبويبها(وهو مايسمى بأخطاء عملية تبويب البيانات) . وهو ما يستلزم تصحيح المقاييس المحسوبة

من التوزيعات التكرارية ولكن هذا التصحيح يتوقف على عوامل متعددة منها شكل التوزيع وطبيعة المقياس 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 الخاصة بالتوزيع 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 لتصبح 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 σ_{cor} نقصد بها الانحراف المعيارى المصحح لشبرد ، (L) تعبر عن طول الفئة الغالب للتوزيع التكرارى .

هذا وتجدر الاشارة الى انه 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 1000 مفرده مثلا) او اذا كان عدد الفئات قليلا (اقل من 20 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

المعـروف بالاسـى السـالب او الاسـى الموجـب) او فـى حالـة التوزيعـات المستطيلة.

<u>**مثال :**</u>

احسب الانحراف المعيارى المصحح للتوزيع الوارد فى المثال السابق جدول (16-4) .

<u>**الحل :**</u>

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 المعيارى المصحح لهذا التوزيع 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 :-

$$\sigma = \sqrt{\frac{\sum X_i^2 F_i}{\sum F_i} - \left(\frac{\sum X_i^2 \ F_i}{\sum F_i}\right)^2}$$

$$= \sqrt{\sum X_i^2 \ \frac{F_i}{\sum F_i} - \left(\sum X_i \frac{F_i}{\sum F_i}\right)^2} \qquad \textbf{(1)}$$

وحيث ان التكرار النسبى ما هو الا تكرار الفئة منسوبا الى مجموع التكرارات اى ان :

$$R_i = \frac{F_i}{\sum F_i} \qquad where \quad i = 1, 2, 3, \dots, n$$

حيث (n) تمثل عدد فئات التوزيع

لذا فأنه بالتعويض عن التكرارات النسبية فى القانون (1) 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 (4 – 18)

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 :

1- **حساب متوسط درجة الحرارة المتوقعة بالمملكة فى هذا اليوم .**

2- **حساب الانحراف المعيارى لدرجة الحرارة المتوقعة فى هذا اليوم .**

<u>الحل :</u>

لاحظ ان التوزيع المعطى فى جدول (4-18) توزيعا تكراريا منتظما لذا يفضل استخدام طريقة الانحرافات المعدلة فى حساب كلا من الوسط الحسابى والانحراف المعيارى والجدول (4-19) يبين الحسابات اللازمة لذلك :

جدول (4-19)

Classes	R_i	X_i	$D_i = \dfrac{X_i - 30}{5}$	$D_i R_i$	$D_i R_i^2$
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 :

قيمة 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 عن قيمة الانحراف المعيارى فهى :

$$\sigma = b\sqrt{\sum D_i^2 R_i - \left(\sum D_i R_i\right)^2}$$

$$= 5\sqrt{1.53 - (0.11)^2} = 6.16 \qquad (degree)$$

ملاحظات حول مقاييس التشتت المطلقة :-

1- **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 Statistical Quality Control**

2- **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 وذلك لأن قيمتة لا تختلف كثيرا من عينة لاخرى مسحوبة من نفس المجتمع محل الدراسة .**

3- رغم ان الانحراف المعيارى ياخذ فى الاعتبارجميع قيم المفردات عند قياس التشتت الا انه يعتبر اكثر تأثيرا بالقيم الشاذة او المتطرفة كما هو الحال بالنسبة للوسط الحسابى كاحد مقاييس المتوسطات الذى يتأثر بدرجة كبيرة بالقيم الشاذة او المتطرفة ، اما 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 الا أنه من الممكن تقليل تأثير القيم الشاذة على الانحراف المتوسط وذلك من خلال استخدام الانحراف المتوسط عن الوسيط بدلا من الإنحراف المتوسط عن الوسط الحسابى فى ايجادة .

4- يعاب ايضا على الانحراف المعيارى انه لا يمكن استخدامه وحسابه فى التوزيعات التكرارية المفتوحة كما لا يمكن حسابه بيانيا وفى هذه الحالة يفضل استخدام نصف المدى الربيعى لدراسة تشتت التوزيع .

5- 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 فى عملية المقارنة وذلك بما يتناسب مع طبيعة البيانات وذلك حتى لا تكون النتائج المتحصل عليها مضللة فى عملية المقارنة .

6- 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 = 1.2 نصف المدى الربيعى

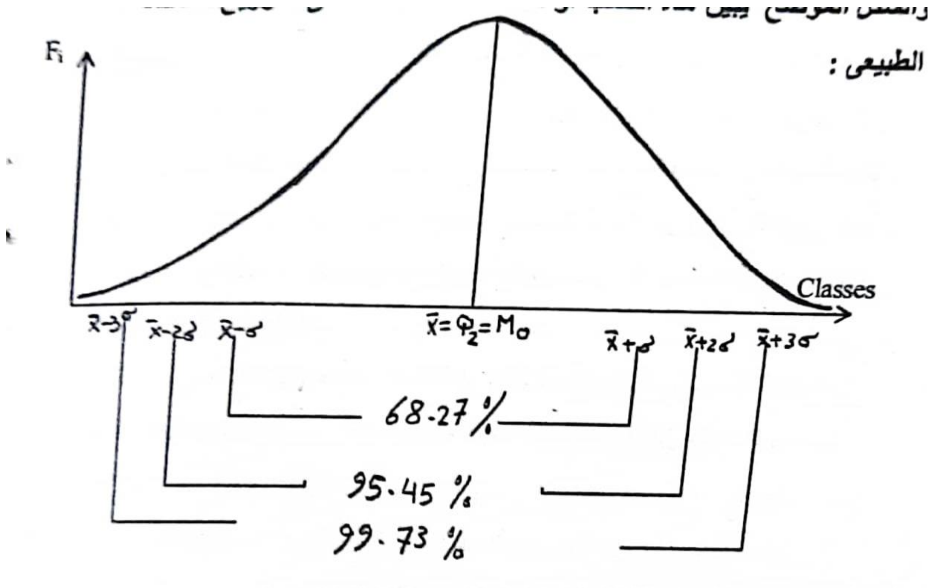
اخيرا فمن اهم خصائص التوزيع المعتاد الطبيعى **Normal Distribution** ما يلى :

* ان 68.27% من جميع المفردات تقع فى المدى ما بين $(\bar{X} \mp \sigma)$.

* ان 95.45% من جميع المفردات تقع فى المدى ما بين $(\bar{X} \mp 2\sigma)$.

* وان 99.73 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 تقترب او توزيعها يقترب من التوزيع المعتاد الطبيعى وذلك اذا كان حجم او عدد المشاهدات كبيرة نسبيا (نظرية الحد المركزية Centeral Limit Theorem) .

<u>ثانيا : مقاييس التشتت النسبية (معاملات الختلاف) :</u>

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 كدراسة مقارنة التشتت فيما بين الاوزان لمجموعة معينة من الأفراد (بالكيلوجرام مثلا) وبين تشتت أعمارهم (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 فى 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 اخر وذلك من خلال استخدام مقاييس نسبية للتشتت تقيس التشتت بوحدات من الوسط الحسابى وهى ما تسمى 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 :

(1):- تتخلص المقاييس النسبية من اثر الإختلافات فى قيمة المتوسطات والتى ربما قد توجد فيما بين المجموعات.

(2):- حيث أن المقاييس النسبية تاتى على شكل نسبة مئوية خالية من التمييز لذا فيمكن استخدامها لمقارنة تشتت مجموعتين سواء متحدتين او مختلفين فى وحدات قياسها .

وبصفة عامة فأن الصورة العامة لمعامل الاختلاف ياخذ الصورة التالية :

$$\text{معامل الاختلاف} = \frac{\text{مقياس للتشتت المطلق}}{\text{مقياس من مقاييس المتوسطات}} \times 100$$

على ان يكون مقياس المتوسطات المستخدمة فى المقام لـه نفس طبيعة (أو طريقة حساب) مقياس التشتت المطلق المستخدمة فى البسط : لذا فأن هناك مقياسين اساسيين من معاملات الإختلاف لدراسة التشتت النسبى وهما :

(1):<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 :

$$\text{معامل الاختلاف النسبى} = \frac{\text{الاختلاف المعيارى}}{\text{الوسط الحسابى}} \times 100$$

أى أن معاما الإختلاف النسبى Cofficient of Variation هو عبارة عن :

$$C_V = \frac{\sigma}{\overline{X}} \times 100$$

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 إلا من خلال وضع إفتراضات معينة على التوزيع التكرارى تؤدى من شأنها إلى غلق الفئات 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 .

(2):معامل الاختلاف الربيعى Quartile Coefficient of Variation

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 :-

$$\text{معامل الربيعى الاختلاف} = \frac{\text{نصف المدى الربيعى}}{\text{الوسيط}} \times 100$$

$$Q.C_V = \frac{Q_1 - Q_1}{2\ Q_2} \times 100$$

هذا وفى حالة ما اذا كان التوزيع محل الدراسة قريبا من التماثل فأن قيمة الوسيط تساوى تقريبا الوسط الحسابى للربيعين الادنى والاعلى .

ومن ثم تكون قيمة الوسيط هى :

$$Q_2 = \frac{Q_1 + Q_3}{2}$$

وبالتعويض عن قيمة الوسيط فى صيغة معامل الاختلاف الربيعى السابقةيكون:

$$Q.C_V = \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$$

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:-

<u>فى حالة المفردات :</u>

$$C_V.M_D = \frac{\frac{\sum|X - \bar{X}|}{n}}{\bar{X}} \times 100$$

or

<u>وفى حالة البيانات المبوبة تكون الصيغة الإحصائية لهذا المعامل على الصورة:</u>

$$= \frac{\frac{\sum|X_i - \bar{X}|}{\sum F_i}}{\bar{X}} \times 100$$

<u>*معامل الاختلاف بدلالة الانحراف المتوسط عن الوسيط وياخذ الصورة التالية:</u>

$$C_V.M_Q = \frac{\frac{\sum|X_i - Q_2|}{n}}{Q_2} \times 100$$

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 فيما بين المجموعات المختلفة .

مثال :

اجريت دراسة لمجموعة مكونة من 80 عاملا باحد المصانع شملت العمر والاجر الاسبوعى وحصلنا على البيانات التالية :

بالنسبة للعمر كأن : الوسط الحسابى = 35 ، التباين = 49

اما بالنسبة للاجر الاسبوعى فكان موزعا على النحو المبين بالجول 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 ؟

الحل : 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 لحساب معامل الإختلاف النسبى للدخل الأسبوعى لذا يتم تكوين الجدول التالى لحساب المجاميع اللازمة لحساب كلا من الوسط الحسابى والإنحراف المعيارى:

جدول (4- 20)

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 :

$$\bar{X} = a + b\,\frac{\sum DiFi}{\sum Fi} = 75 + 4\left(\frac{0}{80}\right) = 75 \quad (L.E)$$

كما أن الإنحراف المعيارى هو :

$$\sigma = b\sqrt{\frac{\sum D_i^2\,F_i}{\sum Fi} - \left(\frac{\sum Di\,F_i}{\sum Fi}\right)^2}$$

$$= 4\sqrt{\frac{80}{80} - \left(\frac{0}{80}\right)^2} = 4(L.E)$$

ومن ثم فأن معامل الاختلاف النسبى للدخل الاسبوعى هو:

$$C_V = \frac{\sigma}{\bar{X}} \times 100 = \frac{4}{75} \times 100 = 5.3\ \%$$

وبمقارنة معاملى الاختلاف للعمر (20%) والدخل الاسبوعى (5.2%) نجد ان توزيع العمر اكثر تشتتا من توزيع الدخل الاسبوعى .

<u>مثال</u> : فيما يلى لديك توزيع درجات مجموعة من الطلاب فى مادة الإحصاء:

Classes	Less than 5o	50-	60-	70-	80 and more
Frequencies	5	15	55	20	5

والمطلوب حساب معامل الاختلاف المناسب لهذا التوزيع .

<u>الحل :</u>

حيث ان الجدول التكرارى الذى يعبر عن توزيع الدرجات جدولا تكراريا مفتوح الطرفين . لذا يتعين حساب معامل الاختلاف الربيعى كذلك فأنه من شكل توزيـع التكرارات يتضح من الجدول ان التوزيـع قريبـا مـن التماثـل . لـذا يمكـن حسـاب الـربيعين الادنـى والاعلـى فقـط دون الحاجـة لحسـاب الوسـيط اى يـتم اسـتخدام صيغة معامل الاختلاف الربيعى المستخدمة للتوزيعات القربية 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 لايجـاد قيمـة كـل من Q_3 , Q_1 :

Classes	F_i	Less then the Class U.L	Asc.C.F
Less than 50	5	Less then 50	5
50-	12	Less then 60	20
60-	55	Less then 70	75
70-	20	Less then 80	95
80 and more	5	Less then the Latest .U.L.	100
∑	100		

وحيث ان :

$$Rank\ of\ Q_1 = \frac{\sum F}{4} \times 1 = \frac{100}{4} = 25 \qquad ,$$

$$Rank\ of\ Q_3 = \frac{\sum F}{4} \times 3 = \frac{100}{4} \times 3 = 75$$

وعليه فأن :

$$Q_1 = 60 + \left(\frac{25-20}{75-20}\right) \times 10$$

$$= 60 + \left(\frac{5}{55}\right) \times 10 = 60.9 \qquad (degree) \qquad ,$$

$$Q_3 = 75 \qquad (degree)$$

لاحظ ان ترتيب Q_3 كان بمثابة احد التكرارت المتجمعة 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 مقارنة لمستوى الاجور الشهرية للعمال فى قطاع صناعة الغزل والنسيج وقطاع الصناعات الدوائية تم اخذ عينة عشوائية من عمال كل قطاع من القطاعين وحصلنا على النتائج التالية لكل منهما :

العينة الاولى (من قطاع الغزل والنسيج)	العينة الثانية (من قطاع الصناعات الدوائية)
$\sum_{i=1}^{30} X_{1i} = 12000$ ومجموع مربعات المفردات هو: $\sum_{i=1}^{30} X\ {}^{i}_{1i} = 4817280$	$\sum_{i=1}^{50} X_{2i} = 22500$ ومجموع مربعات المفردات هو : $\sum_{i=1}^{50} X\ {}^{i}_{2i} = 10141200$

والمطلوب :

(1):احسب الوسط الحسابى والانحراف المعيارى لكل عينة .

(2): قارن بين العينات من حيث التشتت موضحا ايها الاكثر تشتتاً .

(3):اذا تم ادماج العينين فى عينة واحدة فاحسب متوسط العينة الجديدة وكذا معامل الاختلاف لها .

الحل :-

(1):حساب الوسط الحسابى والانحراف المعيارى لكل عينة على حده.

فأنه بالنسبة للعينة الاولى: فإن الوسط الحسابى هو :

$$\bar{X}_1 = \frac{\sum_{i=1}^{30} X_{1i}}{30} = \frac{12000}{30} = 400 \ \ (\mathrm{L.E})$$

اما بالنسبة للانحراف المعيارى للعينة الاولى فهو:

$$\sigma_1 = \sqrt{\frac{\sum x_{1i}^2}{n} - \left(\frac{\sum x_{1i}}{n}\right)^2} = ((4817280/30) - (12000/30)^2)^{0.5}$$

$$= 24 \qquad (\mathrm{L.E.})$$

أما <u>بالنسبة للعينة الثانية</u> :فإن الوسط الحسابى والإنحراف المعيارى هما :

$$\bar{X}_1 = \frac{\sum_{i=1}^{50} X_{2i}}{50} = \frac{22500}{50} = 450 \ \ (L.E)$$

أما الإنحراف المعيارى للأجر فإن:

$$\sigma_1 = \sqrt{\frac{\sum x_{2i}^2}{n} - \left(\frac{\sum x_{2i}^2}{n}\right)^2} = \sqrt{\frac{10141200}{50} - \left(\frac{22500}{50}\right)^2}$$

$$= 18 \quad (L.E)$$

ب- ولمقارنة تشتت المجموعتين أى 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 المقارنة من خلال احد معاملات التشتت النسبى والتى تدرس ماهية وجود إختلافات من عدمها فيما بين عينتى القطاعين إن وجدت . وأنسب مقاييس التشتت النسبى في هذه الحالة هو معامل الإختلاف النسبى والذى يأخذ الصورة التالية :

$$C_V \ for \ the \ first \ sample = \frac{\sigma_1}{\bar{X}_1} \times 100$$

$$= \frac{24}{400} \times 100 = 6\,\%$$

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 الاختلاف النسبى يتضـح لنـا ان اجـور العينـة الاولـى تعتبر الاكثر تشتتا (اى الاقل تجانسا) من اجـور العينـة الثانيـة مستخدمين فى عملية المقارنـة معامل الاختلاف النسبى كاحد مقاييس التشتت النسبى .

جـ - <u>اذا تم دمج عينتيى القطاعين فى عينـة واحدة فأن :</u>

حجم العينـة الجديدة (الكلية) هو :

$$n = n_1 + n_2 = 30 + 50 = 80$$

والوسط الحسابى للعينـة الجديدة هو :

$$\bar{X} = \frac{n_1\,\bar{X}_1 + n_2\,\bar{X}_2}{n_1 + n_2}$$

اما عن تباين العينـة الجديدة (الكلية) فأننا سوف نورد صيغة التباين المشترك للعينتين (دون اثبات) والتى تأخذ الصورة التالية :-

$$\sigma^2 = \frac{n_1\left(\sigma_1^2 + D_1^2\right) + n_2\left(\sigma_2^2 + D_2^2\right)}{n_1 + n_2}$$

حيث D_2, D_1 هى عبارة عن انحراف الوسط الحسـابى لكل عينـة عن الوسط الحسابى العام او الفرق ما بين متوسط العينـة والمتوسط العام (للعنيـتن) اى ان :

$$D_1 = \bar{X}_1 - \bar{X} = 400 - 431.25 = 31.25 \qquad (L.E)$$

ومن ثم فأن :

$$D_1^2 = (-31.25)^2 = 976.5625$$

كذلك فأن انحراف العينة الثانية عن المتوسط العام هو :

$$D_2 = \bar{X}_2 - \bar{X} = 450 - 431.25 = 18.75 \qquad (L.E)$$

وعليه فأن مربع هذا الانحراف هو :

$$D_2^2 = (18.75)^2 = 351.5625$$

ومن ثم فبالتعويض فى صيغة التباين المشترك فيكون :

$$\sigma^2 = \frac{n_1\left(\sigma_1^2 + D_1^2\right) + n_2\left(\sigma_2^2 + D_2^2\right)}{n_1 + n_2}$$

$$= \frac{30\,(576 + 976.57625) + 50(324 + 351.5625)}{30 + 50}$$

$$= 1004.4375$$

وعليه فأن الانحراف المعيارى للعينة الكلية الجديدة هو :

$$\sigma = \sqrt{1004.4375} = 31.693 \qquad (L.E)$$

ومن ثم فأن معامل الاختلاف النسبى للعينة الكلية الجديدة هو:

$$C_V = \frac{\sigma}{\bar{\bar{X}}} \times 100 = \frac{31.693}{431.25} = 7.35\ \%$$

الدرجات المعياريه : Standard Scores

تناولنا فيما سبق كيفية المقارنة فيما 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 من 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 75 درجة بينما كانت درجته فى مادة الاقتصاد هى 80 درجة . فهل هذا يعنى ان 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 80 درجة اكبر من درجته فى مادة الاحصاء 75درجة . 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 فالمسألة 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 كان على درجة عالية من الصعوبة بحيث ان يكون بقية زملائه حصلوا على درجات منخفضة فى تلك المادة وان هذا الطالب يعتبر نسبيا 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 آخر ياخذ فى الاعتبار الاداء النسبى للطالب فى كل مادة من المادتين على حده والمقصود بالاداء النسبى هنا هو مستوى اداء الطالب فى مادة معينة مقارنة 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 المعيارى لدرجات الطلاب فى مادة الاقتصاد هما 71 , 6 درجات على الترتيب . لذا 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 مادة أو 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 $\bar{X}$ ، (σ) فهما 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 ($X - \bar{X}$) 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 المعياريتين 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 انه 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 نؤكد على

ان القيم المعيارية عبارة عن قيم بدلاله وحدات من الانحراف المعيارى لذلك فهى تكون مجردة من وحدات القياس وتستخدم بذلك لمقارنة قيم مختلفة من مجموعات مختلفة وللقيم المعيارية عدة خصائص أهمها ما يلى :

1- **الوسط الحسابى للقيم المعيارية يساوى الصفر دائما:**

فبغض النظر عن الوسط الحسابى للقيم الاصلية يمكن اثبات ذلك أن الوسط الحسابى للقيم المعيارية مساويا الصفر وذلك على النحو التالى :

فحيث ان الدرجة المعيارية للدرجة الطبيعية (x_i) هي (z_i) حيث أن :

$$\bar{Z} = \frac{X_i - \bar{X}}{\sigma}$$

فيكون الوسط الحسابي للدرجات المعيارية اي ($\bar{z}$) عبارة عن :

$$\bar{Z} = \frac{\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\sigma}\right)}{n} = \frac{1}{n\,\sigma}\sum_{i=1}^{n}(X_i - \bar{X})$$

وحيث ان مجموع انحرافات القيم عن وسطها الحسابي يساوي صفرا . لذا فأن

$$\bar{Z} = \frac{1}{n\,\sigma}\ (0) = 0$$

2- **التباين او الانحراف المعياري للقيم المعياريه يساوي الواحد الصحيح**

ويمكن اثبات ذلك علي النحو من خلال مايلى . فحيث أن :

$$\sigma_Z^2 = \frac{\sum_{i=1}^{n}(Z_i - \bar{Z})^2}{n} = \frac{\sum_{i=1}^{n} Z_i^2}{n}$$

وذلك لان $\bar{Z}$ تساوي الصفر (من الخاصية السابقة) . وحيث ان الدرجة المعيارية (z)

هي :
$$\bar{Z} = \frac{X_i - \bar{X}}{\sigma}$$

$$\sigma_Z^2 = \frac{\sum_{i=1}^{n} Z_i^2}{n} = \frac{\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\sigma}\right)^2}{n} = \frac{1}{\sigma^2} \cdot \frac{\sum_{i=1}^{n}(X_i - X)^2}{n}$$
$$= \frac{1}{\sigma^2}\left(\sigma^2\right) = 1$$

ومـن خـلال الخاصـيتين السـابقتين يتضـح لنـا انـه لمعـايرة مجمـوعتين مـن المفردات فان هذا يعني اننا نردها الي مجموعتين لها نفس المتوسط الحسـابي (وهو الصفر) ونفس الانحراف المعياري (وهو الواحد الصحيح) .

مثال :

في إمتحـان الشـهاده الاعداديـة بمحافظـة قنـا وجد احد الطـلاب ان درجاتـه في الرياضيات واللغة العربية واللغة الانجليزية كانت علي الترتيب هي 45 ، 40 ، 42 درجة . فاذا علمت ان متوسط درجات الطالب بالمحافظة في المواد الثلاثـة السـابقة هـو علـي الترتيـب 42 ، 35 ، 40 درجـة . وان الانحـراف المعيـاري للدرجات في تلك المواد الثلاثة بالمحافظة هو علي الترتيب 6 ، 4 ، 5 فمـا هو الحكم الموضوعي علي مستوي هذا الطالب في المواد الثلاثة ؟

الحل :

لكـي يكـون الحكم صـحيحا علـي مستوي اداء الطالـب فـي المـواد الـثلاث يجـب حساب الدرجات المعيارية للطالب فى تلك المواد الثلاث وذلك على النحوالتالى:

فحيث أن الدرجة المعيارية للطالب في مادة الرياضيات ولتكن (Z_1) هي :

$$\bar{Z}_1 = \frac{X_1 - \bar{X}_1}{\sigma_1} = \frac{45 - 42}{6} = 0.5$$

بينما الدرجة 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 الإنجليزية ولتكن (z_3) فهي :

$$\bar{Z}_3 = \frac{X_3 - \bar{X}_3}{\sigma_3} = \frac{42 - 40}{5} = 0.4$$

ومن خلال مقارنة الدرجات المعيارية الثلاث و التى تعطى بيانا موضوعيا عن مستوى الطالب فى المواد الثلاث نجد أن مستوى الطالب فى مادة اللغة العربية أفضل من مستواه فى مادة الرياضيات وكذا أفضل من مستواه فى مادة اللغة الإنجليزية. وهذا الحكم هو الحكم الصحيح على مستوى أداء الطالب فى المواد الثلاث وهو عكس ما توحى به الأرقام المجردة لدرجات الطالب فى المواد الثلاثة الطبيعية .

تمارين الفصل الرابع

(1):الجدول التالي يبين توزيع عدد الاطفال لعدد معين من الاسر:

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 .

(2):البيانات التالية تمثل اعمار مجموعتين من الاشخاص (بالسنوات) :

اعمار المجموعه الاولي هى :

17, 22 , 10, 18 , 27 , 15 , 13 , 30

واعمار المجموعه الثانية هى :

38 , 41 , 33 , 39 , 41 , 73, 36 , 35

والمطلوب :-

أ – قارن بين التشتت في المجموعتين وذلك باستخدام المدى 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 نتائجك في (ا) .

(3): 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 in A	3	61	132	153	140	51	2
No. of families in B	2	14	20	27	28	7	2

والمطلوب :

1-قارن بين متوسط مستوي الانفاق في المدينتين باستخدام المقياس المتوسط الملائم .

2-قارن بين تشتت الانفاق في المدينتين باستخدام مقياس التشتت المطلق والنسبى الملائم .

(4):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 .

ج- معامل الاختلاف المناسب .

(5):فيما يلي التوزيع التكراري المتجمع الصاعد لاعمار عينه من الاشخاص 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 .

ج- اوجد مقياس النزعه المركزية الملاءم .

(6):الجدول التالي يوضح توزيع اجور 160 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

<u>والمطلوب حساب</u> :

1- الوسط الحسابي والوسيط والمنوال .

2- الانحراف المتوسط.

3- الانحراف الربيعي.

4- الانحراف المعياري .

5- معامل الاختلاف

6- دراسة تماثل التوزيع من خلال نتائجك في(1) .

(7):الجدول التكراري التالي يمثل توزيع الاوزان بالكيلو جرام لعينه من 80 شخصا :

Weighted Classes	62-	66-	70-	74-	78-	82-	86-90
Frequency	3	8	20	21	14	10	4

والمطلوب حساب :

1- الوسط الحسابي والانحراف المعياري للتوزيع.

2- الوسيط والانحراف الربيعي .

(8):الجدول التالى يبين توزيع أطوال (بالسنتيمترات) لعينه من 100 شخصا:

Length Classes	150-	155-	160-	165-	170-175
Frequency	8	16	43	29	4

والمطلوب هو حساب الانحراف المتوسط والانحراف المعياري .

(9): الجدول التالي يمثل توزيع درجات 100 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 :

1- حساب الوسيط والمنوال .

2- الوسط الحسابي والانحراف المعياري .

3- الانحراف الربيعي والانحراف المتوسط .

4- معامل الاحتلاف الملائم .

(10):اوجد المدي والانحراف المتوسط والانحراف المعياري للقيم التالية :

(أ): 3 ، 4 ، 12 ، 6 ، 7 ، 4 ، 8 ، 3 ، 5

(ب): 8 ، 8 ، 5 ، 6 ، 6 ، 10 ، 6 ، 8 ، 4 ، 9 ، 4 ، 6

(11):اوجد الانحراف المعياري للبيانات التالية :

أ- 3 ، 6 ، 2 ، 1 ، 7 ، 5

ب- 2 ، 3 ، 6 ، 4 ، 8 ، 2 ، 2 ، 5 ، 4 ، 4

ج- 0 ، 0 ، 0 ، 0 ، 0 ، 1 ، 1 ، 1

(12):الجدول التالي يوضح توزيع تكلفة الاقامة في الغرفة الواحده لعدد 200 سائح في احد الفنادق :

Cost Classes	30-	40-	50-	60-	70-	80-	90-	∑
No. of tourists	16	22	36	55	31	25	10	200

<u>والمطلوب حساب</u> :

1- الوسط الحسابي والوسيط والمنوال .

2- الانحراف المتوسط.

3- الانحراف الربيعي .

4- الانحراف المعياري .

5- معامل الاختلاف .

(13):الجدول التكراري التالي يبين مبيعات أحد منتجات خان الخليلي حسب السعر فكانت كالتالي :

Total Seeling Classes	75-	77-	79-	81-	83-	85-	∑
Frequency	3	23	52	15	7	2	102

والمطلوب حساب :

1- الوسط الحسابي والوسيط والمنوال .

2- الانحراف المتوسط .

3- الانحراف المعياري .

4- معامل الاختلاف .

(14):فيما يلي بيان بدرجات عينية من طلاب معهد السياحه والفنادق بحلوان واخري من طلاب معهد السياحه والفنادق بمدينة السادس من اكتوبر في العام الماضي لمادة 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

والمطلوب :- أحسب الاتي :

1- الوسط الحسابي لدرجة الطلاب والانحراف المعياري ومعامل الاختلاف لعينة طلاب معهد حلوان .

2- الوسيط والربيع الادني والاعلي لطلاب عينة 6 اكتوبر حسابيا وبالرسم .

3- المنوال حسابيا وبالرسم لطلاب عينة حلوان .

4- نصف المدي الربيعي ومعامل الاختلاف الربيعي لطلاب عينة معهد السادس من اكتوبر.

5- قارن بين تشتت التوزيعين .

6- اوجد عدد الطلاب الذين يحصلون علي درجات اقل من 65 درجه لطلاب عينة حلوان ونسبتهم .

7- اوجد نسبة الطلاب اللذين تزيد درجاتهم عن 78 درجة لطلاب مدينة السادس من اكتوبر .

8- اوجد عدد الطلاب اللذين تتراوح درجاتهم فيما بين 58 ، 68 درجة لطلاب عينة حلوان .

9- اوجد الوسط الحسابي والوسيط والمنوال لطلاب مدينة السادس من اكتوبر .

10-هل يمكن القول ان طلاب معهد السياحه والفنادق بمدينة السادس من اكتوبر اكثر كفاءة من طلاب معهد السياحه والفنادق بحلوان .

11-عرض بيانات التوزيعين في شكل مدرج تكراري .

الفصل 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 . هذا ويتم حساب عزوم مجموعة من القوى بانه عبارة عن مجموع حواصل ضرب كل قوة فى ذراع عزمها . والعزم فى الاحصاء ناتج عن تشبيه قيمة الظاهرة بذراع 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 :

(1):العزوم حول الصفر.

(2):العزوم حول الوسط الحسابى.

(3):العزوم حول وسط فرض معين.

(4):العزوم المعيارية .

وبصفة عامة اذا كان لدينا متغيرا يأخذ مجموعة القيم ولتكن (X_i) مجموعة أى أن مجموعة القيم هى X_1 , X_2 ,....., X_n فان العزم ذو الرتبة (r) حول وسط فرضي وليكن (∝) هى عبارة عن متوسط انحرافات القيم (X_i) عن هذا الوسط الفرضى (∝) مرفوعه للقوة (r) . وهذه القوة (r) تفيد فى تحديد قيم العزوم المتتالية الرتبة اى العزم الاول والعزم الثانى و ...... وهكذا لاى رتبة موجبة . ومن ثم فانه يمكن وضع الصيغة العامة للعزم ذو الرتبة (r) حول الوسط الفرضي وليكن (∝)على انه :

$$M_r^{(\propto)} = \frac{\sum_{i=1}^{n}(X_i - \propto)^r}{n} \qquad (1)$$

وهذه الصيغة (1) 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 المعبرة عن مراكز الفئات أي (X_i) ومن ثم فان الصيغة العامة

للعزم ذو الرتبة (r) حول الوسط الفرضى ($\propto$) فى حالـة البيانـات المبوبـة 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 تـم التعويض عن $r = 1$ فـى اى مـن المعـادلتين (1) او (2) نحصـل علـى العزم الاول حول الوسط الفرض ($\propto$) واذا تـم التعويض عن $r = 2$ فـى هـاتين المعادلتين نحصل على العزم الثانى حول الوسط الفرضى ($\propto$)...... وهكذا .

وطبقا للانواع المختلفة من العزوم السـابق الاشـارة اليـها فأنـه اذا تـم التعويض عن قيمـة الوسـط الفرضى ($\propto$) فـى المعـادلتين (1) ، (2) بالقيمـة صفر فاننـا نحصـل علـى العـزوم المتتاليـة حـول الصـفر . امـا عـن العـزوم حـول الوسـط الحسابى فأنـها تنتج حينما يتم التعويض فى هاتين المعـادلتين عن قيمـة الوسط الفرضى على انـها تساوى الوسط الحسابى للتوزيع اى ($\overline{X}$) واخيرا فإن العزوم حول وسط فرضى معين وليكن (a) فأنـها تنتج من خلال التعويض المباشر فى المعـادلتين (1) او (2) عن قيمـة $a = \propto$. امـا العزوم المعياريـة ففيـهـا يـتم ايجـاد البسـط فـى المعـادلتين (1) ، (2) مـن خـلال حسـاب متوسـط الـدرجات المعياريـة وهو مـا سيرد فيمـا بعد الا انـه يمكـن القول ان العزوم حـول الصفر والعزوم حول الوسط تعتبر الاكثر استخداما فى مجـال التحليـل الاحصـائى وفيمـا يلى نتنـاول الانواع المختلفة للعزوم وسوف نكتفى فى كل نوع من أنواع العزوم بالاشارة الى كيفية ايجـاد العزوم الاربعـة الاولـى ويمكـن بـالطبع الحصـول علـى العزوم من قوة اكبر اى بدءا من العزم الخـامس والسـادس و ..... وهكذا .

1- <u>العزوم حول الصفر (العزوم اللامركزية) :-</u>

من خلال المعادلتين (1) ، (2) السابقتين فأنه عند التعويض عن قيمة ∝ على أنه تساوى الصفر أى ∝= صفر فتصبح صيغة العزم ذو الرتبة (r) حول الصفر($M^{(0)}{}_r$)فى حالة البيانات الغير مبوبة على الصورة :

$$M_r = M_r^{(0)} = \frac{\sum_{i=1}^{n}(X_i - 0)^r}{n} = \frac{\sum_{i=1}^{n} X_i^r}{n} \qquad (3)$$

حيث r = 1,2,……,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,….,n كما ان (n) تعبر عن عدد فئات التوزيع فى هذه الحالة .

ومن خلال المعادلات (3) ، (4) يمكن بيان الصيغة الإحصائية للعزوم الاربعة الاولى حول الصفر وذلك فى حالتى البيانات الغير مبوبة والبيانات المبوبة كما هو موضح بالجدول (5-1) التالى :

جدول (5-1)
العزوم الأربعة الأولى حول الصفر

Case **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 فى حالة البيانات المبوبة تكراريا فان العزم الاول حول الصفر ما هو الا الوسط الحسابى اى ان :

$$M_1^{(0)} = \bar{X}$$

كما أن العزم الثانى حول الصفر هو الحد الأول فى صيغة التباين.

2- العزوم حول الوسط الحسابى (العزوم المركزية) :

تكمن اهمية هذا النوع من انواع العزوم فى انها تدرس العزوم حول نقطة مركز التوزيع (وهى بمثابة قيمة الوسط الحسابى) لذا تسمى بالعزوم المركزية .

هذا ويمكن ايجاد العزوم المركزية من خلال التعويض عن $\bar{X} = \propto$ فى المعادلتين (1) ، (2) السابقتين ومن ثم فان صيغة العزم ذو الرتبة (r) حول الوسط 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,……..

حيث (X_i) تعبر عن مركز الفئة (i) ، (F_i) تعبر عن تكرار الفئة (i) i=1,2,……n ، (n) تعبر عن عدد فئات التوزيع التكرارى .

والجدول (2-5) يلخص صيغة العزوم الاربعة الاولى حول الوسط (المركز) وذلك فى حالة البيانات المفردة (الغير مبوبة) وكذلك البيانات المبوبة .

جدول (5- 2)

Case 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 انه من خلال التعريف السابق للعزوم حول الوسط الحسابى فأن العزم الثانى حول الوسط الحسابى ما هو الا تباين التوزيع لذا فهو دائما ابدا مقدارا موجبا فيما عدا فى حالة القيم الثابتة فتباينها او تشتتها صفرا بطبيعة الحال لان الثابت لا يختلف من حيث القيم .

3- *العزوم حول وسط فرضى معين:*

فى المعادلتين (1) ، (2) السابق الاشارة إليهما اذا تم التعويض عن $\propto = a$ حيث (a) عبارة عن مقدار ثابت معين فأننا نحصل على العزوم المختلفة حول

هذا الوسط الفرضى (a) . ومن ثم فان صيغة العزم ذو الرتبة (r) حول وسط فرضى معين وليكن (a) فى حالة المفردات تأخذ الصورة التالية :

$$M_r^{(a)} = \frac{\sum (X_i - a)^r}{n} \qquad \text{where } r = 1,2,3,\ldots\ldots \qquad (7)$$

اما عن صيغة العزم ذو الرتبة (r) حول الوسط الفرضى (a) فى حالة البيانات المبوبة تكراريا فتأخذ الصورة التالية :

$$M_r^{(a)} = \frac{\sum (X_i - a)^r F_i}{\sum F_i} \qquad \text{where } r = 1,2,3,\ldots\ldots \qquad (8)$$

والجدول (3-5) يلخص صورة العزوم الاربعة الاولى حول الوسط الفرضى (a)

جدول (3-5)

Case **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

$$\mathbf{M}_{\mathbf{r}}^{(\mathrm{a})} = \frac{\sum d_i^r}{n}$$

فى حالة البيانات المبوبة

$$= \frac{\sum d_i^r F_i}{\sum F_i}$$

وذلك باعتبار ان $\mathbf{d_i}$ تمثل انحرافات القيم (المفردات فى حالة البيانات الغير مبوبة او مراكز فئات التوزيع (فى حالة البيانات المبوبة)) عن الوسط الفرضي (a) اى ان :

$$\mathbf{d_i = X_i - a}$$

سواء كانت X_i تعبر عن المفرده رقم (i) فى حالة المفردات اوتعبر عن مركز الفئة (i) فى حالة البيانات المبوبة .

4- <u>العزوم المعيارية</u>: Standard Moments

بصفة عامة تأخذ صورة العزم المعيارى ذو الرتبة (r) الصيغة التالية

فى حالة المفردات :

$$M_r^{(z)} = \frac{\sum_{i=i}^{n}\left(\frac{X_i - \bar{X}}{\sigma}\right)^r}{n} = \frac{\sum Z^r}{n} \qquad where\ r = 1,2, \dots \dots \dots \ \mathbf{(9)}$$

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 وذلك لأنه كما سبق وأن أوضحنا أن الوسط الحسابى للقيم المعيارية مساويا للصفر . كما أن العزم الثانى المعيارى دائما ابدا قيمته مساوية للواحد الصحيح وذلك لان العزم الثانى المعيارى ما هو الا تباين القيم المعيارية والتى سبق وان اوضحنا ان قيمة تباينها مساوية للواحد 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> :

1- العزوم الاربعة الاولى حول الصفر.

2- العزوم الاربعة الاولى حول الوسط الحسابى .

3- العزوم المعيارية الاربعة الاولى .

<u>الحل :</u>

1- <u>العزوم الاربعة الاولى حول الصفر تتطلب حساب المجاميع التالية :</u>

$\sum X^4 , \sum X^3 , \sum X^2 , \sum X$ والجدول التالى يبين الحسابات اللازمة لحساب العزوم الاربعة الاولى حول الصفر .

جدول (4-5)

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 هو:

$$M_1^{(0)} = \frac{\sum X}{n} = \frac{25}{5} = 5 = \bar{X}$$

والعزم الثانى حول الصفر هو :

$$M_2^{(0)} = \frac{\sum X^2}{n} = \frac{135}{5} = \frac{135}{5} = 27$$

والعزم الثالث حول الصفرهو :

$$M_3^{(0)} = \frac{\sum X^3}{n} = \frac{575}{5} = 115$$

واخيرا فإن العزم الرابع حول الصفر هو :

$$M_4^{(0)} = \frac{\sum X^4}{n} = \frac{4609}{5} = 921.8$$

2- *العزوم الاربعة الاولى حول الوسط الحسابى (العزوم المركزية) :*

من خلال قيمة العزم الاول حول الصفر والتى تعتبر بمثابة قيمة 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 (5-5)

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
Σ	0	10	0	34

وعليه فان العزوم الاربعة الاولى حول الوسط الحسابى هى :

$$\grave{M}_1 = M_1^{(\bar{X})} = \frac{\sum (X - 5)}{n} = \frac{0}{5} = 0 \quad ,$$

،

$$\grave{M}_2 = M_2^{(\bar{X})} = \frac{\sum (X - 5)^2}{n} = \frac{10}{5} = 2 \quad ,$$

$$\grave{M}_3 = M_3^{(\bar{X})} = \frac{\sum(X - 5)^3}{n} = \frac{0}{5} = 0 \quad ,$$

$$\grave{M}_4 = M_4^{(\bar{X})} = \frac{\sum(X - 5)^4}{n} = \frac{34}{5} = 6.8$$

3- <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 ماهو إلا الوسط الحسابى ($\bar{X}$) وهو فى مثالنا 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 المعيارية الاربعة الاولى كما يوضحها جدول (5-6) وذلك على النحو المبين التالى :

جدول (6-5)

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 المعيارية الاربعه الاولي 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 وذلك ان لم يكن هناك تقريب في قيمة الوسط الحسابي او الانحراف المعياري للمفردات التي تم حساب القيم المعياريه لها . أما العزم الرابع المعيارى فيكون عبارة عن خارج قسمة المقدار 8.494866775 على عدد المشاهدات (5) لينتج أن قيمة هذا العزم مساوية إلى 1.698973355 .

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 يتم حساب المجاميع اللازمة التى تتطلبها صيغ تلك العزوم والموضحة بالجدول 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
13-17	1	15	15	225	3375	50625
Σ	10		86	874	9782	116554

وعليه فان العزوم الاربعه الاولي حول الصفر هي :

$$M_1^{(0)} = \frac{\sum X_i F_i}{\sum F_i} = \frac{86}{10} = 8.6 = \bar{X}$$ ،

$$M_2^{(0)} = \frac{\sum X_i{}^2 F_i}{\sum F_i} = \frac{874}{10} = 87.4$$,

$$M_3^{(0)} = \frac{\sum X_i{}^3 F_i}{\sum F_i} = \frac{9782}{10} = 978.2$$,

$$M_4^{(0)} = \frac{\sum X_i{}^4 F_i}{\sum F_i} = \frac{116554}{10} = 11655.4$$

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 (5 – 7)

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 :

$$M_1^{(\overline{X})} = \frac{\sum (X_i - \overline{X}) F_i}{\sum F_i} = \frac{0}{10} = 0$$ (بالتعريف) ،

$$M_2^{(\overline{X})} = \frac{\sum (X_i - \overline{X})^2 F_i}{\sum F_i} = \frac{134.4}{10} = 13.44 = \sigma^2$$

،

$$M_3^{(\overline{X})} = \frac{\sum (X_i - \overline{X})^3 F_i}{\sum F_i} = \frac{-46.08}{10} = -4.608$$

،

$$M_4^{(\overline{X})} = \frac{\sum (X_i - \overline{X})^4 F_i}{\sum F_i} = \frac{3796.992}{10} = 379.6992$$

- ***العلاقة بين الانواع المختلفة من العزوم :***

تفيد دراسة العزوم كما سبق وأن ذكرنا في معرفة خصائص التوزيعات التكرارية . فيمكن عن طريقة العزوم معرفه أهم مقاييس النزعة المركزية وهو الوسط الحسابي حيث رأينا 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 التباين ومن ثم الانحراف المعياري حيث رأينا 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 أوالقيم 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 :

$$M_3^{(\bar{X})} = M_3^{(0)} - 3\,M_2^{(0)}\,M_1^{(0)} + 2\,(M_1^{(0)})^2$$

$$M_4^{(\bar{X})} = M_4^{(0)} - 4\,M_3^{(0)}\,M_1^{(0)} + 6\,M_2^{(0)}\,(M_1^{(0)})^2 - 3\,(M_1^{(0)})^4$$

- **<u>العزوم المصصححه لشبرد :</u>**

سبق وان ذكرنا انه عند تبويب المفردات او البيانات في صورة جداول تكرارية فاننا نعتمد علي مراكز الفئات عند حساب الكثير من المقاييس الاحصائية . 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 :

$$correction\ of\ \mathbf{M}_2^{(\bar{\mathbf{X}})} = \mathbf{M}_2^{(\bar{\mathbf{X}})} - \frac{{L_i}^2}{12}$$

حيث (L_i) تعبر عن طول الفئة في التوزيع التكراري. كما قام شبرد بتصحيح العزم الرابع حول الوسط الحسابي ليأخذ الصورة المصححة التالية :

$$correction\ of\ \mathbf{M}_4^{(\bar{X})} = \mathbf{M}_4^{(\bar{X})} - \frac{L_i^{\ 2}}{12}\mathbf{M}_2^{(\bar{X})} + \frac{7}{240}L_i^{\ 4}$$

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 نقوم 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 والمبوبة تكراريا لظاهرة ما تأتي الان دراسة ما اذا كان التوزيع الخاص بظاهرة ما متماثلا او ملتويا وما هو شكل ودرجة التواء التوزيع دون ان نقوم 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 الرأسي المقام عند القيمة المتوسطة للتوزيع .

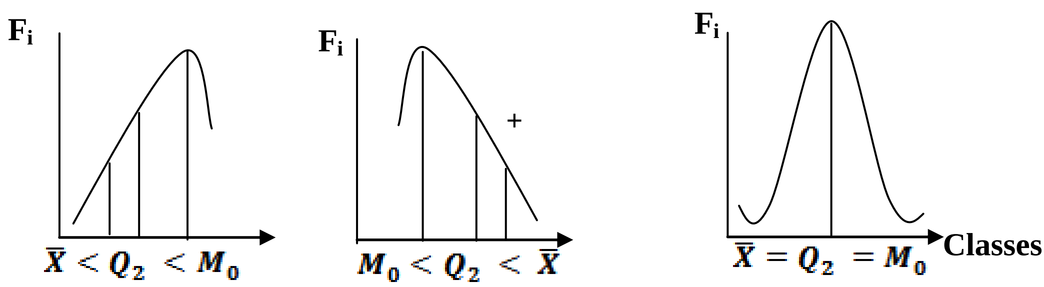
أما التواء التوزيع فنقصد به بعد المنحني عن التماثل اي تركز التكرارات في طرف وقلتها في طرف اخر . حيث يكون التوزيع موجب الالتواء Positive skewness (أي ملتوي جهة اليمين) إذا تركزت التكرارات الكبيرة عند الفئات الدنيا (اي الفئات الاولي) للتوزيع اما التكرارات الصغيرة فتنتشر عند الفئات العليا من التوزيع ويكون ذيل التوزيع جهة اليمين اطول من جهة

اليسـار. والعكـس صـحيح حيـث يكـون التوزيـع سـالب الالتـواء Negative skewness (اي ملتـوي جهـة اليســار) نتيجـة لتركيـز عـدد كبيـر مـن التكرارات عند الفئات العليا للتوزيع بينما تنتشر قليل من التكرارات عند الحدود الدنيا للتوزيع ويكون حينئذ التوزيع ذو ذيل اطول جهة اليسـار (توزيع ذو ذيـل ايسر أو سالب) .

وخلاصة ما سبق هو ان التوزيع التكراري من حيث التماثل او الالتـواء يكـون واحدا من الصور الثلاث : فاما ان يكون التوزيع متماثلا او ملتوي جهـة اليمـين او ملتوي جهة اليسار .

وفـي دراسـتنا لمقـاييس الالتـواء سـوف نميـز بـين نـوعين مـن المقـاييس حيـث يعتمد النوع الاول في ايجادة علي مقاييس النزعـة المركزيـة ومقـاييس التشـتت التـي تعرضـنا لهـا سـابقا . امـا النـوع الثـاني مـن مقـاييس الالتـواء فيعتمـد فـي حسابه علي العزوم .

هـذا ويمكـن مـن خـلال مقارنــة مقـاييس المتوسـطات (الوسـط – الوسـيط – والمنوال) دراسة ماهية تماثل التوزيع من عدمه كما يوضحه شكل (1-5)



Negative skewness
توزيع سالب الالتواء
(C)

Possitive skewness
توزيع موجب الالتواء
(B)

Symmetric distribution
توزيع متماثل
(A)

شكل (1-5)

*** فاذا كان التوزيع متماثلا فان الوسط الحسابى يتساوى مع الوسيط يتساوى مع المنوال اى ان :**

$$\bar{X} = Q_2 = M_0$$

كما هو واضح فى شكل (A-1-5) .

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 فيما 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 (5 – 1 – c).

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 أو حدة الالتواء . هذا ولتحديد درجة الالتواء بالإضافة لنوعه ايضا فيجب حساب أيا من نوعى معاملات الالتواء والتى سترد حالا فى هذا القسم :

1- *<u>معاملات الالتواء بدلاله مقاييس المتوسطات و التشتت : -</u>* 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 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 .**

<u>وفيما يلى اكثر معاملات الالتواء شيوعا :</u>

أ - <u>معامل الالتواء الاول لبيرسون (B_1)</u> وهو عبارة عن :

اى أن معامل الإلتواء الأول لبيرسون هو عبارة عن نسبة الفرق ما بين الوسط الحسابى والمنوال الى الانحراف المعيارى للتوزيع .

$$B_1 = \frac{\bar{X} - M_0}{\sigma}$$

ب – <u>معامل الالتواء الثانى لبيرسون (B_2)</u> وهوعبارة عن :

$$B_2 = \frac{3\,(\bar{X} - Q_2)}{\sigma}$$

اى النسبة بين ثلاثه امثال الفرق بين الوسط الحسابى والوسيط الى الانحراف المعيارى للتوزيع.

وتتراوح قيمتى معاملى الالتواء الاول والثانى لبيرسون(اى B_1 , B_2) فما بين (+3 , 3 -) إلا أن أهم 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 .

هذا 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 :

$$B_3 = \frac{3\,(\mathbf{Q}_3 - \mathbf{Q}_2) - (\mathbf{Q}_2 - \mathbf{Q}_1)}{\mathbf{Q}_3 - \mathbf{Q}_1}$$

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 المتجمع الصاعد او الهابط ثم 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 المئينى نظرا لاعتماده على حساب المئينيات وهو ياخذ الصورة التالية :

$$B_4 = \frac{(\mathbf{P_{90}} - \mathbf{Q_2}) - (\mathbf{Q_2} - \mathbf{P_{10}})}{\mathbf{P_{90}} - \mathbf{P_{10}}}$$

حيث P_{10} , P_{90} تمثل المئين العاشر والمئين التسعون على الترتيب.

-2 <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 أنه اذا كان :

$$\mathbf{M}_3^{(\bar{X})} > 0$$

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 فيما يلى :

(1):فاذا كانت قيمة معامل الالتواء تساوى الصفر يكون التوزيع متماثلا .

(2):واذا كانت قيمة معامل الالتواء موجبا يكون التوزيع ملتويا جهة اليمين أى أن له التواء موجب.

(3):واذا كانت قيمة معامل الالتواء سالبا يكون التوزيع ملتويا جهة اليسار أى ان له التواء سالب.

3- *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 . أى 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 فانه لابد

من وجود مقياس احصائى يدرس شكل قمه التوزيع وهو ما يسمى بمعامل التفرطح والذى يقيس درجة تفرطح منحنى التوزيع اى شكل قمته وهناك معاملين اساسين يمكن استخدامهم فى دراسة شكل قمه المنحنيات التكرارية يسميان بمعاملات التفرطح وهما :

<u>أ – معامل التفرطح بدلاله العزوم :</u>

وهذا المعامل هو خارج قسمة العزم الرابع المعيارى حول الوسط الحسابى على مربع العزم الثانى حول الوسط الحسابى ويرمز له بالرمز ($\propto_1$) أو معامل التفلطح بدلالة العزوم أو معامل التفلطح الأول $\propto_1$ 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 قيمة معامل تفرطح التوزيع بمعامل التفرطح الخاص بالتوزيع المعتدل (اي بالقيمة 3). وهنا نكون بصدد احد الحالات الثلاث التالية :

<u>* اذا كانت قيمة $3 = \propto_1$</u> : فان هذا يعني ان المنحني طبيعي او معتدل اي متوسط التفرطح Mesokurtic كما موضح في شكل (A-2-5) .

<u>* واذا كانت قيمة $\propto_4$ للمنحنى اقل من (3) :</u> فان هذا يعني ان منحني التوزيع مفرطح Platykutric اي ذو قمة عريضة (اي يكون اكثر اتساعا من وسطه وتنخفض قمته عن قمة المنحنى المعتدل) كما هو موضح في الشكل(5- 2- B).

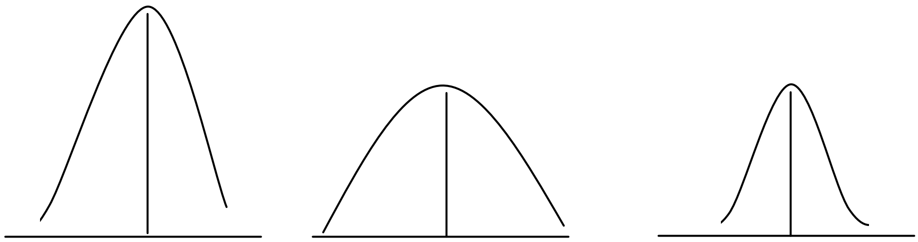
<u>اما اذا كانت قيمة($\propto_4$) اكبر من (3)</u> : فان هذا يعني ان منحني التوزيع مدبب Leptokurtic اي يكون اكثر ضيقا من وسطه وترتفع قمته عن المنحني المعتدل .

<u>ب – معامل التفرطح بدلالة الربيعيين والمئينيات :</u>

وهذا المعامل عبارة عن النسبه ما بين نصف المدي الربيعي وبين الفرق ما بين المئين التسعين والمئين العاشر ويسمى بمعامل التفرطح الثانى 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.



شكل (5-2)

<u>ملاحظات حول الالتواء والتفرطح :-</u>

(1): معاملات الالتواء والتفرطح كلها مقاييس نسبيه ليس لها تمييز . ولذلك فهذه المعاملات تصلح لمقارنة توزيعات سواء متحدة او مختلفة في وحدات قياسها .

(2): يلاحظ ان قيم معاملات الالتواء السابقة قد تختلف بالنسبة للتوزيع الواحد وذلك بسبب اختلاف تعريف كل منها وكذا طريقة أو آلية حسابه . وهذا يعني ضرورة استخدام المعامل نفسه عند مقارنة درجة الالتواء لتوزيعين او اكثر .

(3): تتوقف اشارة اياً من معاملات الالتواء علي اشارة البسط فى قانون هذه المعاملات وذلك لان اشارة المقام دائما ابدا موجبه باستثناء حالة وحيدة وهي حينما تكون قيمه(σ) او الفرق بين الربيعيين الاعلي والادني مساوية للصفر وهي تمثل حالة التجانس الكامل للبيانات (ظاهرة تأخذ قيمة ثابتة دون تغير او اختلاف)

<u>مثال :</u>

ادرس خاصيتي الالتواء والتفرطح للتوزيع التكراري التالي :

Classes	20-	22-	24-	26-	28-	30-32
Frequency	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 الوسط الحسابي . والجدول التالي يبين الحسابات اللازمة لهذه العزوم . حيث يتم حساب الوسط الحسابي أولاً ثم العزوم الأربعة حول الوسط .

Classes	F_i	X_i	$X_i F_i$	$(X_i - 26.2)$	$(X_i - 26.2)F_i$	$(X_i - 26.2)^2 F_i$	$(X_i - 26.2)^3 F_i$	$(X_i - 26.2)^4 F_i$
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

فكانت قيمة الوسط الحسابي كما يلى :

$$\overline{X} = \frac{\sum X_i F_i}{\sum F_i} = \frac{1310}{50} = 26.2$$

وعليه فإن العزوم الأربعة الأولي حول الوسط الحسابى وذلك من خلا المجاميع الموضحة بالجدول السابق هي :

$$M_1^{(\overline{X})} = \frac{\sum (X_i - \overline{X})\, F_i}{\sum F_i} = \frac{\sum (X_i - 26.2)\, F_i}{\sum F_i} = \frac{0}{50} = 0 \quad ,$$

$$M_2^{(\overline{X})} = \frac{\sum (X_i - \overline{X})^2\, F_i}{\sum F_i} = \frac{\sum (X_i - 26.2)^2\, F_i}{\sum F_i} = \frac{336}{50} = 6.72$$

$$M_3^{(\overline{X})} = \frac{\sum (X_i - \overline{X})^3 F_i}{\sum F_i} = \frac{\sum (X_i - 26.2)^3 F_i}{\sum F_i} = \frac{-96}{50} = -1.92$$

$$M_4^{(\overline{X})} = \frac{\sum (X_i - \overline{X})^4 F_i}{\sum F_i} = \frac{\sum (X_i - 26.2)^4 F_i}{\sum F_i} = \frac{5614.08}{50}$$
$$= 112.2816$$

وعليه فأن معامل الألتواء بدلالة العزوم هو :

$$B_4 = \frac{M_3^{(\overline{X})}}{\sqrt{\left(M_2^{(\overline{X})}\right)^3}} = \frac{-1.92}{\sqrt{(6.72)^3}} = \frac{-1.92}{17.420231} = -0.1102$$

وهو ما يعني أن هذا التوزيع وإن كان يقترب من التماثل إلا أنـه يميـل للألتـواء السالب بدرجة ضعيفة.

أما عن معامل التفرطح بدلاله العزوم ($\propto_1$) فهو عبارة عن :

$$\propto_1 = \frac{M_4^{(\overline{X})}}{\left(M_2^{(\overline{X})}\right)^2} = \frac{112.2816}{(6.72)^2} = 2.4864$$

وحيث أن قيمـه معامـل التفـرطح أقل من (3) إلا أنـه يقتـرب بدرجـة كبيـرة من القيمة (3) فهذا يفيد أن هذا التوزيع وإن كان يقتـرب مـن الإعتـدال إلا أنـه يميل بدرجة ضعيفة إلى التفرطح .

مثال:

إذا كـان العزم الاول حـول الـرقم (3) هـو (8) فمـا هـى قيمـة الوسـط الحسـابى لـهذه المفردات .

الحل :

حيث أن صيغة العزم ذو الرتبـة (r) حـول الوسـط الفرضـى (a) ياخذ الصـورة 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 :

$$\frac{\sum X_i}{n} = 11$$

$$i.\ e.,\ \bar{X} = 11$$

أى أن الوسط الحسابى لهذه المفردات هو 11 .

<u>مثال :</u>

اذا كان لديك التوزيع التكرارى المتجمع التالى لتوزيع منتظم لأجر عينة من 50 عاملا (بالجنيه) فى احد المصانع فى اليوم :

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 - حدد نسبه العمال الذين تزيد اجورهم عن 90 جنيها اسبوعيا .

<u>الحل</u> : -

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 55 - 10 = 45 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 التكرار المتجمع الحالى ليعطى تكرار الفئة الحالية كما هو موضح بالجدول التالى :

Classes	F_i	X_i	$D_i = \dfrac{X_i - 80}{10}$	$D_i\ F_i$	$D_i^2\ F_i$
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-	47-43=4	100	2	8	16
105-115	50-47=3	110	3	9	27
$\sum$	50			0	100

وعليه فأن :

$$\bar{X} = a + b\,\frac{\sum D_i\,F_i}{\sum Fi} = 80 + 10\left(\frac{0}{50}\right) = 80 \qquad (L.E)$$

،

$$\sigma = b\sqrt{\frac{\sum D_i^2\ F_i}{\sum F_i} - \left(\frac{\sum D_i\ F_i}{\sum F_i}\right)^2}$$

$$= 10\sqrt{\frac{100}{50} - \left(\frac{0}{50}\right)^2} = 10\sqrt{2} = 14.14 \qquad (L.E)$$

ومن ثم فأن معامل الاختلاف النسبى هو :

$$C_v = \frac{\sigma}{\bar{\bar{X}}} \times 100 \qquad = \frac{14.14}{80} \times 100 = 17.675\,\%$$

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:

$$Q_2 = 75 + 10 \times (\,(25 - 15)/\,(35 - 15)$$

$$= 75 + 10(10/20) = 75 + 5 = 80 \quad (L.E.)$$

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 (معتدل) .

جـ - لإيجاد نسبة العمال الذين تزيد اجورهم عن 90 جنيها فأن يمكن من الجدول التكرارى المتجمع الصاعد معرفة نسبة الاسر الذين تقل أجورهم عن 90 جنيها ثم بطرح النسبة الناتجة من 100% لينتج المطلوب مباشرة بدلا من ايجاد التوزيع التكرارى المجتمع الهابط واستنتاج النسبة المطلوبة .

(على الطالب إستنتاج النسبة المطلوبة والتأكد أن الاجابة هى 22 %) .

تمارين على الفصل الخامس

1- فيمـايلى التوزيـع العمـرى لأفـراد عينـة حجمهـا 60 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 الالتـواء المختلفة وذلـك باستخدام المقـاييس المحسوبة فـى (أولا).

(ب): تطبيـق العلاقــات مـابين الانحـراف المعيــارى وبـين كـل مـن الانحـراف المتوسط ونصف المدى الربيعى والمتعلقة بـالتوزيع المعتـاد مـع ابـداء التعليق المنـاسب .

(ج): حساب نسبة الموظفين الذين تتراوح مرتباتهم مابين :

$$\bar{X} \mp \sigma , \bar{X} \mp 2\sigma , \bar{X} \mp 3\sigma$$

حيث : $\bar{X}$ ، σ همـا الوسـط الحسـابى والانحـراف المعيـارى للمرتبـات علـى الترتيب . قارن بين قيم هذه النسب بمـا يجب ان تكـون عليـه فـى حالـة التوزيـع المعتاد وهى 68.27% , 95.45% , 99.73% علـى الترتيـب مبديا مـا تـراه مناسب من تعلق .

2- فيما يلى التوزيع التكرارى للدخل الأسبوعى لمائة اسرة:

Income Classes	340-	350-	360-	370-	380-	390-	400-	410-
No.0f Families	2	5	9	17	35	16	10	6

والمطلوب :

(أ):رسم المنحنى المتجمع الصاعد ومنه أوجد قيمة الوسيط والربيعين الادنى والاعلى ومن ثم حساب معامل الاختلاف الربيعى .

(ب): قياس الالتواء والتفرطح لهذا التوزيع.

3- استخدام بيانات أجور العمال الاسبوعية فى المصنعين (أ، ب) لصناعة المنسوجات القطنية والواردة فى التمرين الأول من تمارين هذا الفصل فى تحديد كلا من المطلوب التالى:

اولا : باستخدام بيانات المصنع (أ) أحسب العزمين الثانى والثالث حول الوسط الحسابى .

ثانيا : قارن بين التواء توزيع الاجور فى كل من المصنعين وذلك باستخدام مايلي :

(أ):معاملا بيرسون الاول والثانى للا لتواء.

(ب):معامل الالتواء الربيعى.

(ج):معامل الالتواء المئينى .

ثالثا : قارن بين التوزيعين من حيث الاختلاف النسبى .معلقا على ما تحصل عليه من نتائج .

رابعا: قياس التفرطح فى كل من التوزيعين .

4- باستخدام بيانات درجات الطلاب فى كل من كليتى التجارة والتربية والواردة فى تمارين الفصل الثالث .

المطلوب :

(أ):قارن بين تشتت التوزيعين باستخدم المقياسين التاليين:

- نصف المدى الربيعى -الانحراف المعيارى

ب- باستخدام العزوم ، قارن بين تفرطح التوزيعين .

ج- قدر التواء كل توزيع واختلافه النسبى.

5- البيانات التالية تمثل درجات 80 طالبا فى مادة الاقتصاد:

Degree Classes	40-	50-	60-	70-	80-	90-
No.0f stndents	5	12	32	21	7	3

والمطلوب :

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 التفرطح .

ج- احسب نسبة الطلاب الذين تتراوح دراجاتهم مابين : $\bar{X} \mp \sigma$

حيث $\bar{X}$ ، σ هما على الترتيب الوسط الحسابى والانحراف المعيارى لهذا التوزيع . ثم بين كيف يمكن استخدام قيمة هذه النسبة كمؤشر على تماثل هذا التوزيع من عدمه.

6- فيما يلى التوزيع التكرارى لدرجات طلاب كليات التجارة موزعين حسب نظام الدراسة ،وذلك فى احد الامتحانات :

نظام الدراسة / فئات الدرجات	انتظام	انتساب موجه	المجموع
-0	38	40	78
-14	126	77	203
-20	204	107	311
-26	239	57	296
-32	73	5	78
40-36	24	صفر	24
المجموع	704	286	990

والمطلوب :

(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 20 ،30 درجة وذلك فى كل من التوزيعين (انتظام ، انتساب موجه).

7-الجدول التالى يوضح التوزيع التكرارى لدرجات الطلاب المنتظمين بكلية التجارة باحدى الجامعات فى دور مايو 1987 م:

Degree Classes	0-	10-	20-	30-	40-	50-60	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 (بأعتبار أن الطالب الحاصل على درجة أقل من 25 درجة يعتبر راسبا).

ج- عدد الطلاب الذين تتراوح درجاتهم مابين 35 ، 52 درجة .

د – درجة ونوع التواء توزيع الدرجات .

هـ -اذاكانت الدرجة النهائية لامتحان الاحصاء فى هذا الدورهى 60 درجة ، فماهو الحد الأدنى لتقدير جيد إذا علمت ان عدد الطلاب الذين حصلوا على تقدير جيد هو 218 طالبا ،فسر ماقد تجده من اختلاف فى النتيجة التى حصلت عليها.

□

الفصل السادس

الارتباط الخطى البسيط

Simple linear correlation

فى الابواب السابقة اقتصرت دراستنا 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 الدراسات الطبيعية او الاقتصادية او الاجتماعية يكون الاهتمام منصبا على دراسة العلاقة بين ظاهرتين او اكثر بغرض 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 مثلا لدراسة العلاقة ما بين ظاهرة إدمان 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 فيما بين متغيرين او اكثر بعلاقة الارتباط حيث يهتم الباحث بدراسة ما اذا كانت هذه المتغيرات ترتبط بعلاقة خطية أو غير 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 الغرض 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 فيفترض وجود متغيراً واحداً تابعا يتاثر بواحدا او اكثر من المتغيرات المستقلة حيث يتدخل الباحث بتثبيت قيم المتغير او المتغيرات المستقلة عند مستويات معينة (اى انها تقاس بدون اخطاء عشوائية) ثم يشاهد او يسجل القيم التى ياخذها المتغير التابع عند تلك المستويات المحددة للمتغير او مجموعة 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 . اما فى 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 وهو مايسمى بتحليل الإنحدار وهو ما سيرد الفصل التالى لدراستنا .

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. فإذا كان لدينا متغيرين وليكن 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 أوأن نقص احدهما يؤدى الى نقص فى قيمة المتغير الاخر فيقال فى تلك الحالة ان الارتباط بين المتغيرين إرتباطا 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 أو أن نقص احدهما يؤدى الى زيادة فى قيمة المتغير الاخر فيقال فى هذه الحالة ان الارتباط بين المتغيرين إرتباطا عكسيا

(اى سالب) .

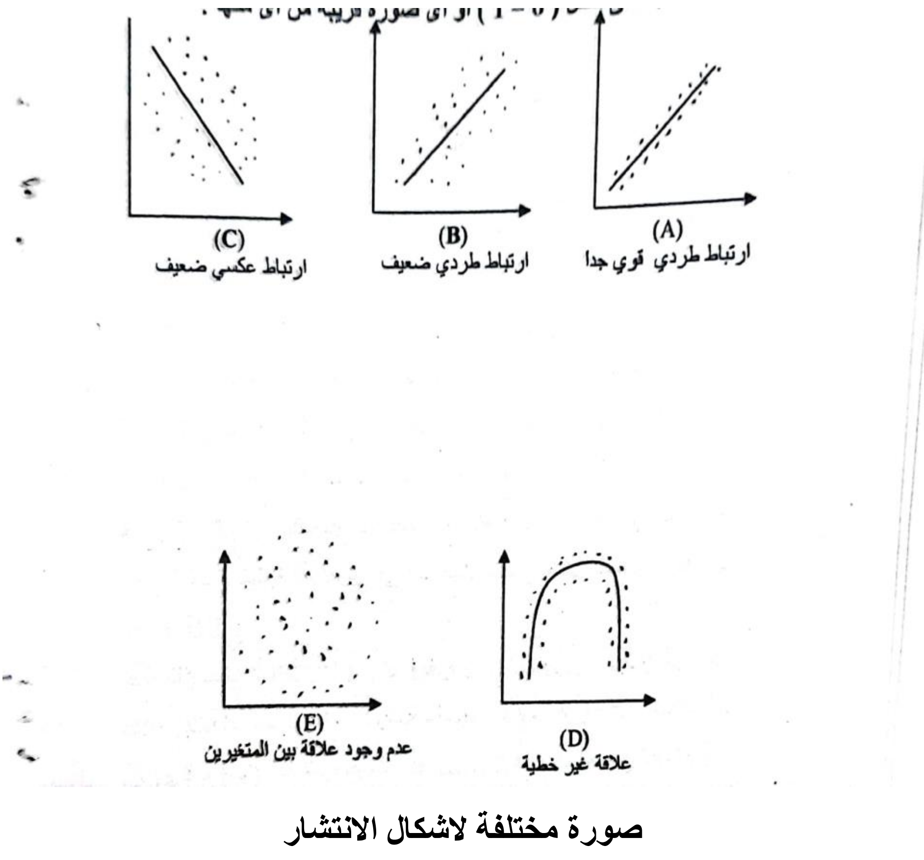
هذا وقبل التعرض لكيفية حساب معامل الارتباط فانه من المفيد ان يتم عرض ازواج القيم المتناظرة للمتغيرين فى شكل بيانى يسمى بشكل الإنتشار . Scatter Diagram

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 i=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 طبيعة العلاقة ما بين المتغيرين

محـل الدراسـة . ويسـاعد شـكل الإنتشـار فـى اختيـار النمـوذج القياسـي الـذى يتناسـب مـع طبيعـة البيانـات موضـع الدراسـة. وهنـاك صـوراً مختلفـة لاشـكال الانتشار نرصد منـها الاشكال التالية والمبينـة فـى شكل (6 – 1) او اى صـورة قريبـة من اى منـها .



صورة مختلفة لاشكال الانتشار

(1-6)

حيث يتضح لنا من خلال اشكال الانتشار الوارده فى الشكل (6-1) ما يلى :

- **بالنسـية للشـكل (A) فـى شـكل (6-1) فانـه يوضـح لنـا وجـود علاقـة طرديـة وخطية قوية جدا ما بين المتغيرين Y,X بمعنى ان اى تغير فى احد المتغيرين**

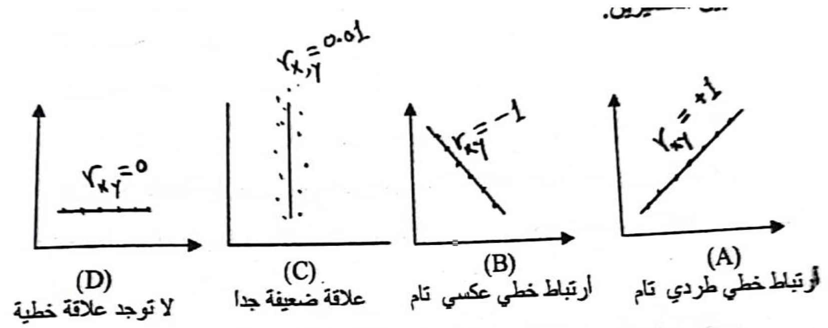
وليكن مثلا (X) يصاحبة تغير فى قيمة المتغير(Y)وفى نفس الاتجاه . وكلما كانت نقاط شكل الانتشار التى تعبر عن ازواج المشاهدات(Y_i , Xi) قريبة من الخط المستقيم والذى يتوسط هذه النقاط فكلما دل ذلك على قوة الارتباط بين المتغيرين . والعكس صحيح فكلما ابتعدت النقاط عن الخط المستقيم الذى يتوسط تلك النقاط فان هذا 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 (A) , (B) , (C) الواردة فى شكل (6-1) تمثل وجود علاقة خطية بين المتغيرين ويطلق على علاقة الارتباط فى هذه الحالة بانها إرتباطا 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 منحنى 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 (6-2) فاذا وقعت جميع نقاط شكل الانتشار على خط مستقيم موجب الميل (اى يصنع زوايه حادة مع الاتجاه الموجب للمحور الافقى) كما هو فى الشكل (6-2-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 (6-2-B) فان هذا يدل على ان الارتباط بين المتغيرين هو ارتباطا خطيا عكسيا تاماً بين المتغيرين.



صور اخرى لاشكال الانتشار

شكل (6-2)

اما اذا تركزت نقاط شكل الانتشار حول خط مستقيم عمودى على أى من المحورين ويوازى المحور الآخر كما هو فى شكل (6-2-C) فان هذا يعنى وجود علاقة ضعيفة جدا بين المتغيرين .

وفى حالة وقوع جميع النقاط على خط مستقيم يوازى احد المحورين وعمودى على المحور الآخر كما هو الحال فى شكل (6-2-D) فان هذا معناه انه لا توجد

علاقة خطية بين المتغيرين (كما هو الحال فى الدوال الثابتة) اى ان معامل الارتباط الخطى بين المتغيرين فى هذه الحالة يكون مساويا للصفر .

واخيرا فانه بالرغم من ان أشكال الإنتشار 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 (A-2-6). 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 (B-6-2) 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 الصحيح سواء كان موجبا او سالبا فان هذا يعنى قوة العلاقة فيما بين المتغيرين . والعكس صحيح فكلما اقتربت قيمة معامل الارتباط من الصفر كلما دل ذلك على ضعف العلاقة الخطية فيما بين المتغيرين . وأخيرا اذا تساوت قيمة معامل الارتباط بالصفر فإن ذلك يعنى عدم وجود علاقة ارتباط او نزعة او تأثير فيما بين قيم احد المتغيرات وقيم المتغير الاخر . اما عن اشارة معامل الارتباط فهى تبين اتجاه العلاقة فيما بين المتغيرين حيث تكون الإشارة موجبة

إذا كانت العلاقة طردية والعكس صحيح تكون الإشارة سالبة اذا كانت العلاقة عكسية فيما بين المتغيرين محل الدراسة .

هذا وعند دراستنا للارتباط فى هذا الباب فاننا سوف نقتصر فى هذه الدراسة على الارتباط الخطى البسيط بين متغيرين فقط .

(6-1) العلاقة بين الظواهر (المتغيرات) :

تأخذ العلاقة فيما بين المتغيرات صورا مختلفة وتتوقف تلك الصورة على طبيعة او نوعية البيانات وكذا عدد المتغيرات المطلوب دراسة العلاقة فيما بينها وتتلخص هذه الصور فيما يلى :

1- علاقة الرتباط :

وهى العلاقة بين المتغيرات التى 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 يلي :-

(أ):الارتباط الخطى البسيط: وهو الذى يستخدم فى قياس قوة العلاقة الخطية فيما بين متغيرين كميين فقط .

(ب):الارتباط المتعددة : وهو يتناول العلاقة بين اكثر من متغيرين كميين .

(ج): الارتباط الجزئي : وهو يهتم بدراسة العلاقة بين متغيرين كميين مع افتراض تثبيت (او استبعاد) تأثير متغيرات أخرى .

2-*الارتباط بين الصفات* : اذا كانت احدى الظاهرتين المطلوب دراسة العلاقة ما بينها عبارة عن صفات ويتعذر قياسها كميا فان قياس ماهية قوة العلاقة فيما 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 الظاهرتين على الأقل 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 خلال ما يسمى بمعامل الارتباط Correlation Cofficient وهناك صورا عديدة لمعاملات الارتباط ومع اختلاف هذه الصور الا انها تتفق جميعا فى الخصائص التالية :

(1): 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.

(2):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 $r_{x,y}= -1$

(3):<u>معامل الارتباط لايتاثر باى من العمليات الجبرية الاربعة</u> (الطرح او الجمع او الضرب او القسمة) اى ان :

$$r_{x,y} = r_{dx\ dy} = r_{Dx\ ,\ Dy}$$

حيث :-

$d_{yi} = y_i - a$ ، $d_{xi} = x_i - a$ 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:

$D_{yi} = \frac{y_i - a}{b}$ ، $D_{X_i} = \frac{X_i - a}{b}$ حيث D_i تعبر عن الانحرافات المعدلة السابق الاشارة اليها .

هذا وسوف نقوم بدراسة معاملات الارتباط التالية :

1- **معامل الارتباط الخطى البسيط لبيرسون.**
2- **معامل إرتباط الرتب لسبيرمان.**
3- **معامل ارتباط كندال.**
4- **معامل ارتباط فهنر.**

1- **<u>معامل الارتباط الخطى البسيط</u> لبيرسون :**

Simple Linear Correlation Coefficient

وينسب هذا المعامل الى كارل بيرسون K. Pearson وهناك اكثر من صورة رياضية لهذا المعامل وهو يقيس ماهية قوة أو ضعف العلاقة الخطية ما بين

متغيرين كميين سواء كانت البيانات مبوبة او غير مبوبة وذلك على النحو التالى :

أولا: فى حالة البيانات الغير مبوية (المفردات) :

فاذا كان لدينا عدد (n) من ازواج المشاهدات للمتغيرين X , Y اىأنه إذا كان لدينا مجموعة ازواج المشاهدات التالية:

(X_1, Y_1) ، (X_2, Y_2)، .... (X_n, Y_n)

وكان الوسط الحسابى للمتغير X هو $\bar{X}$ والوسط الحسابى للمتغير Y هو $\bar{Y}$وكان الانحراف المعيارى للمتغيرين X , Y هما σ_x، σ_yعلى الترتيب فان معامل الارتباط الخطى البسيط لبيرسون (r_{X_y}) هو عبارة عن متوسط حاصل ضرب القيم المعياريه للمتغير ين x , y . أى أن :

$$r_{x,y} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma_x}\right) \left(\frac{y_i - \bar{y}}{\sigma_y}\right)$$

$$= \frac{\sum (X_i - \bar{X})\,(y_i - \bar{y})}{n\,\sigma_x\,\sigma_y}$$

$$= \frac{\sum (X_i - \bar{X})\,(y_i - \bar{y})}{n\sqrt{\frac{\sum (X_i - \bar{X})^2}{n} \cdot \frac{\sum (y_i - \bar{y})^2}{n}}} \qquad (1)$$

$$= \frac{\sum (X_i - \bar{X})\,(y_i - \bar{y})}{\sqrt{\sum (X_i - \bar{X})^2 . \sum (y_i - \bar{y})^2}} \qquad (2)$$

والمقدار $\frac{\sum(X_i - \bar{X})(y_i - \bar{y})}{n}$ يسمى بالتغاير **Covariance** بين المتغيرين X ، y . ومن ثم فإذا كان معامل الارتباط الخطى البسيط هو عبارة عن خارج قسمة التغاير بين المتغيرين على حاصلا ضرب الإنحرافات المعياريه للمتغيرين. هذا وعادة ما تكون قيم الاوساط الحسابية للمتغيرين قيما كسرية مما يؤدى الى صعوبة استخدام صيغة المعادلة (1) فى حساب قيمة معامل الأرتباط بين المتغيرين x ،y وللتغلب على هذه المشكلة فإنه 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 (3) فان يلزم لحساب معامل الارتباط الخطى البسيط حساب المجاميع التالية:

$$\sum x, \sum y, \sum x^2, \sum y^2, \sum xy$$

وإنطلاقا من الخاصية الثالثة لخصائص معامل الارتباط والتى تفيد بان معامل الارتباط لايتاثر بأيا من العمليات الجبريه الاربعة. فإنه يمكن بالاضافة للطريقة المباشرة والمستخدم فيها المعادلة (3)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 (فى حالة المفردات) هى:-

$$r_{x,y} = \frac{\frac{\sum d_x d_y}{n} - \frac{\sum d_x}{n} \cdot \frac{d_y}{n}}{\sqrt{\left[\frac{\sum d_x^2}{n} - \left(\frac{\sum dx}{n}\right)^2\right]\left[\frac{\sum d_y^2}{n} - \left(\frac{\sum dy}{n}\right)^2\right]}}$$

حيث d_x ، d_y تعبر عن إنحرافات قيم ازواج المشاهدات عن وسط فرضى معين سواء كان هذاالوسط الفرضى متساويا للمتغيرين او مختلفا .

- <u>صيغة معامل الارتباط الخطى البسيط لبيرسون فى حالة استخدام طريقة الإنحرافات المعدلة:</u>

$$r_{x,y} = r_{D_x,Dy} = \frac{\frac{\sum D_x D_y}{n} - \frac{\sum D_x}{n} \cdot \frac{\sum D_y}{n}}{\sqrt{\left[\frac{\sum D_x^2}{n} - \left(\frac{\sum Dx}{n}\right)^2\right]\left[\frac{\sum D_y^2}{n} - \left(\frac{\sum Dy}{n}\right)^2\right]}}$$

<u>مثال :</u>

فيما يلى بيان باسعار الفائده على رأس 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

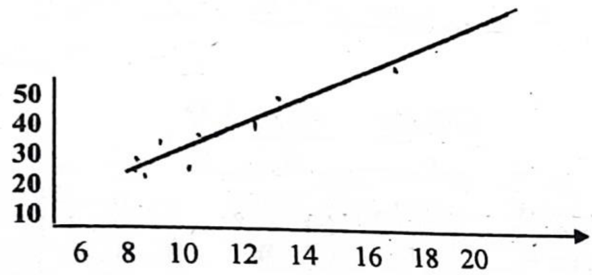
والمطلوب :-

1- ارسم شكل الانتشار وعلق على نتائجك .

2- 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 :

1- رسم شكل الأنتشار لأزواج المشاهدات وذلك على النحو التالى:-



ومن خلال هذا الشكل يتضح لنا ان نقاط ازواج المشاهدات تتركز بدرجه كبيرةحول خط مستقيم ذوميل موجب اى ان العلاقة طردية وقوية جدا فيما بين المتغيرين (x) ، (y) .

2- حساب معامل الارتباط الخطى البسيط :

(أ): باستخدام إنحرافات قيم المتغيرين عن اوساطهما الحسابية اى باستخدام 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 $\bar{x}$ ، $\bar{y}$ فحيث ان :

$$\bar{X} = \frac{\sum X}{n} = \frac{8+9+8+11+\ldots\ldots+18}{7} = \frac{77}{7} = 11 \quad ,$$

$$\bar{Y} = \frac{\sum Y}{n} = \frac{30 + 33 + \dots\dots\dots + 57}{7} = \frac{252}{7} = 36$$

ثم يتم تكوين الجدول التالى لاجراء الحسابات اللازمة لحساب معامل الارتباط فى هذة الحالة :

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	648

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 فيما بين المتغيرين x ، y .

<u>(ب) حساب معامل الارتباط الخطى البسيط باستخدام القيم الاصليه :-</u>

<u>1- 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.

2- <u>باستخدام طريقة الإنحرافات البسيطة :</u> دعنا نقوم بحساب انحرافات قيم المتغير (x)عن الوسط الفرضي وليكن القيمة (10) ونقوم بحساب انحرافات قيم المتغير (y) عن الوسط الفرضى وليكن القيمة (30) . أى أن :

$d_x = x_i - 10$ where $i=1,2,\dots,7$

$d_y = y_i - 30$ where $i=1,2,\dots,7$

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

$$r_{x,y} = r_{D_x,Dy} = \frac{\frac{\sum D_x D_y}{n} - \frac{\sum D_x}{n} \cdot \frac{\sum D_y}{n}}{\sqrt{\left[\frac{\sum D_x^2}{n} - \left(\frac{\sum Dx}{n}\right)^2\right]\left[\frac{\sum D_y^2}{n} - \left(\frac{\sum Dy}{n}\right)^2\right]}}$$

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 فيما بين الدخل والأنفاق الأسبوعى للأسر هو :

$$r_{x,y} = r_{D_x,D_y} = \frac{\frac{\sum D_x D_y}{n} - \frac{\sum D_x}{n} \cdot \frac{\sum D_y}{n}}{\sqrt{\left[\frac{\sum D_x^2}{n} - \left(\frac{\sum Dx}{n}\right)^2\right]\left[\frac{\sum D_y^2}{n} - \left(\frac{\sum Dy}{n}\right)^2\right]}}$$

$$= \frac{\frac{22}{7} - \frac{1}{7} \cdot \frac{3}{7}}{\sqrt{\left[\frac{33}{7} - \left(\frac{1}{7}\right)^2\right]\left[\frac{21}{7} - \left(\frac{3}{7}\right)^2\right]}} = \frac{3.081633}{\sqrt{(4.693878)(2.816326)}}$$

$$\simeq 0.848$$

وهذه النتيجة تفيد أن هناك علاقة طردية قوية فيما بين المتغرين x , y .

ملاحظات على معامل الأرتباط الخطى البسيط لبيرسون :

(1): معامل الارتباط الخطى البسيط لبيرسون يستخدم القيم الفعلية لبيانات المتغيرين فى الحساب ولاتختلف قيمتة باختلاف الطريقة المستخدمة فى حسابه. فكما اتضح لنا من الامثلة السابقة فإن قيمة معامل الارتباط الخطى البسيط لايتأثر بالعمليات الجبرية الاربعة .

(2):معامل الارتباط الخطى البسيط يقيس قوة العلاقة الخطية البسيطة فيما بين المتغرين وهو على هذا الاساس يحدد نسبة التغير فى احد المتغرين والتى يمكن تفسيرها على اساس وجود هذه العلاقة الخطية . ونسبة التغير هذه يتم قياسها عن طريق ما يسمى بمعامل التحديد Coefficient of Determination وهو عبارة عن مربع معامل الارتباط الخطى البسيط لبيرسون اى(r^2) ففى مثالنا السابق مثلا كانت قيمة معامل الارتباط فيا بين الدخل الاسبوعى والانفاق الاستهلاكى الاسبوعى هى 0.848 وهنا تبلغ قيمة معامل التحديد (r^2) القيمة 0.719104 حيث أن :

$$r^2 = (0.8480)^2 = 0.719104 \quad .$$

وهوما يعنى أن 71.9104 % من التغيرات التى تنشأ فى قيمة الانفاق على الاستهلاك الاسبوعى يرجع فى أساسة للتغيرات التى تنشأ على قيمة الدخل الاسبوعى وذلك بموجب العلاقة الخطية التي تربط بينهما وأن 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

الـى عوامـل اخـرى تـؤثر علـى الانفـاق الاسـتهلاكى الاسـبوعى بخـلاف الـدخل الاسبوعى للاسرة.

(3):تتوقف اشـارة معامـل الارتبـاط (اى اتجـاه العلاقـة)علـى قيمـة البسـط فـى صيغة هذا المعامـل وذلـك لأن المقـام والـذى يعبـر عنـه بالإنحرافـات المعياريـة للمتغيرين موجب دائما .

<u>ثانيا:حساب معامل الارتباط الخطى لبسيط لبيرسون فى حالة البيانات المبوبه :</u>

اذا كانت قيم ازواج المشاهدات الخاصة بالمتغيرين او الظاهرتين محل الدراسـة كبيره بحيث يصعب معها اجراء العمليات الحسابية اللازمـه لتحليلها احصائياً فأنه يجب فى هذه الحالة عرض البيانات الخاصة بـالمتغيرين فـي صورة جدول تكـراري مـزدوج وهـو مـا سـبق التعـرض لـه فـى مطلـع دراسـتنا . والجـدول التكراري المزدوج ما هو إلا جدول تكراري فى إتجاهين حيث يتم فيه تخصيص فئات أحد المتغيرين فى الصفوف وفئات المتغير الثانى فى الاعمدة وتمثل القيم الموجـودة ببـاقى خلايـا الجـدول التكـرارات المشـتركة المنـاظرة لفئـات احـد المتغيرين فى مقابل فئات المتغيرالاخر . هذا ويعد إنتشار التكرارات المشتركة داخل الجدول التكرارى المزدوج بأنه يعطى فكرة مبدئية عن طبيعة العلاقة فيما بـين المتغيـرين(y ، x) مـن حيـث كونـهـا علاقـة طرديـة ام عكسـية ، فتركيـز التكرارات حول أحد قطرى الجدول التكرارى المزدوج يفيد بقوة العلاقـة فيمـا بين المتغيرين . اما عن صيغة معامل الارتباط الخطى البسيط فى حالـة البيانـات المبوبة فهى نفس صيغ المعامل فى حالـة البيانـات المفردة مـع تـرجيح مراكـز فئات كل توزيع من التوزيعين بالتكرارات المزدوجة.

وبصفة عامـه فإنـه لحسـاب معامـل الإرتبـاط مـن البيانـات المبوبـة يـتم اشـتقاق التـوزيعين التكـراريين الهامشـيين مـن الجـدول التكـراري المـزدوج . ثـم حسـاب

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 X_i ، Y_j مراكز الفئتين (i),(j) للمتغيرين X ، Y علي الترتيب. أما F_x , F_y فهي تعبر عن التكرارات الهامشيه لكل من x , y

أما F_{xiyj} فهى تعبر عن التكرارات المزدوجة لخلايا الجدول المزدوج والناتجة عن تقاطع الفئتين i ، j للمتغير X ، Y علي الترتيب .

وتمثل الصيغه السابقة استخدام القيم الاصلية دون أى عمليات طرح او قسم من (أو على) المفردات الأصلية أي 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_y , D_x علي الترتيب .

مثال :

الجدول التالي يبين عدد سنوات الخبرة (X) والمرتب الشهري (Y) لعينه من مائة عامل تم اختيارهم بطريقة عشوائية من احدي الشركات:

X \ Y	150-	180-	210-	240-	270 – 300	$\sum$
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
$\sum$	10	15	50	18	7	100

والمطلوب حساب معامل الارتباط الخطي البسيط لبيرسون مفسرا ما تراه مناسبا من تعليق باستخدام معامل التحديد .

الحل :

اذا نظرنا للجدول التكراري المزدوج نجد ان التكرارات المزدوجه تتركز حول القطر الرئيسي للجدول (من أعلي جهة اليسار حتي اسفل جهة اليمين) مما يدل بشكل مبدئي علي وجود علاقة طردية بين المتغيرين x , y وحيث ان أطوال الفئات متساوية لكل من المتغيرين (اي ان التوزيعات الهاشميه منتظمه) لذا يفضل استخدام طريقة الانحرافات المعدلة وذلك علي النحو المبين التالي :

1- ايجاد التوزيعات التكرارية الهامشية وحساب المجاميع اللازمة لحساب الانحراف المعياري لكل توزيع :

***التوزيع التكراري الهامشي 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 $\sum D_x F_x = -5$ ، $\sum d_x^2 F_x = 185$.

*** التوزيع التكراري الهامشي للمتغير (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 : $\sum d_y F_y = -3$, $\sum d^2 y F_y = 101$

*أما لايجاد المجموع المشترك $\sum D_x D_y F_{xy}$: فيتم إستبدال فئات المتغيرين فى الجدول التكراري المزدوج بقيم الإنحرافات المعدلة لهذه المتغيرات ثم حساب إيجاد حواصل ضرب إنحراف الصف (i) أي D_{xi} فى تكرار الخلية المزدوجة أى F_{xy} في إنحراف العمود (j) اي D_{yj} 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 (91) . ومن ثم يكون معامل الارتباط الخطي البسيط لبيرسون هو:-

$$r_{x,y} = r_{D_x,Dy} = \frac{\frac{\sum D_x D_y F_{xy}}{\sum F} - \left(\frac{\sum D_x F_x}{\sum F}\right)\cdot\left(\frac{\sum D_y F_y}{\sum F}\right)}{\sqrt{\left[\frac{\sum D_x^2 F_x}{\sum F_x} - \left(\frac{\sum Dx\, F_x}{\sum F_x}\right)^2\right]\left[\frac{\sum D_y^2 F_y}{\sum F_y} - \left(\frac{\sum Dy\, F_y}{\sum F_y}\right)^2\right]}}$$

$$= \frac{\frac{91}{100} - \left(\frac{-5}{100}\right)\cdot\left(\frac{-3}{100}\right)}{\sqrt{\left[\frac{185}{100} - \left(\frac{-5}{100}\right)^2\right]\left[\frac{101}{100} - \left(\frac{-3}{100}\right)^2\right]}}$$

$$= \frac{0.9085}{\sqrt{[1.8475]\,[1.0091]}} = 0.665$$

اي ان العلاقة طردية متوسطة الي حد ما فيما بين عدد سنوات الخبرة والراتب الشـهري وهـو مـا يؤكـد صـحة التنبـؤ باتجـاه العلاقـه مـن خـلال شـكل تركيـز التكرارات المزدوجة علي جانبي القطر الرئيسي . اما عن معامل التحديد r^2 $= (0.665)^2 = 0.4422$ وهو ما يفيد ان 44.22 % من التغيـرات التـي تنشـأ فـي اي مـن المتغيـرين يرجـع لتغيـر الآخـر والنسـبة الباقيـة ترجـع لعوامـل اخري .

ملاحظات على معامل الارتباط الخطى البسيط:

1- عنـد حسـاب معامـل الارتبـاط الخطـي البسـيط لبيرسـون فـي التطبيقـات العمليـة يجب ان يكون عدد القيم المستخدمة علي الاقل (30) مفرده . لانـه حينمـا تقـل عدد المفردات عن ذلك فإن وجود اي قيمة شـاذه (او حتـي اكثـر مـن قيمـة) او غيـر دقيقـة سيكون لهـا تـأثير كبيـر علـي قيمـة معامـل الارتبـاط الخطـي البسـيط لبيرسـون . لـذا ينصـح بأنـه اذا كـان عـدد القـيم المسـتخدمه فـي حسـاب معامـل الارتبـاط اقل من ثلاثـون مفـردة (والمفـرده هـي زوج مـن المشـاهدات (y, x)) فأن وجود قيمة شاذه سوف يكون لها تاثير كبيرا علي دقة قيمة معامل بيرسون للارتباط وفي هذه الحالة يفضل البحث عن مقياس آخر للارتباط يكون اكثر دقـه لأن معامل الارتباط الخطي البسيط لبيرسون سوف يعطي نتـائج مضللة فـي هـذه الحالـة نتيجة لتأثره بالقيم الشاذه او المتطرفه . فعلى سبيل المثـال لتوضـح تلـك النقطة دعنا نفرض ان لدينا البيانات التالية عن المتغيرين x ، y :

x:	1	2	3	4	5
y:	1	2	3	4	0

فإذا حسبنا معامل الأرتباط الخطى البسيط لبيرسون فيما بين أزواج المشاهدات الخمسة فإننا سنجد أن قيمة هذا المعامل مساوية للصفر وهو مايعنى إنعدام وجود علاقة خطية فيما بين المتغيرين X ، Y تماما . والآن دعنا نعيد حساب قيمة معامل الإرتباط الخطى البسيط لبيرسون وذلك بعد إستبعاد قيمة الزوج الخامس من المفردات (5 ، 0) حينئذ سنجد أن قيمة معامل الإرتباط الخطى البسيط لبيرسون سوف تساوى الواحد الصحيح وهو مايفيد أن العلاقة مابين المتغيرين أصبحت علاقة خطية تامة . وهنا نجد أن إستبعاد الزوج الخامس من المشاهدات قد حول العلاقة الخطية فيما بين المتغيرين من علاقة منعدمة إلى علاقة تامة ويرجع ذلك إلى أن زوج المشاهدات الأخير (الخامس) للمتغيرين على التوالى لا يتسق مع بقية أزواج المشاهدات لأنه يحتوى على قيم شاذة.

2- 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 6-2 / A ،B بينما هي في الشكل (6-1 / A) اكبر منها في الشكل (5-1 / C,B)

3- تتوقف اشارة معامل الارتباط علي اشارة البسط والذي يمثل التغاير بين المتغيرين ، وذلك لان المقام دائما موجب حيث انه حاصل ضرب الانحرافات المعيارية σ_x ، σ_y وكلاهما مقدار موجب .

2- 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 عن عدد الفئات للمتغير الاخر ، كما يمكن ان تكون الفئات متساوية الطول لأحد المتغيرين وغير متساوية الطول للمتغير الاخر .

3- أن إقتراب قيمة معامل الإرتباط الخطي البسيط من الصفر أو حتي مساواتها للصفر وإن 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 المتغيرين x , y في الشكل (D/1-5) تقترب من الصفر في حين ان هناك علاقة غير خطية قوية بين المتغيرين .

4- 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 فى هذه الحالة . 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 .

5- عند حساب معامل الارتباط لبيانات مبوبة فى جدول تكرارى مزدوج فانه اذا كانت فئات كل من المتغيرين متساوية الطول يفضل استخدام طريقة الانحرافات المعدلة .

6- فى حالـة حسـاب معامـل الارتبـاط لبيانـات المتغيـرين x ، y والمبويـة فـى جدول تكرارى مزدوج ، فانه قبل اجراء أية عملية حسابية يمكن الاستدلال على طبيعة العلاقة بين المتغيرين وذلك على النحو التالى :

***اذا كانـت التكـرارات المشـتركة داخـل خلايـا الجـدول تتركـزة حـول قطـر الجـدول الممتد من أعلى ركن من جهة اليسار الى اسفل الجدول جهـة اليمـين فـأن هذا يعنى ان هناك علاقة طردية بين المتغيرين وتزداد قوة هذه العلاقة بزيادة تركـز التكـرارات المزدوجـة حـول هـذا القطـر ، والعكـس صـحيح أى أنـه اذا كانـت التكرارات المشتركة داخل خلايا الجدول تتركزة حول قطر الجدول الممتـد من اعلـى اليمـين الـى ادنـى اليسـار فـإن هـذا يعنـى ان هنـاك علاقـة عكسـية بـين المتغيـرين وتتوقـف درجـة قـوة او ضـعف هـذه العلاقـة علـى مـدى تركـز هـذه التكرارات المزدوجة حول هذا القطر.**

7- تلزم الاشارة الى نقطة هامة وهى : انـه فـى دراستنا للارتبـاط ،يجب علينـا ان نفـرق بـين العلاقـة الاحصـائية – متمثلـة فـى قيمـة معامـل الارتبـاط بـين المتغيـرين – وبـين العلاقـة السـببية . فحـين نجـد ان هنـاك ارتباطـا قويـا بـين متغيـرين (اى قيمـة $|r_{xy}|$ تقتـرب مـن الواحـد الصـحيح) فـان ذلـك لايعنـى بالضرورة ان العلاقة بينهما هى بالضرورة علاقة سببية بمعنى ان التغير فى احد المتغيرين هو السبب فى التغير في قيمة الاخر . فعلى سبيل المثال : لـو كنـا بصدد ايجاد معامل الارتباط بين درجات مجموعة من الطلاب فى مادة الاحصاء وبين درجاتهم فى مادة القانون ، ووجدنا ان معامل الارتباط يسـاوى 0.95 فـان هـذا يعنـى ان هنـاك ارتباطـا قويـا بـين درجـات الطـلاب فـى مـادتى الاحصـاء والقانون ، ولكن هل يعنى ذلك ان ارتفاع (او انخفاض) درجات الطلاب فى مادة القـانون انمـا يرجـع الـى ارتفـاع (او انخفـاض) درجـاتهم فـى مـادة الاحصـاء او

العكس ؟ الاجابة بكل تاكيد هى : لا . وهنا يمكن القول ان هناك علاقة احصائية قوية بين المتغيرين رغم انعدام العلاقة السببية بينها . ومن جهة اخرى قد تنعدم العلاقة الاحصائية – ممثلة فى معامل الارتباط فيما 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 (6-2 ، 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 هو 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 هو أمر يخرج عن نطاق علم الاحصاء والذى يهتم فقط بمجرد قياس العلاقة بين المتغيرات.

8- أنه أحياناً ما نصل إلى نتائج غير منطقية بمعامل الإرتباط الخطي البسيط لقياس ماهية قوة العلاقة الخطية بين المتغيرين . بمعنى 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> : قد يكون هناك متغير ثالث يؤثر على المتغيرين (x ، y) 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 النظرى . 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 يقبله منطق أو يُقرُّه عاقل. إذ أن المنطقي أن العلاقة بين كمية الإنتاج وعدد ساعات تعطل الآلات يجب أن تكون طردية . إذا ما هو تفسير ما حصلناه عليه من نتيجة غير منطقية هنا ؟ وفي حالة كتلك ، يجب أن نفكر فى عامل(متغير) ثالث يؤثر على هذين المتغيرين (كمية الإنتاج ، عدد ساعات تعطل الآلات) وفي نفس الاتجاه . ومن بين التفسيرات المنطقية لذلك هو أن عدد الآلات يزداد عاماً بعد آخر فى المصنع بشكل يؤدي إلى زيادة كمية الإنتاج كما يؤدي أيضاً وفي نفس الوقت إلى زيادة عدد ساعات تعطل الآلات. وبالتالي فإننا نجد أنه خلال العشر سنوات هناك – بشكل عام – زيادة فى كل من كمية الانتاج وعدد ساعات الأعطال ، مما يؤدي فى النهاية إلى أن معامل الإرتباط

الخطي بينهما يكون موجباً وهو ما يخالف المنطق . وفي حقيقة الأمر فإن العلاقة العكسية بين المتغيرين يمكن الحصول عليها إذا ما كان عدد الآلات في المصنع شبه ثابت على مدار السنوات العشر.

<u>ثانيهما</u> : أن العلاقة بين المتغيرين (y،x) موضع الإعتبار هي علاقة ضعيفة ، بمعني أن معامل الإرتباط الخطي بينهما يقترب في قيمته من الصفر . وفي حالات كتلك ، فإنه ليس من المستبعد أن نحصل على ما يفيد بأن هناك علاقة عكسية بين المتغيرين في حين أنها يجب أن تكون طردية. وهذا أمر ممكن ولكن بشرط أن تكون العلاقة الخطية بينهما هي علاقة ضعيفة في جميع الحالات. حيث أنه لا يضر أن نجد أن معامل الإرتباط الخطي يساوي 0.05 في حين أنه يجب أن يكون 0.06- والعكس صحيح . ولكن الأمر غير المنطقي هنا أن تكون العلاقة بين المتغيرين طردية ضعيفة – على سبيل المثال – ونجد أن $r_{xy} = -0.07$ أو العكس.

9- 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.

2 – معامل فهنر للارتباط:

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 فروق المشاهدات لمتغيرين كميين Y , X عن أوساطهم الحسابية ولتكن هذه الازواج من الفروق 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** For i = 1,2,…,n **ى اى يتم حساب قيم:** $(x_i - \bar{x})$, $(y_{i-\bar{y}})$

- يتم حصر عدد الحالات التى تكون فيها ازواج الانحرافات لقيم المتغيرين عن أوساطهم الحسابية 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 لقيم المتغيرين ذات اشارات مختلفة . ونرمز لكل حالة من تلك الحالات بالرمز (d). ومن ثم يكون اجمالي مجموع تلك الحالات وليكن المقدار $\sum d$.

- 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 المتشابهة ،

($\sum$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 سالبة 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 X , Y . فحيث 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 .

3- 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 نادى . 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 موقعها 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 وهكذا إلى أن القيمة الكبرى فى قيم (x) تقابلها القيمة الكبرى للمتغير (y)) فإننا 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 بعد 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- <u>معامل ارتباط الرتب لسبيرمان :</u>

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>أولا: 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 إرتباط الرتب لسبيرمان يتم على النحو الاتي :

(1):ضع 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

بمعنى ان أصغر قيمة فى مفردات المتغير تأخذ الرتبة (1) 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 الوسط الحسابى لتلك الرتب المتتالية لتوضع امام القيم المتساوية من قيم هذا المتغير ولتكن فى النهاية رتب (X) يرمز لها بالرمز (R_x)، وكذلك رتب (Y) يرمز لها بالرمز (R_y).

(2): 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 لوضع الرتب للمتغيرين هو ان يكون المجموع الجبرى لهذه الفروق مساوية للصفر . 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)اى نوجد (D_i^2) ثم بالجمع نحصل على $\sum D_i^2$ والذى يستخدم لبيان مدى الاتفاق والاختلاف بين الرتب المتناظرة .

(3): 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 تكون قيمة المعامل r =1 وهو ما يعنى ان العلاقة طردية تامة بين المتغيرين .

<u>الثانية</u> : اذا كانت الرتب المتناظرة لأزواج المشاهدات عكسية تماما حيث تكون رتب قيم المتغير (x) هى : 1 , 2 ,........, (n-1) , n تقابلها وبصورة عكسية تماما رتب المتغير (y) والتى تأخذ الصورة n , (n-1) ,........, 2 , 1 وهنا تكون العلاقة بين المتغيرين علاقة عكسية تامة .

<u>الثالثة</u>: فى حالة عدم وجود علاقة ارتباط الرتب بين المتغيرين y , x فان قيمة معامل ارتباط الرتب لسبيرمان تكون صفرية .

<u>ملحوظة</u>: ان المؤشر الحقيقى لتمام العلاقة الخطية (سواء طردية تامة او عكسية تامة) هو قيمة معامل الارتباط الخطى البسيط لبيرسون . بمعنى أنه اذا كانت هناك علاقة خطية تامة ما بين المتغيرين y , x فانه يمكن القول بان معامل الارتباط الخطى البسيط سيكون مساويا 1+ , 1- وفى هذه الحالة فان 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 لبيرسون مساوية ايضا ±1 .

هذا بالاضافة الى ان الاختلاف فى قيمة 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 لمجموعة 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 . ولحساب هذا

المعامـل دعنـا نفتـرض ان العمـر هـو المتغيـر (x) والتقـدير هـو المتغيـر (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
$\sum$				0	15.5

لاحظ ان هناك قيم مكرره فى قيم المتغيرين y , x فبالنسبة للمتغير (x) الذى يعبر عن العمر نجد ان القيمـة 17 مكرره مرتـان فلـو لـم تكـن هـاتين القيمتـان مكررتـان لكانـت رتـبهم هـى 3 , 4 لـذا تـم اعتبـار الوسـط الحسـابى للـرتبتين المتتـاليين اى $\frac{3+4}{2} = 3.5$ وكذلك بالنسبة للمتغير (y) الذى يعبر عن التقدير فـان التقـدير good مكـرر ثـلاث مـرات . فلـولا هـذا التكـرار لكانـت رتـب هـذه 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. وعليه فإن قيمة معامل إرتباط الرتب هى :

$$r = 1 - \frac{6\sum D_i^2}{n(n^2 - 1)}$$

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 بأن العلاقة طردية وقوية بين عمر الفرد ودرجة ذكائه .

<u>ثانيا:حساب معامل ارتباط الرتب لسبيرمان فى حالة البيانات المبوبة :</u>

يتم حساب معامل ارتباط الرتب لسبيرمان فى حالة البيانات المبوبة فى جدول تكرارى مزدوج للمتغيرين x , y سواء كانت البيانات كمية أو وصفية حيث يتم إعطاء الفئات رتباً تصاعدية (أو تنازلية) ويشترط أن يكون الجدولين التكراريين منتظمين لكل من X ، Y بمعنى ان يكون كل فئات x متساوية الطول وكذلك كل فئات y متساوية الطول . ويتعذر استخدام هذه الطريقة فى حالة عدم تساوى اطوال الفئات لاحد المتغيرين او كليهما . ويمتاز هذا المعامل بانه يصلح لقياس الارتباط فى حالة الجداول التكرارية المفتوحة .

وجدير بالذكر ان حساب معامل ارتباط الرتب لسبيرمان فى حالة البيانات المبوبة بالجدول التكرارى المزدوج يعطى نفس النتيجة التى تم التوصل اليها باستخدام معامل الإرتباط الخطى البسيط لبيرسون 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 أى لجدول تكرارى مزدوج باتباع الخطوات الاتية :

(1):نعطى فئات المتغير x رتبا تصاعدية تبدأ بالوحدة اى 1 , 2 , ... ,n وكذلك نعطى فئات المتغير y رتبا تصاعدية تبدأ بالوحدة أيضا أى بالارقام : 1 ، 2 ، .... ، m . حيث n ، m يمثلان عدد فئات توزيع المتغير x وعدد فئات توزيع المتغير y على الترتيب .

(2):نرسم اقطار متوازية على الجدول التكرارى المزدوج يتم رسمها 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 y , x لهذا الاقطار المتوازية حيث يكون هذا الفرق مساويا للصفر بالنسبة للقطر الرئيسى ويأخذ قيما مختلفة عن الصفر بالنسبة لباقى الاقطار ويلاحظ ان هذا الفرق بين رتبتى y , x يكون متساويا على جميع اجزاء كل قطر من هذه الاقطار لذلك تسمى هذه الطريقة احياناً بطريقة الاقطار ذات الفروق المتساوية .

(3):نكون جدول تكرارى بسيط لرتب x ونحسب منه الانحراف المعيارى لرتب x اى σ_x وبنفس الطريقة نكون جدول تكرارى بسيط لرتب y ونحسب منه الانحراف المعيارى لرتب y اى σ_y.

(4):نكون جدول تكرارى لفروق الرتب وتكون تكراراتة هى مجاميع تكرارات الاقطار السابقة تقابلها الفروق المختلفة بين رتب x ورتب y ونحسب منه تباين هذه الفروق ونرمز له بالرمز σ_{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 وتباين فروق الرتب بينy , x 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 Classes (y) / Exp. Classes (x)	120-	125-	130-	135-	140-145	∑
280-			5	25	10	40
290-			40	20		60
300-			25	5	20	50
310-		30				30
220-330	10	10				20
∑	10	40	70	50	30	200

والمطلوب حساب معامل ارتباط الرتب لسبيرمان بين الانفاق الاستهلاكى وبين الادخار للاسرة ؟

الحل :

بـالنظر الـى شـكل توزيــع التكـرارات بالجـدول التكـرارى المـزدوج نجـد ان التكرارات تتركز حول القطر العكسى (من اعلى جـه اليمـين الـى ادنـى اليسـار) ممـا يكشـف بصـورة مبدئيـة عـن وجـود ارتبـاط عكسـى بـين المتغيـرين محـل 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 σ_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 يتم 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
Σ 0	200	-120	980

ومن ثم فمن المجاميع المبينة فان تبأين فروق رتب فئات التوزيع 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 متوسطة فيما بين الانفاق الاستهلاكى (x) والادخار (y) للاسرة فى تلك العينة.

<u>ملاحظات حول معامل ارتباط الرتب لسبيرمان :</u>

(1):يستخدم معامل ارتباط الرتب لسبيرمان لقياس العلاقة بين متغيرين سواء كانت القياسات ترتيبية اصلا او كمية تم تحويلها الى قياسات ترتيبية بينما معامل الارتباط الخطى البسيط لبيرسون لا يصلح للاستخدام الا اذا كانت القياسات كمية للمتغيرين او الظاهرتين محل الدراسة .

(2):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 .

(3):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) ترتيبا تنازليا ايضا والعكس صحيح فلا يصح ترتيب احد المتغيرين تنازليا مثلا وترتيب المتغير الاخر تصاعديا فى نفس الوقت .

ب – معامل كندال للارتباط :

إن أهم نقد يوجه لمعامل ارتباط الرتب لسبيرمان هو أنه يعاب عليه عدم وجود معنى طبيعى ومنطقى 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 :

$$(X_1,Y_1)\ ,\ (X_2,Y_2)\ ,\ \ldots\ldots\ldots\ldots\ ,\ (X_n,Y_n)$$

فانه يتم اتباع الخطوات التالية لحساب هذا المعامل :

(1):ضع رتبا تصاعدية لكل من المتغير x , y .

(2):يتم كتابة رتب القيم المتناظرة للمتغيرين على ان تكون رتب احد المتغيرين وليكن (x) مكتوبة فى ترتيبها الطبيعى اى $1 , 2 , 3 ,\ldots, n$.

(3):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 M_t وذلك لنفس الرتبة (t) .**

*يتم حساب الفرق ما بين N_t وبين M_t حيث $(t = 1 , 2,\ldots, n)$ ثم نجمع ذلك بالنسبة لجميع رتب قيم (y) وسنرمز لهذا المجموع بالرمز (C) اى ان :

$$C = \sum_{t=1}^{n} (N_t - M_t)$$

*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 .

<u>مثال:</u>

احسب معامل كندال للارتباط وذلك لبيانات درجات الطلاب العشرة فى مادتى الاحصاء والمحاسبة والذى تم حساب معامل الارتباط لفهنر لهم سابقا .

<u>الحل :</u>

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 ($R_{y)}$) R_x	R_y	N_t	M_t	$C=N_t-M_t$
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 .

3- <u>الارتباط بين الصفات :</u>

<u>(صفات مبوبة تكراريا فى جداول تكرارية مزدوجة):</u>

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 عبارة عن صفات ويتعذر قياسها كمياً (بمعنى ان احد الظاهرتين قد يكون

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 وكليهما يتكون من قسمين فقط 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) First Phenomenon (x)	First part	Second part
First part	F_{11}	F_{12}
Second part	F_{21}	F_{22}

حيث F_{22} , F_{21} , F_{12} , F_{11} تشير الى تكرارات خلايا الجدول التكرارى المزدوج حيث $F_{i\,J}$ تشير الى تكرار الخلية التى تقع فى الصف (i) والعمود (j)حيث $i=1,2$ ، $J=1,2$. فى هذه الحالة فان افضل مقياس لقياس الارتباط بين الظاهرتين x , y هو معامل 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 المتغيرين y , x اقترانا تاما . اما اذا كانت قيمة هذا المعامل قيمة سالبة فإن ذلك يرجع فقط لعملية تبويب البيانات ، لذا فانه يجب اهمال الاشارة السالبة لانه لا يصح ان يقال ان هناك علاقة طردية او عكسية بين ظاهرتين وصفيتين وعندما تنعدم قيمة هذا المعامل اى حينما 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 200 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 .

<u>الحل :</u>

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 خصائص معامل يول للاقتران ما يلى :</u>

*** اذا كانت احد التكرارات الاربعة F_{11}اوF_{12} او F_{21} او F_{22} على الاقل مساوية لصفر فان قيمة معامل يول للاقتران تكون حينئذ مساوية 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 .**

مثال:

فى دراسة لايجاد العلاقة بين الاصابة بمرض القلب (x) والتدخين (y) تم جمع بيانات من 120 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(2 × 2) لذا فانه لقياس قوة 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:

الجـدول التــالى يوضـح البيانــات الخاصــة بتوزيــع 250 طفـلا حسـب الاصــابة 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 الاصابة بمرض الحصبة وعدم التحصين ضده ومن بيانات الجدول المعطى نستطيع ان نستنتج ان الاطفال الذين لـم 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 يأخذ دائما قيمة موجبة وحتى لو أن قيمة المعامل سالبة فإننا نقوم بإهمال إشارة معامل الاقتران مهملين الاشارة السالبة للنتائج 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 تم إعادة ترتيب الجدولين السابقين فانه فى حالة التبديل لصف على آخر او عمود على آخر لنتج لنا معاملى الاقتران مساويا للقيم (0.311 - , 0.94 -) للجدولين على الترتيب 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 .

<u>(ب): معامل التوافق</u>: 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 من أكثر من قسمين . فقد قدم لنا كارل بيرسون K.Pearson مقياسا احصائيا يقيس قوة العلاقة ما بين هاتين الظاهرتين فى تلك الحالة وهو ما 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 :

x \ y	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 :

(1): يتم تربيع تكرار كل خلية اى $(F_{iJ})^2$ حيث $i = 1 , 2 , ... , n$ ، $j=1 , 2 , ... , m$

(2): يتم قسمة مربع تكرار كل خلية على حاصل ضرب مجموع الصف فى مجموع العمود الذى تنتمى اليه الخلية ثم يتم تجميع النواتج لجميع الخلايا وليكن هذا المجموع هو (C) . فيكون معامل التوافق هو :

$$C_{Cont.} = \sqrt{\frac{C-1}{C}}$$

حيث :

$$C = \sum_{i=1}^{n}\sum_{J=1}^{m}\left(F_{iJ}^2 / F_{i}.F_{.J}\right)$$

حيث (Fi.) ، (F.j) عبارة عن مجموع تكرارات الصف(i) ومجموع تكرارات العمود (j) على الترتيب .

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 (دون إثبات) 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	∑
Not Qualified	2	10	12
Preparatory	3	6	9
Secondary	20	3	23
University Qualification and more	35	1	36
∑	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 حساب الاتى :

$$C = \sum_i \sum_J \left(F_{ij}{}^2 / F_{i\cdot} F_{\cdot J}\right)$$

$$= \frac{(2)^2}{12 \times 60} + \frac{(10)^2}{12 \times 20} + \frac{(3)^2}{9 \times 60} + \cdots$$

$$\dots . + \frac{(1)^2}{36 \times (20)} = 1.519$$

$$C_{Cont.} = \sqrt{\frac{C-1}{C}} = ((1.519 - 1) / 1.519)^{0.5} = 0.58$$

وحيث ان الحد الاقصى لمعامل التوافق فى هذا المثال هو :

$$maximum\,Value\ of\ C_{cont.} = \sqrt[4]{\frac{(n-1)(m-1)}{n.m}}$$

$$= \sqrt[4]{\frac{(2-1)(4-1)}{2(4)}} = 0.74$$

ولتحديد ماهية قوة العلاقة يتم حساب النسبة ما بين قيمة معامل التوافق والحد الاقصى له وهى :

$$\frac{0.58}{0.78} = 0.74$$

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 الفصل السادس

(1):البيانات التالية تمثل اعمار مجموعة من 10 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 .

(2):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 .

(3):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 .

(4):اذا كانت x ترمز الى سعر السلعة ، y ترمز الى الكمية المعروضة عند كل ثمن فالمطلوب تحديد قوة 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**

(5):فيما يلى الدرجات التى حصل عليها 10 طلاب فى مادتى الرياضة 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 فيما بين المادتين .

(6):فيما يلى الدرجات التى حصل عليها 10 طلاب فى امتحان مادتى اللغة الانجليزية 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 .

(7): احسب باستخدام البيانات التالية معامل سبيرمان للارتباط بين الظاهرتين 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

(8):فى أحد 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	2	5	1	4	3	6
المجموعة الثانية	5	4	3	2	1	6

والمطلــوب تحديــد قيمــة معامــل الارتبــاط (معامــل الرتــب) بــين افضــلية المجموعتان .

(9):فيما يلى اسعار السلع الجلديـة والخـزف (للوحـدة الواحـدة) خـلال عشـرة ايام فى منطقة خان الخليلى :

السلع الجلدية	5	7	8	4	9	3	2	5	4	3
السلع الجلدية	2	4	5	5	6	5	4	4	3	2

والمطلـوب ايجـاد معامـل بيرسـون وايضـا معامـل ارتبـاط سـبيرمان (الرتـب) ومقارنـة قيمة معاملى الارتباط وبيان سبب الاختلاف فى قيمتها إن وجد .

(10):يوضـح الجـدول التـالى توزيـع عينـة بحسـب فئـات دخـولهم الشـهرية x وإيجار المسكن الذى تقطنه بالجنيهات (y) :

y / x	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 .

(11):اذا كان لدينا البيانات الاتية عن كلا من الظاهرة X ، Y .

$\sum XF_X = 39000$ ، $\sum X^2 F_X = 1605000$

$\sum y_{F_y} = 29950$ ، $\sum_{y^2 F} = 919750$

$\sum_{XYF_{XY}} = 1197250$ ، $\sum_{F_X} = \sum_{F_Y} = \sum_{F_{XY}}$

والمطلوب : ايجاد :

1-الوسط الحسابى والانحراف المعيارى ومعامل الاختلاف للظاهرتين وقارن تشتت الظاهرتين X ، Y .

2-حساب معامل الارتباط بين X ،Y .

(12): 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	∑
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
∑	5	19	18	7	50

والمطلوب :

ايجاد قوة العلاقة ونوعها بين اعمار نزلاء الفندق و عدد الليالى السياحة وعلق على النتيجة .

(13): ارسم الشكل الانتشارى ثم أوجد معامل الارتباط الخطى البسيط بين المتغيرين x , y وذلك من واقع البيانات التالية :

X: 12 ، 18 ، 21 ، 24 ، 28 ، 32 ، 45 ، 56، 62، 71

Y: 11 ، 13 ، 22 ، 23 ، 30 ، 35 ، 42 ، 50 ، 60 ، 65

(14): فيما يلى الدخل الاسبوعى (x) والانفاق الاسبوعى (y) لعينة مكون من 8 أسر :

X: 43 ، 58 ، 55 ، 62 ، 81 ، 72، 98 ، 76 .

Y : 35 ، 55 ، 53 ، 61 ، 79 ، 45 ، 50 ، 62 .

والمطلوب : حساب معاملات الارتباط التالية :

أ-معامل الارتباط الخطى البسيط .

ب-معامل ارتباط الرتب (معامل سبيرمان).

ج-معامل ارتباط كندال .

د-معامل ارتباط فهنر .

(15): فيما يلى بيان بعدد ساعات التدريب (x) التى تلقتها مجموعة من العاملين باحدى الشركات ونتيجة تقييم أدائهم فى نهاية الملائم (y) .

X :	**20**	**42**	**50**	**35**	**40**	**60**	**65**
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 .

(16):البيانات التالية تبين تقديرات نجاح عشر طلاب فى مادتى الاحصاء (x) والرياضيات (y) :

الطالب	تقديرات (X)	تقديرات(Y)
1	مقبول	جيد
2	جيد	جيد
3	مقبول	جيد جداً
4	ممتاز	مقبول
5	جيد جداً	جيد جداً
6	جيد جداً	جيد
7	جيد	مقبول
8	ممتاز	جيد جداً
9	مقبول	ضعيف
10	ضعيف	ضعيف جداً

والمطلوب : تحديد ماهية قوة العلاقة بين تقديرات الطلاب فى مادتى الاحصاء والرياضيات وذلك باستخدام المقياس الاحصائى الملائم .

(17): البيانات التالية تمثل الدخل الاسبوعى (x) والانفاق الاسبوعى (y) وعدد أفراد الاسرة (z) وذلك بالنسبة لعينة مكونة من 10أسر :

X : 80 ، 110 ، 130 ، 75 ، 82 ، 65 ، 76 ، 91 ، 95 ، 100

Y : 75 ، 93 ، 102 ، 60 ، 55 ، 50 ، 63 ، 82 ، 80 ، 94

Z : 5 ، 6 ، 4 ، 3 ، 3 ، 2 ، 3 ، 2 ، 5 ، 6

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 .

(18):من بيانات التمرين السابق ، احسب معاملات الارتباط التالية :

- معامل ارتباط سيبرمان.

- معامل ارتباط فهنر .

- معامل ارتباط كندال.

وذلك لازواج المتغيرات : (Y ، X) ، (Z ، X) ، (Z ، Y) . مع ابداء التعليق المناسب .

(19):الجدول التالى يوضح التوزيع التكرارى المزدوج للمتغيرين X ، Y :

Y \ X	-25	-35	-45	-55	-65	-75	-85	المجموع
-40	3	5	3					11
-50	5	6	1	1				13
-60		1	7	10	1			19
-70			6	9	7	3		25
-80			1	3	8	5	1	18
-90						5	9	14
المجموع	8	12	18	23	16	13	10	100

والمطلوب :

أ-ايجاد معامل الارتباط الخطى البسيط بين X ،Y .

ب-باستخدام معامل التحديد فسر ما تحصل عليه من نتائج.

(20): فيما يلى التوزيع التكرارى المزدوج لدرجات 50 طالبا فى كل من مادتى الاحصاء (x) والرياضيات (y) :

(x) / (y)	-50	-60	-70	-80	-90	المجموع
-60	2	4	3	2		11
-65	4	8	9	5		26
-70	2	4	11	3	2	22
-75		2	15	9	4	30
-80			3	2	2	7
-85				2	1	3
-90					1	1
المجموع	8	18	41	23	10	50

<u>والمطلوب</u> :

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 .

(21): 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	25	8	23
غير ذلك Other	12	15	27
المجموع	37	23	60

<u>والمطلوب</u> : دراسة اثر العوامل الوراثية فى لون العينين وذلك باستخدام المقياس الاحصائى المناسب .

(22): فيما يلى توزيع عينة من 100 شخص حسب درجة تعليمهم (x) وآرائهم تجاه قضية عمل المراة (y) :

x / y	Univ.	Second	Prep And Less	المجموع
مؤيد	5	28	9	42
معارض	50	2	6	58
المجموع	55	30	15	100

<u>والمطلوب</u> : وضح الى اى حد تعتمد وجهة النظر تجاه عمل المراة على درجة التعليم مستخدما فى ذلك المقياس الاحصائى الملائم .

(23):فيما يلى نتيجة امتحان مادة رياضيات الاستثمار(X) فى احدى كليات التجارة حسب حالة الطلاب (منتظمين و منتسبين) (y) وذلك فى دور يناير 1987 :

Y \ X	Failed	Pass	المجموع
إنتظام	192	716	908
إنتساب	144	208	352
المجموع	336	924	1260

<u>والمطلوب :</u> وضح الى اى حد تتوقف نتيجة الامتحان على نظام الدراسة مستخدما فى ذلك المقايس الاحصائى الملائم ؟

(24): الجدول التالى يوضح نوزيع الطلاب المنتظمين و المنتسبين فى احدى كليات التجارة وذلك حسب تقديراتهم فى نتيجة امتحان مادة الاحصاء لدور مايو 1987 :

التقدير نظام الدراسة	ضعيف جداً	ضعيف	مقبول	جيد	جيد جداً	ممتاز	المجموع
إنتظام	166	148	170	167	36	15	702
إنتساب	108	88	59	24	5	1	285
المجموع	274	236	229	191	41	16	987

<u>والمطلوب</u> : باستخدام المقياس الاحصائى الملائم ،حدد درجة قوة وطبيعة العلاقة مابين نظام الدراسة وتقديرات الطلاب فى مادة الاحصاء.

(25):البيانـات التاليـة تمثل درجـات عشـرة طـلاب فـى كـل مـن مـادتى الفلسـفة وعلم نفس :

X	16	10	17	13	12	15	18	16	14	19
Y	14	12	15	14	16	13	17	16	12	15

<u>والمطلوب:</u> حساب معامل الارتباط الخطى البسيط . ثم علق على النتيجة التى تحصل عليها باستخدام معامل التحديد .

(26): فيما يلى التوزيع التكرارى المزدوج للمتغيرين :-

X / Y	-10	-20	-30	-40	60-50	المجموع
-25	2	1	1			4
-30	1	4	2			7
-35	2	3	5	3		13
-40		2	6	6		14
50-45			5	4	3	12
المجموع	5	10	19	13	3	50

<u>والمطلوب</u>

أ- حسـاب معامـل الارتبـاط الخطـى ثـم فسـر النتيجـة التـى تحصـل عليهـا وذلـك باستخدام معامل التحديد.

ب- قارن بين التشتت النسبى لكل من المتغيرين .

ج- ياستخدام معامل بيرسون الاول قارن بين الالتواء فى كل من التوزيعين.

(27):فى دراسة لتحديد العلاقة مابين درحة تعليم الرجل ووجهة نظره تجاه عمل المراة تم استطلاع اراء مائة رجل وكانت آراؤهم على النحو المبين بالجدول التالى :

درجة التعليم وجهة نظر تجاه عمل المراة	جامعى فما فوق	ثانوى	اعدادى فأقل
مؤيد	28	22	5
معارض	5	10	30

المطلوب :

1- تحديد طبيعة ودرجة قوة العلاقة مابين درجة التعليم للرجل ووجهة النظر تجاه عمل المراة وذلك باستخدام المقياس الاحصائى الملائم .

2- لو إفترضنا ان هناك شكوكا حول صحة آراء الرجال الحاصلين على مؤهل إعدادى فاقل وانه تم استبعادهم نهائيا من الدراسة. والمطلوب هنا قياس درجة قوة العلاقة مابين درجة التعليم وتأييد عمل المرأة فى هذه الحالة مستخدما المقياس الاحصائى الملائم .

3- قارن بين النتيجنين فى كل من (أ) ، (ب) مع إبداء التعليق المناسب .

الفصل السابع

الانحدار الخطي البسيط

Simple linear Regression

من خلال دراستنا في الباب السابق تبين لنا 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 لتحديد شكل العلاقة الرياضية التى تربط مابين متغيرين(أو أكثر) 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 1885 علي يد فرانسيس جالتون حينما أهتم بدراسة الصفات الوراثية للابناء والاباء . فقد لاحظ ان الابناء لآباء طوال القامه يميلون في المتوسط الي ان يكونوا قصارا عن آبائهم في حين ان الابناء لآباء قصار القامه يميلون في المتوسط الي ان يكونوا طوالا عن آبائهم . 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 لخطة 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 والآلات 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 . فإذ 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 قيمة المتغير التابع . هذا 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 :

لقد أتضح لنا مما سبق في المقدمة السابقة انه عند دراستنا للانحدار الخطي البسيط لابد من تحديد كل من المتغير المستقل والمتغير التابع . فاذا كان لدينا المتغيران X . Y 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 الخطية التالية :

$$\hat{Y} = a + bX$$

ويسمي هذا النموذج بخط انحدار Y علي X حيث ($\hat{y}$) هي القيم المستنتجه (المتوقعه) والمحسوبة 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) والمحسوبة من معادلة خط انحدار X علي Y وذلك عند قيمY المختلفة .

هذا وكلما زادت تركز نقاط الشكل الانتشاري الذى يبين بصورة مبدئية عن مدى قوة أو ضعف العلاقة بين المتغيرين (شكل الإنتشار السابق الإشارة إليه فى الفصل السابق) حول الخط المستقيم المقدر كلما زادت قوة العلاقة الخطية فيما بين المتغيرين محل الدراسة والعكس صحيح . وكما سبق وان أوضحنا عند دراستنا للارتباط فانه اذا كانت العلاقه بين المتغيرين X ، Y علاقة تامة (اي ان:$r = \pm 1$) فأن جميع ازواج المشاهدات(x ، Y) تحقق ايا من هذين الخطين سواء Y علي X أو X علي Y . وفي تلك الحالة فأن كافة القيم الفعلية تتساوي مع كافة القيم المقدرة المقابلة لها اي أن: $Y = \hat{Y}$ ، $X = \hat{X}$ وذلك لجميع قيم X , Y 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 (أو الفعلية) للمتغيرين X , Y علي الترتيب أما $\hat{Y}$ ، $\hat{X}$ فانهما كما سبق تعريفها يمثلان القيم المتوقعه لكل من Y , X والمقدرة من معادلتي خطي الانحدار Y

علـي X أو X علـي Y علـي الترتيـب . هـذا وكلمـا كـان الفـارق بـين القـيم المشـاهدة والقـيم المتوقعـة صـغيرا كلمـا زادت قـوة العلاقـة الارتباطيـة بـين المتغيرين محل الدراسة والعكس صحيح وسوف نتعرض الان لدراسـة معادلـة خطي الانحدار آخذين في الاعتبار البيانات الغير مبوبة والبيانـات المبوبـة وذلـك علي النحو المبين التالي :

<u>أولا:معادلة خط انحدار Y على X :</u>

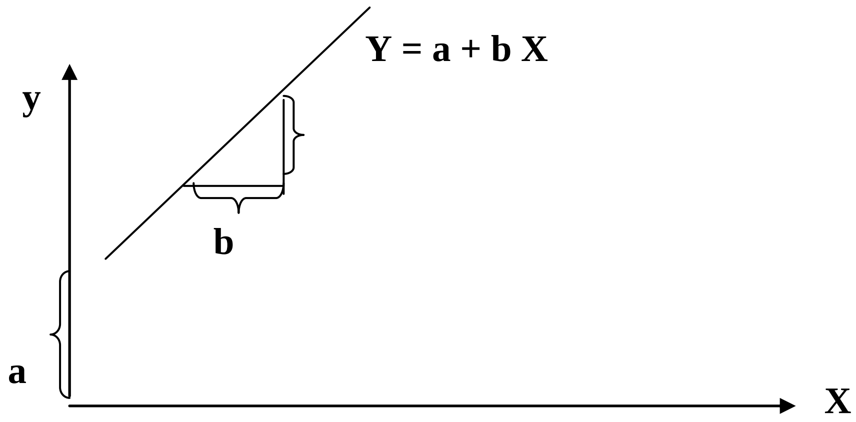
من خلال شكل الإنتشار الذى يشير لشكل العلاقة مـابين متغيـرين فـإذا كـا لـدينا علاقة بين متغيرين أحدهما متغير مستقل وليكن (X) وبين متغير آخر ولـيكن (Y). وبـإفتراض الصـورة الخطيـة لمعادلـة خـط انحـدار Y علـي X تاخـذ الصورة الخطية التالية :

$$\hat{Y} = a + bX \qquad (1\text{-}7)$$

حيث :

$\hat{Y}$ تعبر عن القيمة المتوقعه للمتغير (Y) عند قيم (X) المختلفة ،أما b, a فهمـا مقاديرثـابتـة . وهـذان الثابتـان يحـددان صـورة الخـط المسـتقيم تمامـا لـذا يسمي هذان الثابتان بمعلمتي Parameters خط الانحدار . ومن خلال العلاقـة (1-7) فان الثابت a هو عبارة عن قيمـة Y عندما تنعدم قيمـة المتغيـر المسـتقل X أى حينما X = 0 ، اما عن الثابت (b) فهـو عبـارة معدل التغيـر فـي قيمـة المتغير (Y) بالنسبة الـي (X) . هذا ويسمي الثابت (a) بثابت معادلـة إنحدار Y على X أما (b) فتسمي بمعامل خط انحدار Y علي X .

أمـا X فهـي عبـارة عن القيم المشـاهدة (الفعليـة) للمتغيـر X ويمكـن تمثيـل معادلة خط انحدار(Y) علي (X) بيانيا من خلال الشكل الموضح التالي :



شكل (1-7)

حط إنحدار (y) على (X)

وفي حالة تمام العلاقة بين المتغيرين (Y . X) (اي حينما تكون $r_{xy} = 1\pm$) 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 انعدام ما يسمي بأخطاء التقدير . وخطأ التقدير هو عبارة عن الفرق ما بين القيم الفعلية والقيم المتوقعه للمتغير التابع Y 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 لقيمة خطأ التقدير بالرمز (e_i) فإن :

$$e_i = Y_i - \hat{Y}_i \qquad for\ i = 1, 2, \dots\dots, n$$

حيث أن خطأ التقدير (e_i) عبارة عن الفرق ما بين القيمة المشاهدة والقيمة المتوقعه وهو ما يسمي بالخطأ في تقدير (Y_i) . وهذا الفرق يكون سالبا اذا كانت قيمة Y المشاهدة اسفل خط الانحدار ، ويكون هذا الفرق موجبا اذا كانت قيمة (Y) المشاهدة أعلي خط الانحدار .

وبأستخدام المعادلة (1-7) السابقة فانه يمكن التعبير عن الخطأ في التقدير علي الصورة التالية :

$$e_i = Y_i - \widehat{Y}_i = Y_i - a - bx_i$$

1-ايجاد معادلة خط انحدار (Y) علي (X) للبيانات الغير مبوبة :

اذا كان لدينا عدد (n) من ازواج المشاهدات للمتغيرين X , Y ولتكن :

$$(x_1 , y_1), (x_2 , y_2), (x_3 , y_3) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots, \ (X_n, Y_n)$$

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 (1-7) وهما a , b. وما نعنيه بكلمة افضل تقدير في شأن المعلمتين a , b كما سبق وان ذكرنا في أفضل تقدير لهاتين المعلمتين سوف يجعل الفروق ما بين القيم المشاهدة والقيم المتوقعه للمتغير (Y) أقل مايمكن والتى علي اساسها يمكن الحكم بصورة مبدئية علي مدي جودة تمثيل خط الانحدار للعلاقة بين المتغيرين X ، Y . 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 قيم 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 Square 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 إحصائية مفضلة من بينها خاصية عدم التحيز والكفاءة وهو ما يخرج عن نطاق دراستنا في هذا الكتاب .

<u>*تقدير معلمتي نموذج خط انجدار Y علي X باستخدام طريقة المربعات الصغرى :</u>

وفقا لطريقة المربعات الصغري فانه يمكن تقدير معلمتي معادلة خط انحدار Yعلي X علي النحو التالي :

حيث ان معادلة خط انحدار Yعلي X هي : $\quad Y = a + bx$

فأنه يمكن ايجاد قيمتي a , b والتي تجعل مجموع مربعات الفروق(e_i^2) اقل ما يمكن وذلك من خلال ايجاد النهاية الصغري للدالة (7 – 3) 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 .**

ومـن ثـم فبتطبيـق الشـرط الاول علـي الدالـة (7-3) وذلـك بأعتبـار التفاضـل 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 (7-4) $\sum y_i = na + b \sum x_i$

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

$$\sum x_i\, y_i = a \sum x_i \;\; + b \sum x_{i^2} \qquad\qquad (5\text{-}7)$$

هذا وتسمي المعادلتان (7 -4) و (7 - 5) بالمعادلات الطبيعية للنموذج وبحل هاتين المعادلتين معا يمكن الحصول علي تقدير لمعلمتي نموذج خط الانحدار (a), (b) . و من هاتين المعادلتين يمكن الوصول الي التقدير التالي :-

$$\hat{b} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$

وهناك صورة أخرى وهى :

$$= \frac{\frac{\sum xy}{n} - \frac{\sum x}{n} \cdot \frac{\sum y}{n}}{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2} \qquad\qquad (6\text{-}7)$$

وقد فضلنا الصيغه (7- 6) عن سابقتها وذلك لوجود نوع من التشابه بينها وبين صيغة معامل الارتباط الخطي البسيط لبيرسون . اما قيمة ثابت معادلة الانحدار (a) فيتحدد من خلال الصيغه التاليه :

$$\hat{a} = \bar{y} - \hat{b}\,\bar{x}$$

هذا ويمكن اثبات ان المشتقه الاولي عند النقطة(b , a) تحقق الشرط الثاني لايجاد النهاية الصغري للدالة (7 – 3) حيث تتحدد قيمة (a) بالصيغة (7– 7) بينما تتحدد قيمة معامل الانحدار b من خلال الصيغة (7-6).

هذا ويمكن ان نيسر علي القارئ كيفية تذكر المعادلتين الطبيعتين (7-4) و (7-5) 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 (7-4) وذلك من خلال اعتبار مجموع طرفي معادلة خط إنحدار Y علي X مع اعتبار ($\hat{Y}$) بدلا من (Y)حيث أن مجموع القيم الفعلية يساوي مجموع القيم المتوقعة . اي أن $\sum \hat{y} = \sum y$ اما المعادلة الطبيعية الثانية (7 – 5) 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 :-

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i \qquad (8\text{-}7)$$

ومن العلاقة (7-8) نخلص إلى :

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 فأنه إذا تم إعتبار هذه العلاقة فى المتوسط . فيكون :

$$\bar{y} = a + b\bar{x}$$

$$\hat{y} - \bar{y} + b(x - \bar{x}) \qquad \textbf{(9-7)}$$

ومن المعادلة (7 – 9) يمكن التوصل إلي صورة ثالثة لمعادلة خط انحدار (y) علي X وهي :

$$\hat{y} = \bar{y} + b(x - \bar{x}) \qquad (10\text{-}7)$$

هذا وسوف نقتصر في دراستنا علي استخدامنا للطريقة المباشرة في تقدير معلمتي خط الانحدار .

مثال :

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:-

(1):تقدير معادلة خط انحدار Y علي X بأفتراض انها خطية .

(2):ايجاد حجم المدخرات المتوقع عندما يصل سعر الفائدة الى 20 % .

(3):ايجاد معادلة خط الانحدار حينما تقل قيمة اسعار الفائدة بمقدار 5 % .

الحل

1 – 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 (a) ، (b) فإنه 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 :

$$\hat{b} = \bar{y} - a\bar{x}$$

$$= \frac{105}{7} - (1.524)\left(\frac{77}{7}\right) = -1.769$$ **or = - 1.764**

ومن ثم فان معادلة خط انحدار Y علي X تأخذ الصورة الخطية التالية :

$\hat{Y} = -1.764 + 1.524\ X_i$

1- ولايجاد حجم المدخرات المتوقع اي ($\hat{Y}$) عندما يصل سعر الفائدة اي (X) الي 20 في المائة فأنه يتم التعويض عن قيمة المتغير المستقل (X) في معادلة خط الانحدار المقدرة بالقيمة (20)

اي ان :

$\hat{Y} = -1.764 + 1.524\ (20) = 28.716$ **(million L. E.)**

2- حينما تقل اسعار الفائدة بمقدار (5) في المائة لكافة النقاط المعطاة فأنه بدلا من أيجاد المعادلة المقدرة من خلال طرح 5 في المائة من كافة قيم المتغير

المستقل (x) المعطاة ثم تقدير المعالم من جديد فيمكن إستنتاج المعادلة الجديدة حينما تنخفض أسعار الفائدة بمقدار (5) في المائة بأن يتم التعويض في معادلة النموذج المقدر سابقا (*) عن قيمة (X) على أنها تساوي X مطروحا منها (5) اي ان :

$$\hat{y}\downarrow = -1.764 + 1.524\,(x - 5)$$

X= X - 5

فينتج أن :

$$\hat{y} = -1.764 + 1.524x - 7.62$$

$$\hat{y} = -9.384 + 1.524x$$

اي ان اثر انخفاض سعر الفائدة بمقدار (5) في المائة يقع فقط أثره علي ثابت معادلة الانحدار أي على قيمة (a) أما ميل خط الانحدار (b) فهو لم يتأثر بهذا التغير .

أخطاء التقدير :-

بعد تحديد معادلة خط انحدار Y علي X فأنه يمكن ايجاد القيم المتوقعه للمتغير التابع (Y) المختلفة وذلك كما هو موضح في الجدول التالي بالتطبيق علي المثال السابق :-

ولحساب القيم المتوقعه لـ Y اي $\hat{y}$

وحساب اخطاء التقدير في علاقة إنحدار Y على X : فيتم تكوين الجدول التالي بعد تقدير نموذج خط إنحدار Y على X :

X_i	Y_i	$\hat{y}_i = -1.764+1.524x_i$	$e_i = y_i - \hat{y}_i$
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 العمود 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>2-ايجاد معادلة خط انحدار Y على X للبيانات المبوبة :</u>

لايجاد معادلة خط انحدار Yعلي X للبيانـات المبوبـة من خـلال جدول تكـراري مزدوج نقوم بأتباع الخطوات التالية :-

- اعتبـر (X) المتغيـر الاول (المسـتقل) حيـث X هـي اقسـام هـذا المتغيـر فـي حالة كونه متغيرا منفصلا او مراكز فئاته اذا كان متغيرا متصلا .

- احسـب متوسطات المتغيـر الثـاني (التـابع) Y المقابلـة لاقسـام هـذا المتغيـر المنفصل او مراكز فئات المتغير (X) المتصل . وهذه المتوسطات تعتبر بمثابة قيم المتغير (Y) .

- اوجد معادلـة انحدار Y علي X والتي تأخذ الصيغه (7 – 1) وذلـك مـن خـلال تقدير معالمها من خلال المعادلتين (7-6) و (7 – 7) السابقتين .

والمثال التالي يوضح كيفية تطبيق الخطوات السابقة :-

<u>مثال2 :-</u>

فيما يلي الجدول التالي يبين توزيع 50اسرة حسب فئـات الـدخل (x) والانفـاق (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

<u>والمطلوب</u> : اوجد معادلـة خط انحدار الانفـاق (Y) علـي الـدخل(X) . اي معادلة خط انحدار Y علي X .

<u>الحل :</u>

فيما يلي يتم حساب مراكز فئات الدخل المختلفـة (X) وقيم (Y) المقابلـة لـها .

فمركز الفئة الاولي لــ X =65 فأن قيمة Y المقابلة هي :

$$Y=\frac{2(45)+1(55)}{3}=48.3$$

ومركز الفئـة الثانية لـ X = 75 فان قيمة (Y) المقابلة هي :

$$y=\frac{2(45)+3\,(55)+2(65)+1(75)}{8}=57.5$$

- **ومركز الفئة الثالثة لـ X = 85 فأن قيمة (Y) المقابلة لها هي :**

$$y=\frac{2(55)+5\,(65)+6(75)+\ 1(85)}{8}=69.3$$

- **ومركز الفئة الرابعة لــ X = 95 فأن قيمة (Y) المقابلة لها هي :**

$$y=\frac{1(55)+3\,(65)+4(75)+\ 3(85)+1(95)}{8}=75$$

- **ومركز الفئة الخامسة لـ X = 105 فأن قيمة (Y) المقابلة لها هي :**

$$y=\frac{1(75)+5\,(85)+4(95)}{8}=86.3$$

- **ومركز الفئة السادسة لــ X =115 فأن قيمة (Y) المقابلة لها هي :**

$$y=\frac{2(95)+\ 2\,(105)}{4}=10$$

وأخيرا فإن مركز الفئة السابعة لــ X = 125 فأن قيمة(Y)المقابلة لها هي :

$$y = \frac{1(105)}{1} = 105$$

وبعد حساب كل من مراكز الفئات المختلفة لــــ (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 ميل خط الإنحدار هو :

$$\hat{b} = \frac{\frac{\sum xy}{n} - \frac{\sum x}{n} \cdot \frac{\sum y}{n}}{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

اي ان :

$$\hat{b} = \frac{\frac{54154}{7} - \left(\frac{665}{7}\right)\left(\frac{541.4}{7}\right)}{\frac{65975}{7} - \left(\frac{665}{7}\right)^2}$$

$$\hat{b} = \frac{388.714}{400} = 0.9718$$

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 :

$$\hat{y} = -14.978 + 0.9718x$$

*** العلاقة بين معامل انحدار(Y) علي (X) وبين معامل الارتباط الخطي البسط (r_{XY}) :**

من خلال معادلة ميل خط انحدار Y عليX الواردة بالمعادلة (7- 6) ومعادلة معامل الارتباط الخطي البسيط لبيرسون سنورد العلاقة ما بين معامل خط الانحدارY علي X ومعامل الارتباط الخطي البسيط دون اثبات تسهيلا علي القارئ : حيث أن :-

$$\hat{b} = r_{xy}\left(\frac{\sigma y}{\sigma x}\right) \qquad (11 - 7)$$

هذا ومن خلال المعادلة او العلاقة (7 – 11) 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 (7 – 10) لمعادلة انحدار Yعلي X بدلالة معامل الارتباط الخطي البسيط بين المتغيرين وذلك علي النحو التالي :

فحيث ان

$$\hat{y} = \bar{y} + b(x - \bar{x})$$

وعليه فأن :

$$= \bar{y} + r_{xy}\frac{\sigma y}{\sigma x}(x - \bar{x})$$

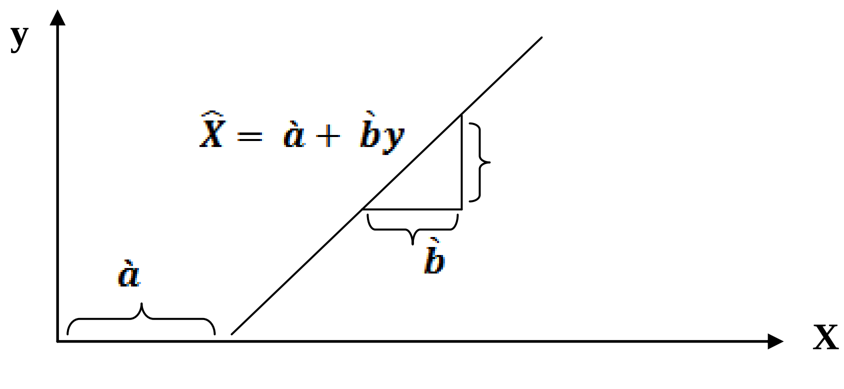
ثانيا : معادلة خط انحدار (X) علي (Y) :-

تاخذ معادلة خط انحدار X علي Y الصورة الخطية التالية :

$$\hat{x} = \grave{a} + \grave{b}y \qquad (12-7)$$

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 X علي Y) وذلك من خلال استبدال Yبــ X والعكس صحيح . ووفقا لذلك فأن (X) في معادلة خط انحدار X علي Y تمثل المتغيرالتابع بينما تمثل (Y) المتغير المستقل . وهنا سوف نتعرض بدون أثبات الي العلاقات المتعلقه بمعادلة خط انحدار X علي Y وذلك علي نحو مماثل لما سبق عرضه في حالة خط انحدار Y عليX . فخطأ التقدير مثلا في قيم (x) يمكن التعبير عنه بالعلاقه

$$\grave{e_i} = x_i - \hat{x}_i \qquad for \quad i = 1, 2, \ldots\ldots, n$$

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 بخط أنحدار X علي Y وذلك علي النحو التالي :

$$\left.\begin{aligned} \sum x &= na + \grave{b}\sum y \\ \sum xy &= \grave{a}\sum y^2 + \grave{b}\sum y \end{aligned}\right\} \quad (13-\ \ 7)$$

وهاتين المعادلتين تحققان شرط ان مجموع مربعات اخطاء الفروق أو مربعات الأخطاء $\sum \grave{e_i}^{\,2}$ يكون اصغر ما يمكن . ولعله من المفيد أن نسهل علي القارئ كيفية تذكر المعادلتين الطبيعيتين الواردتين في العلاقه (7 -13) وذلك علي النحو التالي :

حيث يمكن الحصول علي المعادلة الاولي بأخذ مجموع الطرفين للمعادلة (12-7) وذلك بأعتبار ان (X) بدلالة من $(\hat{x})$. وأما المعادلة الثانية فيمكن الحصول عليها من خلال ضرب طرفي المعادلة (12-7) فى المتغير المستقل (Y) فى هذه الحالة ثم إعتبا رالمجموع بعد الضرب (مع أعتبار أن X بدلا 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 :

$$(x - \bar{x}) = \hat{b}\,(y - \bar{y}) \qquad \textbf{(15 -7)}$$

أو

$$x = \bar{x} + \hat{b}\,(y - \bar{y})$$

(16 – 7)

كما أن العلاقة ما بين خط انحدار X علي Y ومعامل الارتباط الخطي البسيط لبيرسون r_{xy} تاخذ الصورة التالية :

$$\hat{b} = r_{xy}\left(\frac{\sigma\bar{x}}{\sigma y}\right) \qquad \textbf{(17-7)}$$

ومن ثم يمكن كتابة المعادلة (16-7) علي الصورة :

(18-7)

$$X = \bar{x} + r_{xy}\left(\frac{\sigma\bar{x}}{\sigma y}\right)(y - \bar{y})$$

كما أن مجموع القيم الفعلية للمتغير التابع X تساوي مجموع القيم المتوقعه $\hat{x}$

اي ان : $\sum X = \sum \hat{x}$

مثال (3) :

البيانات التالية تمثل تكاليف الاعلان (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 :-

(1):إيجاد معادلة خط انحدار تكاليف الاعلان (X) علي قيمة المبيعات (Y) .

(2):إذا بلغت قيمة المبيعات في أحد السنوات 160 الف جنيه فما هي القيمة المتوقعه للمنصرف علي الاعلان في هذا العام .

(3):حساب اخطاء التقدير لقيم (X) المختلفة مع التعليق علي ما تحصل عليه من نتائج .

الحل :

(1):بافتراض أن معادلة خط انحدار تكاليف الاعلان (X) علي قيمة المبيعات (Y) تأخذ الصورة الخطية التالية .

$$\hat{x} = \hat{a} + \hat{b}\, y$$

ولتقدير معلمتي خط الانحدار $\hat{a}$, b يتم إيجاد المجاميع الموضحه في الجدول التالي : فحيث ان:

X	y	X y	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

$$\hat{b} = \frac{\frac{\sum xy}{n} - \frac{\sum x}{n} \cdot \frac{\sum y}{n}}{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

$$= \frac{\frac{25680}{10} - \left(\frac{180}{10}\right)\left(\frac{1370}{10}\right)}{\frac{192100}{10} - \left(\frac{1370}{10}\right)^2}$$

$$\hat{a} = \bar{X} - \hat{b}\,\bar{y}\ \left(\frac{102}{441}\right) = 0.23\ \left(\frac{1370}{10}\right)\ = \left(\frac{180}{10}\right) - 0.23\ \left(\frac{1370}{10}\right)$$

$= - 13.51$

ومن ثم فأن خط انحدار x علي y تأخذ الصورة التالية

$$\hat{x} = -13.51 + 0.23y$$

(2): بعد إيجاد معادلة خط انحدار X علي Y فأن يمكن حساب القيمة المتوقعه لـ X عندما تساوي Y القيمة 160 وذلك علي النحو التالي :

$$\hat{x} = -13.51 + 0.23(160) = 23.29$$

أي أنه يمكن القول أنه إذا بلغت قيمة المبيعات في أحد الاعوام 160 ألف جنيه فأنه من المتوقع أن يكون المنصرف علي الاعلان في هذا العام هو 23.29 الف جنيه .

(3):الجدول التالي يوضح كيفية حساب القيم المتوقعه لـ X وأخطاء التقدير المناظرة لها :

x_i	y_i	$\hat{x}_i = -13.51 + 0.23\, y_i$	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:

$$\sum x_i = \sum \hat{x}_i = 180 \ \sum e_i = 0$$

- ***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 :

(1):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 .

(2):يتم حساب متوسطات قيم المتغير الثاني (التابع) المقابلة لمراكز فئات المتغير (Y) أو المقابلة لاقسام هذا المتغير . وهذه المتوسطات تمثل قيم المتغير (X) .

(3):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 (2) 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 :

1-مركز الفئة الاولي لـ Y = 45 يقابله قيمة X المقابلة لها هي:

$$x = \frac{2(65)+2(75)}{4}$$

2-مركز الفئة الثاني لــ Y = 55 يقابله قيمة X المقابلة لها هي :

$$X = \frac{1(65)+3(75)+(85)+1(95)}{7} = 79.3$$

3-مركز الفئه الثالثة لـ Y = 65 يقابله قيمة X المقابلة لها هي :

$$x = \frac{2(75)+5(75)+(85)+3(95)}{10} = 86$$

4-مركز الفئة الرابعة لـ Y = 75 يقابله : قيمة X المقابلة لها هي :

$$x = \frac{1\,(75) + 6(85) + 4\,(95) + 1(105)}{12} = 89.2$$

5-مركز الفئة الخامسه لـ Y = 85 يقابله : قيمة X المقابلة لها هي :

$$x = \frac{1\,(85) + 3(95) + 5(105)}{9} = 99.4$$

6-مركز الفئة السادسة لـ Y = 95 يقابله : قيمة X المقابلة لها هي :

$$x = \frac{1\,(95) + 2(105) + 2(115)}{5} = 107$$

7-مركز الفئة السابعة 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 والقيم المقابلة لها فأنه يتم إيجاد المجاميع اللازمة لحساب معلمتي خط الانحدار X علي 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 :

$$\hat{b} = \frac{\frac{\sum xy}{n} - \frac{\sum x}{n}.\frac{\sum y}{n}}{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2} = \frac{\frac{50827}{7} - \left(\frac{649.2}{7}\right)\left(\frac{525}{7}\right)}{\frac{42175}{7} - \left(\frac{525}{7}\right)^2}$$

$$= \frac{305.29}{400} = 0.76$$

وكذلك فأن :

$$\hat{a} = \bar{x} - \hat{b}\hat{y} \qquad = \frac{649.2}{7} - (0.76)\left(\frac{525}{7}\right) = 92.24 - 54$$

$$= 35.74$$

وعليه تأخذ صورة معادلة خط إنحدار x علي y الصورة التالية :

$$\hat{x} = 35.74 + 0.764\ y$$

أما عن العلاقة ما بين ميلي خطي الانحدار اي $\hat{b}$, $\hat{\mathrm{b}}$ ومعامل الارتباط الخطي البسيط لبيرسون r_{xy} فحيث أن :

$$b = r_{xy}\ (\sigma_x / \sigma_y)$$

$$\hat{b} = r_{xy}\frac{\sigma\bar{x}}{\sigma y}$$

$$\hat{b}\hat{b}\ r_{xy}^2$$ لذا فإن

وعليه فإن : (19 -7) $r_{xy} = \sqrt{\hat{b}\hat{\hat{b}}}$

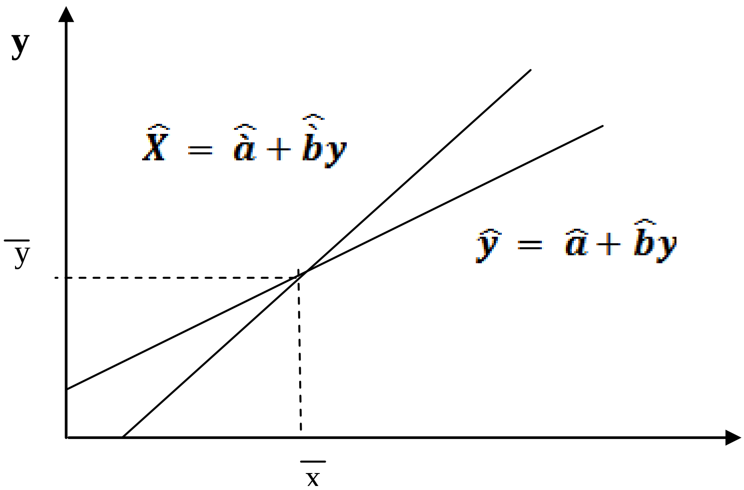
ويأخذ معامل الارتباط الخطي البسيط لبيرسون إشارة أيا من معاملي الانحدار. هذا ويمكن ملاحظة ذلك من خلال العلاقة (7- 19) . ومن خلال تلك العلاقة فإنه يمكن القول أن :

$$\left.\begin{aligned} \hat{b} &= \frac{r_{xy}^2}{\hat{\hat{b}}} \\ \hat{\hat{b}} &= \frac{r_{xy}^2}{\hat{b}} \end{aligned}\right\} \quad (20 - 7)$$

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 $r_{xy} = \pm 1$. 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 $\hat{y}$ حيث 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 الخطين من بعضها علي القيمة المطلقة لمعامل الإرتباط الخطى البسيط لبيرسون r_{xy} فكلما اقتربت هذه القمة من الواحد الصحيح كلما زادت درجة اقتراب العلاقة 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 الفصل السابع

(1)- اذا كانت :

$\bar{X} = 52$, $\bar{Y} = 7$, $r_{xy} = 0.4$, $\sigma_x = 2$, $\sigma_y = 3$

اوجد معادلة انحدار Y علي X .

(2)- في تحليل 60 مفرده ، وجد أن معادلتي خطي الإنحدار هي كالاتي :

$$1000Y = 768X - 3608$$

$$500X = 64 + 24Y$$

والمطلوب تحديد قيمة معامل الارتباط الخطى البسيط لبيرسون مابين X ، Y .

(3)- الجدول التالي يبين الارقام القياسية للانتاج (X) والدخل القومي (Y) بالملايين في الفترة من 1941حتي 1947 في احدي الدول .

الدخل القومي(Y)	الرقم القياسي (X)	السنةYear
500	100	1941
520	105	1942
580	107	1943
600	98	1944
610	109	1945
600	110	1946
650	114	1947

المطلوب :

1- ايجاد معامل الارتباط بين الظاهرتين .

2- ايجاد معادلتي خطى الانحدار (إنحدارY علي X ، انحدار X علي Y).

1- الحصول علي القيمة المتنبأ بها للدخل القومي عندما يكون الانتاج 120 .

2- الحصول علي القيمة المتنبأ بها للانتاج عندما يكون الدخل القومي 700 مليون جنية .

(4)- البيانات التالية تمثل الكميات المعروضة (X)عند الاسعار المناظرة(Y): والمطلوب ايجاد معادلتي خطي الإنحدار .

سعر السلعة	5	28	10	9	20	15	12	26	18	12
الكميـــــــة المعروضة	90	409	120	155	37	153	130	452	250	199

(5)- اذا فرضنا أن لديك البيانات التالية : المتغير (X) يعبر عن أعمار الرجال المتزوجين ، (Y) تمثل عدد مالديهم من أطفال . وكانت لديك البيانات التالية:

$\sigma y = 1.3535$, $\sigma x = 5.5605$, $\overline{Y} = 2.09$ ، $0.484 = r_{xy}$,

$\overline{X} = 33.175$

<u>والمطلوب</u> :

- **حساب معادلة خط انحدار Y علي X .**
- **حساب معادلة خط إنحدار X علي Y .**
- **تقدير ماهو عدد الاطفال لرجل عمره 30 سنة .**
- **تقدير ماهو العمر لرجل لديه طفلان .**

(6)- الجدول التالي يبين دخل مجموعة من 8 أسر في اليوم (X) وما تنفقه فى خلال نفس الفترة (Y) :

X	64	52	84	64	76	56	68	54
Y	52	40	60	52	60	42	50	25

<u>والمطلوب</u> :

1: تقديرمعادلة خط انحدار الدخل علي الانفاق .
2:تقدير معادلة خط انحدار الانفاق علي الدخل .
3:قدر إنفاق الاسر التي يبلغ دخلها 70 جنيه في الاسبوع .
4:قدر دخل الأسرة التي تنفق 48 جنبه في اسبوع .
5:ماهو معامل الارتباط بالطريقة العادية والتحقق من النتائج عن طريق معاملي خطي الإنحدار .
6:إيجاد معادلتي خطى الإنحدار بمعلومية معامل الارتباط وقارن النتائج التي حصلت عليها في (1) ، (2) .

7:عرف كل من الارتباط والانحدار موضحا الفرق بينهما .

(7): فيما يلى لديك الجدول التالى يوضح بيانات الانتاجية في العمل(Y) والخبرة بالسنوات (X) لعينة من 7 عمال :

الخبرة بالسنوات X	الانتاجية Y
3.9	70
3.1	49
4.9	97
3.2	71
1.8	45
2.9	61
2.1	39

من البيانات السابقة احسب :

1. **الوسط الحسابي لكل من المتغيرين .**
2. **معامل الارتباط بين المتغيرين .**
3. **ماذا تستخلص من قيمة معامل الارتباط . وهل تحتاج لحساب معادلة خط الانحدار ؟**

(8) -الجدول التالي يمثل توزيع عينة من 500 شخص تبعا للطول (Y) والوزن (X) :

	اقل من 160	-160	-170	-180	190 فأكثر	مجموع
اقل من 6	15	20	15	-	-	50
-60	10	50	40	-	-	100
-65	-	50	40	10	-	100
-70	-	30	40	30	-	100
-75	-	-	40	40	20	100
-80 فاكثر	-	-	10	15	25	50
	25	150	185	95	45	500

<u>والمطلوب</u> :

أ- إحسب معامل الارتباط الخطى البسيط بين الوزن والطول لهذه البيانات.

ب- استنتج معادلة الانحدار الخطي للطول علي الوزن بدلالة معامل الارتباط .

ج- إستخدم المعادلة التي حصلت عليها في تقدير طول أحد الاشخاص وزنه 85 كيلو جرام .

د- إستنتج معادلة إنحدار الوزن علي الطول بدلالة معامل الارتباط للبيانات المبوبة .

(10) -البيانات التالية توضح العلاقة بين اطوال وأوزان عشرة أشخاص :

66 ، 64 ، 64 ، 76 ، 69 ، 68 ، 59 ، 74 ، 62 ، 72 : Weight (X)

143 ، 134 ،159 ،179، 168 ،156 ،134، 181 ، 149 ،157 : *Length* (Y)

<u>والمطلوب</u> :

أ – إيجاد معادلة خط انحدار الطول علي الوزن .

ب- قدر القيمة المتوقعة لــ Y عند X = 80.69 .

(11): البيانات التالية توضح الدخل الاسبوعي (X) والاتفاق الاسبوعي علي الطعام (Y) وحجم المدخرات الاسبوعية (Z) وذلك لعينة مكونه من عشرة اسر :

78 ، 65 ، 90 ، 83 ، 67 ، 44 ، 53 ، 71 ، 82 ، 98 : X

52 ، 45 ، 70 ، 49 ، 32 ، 28 ، 25 ، 38 ، 42 ، 53 : Y

8 ، 10 ، 5 ، 13 ، 8 ، 2 ، 12 ، 3 ، 7 ، 10 : Z

<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 (أ) ، (ب) اوجد :

*** معادلة خط انحدار حجم المدخرات الاسبوعي للاسرة علي الدخل الاسبوعي لها(أي معادلة خط انحدار Z علي X) .**

*** معادلة خط انحدار الانفاق الاسبوعي للاسرة على الطعام علي الدخل الاسبوعي (أي معادلة خط انحدار Y علي X).**

*** بين ايهما أكثر تأثرا بالدخل الاسبوعي ، الانفاق الاسبوعي علي الطعام أم حجم المدخرات الاسبوعيه .**

(د):من نتائجك في (ج) إحسب ما يلي :

***أخطاء التقدير في حساب كل من Y , Z وذلك عند القيم التالية لــ X : 82 ، 44 ، 20**

***إحسب القيم المتوقعه لكل من Y , Z وذلك عند اليم التالية لــ X : 75، 120 ، 40**

(ه):ما هو الاثر الناتج علي حجم المدخرات الاسبوعية إذا نقص دخل الاسرة بمقدار عشرة جنيهات اسبوعيا ؟

(12) البيانات التالية توضح تكاليف الاعلان وقيم المبيعات (بآلاف الجنيهات) وذلك لإحدي الشركات في ثماني شهور متتالية :

الاعلان (X) :	28	51	20	46	53	35	20	25
المبيعات(Y) :	104	130	136	113	150	175	127	108

<u>والمطلوب</u> :

(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 .

(13):استخدام البيانات التالية عن الظاهرتين : X ، Y وذلك لايجاد معادلتيي خط انحدارY علي X وانحدار X علي Y :

$\Sigma x = 50$, $\Sigma y = 20$, $r_{xy} = .1037$, $\sigma x = 1.392$, $\sigma y = .866$, $n = 8$

(14) استخدام البيانات التالية عن الظاهرتين X، Y لاجراء مقارنة بين التشتت النسبي في كل منها :

b (معامل انحدار Y علي X) = 1.8

$$r_{xy} = 0.72 \qquad \frac{\Sigma y}{\Sigma x} = 2.5$$

(15) :اثبت أن معامل الارتباط الخطي بين X ، Y أى(r_{xy}) يمكن كتابته علي الصورة التالية

$$r_{xy} = b \; \frac{\sigma y}{\sigma x}$$

حيث b هي معامل انحدار Y علي X ، σx ، σy هما الانحرافات المعيارية للمتغيرين X و Y علي الترتيب .

(16): بافتراض أن العلاقة بين المتغيرين بين X ، 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 :

y / X	95	105	115	125
52.5	4	7	5	2
57.5	6	10	7	4
62.5	6	12	10	7
67.5	3	8	6	3

(17) البيانات التالية تمثل درجات عشرة طلاب في كل مادتي الرياضيات (X) والمحاسبة (Y) :

10	9	8	7	6	5	4	3	2	1	Student
12	14	16	13	20	15	16	10	11	18	X
15	12	12	17	18	13	10	14	8	13	Y

والمطلوب :

أ- حساب معامل الارتباط الخطي فيمابين المتغيرين X ، Y .

ب- إيجاد معادلة انحدار Y علي X ثم حساب خطأ تقدير Y عند القيمة = 15 X .

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) ؟

(18) البيانات التالية توضح الدخل اليومي (X) والانفاق اليومي (Y) وذلك لعينة مكونه من سبع أسر :

19	21	15	11	9	13	17	Incomes(X)
9	10	7	5	4	6	8	Expedition (Y)

والمطلوب

أ – أوجد معادلة خط انحدار Y علي X .

ب- إحسب 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 اي من المتغيرين X ، Y هو الاكبر تشتتا .

هـ- من نتائجك في أ ، ج أوجد معادلة خط انحدار X علي Y .

(19): (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

$$Y = -5.9 + 1.35\,X$$
$$X = 4.51 + 0.73\,Y$$

والمطلوب :
أ – احسب معامل الارتباط الخطي بين X ، Y .
ب – مستعينا بنتائجك في (أ) ، حدد أيهما الأكبر من حيث التشتت النسبي درجات الاحصاء(X) أم درجات المحاسبة (Y) ؟
ج – إذا حصل طالب معين من بين محموعة الطلاب علي الدرجة 15 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) أحسب خطأ تقدير درجته في كل من مادتي الاحصاء والمحاسبة ، ثم علق علي ضوء نتائجك في (أ) .

(20): البيانات التالية تعطي التوزيع التكراري المزدوج لخمسين من الازواج موزعين حسب عمر الزوج (X) وعمر الزوجه (Y) :

Y \ X	25-	35-	45-	55-	65-75	$\sum$
25-	5	4	3			12
35-	3	6	5			14
45-			6	8	2	16
55-65				2	6	8
$\sum$	8	10	14	10	8	50

المطلوب :
أ- حساب معامل الارتباط الخطى فيما 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 40سنة.
جـ- ما هو الاثر الناتج علي عمر الزوجة إذا نقص عمر الزوج بمقدار خمس سنوات
د- إستخدم أيا من معاملي بيرسون الاول أو الثاني للالتواء وذلك لتحديد شكل كل توزيع من حيث نوع ودرجة الالتواء .

" المراجع "

أولا:مراجع باللغة العربية :

1-د.إبراهيم موسى عبدالفتاح "مبادئ الإحصاء الوصفى والتحليلى" المكتبة العلمية بالزقازيق ، 2005 .

2-د.أبوبكرعبدالرحمن عبدالمتعال "مبادئ الإحصاء" ،كلية التجارة ،جامعة جنوب الوادى 2008 .

3-د.أمانى موسى محمد ، د. محمد محمود خطاب "الإحصاء فى: إتخاذ القرارات- مدخل حديث "،كلية التجارة بسوهاج، جامعة سوهاج ، 2008 .

4-د.صلاح مهدى محمد " الإحصاء " ،كلية التجارة ، جامعة عين شمس ، 2008 .

5-د.محمد جلال الدين أبو الدهب وآخرون " مقدمة فى الإحصاء للتجاريين- بين النظرية والتطبيق:، مكتبة عين شمس ،1991-1992 .

6-د.مصطفى جلال مصطفى وآخرون "أساسيات الطرق الإحصائية" ،كلية التجارة ،جامعة عين شمس ، 2015 .

7-د.نادية مكارى "محاضرات فى مبادئ الإحصاء"، كلية الإقتصاد والعلوم السياسية، جامعة القاهرة ، 1977 .

ثانيا: **المراجع الأجنبية**:

1-Boerman Bruce L. & O`Connell Richard T., " Applied Statistics Improving Business Processes", Richard D. Irwin, Inc. Company, U.S.A., 1997.

2-Frank Harry & Althoen Steven C.," statistics Concepts and Applications "Cambridge University Press. Great Britain ,1995 .

3-Levine David M., Berenson Mark L. & Steven David, " Statistics for Managers using Microsoft Excel", Prentice – Hall Inc., U.S.A., 1977.