



"الإحصاء الوصفي"

لطلاب الفرقة الثانية

إعداد وتأليف

د. محمد عبدالحميد الربيع

كلية التجارة

قسم الأساليب الكمية

العام الجامعي

2024 - 2023

مقدمة

مما لا شك فيه أن علم الإحصاء أصبح اليوم ركيزة أساسية في جميع فروع المعرفة . فلقد بات هذا العلم لازماً في مختلف مجالات العلوم الاجتماعية والاقتصادية والهندسية والطبيعية .

هذا ولقد تطور علم الإحصاء تطوراً كبيراً في الآونة الأخيرة حيث تغيرت صورته القديمة والعالقة في أذهان الناس على أنه علم العد وجمع البيانات وعرضها بيانياً إلى كونه الآن علم رسم السياسات واتخاذ القرارات في شتى مجالات المعرفة على أساس علمي سليم رغم عدم توافر المعلومات الكافية فأتسع لذلك نطاق استخدامه وازداد الاهتمام بأساسياته وأساليب التحليلية . وأصبح لعلم الإحصاء في الوقت الحاضر دوراً هاماً 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 تعقيداً 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 .

• الإحصاء الرياضي أو الاستدلال الإحصائي :

Mathematical Statistics or Statistical Inference

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

هذا ولقد جاء هذا الكتاب باللغة العربية في كل محتوياته فيما عدا القوانين والرموز المستخدمة فقد جاءت باللغة الإنجليزية (وبعضها حروفاً أو رموزاً 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 مبسطاً 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 ودارسي علم الإحصاء في كافة فروع المعرفة مادة علمية تساعدهم في دراسة وتحليل الظواهر محل الاهتمام .

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

دكتور

محمد عبد الحميد الربع

الهدف من تدريس مقرر الإحصاء الوصفي :

يهدف هذا المقرر الى:

- (1): اكساب الطالب بمعارف ومفاهيم علم الاحصاء وفروعه وكيفية اتمام عملية جمع البيانات وعرضها بيانيا وجدوليا.
- (2): تنمية قدره الطالب على كيفية تلخيص الظواهر الكمية والوصفيه من خلال حساب مؤشرات رقميه تلخص طبيعه تلك الظواهر.
- (3): تنمية ورفع قدره الطالب على اتمام البحوث العلميه ومعالجه المشاكل باسلوب علمى احصائى يتمتع بدرجة عاليه من الدقة والثقة .
- (4): تنمية قدره الطالب الاحصائيه على معرفة ودراسه قوة العلاقة فيما بين المتغيرات او الظواهر بانواعها المختلفه وكيفية تحديد شكل العلاقة الخطية فيما بين متغيرين كميين.
- (5): تزويد الطالب بمعرفه عناصر المشكله السكانيه وطبيعتها وكيفية حساب معدلات النمو السكانى.
- (6): تنمية قدره الطالب على كيفية دراسه التغيرات فى اسعار وكميات وقيم السلع الاقتصاديه من خلال دراسة الارقام القياسية .

محتويات المقرر

رقم الصفحة	بيان
1	<u>أولاً: الجزء الأول من المقرر في الإحصاء الوصفي.</u>
1	- صفحة العنوان "الإحصاء الوصفي"
2	مقدمة الكتاب
4	الهدف من تدريس مقرر الإحصاء الوصفي .
5	محتويات المقرر .
8	الفصل الأول: " المفاهيم الأساسية في الإحصاء"
8	- مفهوم علم الإحصاء وإستخداماته المختلفة.
10	- الهدف من دراسة علم الإحصاء.
11	-المجتمع الإحصائي والعينة
13	-خطوات جمع البيانات الإحصائية .
16	-- وسائل جمع البيانات .
18	-- أساليب جمع البيانات
22	-الفرق بين أسلوب الحصر الشامل والعينة .
22	-خطوات (منهج) البحث الإحصائي .
27	-التحليل الإحصائي وإستخدام الحاسب الآلى .
27	-خطوات التحليل الإحصائي بإستخدام الحاسب الآلى .
32	-- البرمجيات الإحصائية الجاهزة .
34	-- أنواع العينات .
40	الفصل الثانى: " أساليب عرض البيانات الإحصائية "
43	* العرض الجدولى للبيانات الإحصائية .
44	- مراحل العرض الجدولى للبيانات الوصفية .
47	- مراحل العرض الجدولى لبيانات الظواهر الكمية .
57	- الجداول أو التوزيعات التكرارية المفتوحة .
59	- التوزيعات التكرارية النسبية .
63	- التوزيعات أو الجداول التكرارية المزدوجة .
74	- الجداول (التوزيعات) التكرارية المتجمعة .
96	* العرض البيانى للبيانات الإحصائية .
97	- العرض البيانى للبيانات الوصفية .
109	- العرض البيانى للظواهر الكمية المبوبة تكراريا .
125	- أشكال المنحنيات التكرارية .

131	- المنحنيات التكرارية المتجمعة .
139	- مخطط الساق والورقة .
145	- الأشكال البيانية الخاصة بالسلاسل الزمنية .
150	- تمارين الفصل الثاني .
163	الفصل الثالث: "مقاييس النزعة المركزية (المتوسطات) .
165	-الوسط الحسابي .
197	- الوسيط .
208	-شبيهات الوسيط .
224	-المنوال .
246	-العلاقة بين الوسط الحسابي والوسيط والمنوال .
252	-الوسط الهندسى .
256	-الوسط التوافقي .
267	-تمارين على الفصل الثالث .
279	<u>ثانيا : الجزء الثانى من المقرر فى الإحصاء الوصفى .</u>
279	الفصل الرابع: "مقاييس التشتت " .
283	* مقاييس التشتت المطلقة
283	-المدى .
287	-شبيهات المدى .
289	-نصف المدى الربيعى .
295	-الانحراف المتوسط .
305	-التباين والانحراف المعياري .
331	-الانحراف المعياري المصحح لشبرد .
341	*مقاييس التشتت النسبية (معاملات الاختلاف) " .
343	-معامل الاختلاف بدلالة الوسط الحسابى والانحراف المعياري .
343	- معامل الاختلاف الربيعى .
354	- الدرجات المعيارية .
361	- تمارين الفصل الرابع .
368	الفصل الخامس: " العزوم والالتواء والتفرطح " .
368	* العزوم .
371	-العزوم حول الصفر (العزوم اللامركزية).
372	-العزوم حول الوسط (العزوم المركزية).

374	- العزوم حول وسط فرضى معين .
376	- العزوم المعيارية .
384	- العلاقة بين الأنواع المختلفة من العزوم .
385	- العزوم المصححة لشبرد .
386	* الإلتواء .
388	-معاملات الإلتواء بدلالة مقاييس المتوسطات والتشتت .
391	-قياس الإلتواء باستخدام العزوم المركزية .
392	* التفرطح .
392	- معامل التفرطح بدلالة العزوم .
393	- معامل التفرطح بدلالة الربيعين والمئينيات .
401	- تمارين الفصل الخامس .
406	الفصل السادس : " الإرتباط الخطى البسيط
413	-العلاقة بين الظواهر (المتغيرات) .
414	* دراسة علاقة الإرتباط
415	- معامل الإرتباط الخطى البسيط لبيرسون .
438	- معامل فهنر للإرتباط .
442	- معامل إرتباط الرتب لسبيرمان .
455	- معامل كندال للإرتباط .
457	* الإرتباط بين الصفات .
458	- معامل الإقتران .
463	- معامل التوافق .
468	- تمارين على الفصل السادس .
480	<u>الفصل السابع</u> : " الإندار الخطى البسيط "
484	- معادلة خط إندار Y على X .
499	- معادلة خط إندار x على y .
510	تمارين الفصل السابع .
510	- المراجع .

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 الحياة حيث يعتبر هذا العلم بمثابة مدخلا كيميا للعلوم فى إتخاذ القرارات . ويمكن النظر إليه باعتباره أحد فروع الرياضة التطبيقية . كما يمكن تعريف 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 بناءً 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 و تطورها فأصبحت بمثابة الأساس العلمي للأساليب الإحصائية مما أدى إلى ظهور الإحصاء الرياضي والتحليلي أو الاستدلالي وتطبيقاته في جميع مجالات أنواع المعرفة في الحياة وهو ما ساهم في دراسة خصائص المجتمعات والتي لا بد

فيها من اتخاذ قرار محدد بالرغم من عدم توافر بيانات كاملة عن هذه المجتمعات .

هذا وبعد التطور المتسارع في علم الإحصاء وأساليبه المختلفة فقد أصبح لهذا العلم أهمية كبرى في كثير من المجالات الحيوية المختلفة وذلك لأن الأسلوب الإحصائي أصبح قادراً على بحث وتقدير المخاطر التي تنجم عن اتخاذ قرار معين. ففي الدراسات الطبية والدوائية مثلا يستخدم الاسلوب الإحصائي لدراسة وتحديد أسباب انتشار الأمراض المختلفة ومدى فاعلية الأدوية المختلفة في علاجها ، هذا بالإضافة إلى دراسة العوامل الوراثية وتأثيرها ...

وفي الصناعة مثلا يتم استخدام التحليل الإحصائي في عملية مستوى جودة المنتج والرقابة على هذا المستوى - مراقبة جودة المنتج Statistical Quality Control وكذلك في دراسة التنبؤ باتجاهات التغيير في الطلب 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 السماد أو المبيدات الحشرية الواجب استخدامها أو أثر اختلاف نوع التربة وطريقة الري على كمية المحصول من أحد المنتجات ...

وكذلك نجد أن التحليل الإحصائي هو الأساس الكبير في الدراسات الطبيعية والكيميائية وأبحاث الفضاء والذرة .. إلى غير ذلك من كافة فروع أنواع المعرفة الحديثة . هذا ويضاف إلى تلك الزيادة الملحوظة لاستخدامات الإحصاء في العلوم الاجتماعية كالدراسات الاقتصادية والسياسية واستطلاعات الرأي العام وأبحاث الاجتماع وعلم النفس ، كما يجب أن لا نتجاهل الاستخدام

التقليدي للبيانات الإحصائية العامة في رسم السياسات الحكومية سواء المرتبطة بالتعليم أو الخدمات الصحية أو خدمات الأمن
والآن وبعد تناولنا لمفهوم علم الإحصاء ومجالات استخدامه في فروع العلوم المختلفة دعنا نعرض ما هو الهدف من دراسة علم الإحصاء ؟

*** الهدف من دراسة علم الإحصاء :**

الهدف الأول: " عملية جمع البيانات الإحصائية "

تعتبر عملية جمع البيانات الخاصة حول أى مشكلة موضع الدراسة 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 جدولياً أو بيانياً أو من خلال حساب بعض المقاييس الإحصائية المستخلصة من البيانات يجب تحليل هذه البيانات تحليلاً علمياً دقيقاً يؤدي إلى كيفية التنبؤ المستقبلي واتخاذ القرارات الملائمة بصدد تلك الظواهر محل الدراسة ودراسة الطرق واختبارات الفروض الإحصائية حول المقاييس المستخلصة من بيانات تلك الظواهر .

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

أولاً: المجتمع الإحصائي والعينة : Statistical Population & Sample

تختلف معنى كلمة المجتمع عند الإحصائي عن المعنى الشائع الاستخدام . فهو لا يشير إلى تجمع الأفراد في منطقة معينة وإنما المقصود هو ذلك الكل الذي نرغب في دراسته ومعرفة المعلومات التفصيلية عن خصائصه المميزة. وعلم الإحصاء يهتم بدراسة المجتمعات الإحصائية المختلفة .

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

فمثلاً إذا كان هدف الدراسة الإحصائية هو تحديد متوسط أجر ساعة العامل المصري . فإن المجتمع الإحصائي هو عبارة عن أجور جميع العاملين المصريين في الساعة . وإذا كان هدف الدراسة هو تحديد نسبة الوحدات المعيبة في هذا الخط الجديد من خطوط الإنتاج فالمجتمع الإحصائي هو عبارة عن كل الوحدات المنتجة في هذا الخط الجديد. وكذلك إذا كان هدف الدراسة الإحصائية هو تحديد نسبة البطالة في مصر فإن المجتمع الإحصائي هو جميع الأفراد من سكان ج . م . ع في سن العمل والقادرين على العمل وبيحثون عنه.

فالخلاصة أنه يمكن تعريف المجتمع الإحصائي بأنه عبارة عن مجموعة المفردات أو العناصر المحددة والمعرفة بحدود زمنية ومكانية معينة وتشترك في صفات وخصائص محددة يمكن ملاحظتها وقياسها كميًا . ومهما تعددت المجتمعات الإحصائية حسب تنوع الظواهر والمتغيرات في مجالات كافة أنواع

المعرفة فإن يمكن تقسيمها حسب تنوع الظواهر أو المتغيرات في مجالات كافة أنواع المعرفة إلى نوعين أساسيين : -

الأول : مجتمع إحصائي محدود : Limited Population

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

الثاني : مجتمع إحصائي غير محدود Unlimited 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 للعد أو لتحديد الحجم . مثال ذلك عدد كرات الدم الحمراء أو البيضاء في دم العنصر البشري . فهذا العدد غير قابل للحصر أو للعد لذا يصنف المجتمع في تلك الحالة على أنه مجتمع إحصائي غير محدود

هذا وتسمى الخصائص الدالة على أي مجتمع إحصائي والتي يمكن ملاحظتها أو قياسها كميًا باسم معلمات المجتمع Parameters . وغالباً ما تكون القيم الحقيقية لمعلمات المجتمع الإحصائي مجهولة (غير معلومة). إلا أنه يمكن تقديرها من خلال الاستدلال عنها وذلك من خلال سحب عينة عشوائية بطريقة إحصائية سليمة لكي تكون ممثلة لهذا المجتمع تمثيلاً تاماً (جيداً) . وتجري الدراسة الإحصائية على تلك العينة لإستخلاص مايسمى بإحصاءات العينة ثم بإستخدام طرق أو أساليب الإحصاء التحليلي (الاستدلالي) يتم

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

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

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

1-تحديد هدف ومشكلة مجتمع الدراسة : فإن تحديد هدف البحث هو الخطوة الأولى للقيام بأي بحث أو دراسة إحصائية. فبناءً على هذا الهدف يتحدد "المجتمع" محل البحث وكذلك تتحدد "المفردة" التي يجب جمع البيانات عنها ونوع ومصدر البيانات والأسلوب الذي يجب استخدامه في جمع تلك البيانات .
فمثلاً إذا كان الهدف من بحث ما هو المقارنة ما بين كيفية إنفاق دخل الأسرة في ريف ج.م.ع بكيفية إنفاق دخل الأسرة في حضر ج.م.ع خلال فترة زمنية معينة . فإن المجتمعات محل الدراسة تكون هي كلاً من ريف وحضر ج.م.ع خلال هذه الفترة الزمنية المحددة . كما أن المفردة التي تتكون منها هذه المجتمعات والتي يجب أن تُجمع عنها البيانات هي "الأسرة" والبيانات المطلوبة هي دخل الأسرة في تلك الفترة الزمنية المحددة وكيفية توزيعه على مصادر الإنفاق المختلفة .

فالمجتمع الإحصائي هو كما ذكرنا سابقاً عبارة عن مجموع المفردات التي يجب أن تُجمع عنها البيانات . كما أن المفردة محل البحث هي وحدة جمع البيانات والمفردة محل البحث قد تكون سلعة منتجة بطريقة معينة أو هي القرية ذات المساحة المحدودة أو هي الطفل في سن معين إلخ . فمن

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

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

أ- **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 تاماً كما يجب دائماً 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 أيضاً 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 مسحوية 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 على كافة مفردات المجتمع محل الدراسة والتي تسمى بالإطار والذي يجب أن يتوافر فيه الدقة والشمول وعدم التكرار . ثم تجري عملية سحب متتالي لمفردات العينة من مجتمع الدراسة سواء يدويًا أو آليًا أو باستخدام جداول الأرقام العشوائية .

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 صحيفة الاستقصاء (Questionnaire) ويجب أن تراعي بعض القواعد عند وضع هذه الأسئلة . فمثلاً يجب أن تكون مختصرة بقدر الإمكان وأن لا تتضمن أسئلة مكررة . كما يجب أن تتجنب الاستمارة الأسئلة ذات الإجابات المعتمدة على الحكم الشخصي للفرد (فمثلاً بدلاً من سؤال الفرد المبحوث عن مستوى تعليمه يمكن أن يطلب منه وضع علامة أمام أحد الحالات التعليمية المعروفة : لا يقرأ ولا يكتب ، يقرأ فقط ، حاصل على شهادة متوسطة ، إلخ...) .

كما يجب وضع الأسئلة في صورة يسهل الإجابة عليها بطريقة واضحة . هذا وبعد التأكد من صلاحية استمارة البحث وبعد أن تتم عملية تهيئة المجتمع لعملية جمع البيانات (وذلك بالإعلان عن أهمية البحث وضرورة الدقة في الإجابة ومدى سرية البيانات) وبعد أن يتم تدريب الأفراد اللذين سيقومون بجمع البيانات يبدأ الجمع الفعلي للبيانات من عملية ميدانية تسمى عادة

بالدراسة الاستكشافية Pilot Study والتي تجرى على عينة مماثلة إلى حد كبير للعينة أو المجتمع الأصلي . والهدف من تلك الدراسة المسبقة يتلخص في ثلاث أسباب رئيسية . فأولاً تساعد هذه المعاينة المصغرة على التعرف على الصعاب التي يمكن أن تواجه البحث الأصلي في الميدان مقدماً فيمكن بذلك عمل أية تعديلات سواء في الاستمارة أو في طريقة المقابلة أو غير ذلك من التعديلات خصوصاً ما يتعلق منها بنظام سير العمل في الميدان . وثانياً فإن هذه الدراسة الاستكشافية تساعد على التعرف على مدى فهم جامعي البيانات للمصطلحات الواردة بالاستمارة وتقدم بذلك فرصة طيبة لتدريبهم وزيادة خبرتهم في الميدان . وأخيراً 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 أيضاً على التعرف على مدى وجود بدائل جديدة للإجابات يمكن أن تضاف للاختبار المستجيب . وعموماً فالدراسة الاستكشافية ضرورية للاطمئنان على صحيفّة 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 أحياناً وسيلة معينة دون أخرى . هذا ويمكننا تقسيم وسائل جمع البيانات إلى ما يلي :

* وسائل مباشرة * وسائل غير مباشرة

أ- وسائل جمع البيانات المباشرة :

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 أو المجلات . ويكون على الفرد المبحوث أن يستوفي البيانات المطلوبة ويعيدها إلى الهيئة القائمة على البحث . وعادة ما تُرسل صحيفة البحث إلى الفرد المبحوث مرفقاًً بها خطاباًً عن أهمية البحث وضرورته . وكذلك يرفق به مظروف بعنوان الجهة أو الهيئة القائمة على البحث ملصقاًً عليه طابع البريد حتى لا يتكلف الفرد المستجيب مالاً أو وقتاً . كما يجب أن يضمن الباحث للمستجيبين السرية الكاملة للبيانات كما يحق للفرد المستجيب عدم ذكر اسمه عند الرد .

أما وسيلة الاتصال التليفوني فهي تعتمد في جمعها للبيانات على الاتصال التليفوني بالمفردات محل البحث لاستيفاء البيانات المطلوبة . حيث يحصل الباحث على أرقام تليفونات المستجيبين ويتصل بهم لملأ صحيفة البحث من

واقع إجاباتهم . وهذه الطريقة تتميز بالبساطة والسهولة والسرعة وقلة التكاليف . لكن أهم ما يعيبها أنه ربما تتضمن العينة محل البحث مجموعة من المفردات ليس في حوزتهم تليفونات .

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

ب- وسائل جمع البيانات الغير مباشرة:

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

(4-1): أساليب جمع البيانات :-

عند جمع البيانات الاحصائية يمكن استخدام إما اسلوب الحصر الشامل أو اسلوب العينة (المعاينة) . فاستخدام اسلوب الحصر الشامل لجمع البيانات معناه جمع البيانات من جميع مفردات المجتمع . أما اسلوب العينة فهو عبارة عن جمع البيانات من بعض مفردات المجتمع فقط . وفي بعض الأحيان تجد

أن طبيعة المجتمع محل الدراسة وطبيعة البيانات المطلوبة قد يفرضان على الباحث استخدام أحد الأسلوبين 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 لا بد من استخدام أسلوب العينة . وبلغت أخرى إذا كانت البيانات المطلوبة تؤدي إلى إهلاك المفردة محل البحث مثل حساب متوسط عدد ساعات إضاءة المصابيح الكهربائية التي ينتجها أحد المصانع أو متوسط عدد كرات الدم البيضاء أو الحمراء في المليمتر الواحد المكعب وكذلك البيانات الخاصة بالعمالة والأجور فإنه أيضاً لا بد من استخدام أسلوب 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 لا بد من استخدام أسلوب 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 الأسلوبين سواء الحصر الشامل أو العينة . وفي تلك الحالة فإن المقارنة بين الأسلوبين تعتمد أساساً على مدى الإمكانات المادية والفنية المتاحة للبحث ومدى الدقة التي يمكن أن يحققها أحد الأسلوبين في عملية جمع البيانات . فجميع البيانات باستخدام أسلوب الحصر الشامل يحتاج إلى الكثير من النفقات والوقت بينما نجد أن أسلوب 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 أسلوب العينة بطريقة سليمة يتطلب توافر إطار يشمل جميع مفردات المجتمع حتى يمكن استخدامه عند اختيار تلك

المفردات التي ستظهر في العينة . كما يجب مراعاة 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): خطأ التحيز :- وينشأ هذا النوع من الخطأ نتيجة تحيز الأفراد - لا شعورياً - لبعض القيم أو البيانات . كما أنه قد ينشأ نتيجة الإهمال في جمع البيانات أو عدم دقة الرقابة أو عدم وضوح الأسئلة الموجودة باستمرار 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 بأسلوب 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): خطأ الصدفة العشوائي (الخطأ العشوائي) :- ويظهر هذا الخطأ بدرجة أكبر عند استخدام أسلوب الحصر الشامل مقارنة بأسلوب العينة . فهو ينجم عن عملية عدم الدقة عند تسجيل أو جمع البيانات . وهذا النوع من الأخطاء ليس له اتجاه ثابت ويمكن التحكم فيه وتقليل آثاره وذلك إذا ما تم تحرى الدقة عند جمع البيانات بطريقة علمية أو إحصائية سليمة . وكذلك فإنه يمكن

استخدام الأساليب الرياضية لتقدير مدى تأثير هذا النوع من الأخطاء على نتائج الدراسة .

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

(1): نوع العينة : فيفضل العينات العشوائية الإحصائية البسيطة والتي يكون فيها لكل مفردة في مجتمع الدراسة لها نفس فرصة الاختيار والتمثيل ضمن مفردات العينة . وفيها يمكن التحكم وقياس أخطاء المعاينة العشوائية على خلاف العينة العمدية **Judgement Sample** والتي يختار الباحث مفرداتها بشكل شخصي (قصدي) من مجتمع مشكلة البحث .

(2): حجم العينة :- حيث توجد علاقة عكسية فيما بين حجم أخطاء المعاينة وبلغة أخرى درجة الدقة الإحصائية وبين حجم العينة وحجم الميزانية المرصودة للبحث .

(3): تباين المجتمع :- فكلما كان تباين المجتمع كبير فإن ذلك يؤدي إلى زيادة حجم تمثيل عينة جيدة لمجتمع مشكلة البحث ومن ثم أخطاء المعاينة العشوائية . وإن كان نوع وحجم العينة المسحوبة يؤديان إلى تحسين عملية الحصول على عينة إحصائية جيدة .

وفيما يلي الجدول التالي يبين الفرق بين أسلوب 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 دقة عالي - الحصر الشامل يحتاج إلى وقت وجهد وتكاليف أعلى	العينة تتميز بمستوى دقة أقل العينة تحتاج إلى وقت وتكاليف وجهد أقل .
		- الحصر الشامل يعطي نتائج مباشرة وكلية عن المجتمع محل الدراسة .	العينة لا تعطي نتائج مباشرة بل تقديرية عن المجتمع كله .
		- الحصر الشامل يتعرض لخطأ الصدفة فقط	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 الظاهرة أو سلوك أو حالة ما يراد علاجها . وفي تلك المرحلة من خطوات البحث لابد أن تتم عملية التحديد الواضح والدقيق للمشكلة موضع الدراسة كخطوة أولى في سبيل تحديد عناصر حل تلك المشكلة .

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 فروضاً موضوعية (بعيدة عن الميل الشخصي للباحث) ، وأن تصاغ بسهولة ووضوح ويتم ترتيبها حسب درجة أهميتها (أولويتها) .

الثالث : جمع البيانات :

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

الرابع : تحليل البيانات :

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

الخامس : تفسير النتائج :

تعتبر عملية اختبارات صحة الفروض الإحصائية الموضوعية على ضوء نتائج مخرجات عمليات التحليل الإحصائي المستخدمة ، وهي احتمالية (أي إمكانية أن تكون صواباً أو خطأ بمستويات ثقة محددة) من أهم آليات تفسير النتائج. وتتم عملية التفسير لمخرجات عمليات التحليل الإحصائي على مستويين ، وهما :

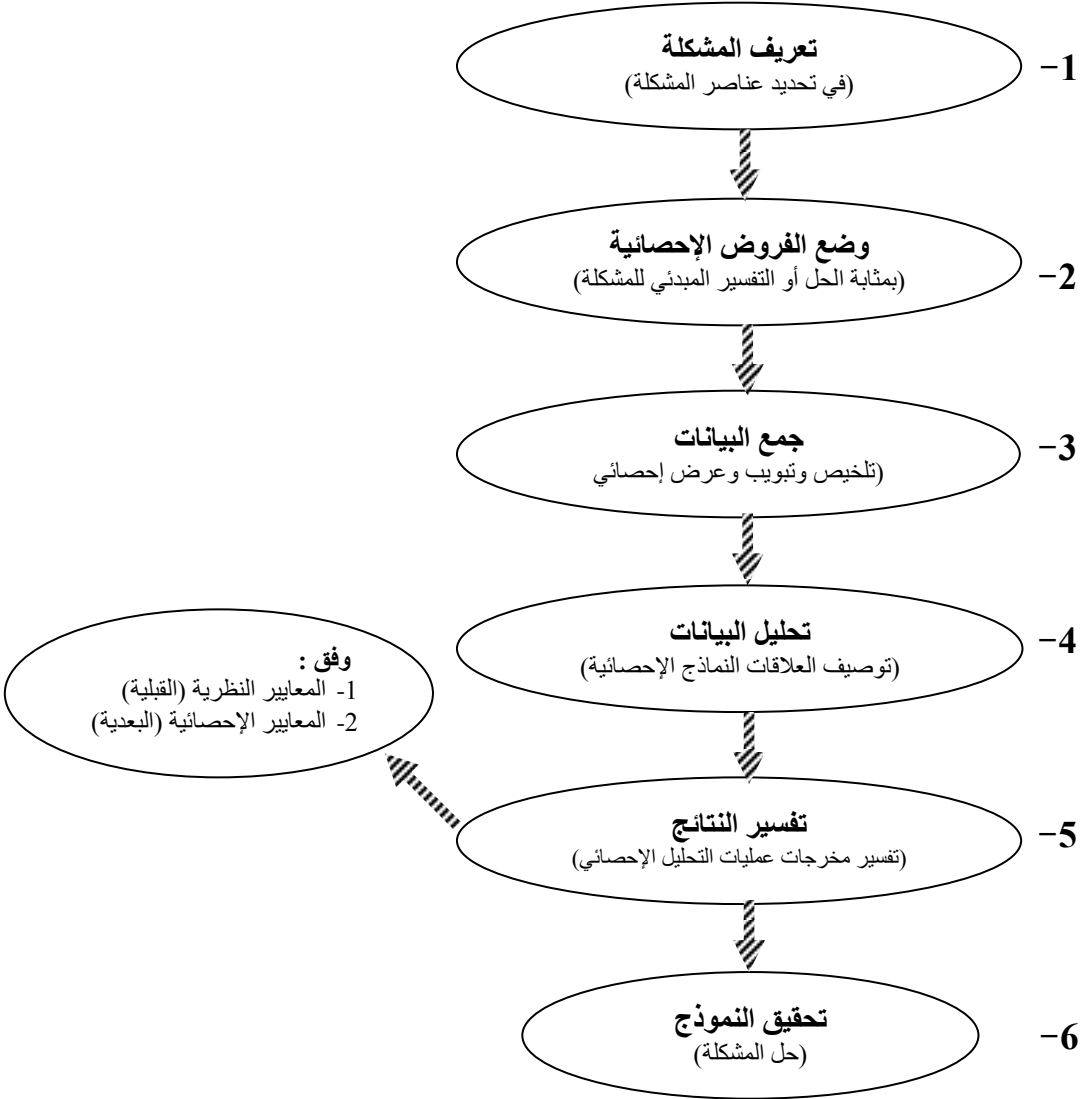
(1): معيار قبلي : Prior : يسترشد الإحصائي بالمعايير النظرية الحاكمة لعلاقات الظاهرة أو المتغيرة تحت المعالجات الإحصائية ، ويقع عبء هذا الدور أساساً على المتخصص في مجال الدراسة ، فقد يكون : ديموجرافي أو اجتماعي أو اقتصادي أو سياسي أو نفسي أو تربوي أو طبي أو غيره ، في مجالات العلوم وأفرع المعرفة المختلفة .

(2): معيار بعدي Posterior :

وهنا يجب التنويه إلى أن المعايير الإحصائية (البعديّة) هي داعم أو مؤكد للمعايير النظرية القلبية ولكن لا تجبها .

السادس : تحقيق النموذج :-

يأتي علاج المشكلة تحت الدراسة بتحقق صحة الفروض النظرية الموضوعية باحتمال على ضوء مخرجات عمليات التحليل الإحصائي المستخدمة ، وغالباً ما تصاغ في صورة علاقات إحصائية تتوافر فيها القدرة على تفسير سلوك الظاهرة تحت التفكير الإحصائي في الماضي والحاضر والمستقبل ، أما إذا لم يتحقق ذلك يبدأ الشك في النموذج الإحصائي المقترح ، ويجب على الباحث إعادة النظر في صياغة فروض نظرية جديدة ، والعودة إلى الخطوة الثالثة ، ويمكن إيجاز خطوات التفكير الإحصائي ، في الشكل 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 منفعة - فالحاسب الآلي لا يوفر لمستخدمه التفكير الاحصائي عند التعامل مع المشاكل تحت المعالجة الاحصائية ، فالحاسب الآلي لا يختار للاحصائي الاسلوب أو الطريقة الاحصائية واجبة الاستخدام أو يفسر له مخرجات عمليات التشغيل الاحصائي المختلفة ، وإنما يقع عبئه على الاحصائي المتخصص ، والأمر هنا يتوقف على مدى كفاءته الاحصائية ، وبما يتوافر لديه من مخزون معرفي بالنظرية الاحصائية ، ومستحدثاتها والتي تشتمل على فروع متنوعة صارت مجالات تخصص مختلفة تحت مظلة علم الاحصاء .

(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 فيما يلي :

الأول: تفرغ البيانات: وفيها تتم عملية تفرغ 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 : عند التعامل مع الظواهر أو المتغيرات الوصفية مثال : الإجابات ، وبمستوياتها على مستوى السؤال الواحد ، في استمارة الاستبيان .

(2): القيم المفقودة : مثال ذلك حالة النسيان أو الرفض أو الغير صحيحة لبعض إجابات استمارة الاستبيان ، وتعالج بالاستبعاد أو بإعطائها رمز خاص ، ويجب التعامل معها بطريقة ملائمة لموضوع الدراسة . كما يجب التأكيد على ضرورة المراجعة الدقيقة لعملية تفرغ 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 إلى الحاسب في ملف يسمى بملف البيانات وفق قواعد البرنامج الاحصائي الجاهز المستخدم ، مع التأكيد على أهمية مراجعة عملية إدخال البيانات للتأكد من سلامتها ، خاصة والبرمجيات الاحصائية الجاهزة تعطي إمكانية وضع حدود لعدد الحقول عند إدخال البيانات ، فإذا تم إدخال بيانات أكبر من عدد الحقول المحددة مسبقاً يعطي مثلاً رمز (**) بديلاً عن القيم في ملف البيانات .

الثالث: وصف البيانات: تبدأ عملية التحليل الاحصائي للبيانات بخطوة وصف البيانات ، مستخدمين الجداول التكرارية والرسوم البيانية والمقاييس الاحصائية الأولية

(المتوسطات ، التشتت ، الالتواء ، التفلطح) ، بهدف التعرف على الوصف الاحصائي للبيانات ، لحسن التعامل معها ، واختيار الطرق والأساليب الاحصائية الملائمة .

الرابع: توصيف العلاقات (أو النماذج) الاحصائية : تحكم عملية اختيار الأساليب والطرق الاحصائية الملائمة ، كل من نوع وخصائص البيانات ، وهدف الدراسة تحت التحليل الاحصائي ، وفيها يتم الآتي:

(1):اختبارات الفروض الاحصائية :

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

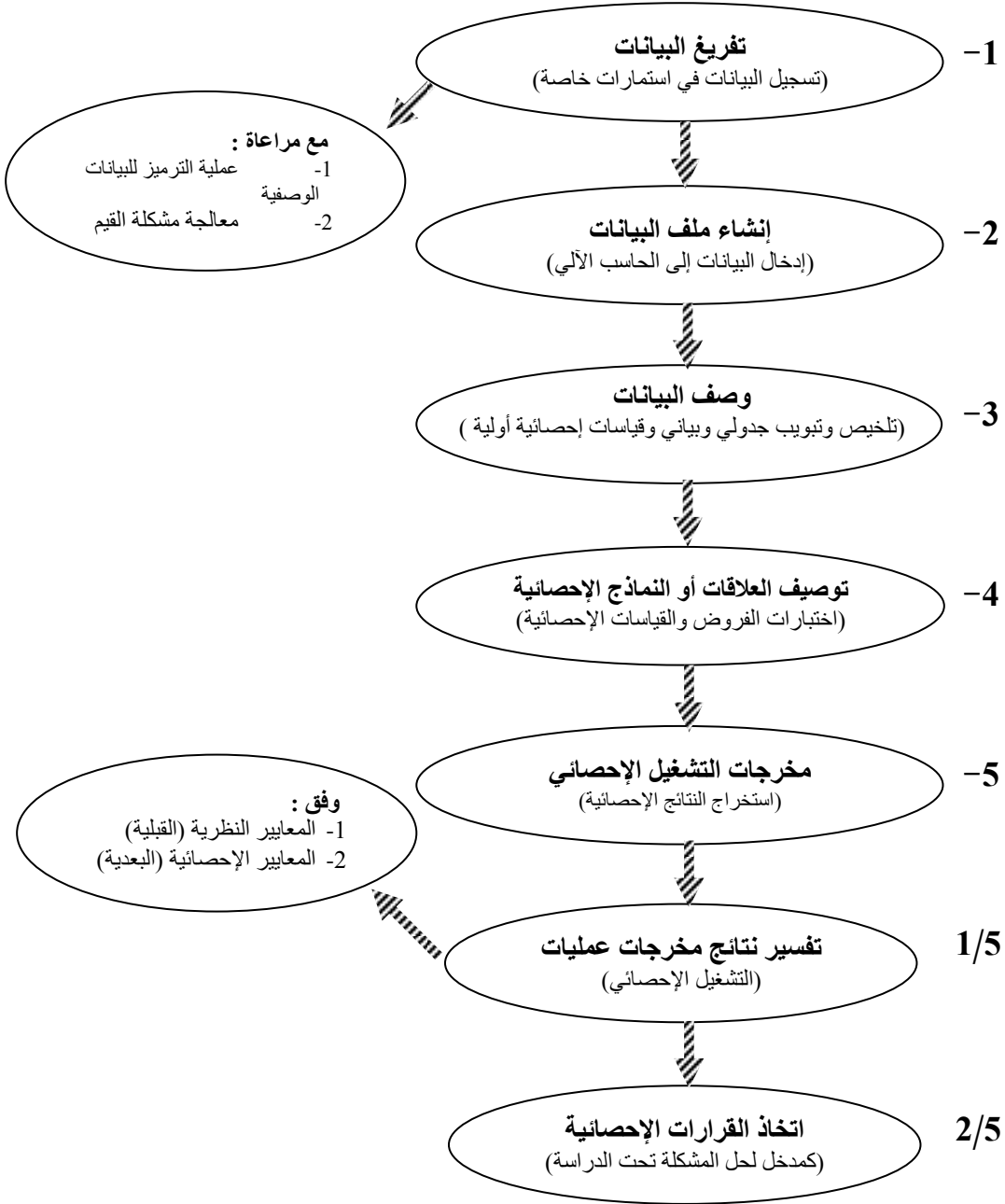
(2):قياس العلاقات (أو النماذج) الاحصائية :

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 ، ومستخدمين من الأساليب والطرق الاحصائية المتنوعة ، مثال : تحليل الانحدار (البسيط ، المتعدد) ، تحليل التباين(في اتجاه واحد أو أكثر) ، تحليل التغاير ، التحليل العامل ، تصميم التجارب ، وغيرها ، مع

- ملاحظة أنه يمكن عامة تقسيم مخرجات عمليات التحليل الاحصائي إلى ثلاث مجموعات ، وهي :
- أ - وصف البيانات .
- ب- اختبارات الفروض الاحصائية (المعلمية - اللامعلمية) .
- ج - قياس العلاقات (أو النماذج) الاحصائية .

الخامس : مخرجات عملية التشغيل الاحصائي :

- ويتم التعامل مع مخرجات التشغيل الاحصائي المختلفة على مستويين :
- (1):تفسير نتائج مخرجات عمليات التشغيل الاحصائي (وصف البيانات ، اختبارات الفروض ، قياس العلاقات) ، على ضوء : المعايير النظرية (القبليّة) الحاكمة لعلاقات النظرية للظاهرة أو المتغيرة محل التحليل الاحصائي ، وبدعم المعايير الاحصائية الحاكمة للطرق والأساليب الاحصائية المستخدمة .
- (2):اتخاذ القرارات الاحصائية على ضوء التفسير السابق كمدخل لحل المشكلة تحت التحليل الاحصائي .
- ويمكن إيجاز خطوات التحليل الاحصائي باستخدام الحاسب الآلي في الشكل التالي :



شكل (2-1)

توضيحي عام خطوات التحليل الإحصائي باستخدام الحاسب الآلي

(1-5-2): البرمجيات الإحصائية الجاهزة :

ويمكن تناول تعريف البرمجيات الإحصائية الجاهزة تحت نوعين رئيسيين :

أ- البرمجيات الإحصائية البسيطة :

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

ب- البرمجيات الإحصائية الأكثر تعقيداً :

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

Sas , Spss , Minitab , Bmdp وغيرها .

تعريف البرنامج الإحصائي الجاهز Spss :

وهو برنامج إحصائي متكامل يقوم بالعديد من عمليات التحليل الإحصائي 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 Sps:

- (1) Reliability Analysis (لقياس ثبات ، ومصداقية البيانات) .
- (2) Store (لترتيب التصاعدي أو التنازلي للبيانات) .
- (3) Frequencies (لعمل الجداول التكرارية ، العرض البياني ، حساب مقاييس إحصائية أولية [الموضع ، التشتت ، الإلتواء والتقاطع) .
- (4) Cross tabs (لعمل الجداول التكرارية المزدوجة) .
- (5) Descriptive (لحساب العديد من المقاييس الإحصائية الأولية للبيانات) .
- (6) Tests of Hypotheses (لإجراء اختبارات الفروض الإحصائية المعلمية) .
- (7) NPAR (لإجراء اختبارات الفروض الإحصائية اللامعلمية) .
- (8) Graph (لرسم أشكال الانتشار بين المتغيرات) .
- (9) Correlation (لقياس معاملات الارتباط بين المتغيرات الكمية أو الوصفية) .
- (10) Regression (لقياس الانحدار الخطي البسيط ، المتعدد ، غير الخطي بين المتغيرات) .
- (11) Cluster Analysis (التحليل العنقودي) .
- (12) Discriminant Analysis (تحليل التمايز) .
- (13) Factor Analysis (التحليل العاملي) .
- (14) Logistic Regression (الانحدار اللوجستي) .

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

مع ملاحظة : أهم العمليات على البيانات :

(1): يمكن إجراء بعض التحويلات 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 بأسلوب إحصائي سليم فإن بياناتها تتسم بالدقة وقلة التكلفة والمشقة عن بيانات الحصر الشامل . ويمكن بدراسة خصائص العينة (أو ما يسمى بإحصائي العينة) استنتاج الاحصائية المختلفة للعينة والتي يمكن تعميمها بالطرق الاحصائية لاستنتاج خصائص المجتمع المسحوبة منه تلك العينة . فيجب أن تتصف العينة بجميع خصائص المجتمع التي تمثله . لذا يلزم اختيار مفردات العينة بإحدى الطرق التي تضمن صحتها على قدر الإمكان . وتتلخص أهم العينات التي يمكن الحصول عليها بطريقة إحصائية جيدة من مجتمع الدراسة فيما يلي :-

1- العينة العشوائية البسيطة Simple Random Sample

وهي أبسط أنواع العينات وتتميز بأنها تعطي فرصاً متساوية لجميع مفردات المجتمع للظهور في العينة كما أن عامل الصدفة يكون هو العامل الوحيد المحدد للمفردات التي ستظهر بالعينة. ولذلك فإن اختيار مفردات العينة يتم بأي طريقة عشوائية : (أي بطريقة تضمن تساوي فرص الظهور في العينة أمام جميع مفردات المجتمع) . فمثلاً عند اختيار طالبين من خمسة عشر طالباً بطريقة عشوائية يمكن استخدام 15 بطاقة متماثلة تماماً يكتب على كل منها اسم أحد الطلبة ثم تخلط البطاقات جيداً وتسحب إحداها فيكون الاسم المدون بها هو اسم أول طالب في العينة ثم تخلط البطاقات الباقية جيداً ويسحب منها بطاقة أخرى ويكون الاسم المدون بها هو اسم الطالب الثاني في العينة ، ويمكن الاستعانة بجدول الإعداد العشوائية لتبسيط عملية الاختيار العشوائي من المجتمعات الكبيرة .

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

2- العينة العشوائية المنتظمة Systematic Random Sample

والهدف من استخدام هذه العينة هو تجنب التكرار المطول الذي يظهر عند استخدام العينة العشوائية البسيطة . ويعتمد اختيار مفردات هذه العينة أساساً على تقسيم المجتمع إلى فئات متساوية (عددها يساوي عدد مفردات العينة)

ثم اختيار مفردة واحدة من الفئة الأولى بطريقة عشوائية واختيار باقي المفردات من الفئات المتتالية بطريقة منتظمة .

فمثلاً : إذا كان المجتمع مكون من 480 مفردة وإذا كان المطلوب اختيار عينة حجمها 60 مفردة ، فإننا نقسم المجتمع إلى 60 فئة يتكون كل منها من 8 مفردات : $(8 = \frac{480}{60})$ وتكون أرقام مفردات الفئة الأولى هي : 1 ، 2 ، 8 ، 0000 وأرقام مفردات الفئة الثانية هي : 9 ، 10 ، .. ، 16 وهكذا . ثم نبدأ باختيار أحد مفردات الفئة الأولى بطريقة عشوائية ولنفرض أنها كانت المفردة رقم 4 ، فتكون الخطوة التالية هي اختيار مفردة من الفئة الثانية بطريقة منتظمة فيكون رقمها هو $4 + 8 = 12$ ثم نختار مفردة من الفئة الثالثة بنفس الطريقة فيكون رقمها $4 + 16 = 20$ أو $12 + 8$ وهكذا 0000 فنجد أن المفردات التي تظهر في العينة هي المفردات المناظرة للأرقام : 4 ، 12 ، 20 ، 28 ، 26 ، 0000 ، 468 ، 476 .

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

3- العينة الطبقيّة Stratified Sample

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

- ويتوقف حجم العينة التي يتم اختيارها من كل طبقة على أحد العوامل التالية:
- أ - **حجم الطبقة** : فحجم العينة من طبقة معينة يتناسب طردياً مع حجم هذه الطبقة .
- ب- **مدى التجانس بين مفردات الطبقة** : فكلما ازداد التجانس بين مفردات الطبقة الواحدة كلما أدى ذلك إلى تخفيض حجم العينة المختارة من هذه الطبقة .
- ج - **تكاليف الحصول على بيانات المفردة من كل طبقة** : فكلما زادت نفقات جمع البيانات من المفردة في طبقة معينة كلما أدى ذلك إلى تخفيض حجم العينة من هذه الطبقة .

4- العينة المتعددة المراحل Multi-stage Sample

إذا كان المجتمع محل الدراسة كبيراً و به الكثير من الاختلافات والعوامل المؤثرة على المفردات . وإذا كان من المرغوب فيه اختيار عينة ممثلة لهذا المجتمع بما فيه من اختلافات دون أن تكون نفقاتها مرتفعة فإنه يمكن اختيار العينة على مرحلتين أو أكثر . مثال ذلك إذا كان المطلوب هو استخدام أسلوب العينة لجمع بيانات عن كيفية انفاق دخل الأسرة في الجمهورية العربية ، فإنه يمكن - كمرحلة أولى - تقسيم الجمهورية إلى مجموعات متشابهة من المحافظات (مثلاً وجه قبلي ، وجه بحري ، الواحات ، 000) . ويتم اختيار عينة من كل مجموعة ، وتبدأ المرحلة الثانية باختيار مجموعة من مدن وقرى المحافظات التي ظهرت في المرحلة الأولى . ثم تأتي المرحلة الثالثة وهي اختيار مجموعة من أسر القرى والمدن التي ظهرت في المرحلة الثانية . وبذلك نكون قد استخدمنا عينة من ثلاث مراحل وتمتاز هذه العينة بأن الأسر التي تظهر في المرحلة الأخيرة تكون مركزة في مجموعة محدودة من المدن والقرى وبالتالي

فإن نفقات جمع البيانات لن تكون مرتفعة . كما أنه نتيجة الطريقة المتبعة في الاختيار أصبحت العينة ممثلة لجميع أنحاء الجمهورية دون أن يكون حجمها مبالغاً فيه .

د- الجداول العشوائية وكيفية استخدامها :

الجدول العشوائية ما هي إلا الأرقام 0 ، 1 ، 2 ، ، 9 مرتبة في أعمدة متتالية تبعاً لترتيب الحصول عليها بإحدى الطرق العشوائية .

ولتوضيح كيفية استخدام هذه الجداول لنفرض أننا نريد اختيار عينة عشوائية بسيطة مكونة من أربع عمال شركة إنتاجية معينة وعددهم 650 عاملاً . في هذه الحالة نبدأ بتكوين "الإطار" وهو قائمة تشمل أسماء جميع العمال بالشركة (دون حذف أو تكرار) وتعطى لكل منهم رقماً مسلسلاً 1 ، 2 ، ... وكذا حتى 650 . وبذلك نعتبر أن كل عامل يمثل عدد واحد (مكون من ثلاث خانات) فنختار ثلاث أعمدة من جداول الأعداد العشوائية ونقرأ أربعة أعداد متتالية وكل منهم يمثل أحد العمال اللذين يتم اختيارهم في العينة . فمثلاً إذا اخترنا الأعمدة الثلاث الأخيرة من الجدول المعطى في نهاية هذا الكتاب نجد أن الأربعة أعداد التي نحصل عليهم هي : 503 ، 529 ، 102 ، 113 وكل منها يمثل أحد العمال في العينة (يلاحظ أننا تجاهلنا الرقم 702 لأنه أكبر من 650 ، أي لا يوجد عامل في القائمة التي لدينا يناظر هذا العدد) .

وإذا كان المطلوب هو اختيار عينة عشوائية بسيطة من مجتمع حجمه 100 مفردة فإنه يمكن توفير الكثير من الجهد المرتبط بعملية الاختيار العشوائي إذا تم إعداد القائمة باستخدام الأرقام 00 ، 01 ، 02 ، ... ، 99 أو الأرقام 01 ، 02 ، ... ، 99 ، 00 (أي .. بدلاً من 100) لأنه في هذه الحالات يتم استخدام عمودين فقط من أعمدة الجدول العشوائي بدلاً من استخدام ثلاثة

(إذا استخدمت الأعداد 1 ، 2 ، 00 ، 100 فإنه يجب أن نعتبر كل منها مكون من ثلاث خانوات أي : 001 ، 002 ، 000 ، 099 ، 100 حتى يكون لجميع الأعداد نفس فرصة الظهور في العينة) .

كذلك إذا كان حجم المجتمع هو 170 مثلاً فيجب استخدام ثلاثة أعمدة ولكن بدلاً من إهمال جميع الأعداد من 170 أو حتى 999 فإنه يمكن اعتبار الرقم المسلسل 1 مقابلاً للأعداد 001 ، 201 ، 401 ، 601 ، 801 . والرقم المسلسل 2 مقابلاً للأعداد 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 وبمبسطة يسهل فهمها .

(2): عرض البيانات الإحصائية بيانياً (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 قيماً - أو أوجهها - مختلفة

تسمى بعناصر نطاق المتغير . والمتغير الإحصائي يتم تقسيمه إلى قسمين رئيسيين هما :

الأول : المتغير الوصفي: Qualitative Variable

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

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

الثاني : المتغير الكمي: Qualitative Variable

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

(1):متغير كمي منقطع (منفصل): 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 عدد الوحدات المنتجة من الثلاجات أو الغسالات أو إلخ .

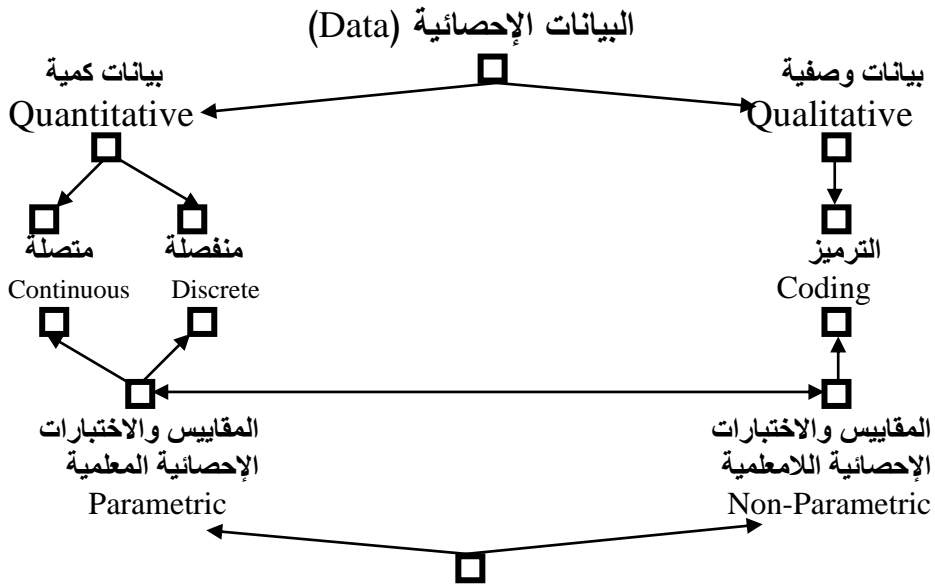
(2): متغير كمي متصل (مستمر) : Continuous

وهو متغير كمي عناصر نطاقه قيمًا عددية صحيحة أو كسرية . مثال ذلك المتغير الذي يعبر عن درجات الطلاب أو وزن مجموعة من الطلاب أو المتغير الذي يعبر عن الطول أو المساحة أو درجات الحرارة المسجلة في عواصم العالم أو الدخل أو الأجر اليومي أو الأسبوعي أو الشهري أو... إلخ . هذا ويجب التنويه إلى أن البيانات الإحصائية غالبًا ما تكون غير كافية بحد ذاتها بل تستخدم للتوصل إلى استنتاجات معينة وهو ما يعرف بعملية التحليل أو المعالجة الإحصائية للبيانات . ويقوم علم الإحصاء بالدور الكبير في هذا الخصوص .

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

ويمكن عرض مجمل كافة المقاييس والاختبارات الإحصائية المستخدمة في عمليات التحليل الإحصائي للبيانات تحت فرعين أساسيين : المقاييس والاختبارات الإحصائية المعلمية Parametric : ولها أولوية الاستخدام (ما دامت الشروط الإحصائية متحققة) ، وإلا استخدمت المقاييس الإحصائية اللامعلمية Non-Parametric عند التعامل مع البيانات الكمية وتكون لها أولوية الاستخدام ما دامت الشروط الإحصائية غير متحققة لإستخدام المقاييس المعلمية هذا بالإضافة لإستخدامها عند التعامل مع البيانات

الوصفية . غير أن اتساع دوائر استخدامات الحاسب الآلي ، ومع التطوير الموازي والمتسارع في تطبيقات البرامج والحزم الإحصائية الجاهزة ، قد سهل الأمر على كافة المستخدمين من غير الإحصائيين عند التعامل مع المجموعات الكبيرة من البيانات على مستوى العرض الجدولي والبياني والتحليل الإحصائي بسرعة ودقة كبيرتين (وإن كان يفضل دائماً عند التعامل مع التحليل الإحصائي للبيانات أن تتم بمعرفة أو تحت إشراف إحصائي) . ويمكن تقديم عرض تلخيصي لمجمل موضوع البيانات ومعالجتها الإحصائية في الشكل التوضيحي التالي .



شكل (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 جدولياً .

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

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

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

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 مفردة من مفردات الظاهرة محل الدراسة . ثم يلي عملية تحديد أقسام

(أو فئات) الظاهرة عملية التفريغ في الجدول ويليها أخيراً عملية العد كما توضحه الأمثلة التالية .

مثال (1) : فيما يلي تقديرات نجاح 20 طالباً في مادة الإحصاء :

جيد جداً - جيد - مقبول - جيد - ممتاز - جيد - جيد جداً - جيد - جيد - جيد جداً - مقبول - جيد جداً - جيد جداً - مقبول - جيد - ممتاز .

الحل :

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 جداً ، ممتاز . لذا فإنه يتم رسم جدول من ثلاثة أعمدة ، تعكس هذه الأعمدة المراحل الثلاث لتبويب هذه البيانات جدولياً .

حيث يتم سرد أقسام الظاهرة الأربعة في العمود الأول تحت اسم فئات التقدير (أو تقديرات النجاح) ويفضل وصفها في صورة مرتبة سواء تنازلياً أو تصاعدياً .

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 جداً في العمود الثاني تعبر تلك العلامة عن تقدير نجاح

الطالب الأول في هذه المجموعة ، ثم نبدأ في وضع علامة مناظرة أمام

التقدير جيد تعبر عن تقدير نجاح الطالب الثاني ، ثم علامة مناظرة أمام
التقدير مقبول ، ثم علامة مناظرة أمام التقدير جيد وهكذا حتى يتم تفرغ كل
المفردات المعبرة عن التقديرات والتي تنتهي بتقدير ممتاز . مع مراعاة أنه
لتسهيل عملية العد فإنه أُنْفِق على أن تكون العلامة المعبرة عن المفردة عبارة
عن شرطة مائلة (/) وأن تظهر العلامات في شكل حزمة تحتوي على خمسة
علامات . ولتجنب الوقوع في أخطاء عند العد فإن العلامة الخامسة - إن
وجدت - توضع بطريقة عكسية لوضع العلامات الأربعة التي تسبقها . وبذلك
تأخذ الحزمة الصورة (////) . وفي واقع الأمر فإن الحزم تساعد كثيرًا في
تسهيل عملية العد النهائية حيث يكون معروفًا ان كل حزمة تحتوي على
خمس مفردات . هذا ويجب أن يراعى عند تسطير الجدول الذي ستنتم فيه
عملية التبويب أن يكون العمود الثاني متسعًا إلى حد ما حتى يتسنى ظهور
الحزم بشكل واضح وذلك من خلال ترك فراغ ملائم فيما بين كل حزمة وأخرى .
ومما لا شك فيه أن درجة اتساع هذا العمود تتوقف على عدد المفردات أو
حجم البيانات المطلوب تبويبها . والجدول التالي يلخص المراحل الثلاث لعملية
تبويب بيانات تقديرات نجاح الطلاب جدولياً :

جدول (1-2)

عملية تفريغ تقديرات نجاح عشرون طالباً في مادة الإحصاء

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

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

**** مراحل العرض الجدولي لبيانات الظواهر الكمية :**

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

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

بصفة عامة فإن العرض الجدولي للبيانات الإحصائية الكمية المتقطعة لا يختلف عن العرض الجدولي للظواهر الوصفية سائلة 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 ، 4 ، 5 ، 2 ، 4 ، 2 ،

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

والمطلوب تبويب هذه الظاهرة (تفريغها) جدولياً .

الحل: لاحظ أن الظاهرة محل الدراسة في هذا المثال هي حجم الأسرة وهذا المتغير (حجم الأسرة) لا شك أنه يعبر عن متغير كمي متقطع (لأنه يأخذ قيمًا منفصلة ولا يمكن أن يقبل التقسيم أو التجزئة) لذا يتم تفريغ هذه البيانات من خلال استعراض الأحجام المختلفة لمجموعة الأسر المعطاة . أي أن أقسام الظاهرة عبارة عن مجموعة الأرقام التالية 1 ، 2 ، 3 ، 4 ، 5 ، 6 . فإذا بدأنا بقراءة أحجام الأسر واحدًا تلو الآخر ملتزمين بترتيب المفردات كما هي واردة بالمثال (2) فأننا نبدأ بوضع علامة مناظرة للحجم (5) في العمود الثاني بالجدول (2 - 2) ثم علامة مناظرة للحجم (4) ثم علامة مناظرة للحجم (3)

.... وهكذا حتى يتم تفريغ الجدول حسب أحجام الأسر والذي يوضحه الجدول التالي :

جدول (2-2)

تفريغ أحجام 30 أسرة

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

(2) مراحل العرض الجدولي للبيانات الكمية المتصلة (المستمرة) :

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 لكي يتم تحديد عدد الفئات والحد الأدنى والأعلى لكل فئة فإنه يتم اتباع الخطوات التالية :

1- تحديد المدى (Range)

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

$$\text{المدى} = \text{أكبر قيمة} - \text{أصغر قيمة} + 1$$

أي أن :

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

2- تحديد عدد الفئات: The Number of Intervals or Classes (NI) or (C):

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

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

وهذه القاعدة تعطي نتائج مناسبة لأي 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 صغيرًا لأن

ذلك يؤدي إلى طمس معالم التوزيع . وفي النهاية فإن تحديد عدد الفئات المناسب قد يكون اختياريا متروكاً للتقدير الشخصي لمن يقوم بتفريغ وأعداد الجداول التكرارية . وعموماً 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 :

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

أي أن :

$W = (\text{Range } (R)) / C = (\text{Maximum value} - \text{Minimum value} + 1) / C$
وينتج في تلك الحالة طول فئة ثابت لكل فئات التوزيع أي يعطي جدولاً تكرارياً متساوياً من حيث أطوال فئاته . هذا ويفضل أن تكون أطوال فئات الجدول التكراري متساوية حتى يسهل معها عرض البيانات سواء جدولياً أو بيانياً كما سنرى فيما بعد . مع مراعاة في حالة ما إذا كان ناتج عملية حساب طول الفئة رقماً كسرياً فيفضل تقريب الناتج إلى أقرب عدد صحيح . فمثلاً إذا كان المدى المحسوب في الخطوة الأولى وليكن 200 وكان عدد الفئات هو 7 فئات فإن ناتج القسمة هو 28.571 يتم تقريبه ليصبح طول الفئة 29 وحدة قياس تابعة لنفس وحدة قياس الظاهرة أو المتغير محل الدراسة .
هذا ويفضل اختيار عدد الفئات رقماً بحيث يكون ناتج قسمة المدى عليه يعطي طول الفئة مساوياً للرقم (5) أو (10) أو مضاعفاتهم .

هذا وتجدر الإشارة إلى أنه في بعض الحالات تستلزم طبيعة البيانات استخدام فئات غير متساوية من حيث أطوالها وذلك نظرًا لقلّة التكرارات المقابلة لهذه الفئات . أو قد تكون فئات الجدول أو التوزيع التكراري محددة سلفاً في صورة فئات غير متساوية من حيث أطوالها . وبصفة عامة يطلق على الجدول أو التوزيع التكراري المتساوي من حيث أطوال فئاته بالتوزيع المنتظم . أما الجدول التكراري الذي تشذ فيه طول أحد فئاته على الأقل عن باقي فئات التوزيع فيسمى بالجدول التكراري الغير منتظم .

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

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

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

أما عن الحد الأعلى لكل فئة من فئات التوزيع فلتجنب أي لبس فقد أُتفقَ 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- فنقرأ : من 20 جنييه حتى أقل من 30 جنييه .
- 30- وتقرأ : من 30 جنييه حتى أقل من 40 جنييه .
- 40- وتقرأ : من 40 جنييه حتى أقل من 50 جنييه .
- 50- وهكذا

أي أنه وفقاً لهذه الطريقة فإننا نكتب الحد الأدنى لكل فئة فقط تاركين الحد الأدنى ليتحدد ضمناً بقيمة أقل من الحد الأدنى للفئة التالية لها . وعلى ذلك فالعامل الذي يتقاضى أجرًا يوميًا وليكن 29.5 جنييه 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 بالجنييه 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 تفرغ هذه البيانات في صورة جدول تكراري .

الحل : لتفرغ بيانات أجور العاملين المعطاه يتم الآتي :

(1): حساب المدى :

$$R = \text{Max} - \text{Min} + 1 \quad \text{حيث أن}$$

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

(2): تحديد عدد الفئات :

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

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

(3): تحديد طول الفئة : فحيث أن:

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

ومن ثم فإن فئات التوزيع تأخذ الصورة التالية :

$$-30 \quad , \quad -34 \quad , \quad -38 \quad , \quad -42 \quad , \quad -46 \quad , \quad -50 \quad -54$$

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

(4): تحديد الحد الأدنى والأعلى لفئات الأجور الأسبوعية : وهنا يتم تفرغ

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 الأجر المعطاة على فئات الأجر التي تنتمي إليها . فنحصل على الجدول التكراري (3-2) . فبقراءة المفردات المعبرة عن أجر العمال الأسبوعية المفردة تلو الأخرى وذلك بنفس الترتيب الوارد في بيانات مثال (3) فأنا 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 . أما المفردة 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 الأجر 42 جنيه وهي تنتمي للفئة (42-) فيتم وضع علامة أمامها ومن ثم تنتهي عملية التفريغ . فيتم ترجمة تلك العلامات إلى أرقام تعبر عن التكرارات المقابلة لكل فئة من فئات الأجر الأسبوعية .

هذا ولكي نتأكد من أن عملية التفريغ شملت كافة المفردات الموضحة بالمثال يجب أن يكون مجموع التكرارات الموضحة في العمود الثالث مساوياً لعدد المفردات الإجمالي المطلوب تبويب بياناتها وهي (40) مفردة في مثالنا. والجدول التالي (2-3) يلخص مجموعة الخطوات السابقة .

جدول (2-3)

تفريغ جدول التوزيع التكراري للأجر الأسبوعي لـ 40 عاملاً (بالجنبيه)

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

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

(5):مرحلة العد:

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

جدول (2-4)

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

وفي هذا التوزيع التكراري الوارد بالجدول (2-4) نجد أن الحد الأدنى للفئة الأولى في التوزيع معلوم (30) وأن الحد الأعلى للفئة الأخيرة (54) كذلك معلوم لذا يسمى هذا التوزيع التكراري بأنه توزيع مغلق (مقفّل) من الطرفين. وفي حالة غياب أحد هذين الحدين يسمى التوزيع التكراري مفتوح سواء من طرف واحد أو من الطرفين . فمثلاً إذا كان التوزيع غير معلوم الحد الأدنى للفئة الأولى ومعلوم الحد الأعلى للفئة الأخيرة فيسمى التوزيع بأنه مفتوح من أعلى والعكس صحيح إذا كان التوزيع التكراري معلوم الحد الأدنى للفئة الأولى وغير معلوم الحد الأعلى للفئة الأخيرة يسمى التوزيع التكراري بأنه مفتوح من أسفل . أما في ظل غياب أو عدم معلومية الحدين الأدنى للفئة الأولى والأعلى للفئة الأخيرة فيسمى التوزيع التكراري مفتوح الطرفين .

الجدول أو التوزيعات التكراري المفتوحة : في بعض الأحيان قد نضطر إلى اللجوء إلى الجداول التكرارية المفتوحة وذلك في حالة تعذر تحديد الحد الأدنى للفئة الأولى أو الحد الأعلى للفئة الأخيرة أو كليهما معاً . والجدول التالي (جدول (2 - 5)) يبين توزيعاً تكرارياً مفتوح الطرفين .

جدول (2-5)

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 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 (2 - 6)) يعطي صورة واضحة عن الجداول التكرارية الغير منتظمة .

جدول (2-6)

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 (2-6) أن أطوال فئاته الستة الأولى هي 10 جنيه أما الفئتين الأخيرتين (-290) ، (310-340) فإن أطولهما هما 20 ، 30 جنيه على الترتيب . لذا فإن هذا التوزيع التكراري توزيعاً غير متساوي أطوال فئاته (توزيعاً غير منتظماً) .

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

ملحوظة: بعد إتمام عملية تفريغ البيانات والحصول على الجداول التكرارية كما هو وارد عل سبيل المثال في جدول (2-5) أو (2-6) فإن الجدول

التكراري البسيط (الذي يحوي ظاهرة واحدة أو متغير واحد محل الدراسة) يفيد في تحديد عدد التكرارات التي تنتمي لكل فئة من فئات التوزيع التكراري أي عدد التكرارات التي تقع ما بين الحدين الأدنى والأعلى لكل فئة من فئات التوزيع بصورة صريحة. فمثلاً في الجدول (5-2) فإن هناك سبعة طلاب درجاتهم تقل عن عشرة درجات وكذلك فإن هناك تسعة طلاب تتراوح درجاتهم ما بين عشرة درجات وأقل من 20 درجة وهناك 15 طالباً تتراوح درجاتهم ما بين عشرين درجة وأقل من 30 درجة وهكذا .

هذا ويمكن من خلال تلك الجداول التكرارية البسيطة الحصول على نسبة التكرارات التي تنحصر ما بين حدي كل فئة من فئات التوزيع التكراري وهو ما يسمى بالجداول التكرارية النسبية والتي سيتم تناولها في الجزء التالي . وهنا نود أن ننوه بأنه تصح مقارنة توزيعين تكراريين بسيطين متى تساوت مجموع التكرارات الأصلية لهما أما في حالة إختلاف مجموع التكرارات للتوزيعين فلا بد من الحصول على التكرارات النسبية لكل توزيع 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 (Σ F_i) :
أي أن :

$$R_i = (F_i / \Sigma F_i) , \text{ where } i = 1, 2, \dots, r$$

حيث:

(r) تعبر عن عدد فئات التوزيع التكراري وعلامة (Σ) تعبر عن المجموع.

Or:

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

والفرق ما بين الصيغتين أن الأولى تعطي التكرار النسبي كسرًا عشريًا وفي هذه الحالة يكون مجموع عمود التكرارات النسبية مساويًا للواحد الصحيح أما الصيغة الثانية فهي نسبة مئوية وباستخدامها يكون مجموع عمود التكرارات النسبية مساويًا 100% .

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

عمود من عمودي التكرارات النسبية يكون مساوياً للواحد الصحيح ومن هنا تم التخلص من أثر عدم التساوي في تلك الحالة ومن ثم تصح عملية المقارنة .

مثال (4) :

الجدول التالي يوضح التوزيع التكراري لدرجات مجموعة الطلاب والطالبات .

جدول (7-2)

التوزيع التكراري لدرجات مجموع من الذكور والإناث

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

والمطلوب إجراء المقارنة بين التوزيعين باستخدام التوزيع التكراري الملائم.

الحل : بالنظر لمجموع تكرارات كلاً من الذكور والإناث لنفس فئات التوزيع نجد أن هناك اختلاف في مجموع التكرارات الأصلية للذكور عن الإناث (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 الذكور ولتكن (R_iM) وكذلك للإناث ولتكن (R_iF) وذلك كما يوضحها الجدول (8-2) التالي :

جدول (2-8)

التوزيع التكراري الأصلي والنسبي لدرجات الذكور (M) والإناث (F)

Classes	التكرارات الأصلية		التكرارات النسبية (%)		Remarks
	F _i M	F _i F	R _i M	R _i F	
Less than 10	6	15	$\frac{6}{60} \times 100$ = 10	$\frac{15}{100} \times 100$ = 15	R ₁ M < R ₁ F
10 -	12	17	20	17	R ₂ M > R ₂ F
20 -	12	18	20	18	R ₃ M > R ₃ F
30 -	9	18	15	18	R ₄ M < R ₄ F
40 -	12	10	20	10	R ₅ M > R ₅ F
50 -	6	9	10	9	R ₆ M > R ₆ F
60 and more	3	13	5	13	R ₇ M < R ₇ F
Σ	60	100	100 %	100 %	

لاحظ أنه مجموع عمودي التكرارات النسبية سواء للذكور أو الإناث مساوياً 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 .

*مذلك 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 يحدده التكرار النسبي وهو الأصح في هذه الحالة .

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

التوزيعات أو الجداول التكرارية المزدوجة : 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 رغبتنا في عمل دراسة لمعرفة ما إذا كان هناك علاقة من عدمه فيما بين هاتين الظاهرتين فإنه يصعب إجراء مثل تلك الدراسة إذا ما وضعنا بيانات كل ظاهرة في جدول تكراري بسيط . لكن يمكن إجراء مثل تلك الدراسات إذا ما تم وضع بيانات الظاهرتين (أيًا كانت الظاهرتين وصفيتين 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) أي إلى مجموعة من الصفوف والأعمدة . حيث يتم تمثيل فئات أحد الظاهرتين في الصفوف والأخرى في الأعمدة . فالتوزيعات التكرارية المزدوجة تستخدم في تلخيص بيانات أزواج من المشاهدات أو القيم بغرض دراسة نوع ودرجة قوة العلاقة فيما بين الظاهرتين أو المتغيرين محل الدراسة أو ما يسمى بعلاقة الارتباط وهو ما سيرد دراسته فيما بعد في هذا الكتاب .

هذا ولتكوين الجدول التكراري المزدوج فإننا نتبع نفس الخطوات السابق شرحها في حالة الجداول التكرارية البسيطة وذلك من خلال تطبيقها على الظاهرتين أو المتغيرين محل الدراسة كل على حده . فبالنسبة لكل ظاهرة أو متغير يتم تحديد أقسام الظاهرة أو المتغير محل الدراسة (الفئات) كما سبق ثم يتم تصوير

جدول تكراري مزدوج حيث يتم تمثيل فئات الظاهرة الأولى في الصفوف والثانية في الأعمدة . ثم يلي ذلك عملية تفرغ بيانات الظاهرتين معاً في هذا الجدول التكراري المزدوج (مرحلة التفرغ) . 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 كلياً على حدة (فئات كل ظاهرة على حدة) .
حيث يتم تحديد أقسام كل ظاهرة على حدة وذلك باتباع نفس خطوات الأسلوب السابق شرحه في حالة الجدول التكراري البسيط .

(2):عملية التفرغ : (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 التفرغ 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 تفرغ 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 وما يناظرها 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	10	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 الدخل (x) والإنفاق (y) لتلك الأسر .
- 2- من خلال نتائجك في (1) استنتج التوزيع الهامشي للدخل الأسبوعي (x) للأسر وكذا التوزيع الهامشي للإنفاق الأسبوعي (y) للأسر .
- 3- من خلال نتائجك في (2) اوجد التوزيع التكراري النسبي لكل من الدخل الأسبوعي (x) والإنفاق 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)

- $NI \text{ (or) } C = 1 + 3.3 \text{ Log } (n)$
 $= 1 + 3.3 \text{ Log } (50)$
 $= 6.6066 = 7 \text{ Intervals (classes)}$
- $LI \text{ (or) } W = \text{Range} \div NI$
 $= 66 \div 7 = 9.43 \quad (\text{L.E})$

ويتم تقريب طول الفئة بالزيادة فتصبح طول الفئة مساوياً عشرة جنيهاً تقريباً. 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 وخمسة وعشرون جنيهاً لذا فإن فئات الدخل (x) تكون عبارة عن : 60- , 70- , 80- , 90- , 100- , 110- , 120-130

ثانياً: بالنسبة لبيانات الإنفاق (y) :

- $\text{Range for } (y) = 108 - 41 + 1 = 68 \quad (\text{L.E})$
- $NI \text{ (or) } C = 7 \text{ Intervals (classes)}$

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

- $LI \text{ (or) } W = 68 \div 7 = 9.71 = 10 \quad (\text{L.E})$

وحيث أن أصغر قيمة في الإنفاق هي إحدى وأربعون جنيهاً. لذا فإن من المناسب أن يكون الحد الأدنى للفئة الأولى هو أربعون جنيهاً. وحيث أن أكبر قيمة للإنفاق هي مائة وثمانية جنيهاً. لذا فإن فئات الإنفاق (y) تأخذ الشكل التالي : 40- , 50- , 60- , 70- , 80- , 90- , 100- 110

ومن خلال نتائجننا عن فئات كل من المتغيرين y, x يتم تصوير جدول تكراري مزدوج حيث يتم تمثيل فئات الدخل (x) وليكن في الصفوف وفئات الإنفاق (y) في الأعمدة فيعطى الصورة الموضحة في الجدول (9 - 2) على النحو التالي :

جدول (9-2)

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

$x \backslash y$	40-	50-	60-	70-	80-	90-	100-110	Σ
60-								
70-								
80-								
90-								
100-								
110-								
120-130								
Σ								50

أما عن مرحلة تفرغ 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 ببيانات 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

(/) في هذه الخلية . أما الزوج الثاني من المشاهدات [وهو $x = 83$ ، $y = 75$] فإن ينتمي إلى الخلية الناشئة من تقاطع الصف الثالث الممثل للفئة (80-) مع العمود الرابع الممثل للفئة (-70) . لذا يتم وضع علامة (/) في الخلية الناشئة عن تقاطع الصف الثالث مع العمود الرابع .

وهكذا بالنسبة لبقية أزواج المشاهدات حتى نصل إلى الزوج الأخير من المشاهدات [وهو $x = 119$ تقابلها $y = 108$] والذي ينتمي للخلية الناشئة عن تقاطع الصف السادس الممثل للفئة (-110) مع العمود السابع الممثل للفئة (100 - 110) . لذا يتم وضع شرطة في الخلية الناشئة عن تقاطع الصف السادس مع العمود السابع كما هو موضح في جدول التفريغ (10-2) والذي يبين العلامات التي تعكس ازدواج المشاهدات .

جدول (10-2)

تفريغ مزدوج للدخل الأسبوعي (x) والإنفاق الأسبوعي (y) لخمسين أسرة

$y \backslash x$	40-	50-	60-	70-	80-	90-	100-110	Σ
60-	//	/						3
70-	//	///	//	/				8
80-		//	///	/// /	/			14
90-		/	///	////	///	/		12
100-				/	///	//		8
110-						//	//	4
120-130							/	1
Σ	4	7	10	12	9	5	3	50

وبترجمة العلامات الواردة بخلايا جدول التفريغ المزدوج المبين في الجدول (2-10) إلى أرقام نحصل علي الجدول التكراري المزدوج الذي تحتوي خلاياه على التكرارات المختلفة كما هو موضح في الجدول (2-11) التالي :

جدول (2-11)

جدول تكراري مزدوج يوضح العلاقة بين الدخل

الأسبوعي (x) والإنفاق الأسبوعي (y) لخمسين أسرة

x \ y	40-	50-	60-	70-	80-	90-	100-110	Σ
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
Σ	4	7	10	12	9	5	3	50

هذا ويمكن من الجدول (2-11) القول أن تكرار فئة معينة من فئات الدخل (x) هو عبارة عن مجموع التكرارات المشتركة في صف هذه الفئة . فمثلاً يمكن القول من خلال بيانات الصف الأول في هذا الجدول (وهو صف فئة الدخل 60 جنيه) أن هناك ثلاث أسر يتراوح دخلها ما بين ستون وأقل من سبعون جنيهاً . وكذلك هناك ثماني أسر يتراوح دخلها ما بين سبعون جنيهاً

وأقل من ثمانون جنيهاً . وكذلك الحال بالنسبة لفئات الإنفاق . حيث يمكن القول أن هناك أربعة أسر يتراوح إنفاقها الأسبوعي ما بين أربعون وأقل من خمسون جنيهاً وكذلك هناك سبعة أسر يتراوح إنفاقها الأسبوعي ما بين خمسون وأقل من ستون جنيهاً وهكذا .

هذا ويجب أن ننوه إلى أنه يعتبر من المألوف أن نجد عدداً من الخلايا في الجدول التكراري المزدوج لا يحتوي على تكرارات وهو ما يعني 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 (11-2) يمكن استنتاج التوزيع التكراري الهامشي للدخل الأسبوعي (x) وهو عبارة عن فئات الدخل (x) المختلفة وما يقابلها من مجاميع تكرارات الصفوف الواردة بالعمود الأخير في الجدول (11-2) دون أخذ الإنفاق الأسبوعي (y) في الاعتبار . والجدول (12-2) يبين التوزيع التكراري الهامشي للدخل الأسبوعي (x) .

جدول (2-12)

التوزيع التكراري الهامشي للدخل الأسبوعي
للعيينة المكونة من خمسين أسرة (x)

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

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

Weekly Expenditure classes (y)	40-	50-	60-	70-	80-	90-	100- 110	Σ
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 (2-14) ، (2-15) 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	Σ
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 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	Σ
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 وأقل من تسعون جنيهًا وهكذا .

هذا وفي نهاية دراستنا للجدول التكرارية المزدوجة تجدر الإشارة إلى ما يلي :
1- ليس بالضرورة أن يتساوى عدد الفئات أو طول الفئة للمتغيرين في الجدول التكراري المزدوج . إذ أن عدد الفئات أو طول الفئة لكل متغير يتوقف على المدى المطلق (Range) لهذا المتغير .

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

3- التوزيع التكراري الهامشي لمتغيراً ما هو عبارة عن التوزيع التكراري البسيط لهذا المتغير . والتوزيع التكراري البسيط يبين عدد التكرارات التي تقع ما بين الحد الأدنى والأعلى لكل فئة من فئات التوزيع .

الجدول (التوزيعات) التكرارية المتجمعة : Cumulative Frequency
سبق وأن ذكرنا حالاً أن التوزيع التكراري البسيط يعطي عدد التكرارات التي تقع ما بين الحد الأدنى والأعلى لكل فئة من فئات التوزيع بصورة صريحة . لكن في كثير من الأحيان قد يهتم الباحث بمعرفة عدد التكرارات أو المفردات التي تأخذ قيمة معينة سواء كانت تلك القيمة المعنية تمثل الحد الأدنى أو الأعلى لأي فئة من فئات التوزيع أو تقع ما بينهما . فمثلاً قد يكون الباحث في حاجة لأن يعرف ما هو عدد الطلاب اللذين تقل درجاتهم عن 90 درجة لتوزيع تكراري بسيط معطي . وفي أحيان أخرى قد يهتم الباحث بمعرفة عدد المفردات أو التكرارات التي تأخذ قيمة تزيد على أو تساوي قيمة معينة سواء كانت تلك القيمة تمثل الحد الأدنى أو الأعلى لأي فئة من فئات التوزيع

أو أي قيمة تقع ما بينهما . وليكن على سبيل المثال ما هو عدد الموظفين اللذين تبلغ مرتباتهم الأسبوعية 95 جنيهاً على الأقل (95 جنيهاً فأكثر) .
ففي واقع الأمر فإن الجداول التكرارية البسيطة لا تلبى - بطريقة مباشرة - هذه الحاجات . ولكي يصل الباحث إلى ما يريد فإنه يقوم بتجميع التكرارات تصاعدياً في الحالة الأولى بينما يقوم بتجميعها تنازلياً في الحالة الثانية .
ويسمى الجدول في الحالة الأولى بالجدول التكراري المتجمع الصاعد

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

بينما يسمى الجدول في الحالة الثانية بالجدول التكراري المتجمع الهابط أو النازل (Dsc.C.F.T)
هذا ويتكون كلاً من الجدول التكراري المتجمع الصاعد أو الهابط من خلال الجدول التكراري البسيط للظاهرة موضع الدراسة . ويستخدم الجدول التكراري المتجمع الصاعد (Asc.C.F.T) في تحديد عدد المفردات التي تأخذ قيمة أقل من قيمة معينة . كما يستخدم في تحديد القيمة التي يقل عنها عدداً معيناً أو نسبة من المفردات . أما الجدول التكراري المتجمع الهابط (Dsc.C.F.T) فإنه يستخدم في تحديد عدد المفردات التي تأخذ قيمة تساوي أو تزيد عن قيمة معينة كما يستخدم في إيجاد القيمة التي يزيد عنها أو يساويها عدداً معيناً أو نسبة معينة من مفردات التوزيع التكراري .

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

(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 ما بين حدي فئة معينة فلاحظ أن هناك مرادفات لغوية مكافئة لكلمة أقل من حيث أن :-

أقل من \equiv يقل عن \equiv لا تزيد عن \equiv على الأكثر \equiv الحد الأقصى
فكل هذه المرادفات أو ما يعادلها لغوياً يفيد ضرورة تكوين الجدول التكراري
المتجمع الصاعد (Asc.C.F.T) واستنتاج المطلوب منه .

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

تزيد عن \equiv فأكثر \equiv لا تقل عن \equiv على الأقل \equiv الحد الأدنى .
فكل هذه المرادفات وما يعادلها لغوياً يفيد ضرورة تكوين التكراري 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 الأصلي
على الأقل (أو على الأكثر) فإنه يمكن استنتاج المطلوبين معاً أما من خلال
التوزيع التكراري المتجمع الصاعد أو من خلال التوزيع التكراري المتجمع
الهابط كما سنرى في الأمثلة فيما بعد .

مثال (6) :-

فيما يلي لديك التوزيع التكراري للأجور الأسبوعية لعينة من خمسين
عاملاً (بالجنيه) كما يوضحها الجدول (16-2) التالي :

جدول (16 - 2)

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

والمطلوب :-

أولاً: إيجاد التوزيع التكراري للمتجمع الصاعد وكذا التوزيع التكراري النسبي للمتجمع الصاعد .

ثانياً: إيجاد التوزيع التكراري للمتجمع الهابط وكذا التوزيع التكراري النسبي للمتجمع الهابط .

ثالثاً: باستخدام 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 العمال على 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 .

الحل :-

أولاً: وفقاً لتعريف التوزيع التكراري للمتجمع الصاعد . فإن التكرار للمتجمع الصاعد (Asc.C.F.T) المناظر لأي فئة من فئات التوزيع الأصلي ما هو إلا عبارة عن تكرار هذه الفئة مضافاً إليه تكرارات الفئات السابقة . ويتكون هذا

الجدول من عمودين يتم كتابة الفئات التكرارية المتجمعة الصاعدة في صورة (أقل من الحد الأعلى لكل فئة في فئات التوزيع) في العمود الأول أما التكرارات المتجمعة الصاعدة (Asc.C.F.T) فيتم تحديدها في العمود الثاني من هذا الجدول . أما التكرارات النسبية المتجمعة الصاعدة (R_i.Asc.C.F) للفئة (i) فهي عبارة عن التكرارات المتجمعة الصاعدة للفئة (i) مقسومة على مجموع التكرارات سواء ككسر عشري أو كنسبة مئوية . أي أن :-

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

حيث (r) تمثل عدد فئات التوزيع، حيث Asc.C.F (i) هو التكرار المتجمع الصاعد للفئة (i) . والجدول (2-17) يبين حسابات التوزيع التكراري المتجمع الصاعد والمتجمع النسبي 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
∑	50	less than 100	50	100

لاحظ في الجدول (2-17) ما يلي :-

(1): قمنا بافتراض أنه توجد فئة وهمية (افتراضية) تسبق الفئة الأولى فيكون الحد الأعلى لتلك الفئة الوهمية ما هو إلا الحد الأدنى للفئة الأولى (30جنيه) فيكون التكرار المتجمع الصاعد المقابل للحد الأدنى للفئة الأولى (وهو الحد الأعلى لتلك الفئة الوهمية يكون دائماً أبداً مساوياً للصفر حيث لا توجد هناك تكرارات سابقة لتلك الفئة الأولى) . هذا ويمكن الاستغناء عن السطر الأول في التوزيع المتجمع الصاعد . وفي هذه الحالة نبدأ حسب تمييز العمود الأول من التوزيع المتجمع الصاعد وهو أقل من الحد الأعلى للفئة الأولى والذي يبلغ 40 جنيه . فيكون عدد العاملين اللذين يقل أجرهم عن أربعون جنيهاً هو عامل وحيد وهو بمثابة تكرار الفئة الأولى في التوزيع التكراري الأصلي . يلي ذلك السطر الثاني في الجدول المتجمع الصاعد فإن عدد العاملين اللذين يقل أجرهم عن الحد الأعلى للفئة الثانية (والذي يبلغ 50 جنيه) وهو عبارة عن 5 عمال وهذا التكرار المتجمع بمثابة تكرار الفئة الأولى مضافاً إليه تكرار الفئة الثانية في التوزيع التكراري الأصلي (أي أن : $5 = 1 + 4$) وبذلك يكون التكرار المتجمع الصاعد المقابل للحد الأعلى للفئة الثانية ما هو إلا تكرار الفئة الأولى مضافاً إليه تكرار الفئة الثانية في التوزيع التكراري الأصلي . وهكذا يتم إيجاد التكرار المتجمع الصاعد لبقية الفئات فئة تلو الأخرى حتى نصل إلى التكرار المتجمع الصاعد للحد الأدنى للفئة الأخيرة (وهو بمثابة الحد الأدنى لفئة وهمية تلي الفئة الأخيرة والذي يعتبر بمثابة الحد الأعلى للفئة الأخيرة بالجدول التكراري الأصلي) والذي يجب أن يكون مساوياً للمجموع الكلي للتكرارات أي خمسون عاملاً في هذا المثال .

(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 التكراري الأصلي . 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 (أقل من 70 جنيه) مضافاً إليه تكرار الفئة من 70 حتى أقل من 80 جنيه أي أن (26 + 11 = 37) .

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

(4): الجدول التكراري المتجمع الصاعد النسبي ما هو إلا صورة من الجدول التكراري المتجمع الصاعد لكن بعد قسمة عمود التكرارات المتجمعة الصاعدة بالكامل على مجموع تكرارات التوزيع الأصلي وهي القيمة 50 عامل في مثالنا محل الدراسة .

(5): التكرار المتجمع الصاعد أما أن يبدأ بالصفر (وذلك في حالة افتراض أن هناك فئة وهمية تسبق الفئة الأولى فيكون الحد الأعلى لتلك الفئة الوهمية بمثابة الحد الأدنى للفئة الأولى) أو يبدأ بتكرار الفئة الأولى (وذلك في حالة بداية الفئات المتجمعة بالحد الأعلى للفئة الأولى) وينتهي بالمجموع الكلي لتكرارات التوزيع الأصلي .

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

ثانياً: إيجاد التوزيع أو الجدول التكراري المتجمع الهابط (Dsc.C.F.T) وكذا التوزيع التكراري المتجمع الهابط النسبي (R. Dsc.C.F.T) . فلتكوين الجدول التكراري المتجمع الهابط فإنه يعاد تعريف الفئات كما هو موضح في الجدول (18 - 2) حيث يتم كتابة فئات التوزيع التكراري المتجمع الهابط على صورة " الحد الأدنى للفئة الأصلية فأكثر . ومن ثم فبالاستعانة ببيانات الجدول (16 - 2) سوف نوضح كيفية تكوين الجدول التكراري المتجمع الهابط وكذا الهابط النسبي وهوما يوضحه الجدول (18-2) .

جدول (18 - 2)

التوزيع التكراري المتجمع الهابط (Dsc.C.F.T) وكذا التوزيع
التكراري المتجمع الهابط النسبي (R.Dsc.C.F.T) لأجور عينة من 50
عامل

Classes	F _i	More than Lower Limit	DscCFD	R _i DCFD (%)
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): المتجمع الهابط المقابل لفئة معينة (للحد الأدنى لفئة معينة) من فئات التوزيع الأصلي ما هو إلا عبارة عن مجموع التكرارات مطروحاً منه تكرارات الفئات السابقة لتلك الفئة . فعلى سبيل المثال فإن التكرار المتجمع الهابط للفئة الأولى ما هو إلا عبارة عن المجموع الكلي للتكرارات (50 عاملاً) حيث أنه لا توجد تكرارات سابقة لهذه الفئة . 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 $50 - 1 = 49$. وهكذا بالنسبة لبقية الفئات .

(2): يمكن تكوين الجدول التكراري المتجمع الهابط بطريقة أسهل عن طريق اتباع نفس الأسلوب المستخدم في حالة الجدول التكراري المتجمع الصاعد. وذلك من خلال بداية عملية التجميع المتتالية من أسفل لأعلى على عكس التوزيع التكراري المتجمع الصاعد . فمثلاً التكرار المتجمع الهابط المقابل للحد الأدنى للفئة الأخيرة (وهو 90 جنيه) يقابله التكرار المتجمع الهابط (5 عامل) وهو نفس تكرار الفئة الأخيرة في التوزيع التكراري الأصلي . بينما يكون التكرار المتجمع الهابط للفئة قبل الأخيرة أو الحد الأدنى للفئة قبل الأخيرة فهو عبارة عن تكرار الفئة الأخيرة مضافاً إليه تكرار الفئة قبل الأخيرة وذلك من التوزيع التكراري الأصلي أي $5 + 8 = 13$ وهكذا إلى أن نصل إلى التكرار المتجمع الهابط المقابل للحد الأدنى للفئة الأولى والذي يجب أن يكون مساوياً للمجموع الكلي للتكرارات . أو يمكن القول أن التكرار المتجمع الهابط لفئة ما هو إلا التكرار المتجمع الهابط للفئة السابقة مطروحاً منه تكرار الفئة السابقة من التوزيع الأصلي . فمثلاً التكرار المتجمع الهابط للفئة الأولى هو (50) نظراً لعدم وجود فئات سابقة للفئة الأولى . أما التكرار المتجمع الهابط للفئة الثانية فما هو إلا التكرار المتجمع الهابط للفئة الأولى مطروحاً منه التكرار الأصلي للفئة الأولى (1 عامل) . كذلك فإن التكرار المتجمع الهابط للفئة الثالثة ما هو إلا التكرار المتجمع الهابط للفئة الثانية (49 عاملاً) مطروحاً منه التكرار الأصلي للفئة الثانية (4 عامل) أي أن $49 - 4 = 45$. وهكذا إلا أن نصل إلى التكرار المتجمع الهابط للفئة الأخيرة (90 فأكثر) فيكون عبارة عن التكرار المتجمع الهابط للفئة قبل الأخيرة (80 فأكثر) أي (13 عاملاً) 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 الخصوص

يفترض 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 100 على حسب بيانات هذا المثال .

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

(4): الجدول التكراري المتجمع الهابط النسبي ما هو إلا صورة من الجدول التكراري المتجمع الهابط مع الاختلاف الوحيد وهو أن التكرارات النسبية المتجمعة الهابطة النسبية ($R_i.Dsc.C.F.T$) ما هي إلا التكرارات المتجمعة الهابطة منسوبة إلى أو مقسومة على مجموع التكرارات .

(5): من الجدول التكراري المتجمع الهابط يمكن الحصول على كثيرًا من المعلومات والتي لا يمكن الحصول عليها بطريقة مباشرة من خلال 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 تزيد على أو تساوي قيمة معينة لا توجد بصورة صريحة في الجدول المتجمع . وكذلك يمكن تحديد قيمة الأجر الذي يحصل عليه أو أكثر منه عدداً معيناً من العمال حتى لو كانت هذه القيم غير موضحة صراحة في الجدول . وهو ما سنراه باستخدام طريقة الاستكمال العددي أو النسبة والتناسب كما سنراه في

نهاية هذا المثال أو أثناء تناولنا لدراسة الوسيط كأحد مقاييس المتوسطات أو مقاييس النزعة المركزية وهو ما سيرد فيما بعد في أبواب لاحقة بمشيئة الله .

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

ثالثاً: باستخدام التوزيع التكراري الملائم تحديد :-

(أ) - عدد العمال اللذين تتراوح أجورهم ما بين 80 جنيه حتى أقل من 90 جنيه وكذلك نسبة هذا العدد من إجمالي التكرارات .

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 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 :

$$\text{نسبة العمال اللذين تتراوح أجورهم ما بين 80 أقل من 90 جنيه} \\ = (8 / 50) \times 100 = 16\% .$$

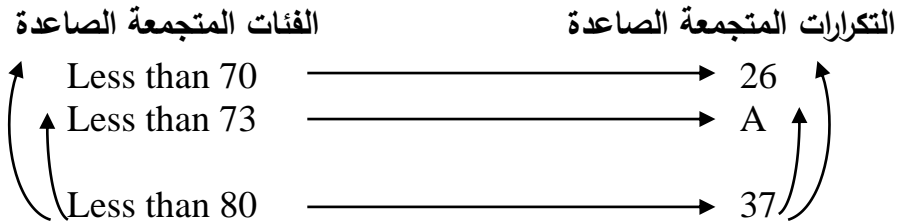
أي أن هناك 16% من إجمالي عدد العاملين تتراوح أجورهم ما بين 80 وأقل من 90 جنيه اسبوعياً .

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

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

في مثل هذه الحالات لاحظ 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 . وهنا يجب التنويه أن فكرة إنشاء الجداول التكرارية البسيطة وكذا المتجمعة تبنى على أساس أن التكرارات تتوزع بانتظام

(بالنسبة والتناسب) على مدار طول الفئة الواحدة . لذا فإن قيمة الأجر الأسبوعي 73 جنيه تقع ما بين الأجر 70 جنيه ، 80 جنيه اللذان يقابلهم تكرارات متجمعة صاعدة كما يوضحها 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 تصوير الشكل الكروي التالي :-



وباستخدام فكرة النسبة والتناسب فإن :

(الحد الأخير - الحد الأول) / (الحد الأخير - الحد الأوسط) للفئات المتجمعة = (الحد الأخير - الحد الأول) / (الحد الأخير - الحد الأوسط) للتكرارات المتجمعة. أي أن :

$$(73 - 80) / (70 - 80) = (A - 37) / (26 - 37) \text{ . أي أن :}$$

$$7 / 10 = (A - 37) / 11 \text{ . وعليه فإن :}$$

$$\text{i.e., } 10(37 - A) = 77$$

$$\text{i.e., } 370 - 77 = 10A$$

$$A = 293 \div 10 \text{ فتكون قيمة } A \text{ هي :}$$

$$= 29.3 \approx 29 \text{ (Persons)}$$

ومن ثم فإن عدد العمال اللذين تقل أجورهم عن 73 جنيه ما هو إلا 29 عاملاً . لاحظ أن الناتج لابد أن يقع ما بين التكرارين المتجمعين الصاعدين أي ما بين 26 ، 37 عاملاً وهو معياراً لدقة الحسابات . ولاحظ أيضاً أن

الناتج (29 عاملاً) يقترب من التكرار المتجمع الصاعد (26) حيث أن المقابل على مستوى الفئات المتجمعة نجد أن الأجر 73 جنيه يقترب أيضاً من 70 جنيه وليس 80 جنيه . أما عن النسبة فإن نسبة العمال اللذين تقل أجورهم عن 73 جنيه (أو تبلغ أجورهم 73 جنيه على الأكثر) 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 على الأكثر .

(ج) لتحديد قيمة الأجر الذي يتقاضاه 65 % من إجمالي عدد العمال على الأقل . لاحظ أولاً أن اللفظ على الأقل = القيمة فأكثر . لذا فإن الإجابة 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 :

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

أي أن : $(X - 70) / 10 = (9 -) / (14 -)$. أي أن :

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

$$14x = 890$$

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

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 الجدول التكراري المتجمع الهابط . وبملاحظة العدد 47 عاملاً نجد أنه يقع في عمود التكرارات المتجمعة الهابطة (وليس الفئات) ما بين التكرارين المتجمعين الهابطين 49 ، 45 . وبافتراض أن الأجر المطلوب وليكن (y). فلاستنتاج قيمة (y) يتم تكوين الرسم الكروكي التالي :

استخدام قاعدة النسبة والتناسب كما يلي :

More than 40 —→ 49
 More than (y) —→ 47
 More than 50 —→ 45

فيكون :-

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

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

i.e. ,

$$20 = 4 (50 - y)$$

$$4 y = 200 - 20$$

i.e. ,

$$y = 180 \div 4 = 45 \quad (\text{L.E})$$

أي أن قيمة الأجر الذي يتقاضاه 47 عاملاً كحد أدنى ما هو إلا 45 ج.
ملحوظة : لاحظ من النتائج السابقة أنه حينما كان عدد العمال 47 يقع في منتصف التكرارات المتجمعة الهابطة ما بين 49 ، 45 عاملاً بالضبط وهي بالضبط فإن الناتج وهو قيمة (Y) القيمة 45 جنية تقع في منتصف الفئة أي في نقطة المنتصف ما بين 40 ، 50 جنيه .
 وبلغت أخرى : فحيث أن :

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

لذا فإن :

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

وهي نفس النتيجة السابقة.

(هـ): لتحديد نسبة العمال الذين تتراوح أجورهم ما بين 55 ، 87 جنيه. فحيث أنه لا توجد فئة صريحة من فئات التوزيع التكراري الأصلي الواردة بالجدول (2-16) تعطي صورة مباشرة عن تلك الفترة أو هذه الفئة من (87-55) لذا يجب استخدام أحد الجداول التكرارية المتجمعة أما الصاعد أو الهابط. والآن دعنا نستنتج المطلوب من التوزيع التكراري 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 = $(12+5) \div 2$ أي $7.5 \approx 8$ عاملاً .

هذا وبافتراض أن عدد العمال اللذين يقل أجرهم الأسبوعي عن 87 جنيه من الجدول التكراري المتجمع (B) عامل. فإنه لاستنتاج قيمة (B) وليكن الصاعد الوارد في الجدول (2-17) فإن قيمة الأجر 87 جنيه تقع في الفئة من 80 إلى 90 وعليه يكون لدين الشكل الكروكي التالي :

Less than 80 —————> 37
 Less than 87 —————> B
 Less than 90 —————> 45

فيكون :

$$\frac{90-80}{10} = \frac{45-37}{8} \quad \text{i. e. ,}$$

$$\frac{3}{24} = \frac{45-B}{10(45-B)}$$

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

$$\text{i.e. , } B = 42.6 \approx 43 \quad (\text{Persons})$$

والآن دعنا نلخص النتائج المستخلصة حتى الآن :-

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

Less than 30 L.E	→	0	Persons
Less than 55 L.E	→	9	Persons
Less than 87 L.E	→	43	persons

فيكون عدد العمال اللذين تتراوح أجورهم الأسبوعية ما بين 55 جنيه ، 87 جنيه هو بمثابة الفرق ما بين 43 ، 9 عمال أي $34 = 9 - 43$ عاملاً . وعليه فإن نسبة العمال اللذين تتراوح أجورهم الأسبوعية ما بين 55 ، 87 جنيه هي :

$$R_i = \frac{F_i}{\sum F_i} \times 100 = \frac{34}{50} \times 100 = 68 \%$$

أي أن هناك 68 % من إجمالي عدد العمال يتراوح أجرهم الأسبوعي ما بين 55 ، 87 جنيهاً .

(و): حيث أنه لا توجد فئة من فئات التوزيع التكراري الأصلي الوارد بالجدول (16-2) تبدأ بالقيمة 53 جنيه . لذا فإن هذا المطلوب لا يمكن استنتاجه من التوزيع التكراري الأصلي . ويمكن استنتاج هذا المطلوب من أحد التوزيعين التكرارين المتجمعين أما الصاعد وأما الهابط .

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

$$50 \times 40\% = 20 \text{ عامل} . \text{ وحيث أن عدد العمال اللذين يقل أجرهم}$$

عن 53 جنيه يمكن استنتاجه من خلال افتراض أن هذا العدد وليكن (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 :

$$(Z - 12) / (5 - 12) = (53 - 60) / (50 - 60)$$

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

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

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

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

* هناك 7 عامل يقل أجرهم الأسبوعي عن 53 جنيه .

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

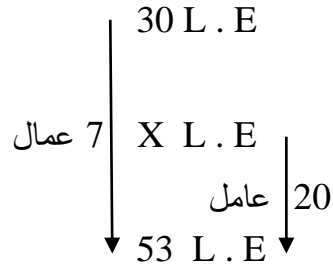
ولاستنتاج قيمة الدخل x دعنا نفرض الفرضين التاليين :-

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

والثاني : ان قيمة الدخل $(X) < 53$ جنيه.
 هذا وإحتبار مدى صحة هذين الفرضين فيكون ذلك من خلال الرسم الكروي
 التالي للحالتين السابقتين :

I - في حالة افتراض أن :

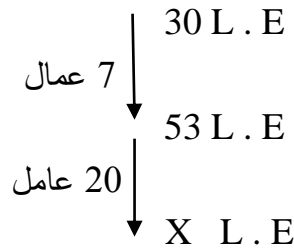
$x > 53$ جنيه فيكون :



وهذا مرفوض لأن القيمة (20)
 عامل لا يمكن أن تكون جزءاً من
 الكل وهو (7) عمال .

II - في حالة افتراض أن :

$x < 53$ جنيه فإن :



ونظراً لاستقلال السهمين فإن هذا
 الاتجاه مقبول رياضياً .

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

عدد العمال اللذين يقل أجرهم عن (x) جنيه = 7 + 20 = 27 عامل. ومن ثم فالمطلوب الآن هو تحديد قيمة الأجر (x) الذي يقل عنه 27 عاملاً . وبالرجوع للجدول التكراري المتجمع الصاعد الوارد بالجدول (17- 2) نجد أن التكرار المتجمع الصاعد 27 يقع ما بين التكرارين المتجمعين الصاعدين 26 ، 37 المقابلين للفئتين المتجمعتين 70 ، 80 جنيه على الترتيب ومن ثم يصبح لدينا الشكل الكروكي التالي :

Less than 70 → 26

Less than x → 27

Less than 80 → 37 فيكون :

80 - 70 37 - 26

_____ = _____

80 - x 37 - 27
10 11

_____ = _____ i.e., 100 = 11(80-x)
80 - x 10

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

فتكون قيمة x هي :

X = 780 ÷ 11 = 70.909 (L.E)

أي أن قيمة الدخل (x) هو 70.909 جنيه .

ملحوظة :- في المطلوب الأخير من هذا المثال لاحظ أنه تم رفض الفرض الأول الذي افترض أن قيمة x تقل عن 53 جنيه نظراً لأن هذا الفرض يتنافى مع المنطق الرياضي الذي يحكم بضرورة أن يكون الجزء (وهو الرقم الذي يدل عن 20 عامل في الرسم الكروكي) أقل من الكل (وهو الرقم 7

عامل في نفس الرسم) . أي أن هذا الفرض سيتم رفضه في الحالات التي يتنافى فيها مع المنطق الرياضي . لكن إذا كانت معطيات التمرين وما تم استنتاجه أفادت أن الجزء أقل من الكل فسوف تكون النتيجة هي قبول هذا الفرض أيضاً . وفي هذه الحالة يتم طرح التكرارين المتجمعين اللذين تم استنتاجهم بدلاً من الجمع كما هو في الفرض الثاني من هذا المثال وخلاصة ما سبق هو أنه في حالة قبول الفرض الأول فستكون هناك إجابتين صحيحتين لقيمة الأجر (x) . حيث أن الاتجاه الثاني سيكون مقبول دائماً نظراً للاستقلال دائماً فيما بين الجزئين اللذين تم استخلاصهم من المعطيات ونتائج الحل .

ثانياً : - العرض البياني للبيانات الاحصائية :

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

وخلاصة الأمر هو أنه رغم أن عملية عرض البيانات جدولياً تيسر على القارئ فهم هذه البيانات كما تساعد على استخلاص النتائج والعلاقات التي تربط بين الظواهر المختلفة إلا أن هذه العملية (العرض الجدولي) قد لا تلقي قبولاً من القارئ العادي الغير متخصص حيث يجد صعوبة في عملية استنباط النتائج من واقع بيانات الجداول التكرارية . ولعل العرض البياني

للبيانات يكون أكثر جاذبية للقارئ وذلك لما يعكسه من سرعة في فهم الظاهرة موضع الدراسة . إذ أنه من اليسير على القارئ حتى غير المتخصص - فهم طبيعة تطور الظاهرة بمجرد النظر للشكل البياني . أضف إلى ذلك أن العرض البياني للبيانات كثيراً ما يستخدم في إجراء عمليات المقارنة والتي يمكن إدراك نتائجها بشكل أيسر وأسرع مما لو استخدمنا العرض الجدولي لهذه البيانات . هذا بالإضافة إلى أن العرض البياني يمكننا من تحديد أشكال التوزيعات التكرارية المختلفة . وبصفة عامة وكما هو الحال في العرض الجدولي فإن العرض البياني للظواهر يعتمد على طبيعة البيانات أي طبيعة المتغيرات محل الدراسة من حيث كونها وصفية أم كمية بنوعيتها المنفصل أو المتصل . أي أننا سنتناول بشيء من الإيجاز والتوضيح العرض البياني للظواهر أو المتغيرات المختلفة :-

أ - الظواهر أو المتغيرات الوصفية .
 ب- الظواهر أو المتغيرات الكمية بنوعيتها المنفصل والمتصل والمبوبة تكرارياً.

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

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

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

Bar Charts

1- الأعمدة البيانية :-

وهي عبارة عن مجموعة من الأعمدة الرأسية (مستطيلات) قواعدها متساوية (سواء منفصلة أو متصلة أو مجزأة) تتناسب في ارتفاعاتها مع

قيم الظاهرة التي تعكسها تلك الأعمدة بيانياً . وفي هذه الحالة يتم تمثيل أقسام أو فئات الظاهرة على المحور الأفقي بينما يتم تخصيص المحور الرأسي كقياس لتكرار 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 في التعرف على التطور التاريخي لظاهرة ما أو أكثر خلال فترات زمنية مختلفة. ولرسم الشكل البياني باستخدام الأعمدة البيانية يتم الآتي :

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 ثلثي 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	Σ
Frequency	6	5	3	1	15

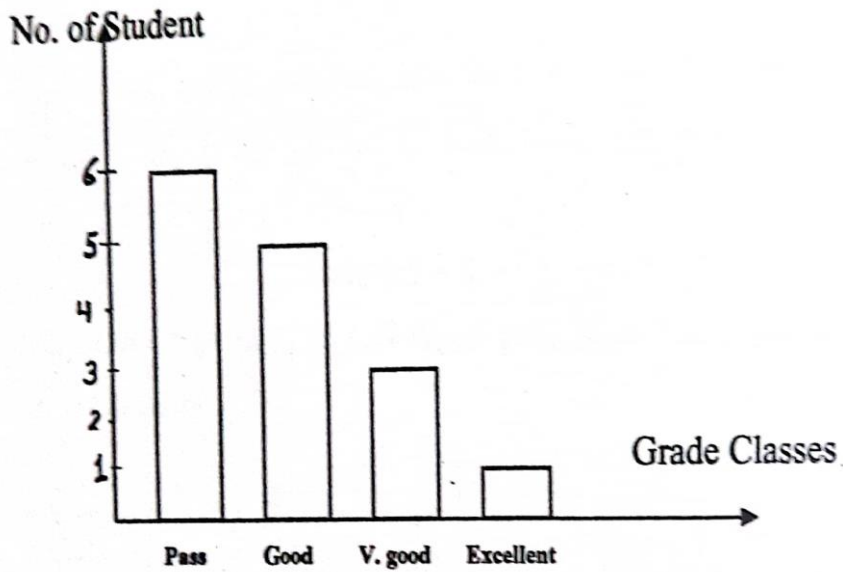
والمطلوب تمثيل بيانات هذا الجدول باستخدام الشكل البياني الملائم .

الجل :-

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

شكل (1-2)

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** للترقية 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)

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

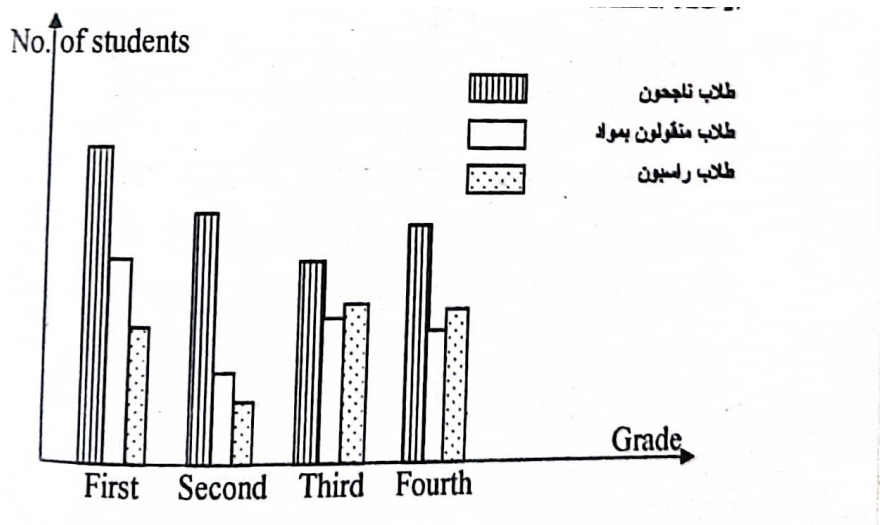
والمطلوب: استخدام الشكل البياني الملائم لعرض بيانات هذا الجدول .

الجل :-

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)

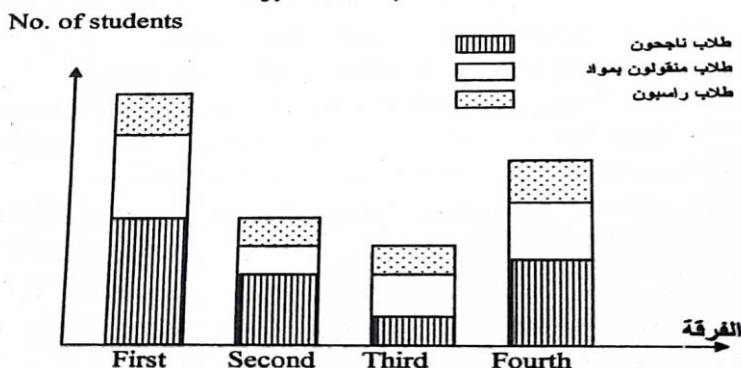
تمثيل بيانات نتيجة الفرق الدراسية الأربعة لإحدى كليات التجارة باستخدام الأعمدة الملتصقة .



حيث يتم تمثيل كل فرقة دراسية بثلاثة أعمدة ملتصقة تبين بالترتيب الطلاب الناجحون والمنقولون بمواد والراسبون أما ارتفاع 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 المجزأة وذلك على أساس نسبي متبعين في ذلك نفس الأسلوب السابق . حيث يكون في تلك الحالة ارتفاعات الأعمدة تتناسب في ارتفاعاتها مع نسبة تكرارات كل ظاهرة أو كل قسم من أقسام الظاهرة . ومن أهم مزايا هذا الأسلوب أنه يسهل عمليات المقارنة فيما بين أقسام الظاهرة في السنوات المختلفة وذلك على أساس نسبي وليس مطلق . ويكون ذلك بصفة خاصة إذا ما اختلفت مجاميع التكرارات كما سبق ذكره .

ملاحظات: عند استخدام الأعمدة البيانية كأسلوب من أساليب العرض

البياني فإنه يجب مراعاة ما يلي :-

1- ان تكون قواعد الأعمدة متساوية.

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

ولما كانت قواعد الاعمدة متساوية فإن مساحة الاعمدة تتناسب مع ارتفاعها والتي 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 باقي القيم فإنه يمكن كسر العمود نفسه في هذه الحالة مع كتابة قيمة البيان أعلى العمود كما سيرد في المثال التالي .

3- ان تكون المسافات فيما بين الاعمدة متساوية.

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

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

جدول (21-2)

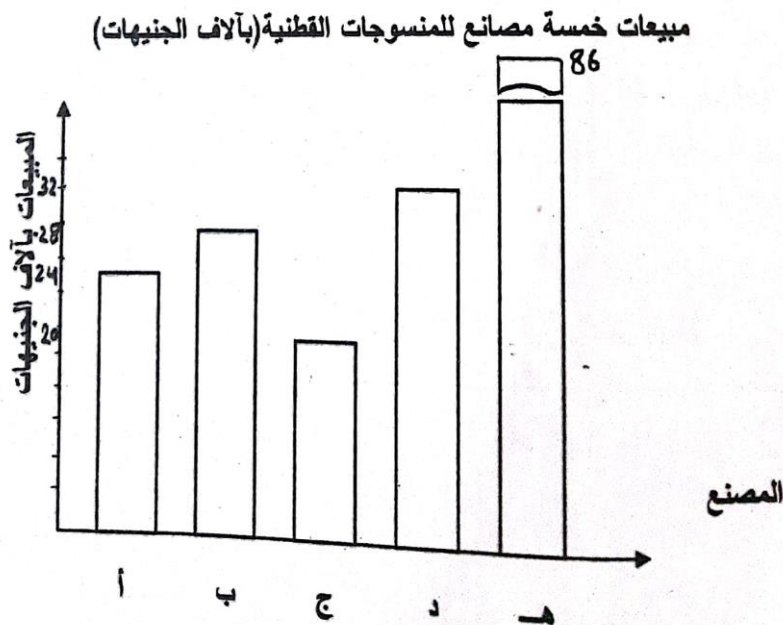
المصنع	أ	ب	ج	د	هـ
قيمة المبيعات	24	28	20	32	86

والمطلوب : عرض هذه البيانات باستخدام الاعمدة البيانية .

الحل : لاحظ ان قيمة المبيعات للمصنع (هـ) كبيرة جدا بالمقارنة بباقي 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 (3-2) .

شكل (3-2)



2- الرسم الدائرية :- pie charts

تعتبر الدوائر من الاشكال الهندسية الشائعة لتمثيل الظاهرة محل الدراسة بيانيا وبصفة خاصة إذا كانت تلك الظاهرة مقسمة الى مجاميع جزئية فإذا كانت لدينا ظاهرة واحدة فإنها تمثل بدائرة مناسبة دون أية قيود لتحديد نصف قطر تلك الدائرة . ثم يتم تقسيم تلك الدائرة او بمعنى زاوية مركز الدائرة الى قطاعات دائرية تتناسب مساحتها مع النسب المئوية للمكونات الجزئية للظاهرة. ولما كانت الزاوية المركزية للدائرة مساوية 360° فإن 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 إذا كنا نود إجراء المقارنة فيما بين ظاهرتين كل واحدة تحتوى على مجموعة من الأوجهه، ففي هذه الحالة يتم تخصيص دائرة لتمثيل كل ظاهرة على حده. ولكن يجب فى تلك الحالة رسم مساحة دائرتين متناسبتين مع قيمة الظاهرتين ، وحيث ان:

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

$$\text{اي ان :}$$

$$(ق1 / ق2) = (طنق1^2 / طنق2^2) = (نق1^2 / نق2^2)$$

حيث نق1 ، نق2 هما نصفى قطر الدائرة الاولى والثانية على الترتيب
وباعتبار الجذر التربيعى للعلاقة السابقة فيكون :-

$$\frac{\text{القيمة الاجمالية للظاهرة الاولى (ق1)}}{\text{القيمة الاجمالية للظاهرة الثانية (ق2)}} = \frac{1 \text{ نق}}{2 \text{ نق}}$$

فمثلا اذا كانت اجمالى قيم اقسام الظاهرة الاولى هو 1600 واجمالى اقسام الظاهرة الثانية هو 900 فان :-

$$\frac{1600}{4} = \frac{900}{3} \quad \text{1 نق}$$

$$\frac{1600}{3} = \frac{900}{4} \quad \text{2 نق}$$

فاذا افترضنا ان الدائرة الاولى نصف قطرها 6 سم فان:

$$\text{نصف قطر الدائرة الثانية} = 6 \times (\text{نق2} / \text{نق1}) = 6 \times (4/3)$$

$$= 4.5 \text{ سم}$$

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 :-

$$\text{زاوية اى قطاع} = (\text{قيمة الظاهرة فى هذا القطاع} / \text{مجموع قيم الظاهرة}) \times 360^\circ$$

مثال (9) : فيما يلى البيانات التالية تمثل المبالغ المنفقة (بالمليار جنيهه)

على القطاعات المختلفة فى احدى الدول عام 1994م :

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

والمطلوب استخدام الدوائر لتمثيل هذه البيانات بيانيا .

الحل:-

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

$$\text{الزاوية التى تمثل اى قطاع} = (\text{قيمة القطاع} / \text{اجمالى القطاعات}) \times 360^\circ$$

اي ان :

$$\text{الزاوية التى تمثل القطاع الصناعى} = 360 \times (48/20) = 150^\circ$$

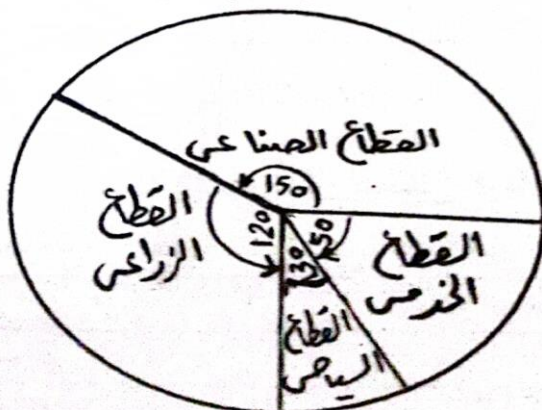
والزاوية التي تمثل القطاع الزراعي $120^\circ = 360 \times (48 / 16) =$

والزاوية التي تمثل القطاع السياحي $30^\circ = 360 \times (48 / 4) =$

والزاوية التي تمثل القطاع الخدمي $60^\circ = 360 \times (48 / 8) =$

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

شكل (4-2)



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

جدول (2-22)

اعداد الطلاب ذكورا واناثا في احدى الكليات

خلال السنوات 1960-1965

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 بنوعيههم بيانيا باستخدام الدوائر.

الحل :

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

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

$$\begin{array}{l} \text{نق 1} \quad \text{الجذر} \quad \text{مجموع الظاهرة خلال عام 1960} \\ \text{---} = \text{التربيعى} \quad (\text{-----}) \\ \text{نق 2} \quad \text{للمقدار} \quad \text{مجموع الظاهرة خلال عام 1965} \\ \qquad \qquad \qquad 210 \quad 90+120 \\ 0.65 = \frac{\quad}{495} = \frac{\quad}{175+320} = \end{array}$$

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 للقيمة 0.65 فمثلا يمكن ضرب كلا من نق 1 ، نق 2 فى اى مقدار ثابت وليكن مثلا (5) فنجد ان:

$$\text{نق 1} = 0.65 \times 5 = 3.25 \text{ سم} , \text{ نق 2} = 1 \times 5 = 5 \text{ سم}$$

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

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

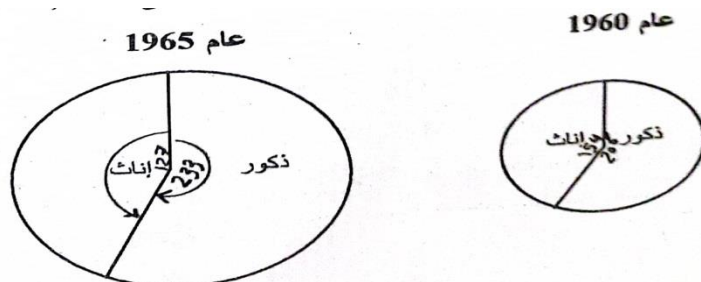
وبذلك يمكن رسم الدائرتين مع تحديد القطاعات داخل كل دائرة كما يوضحه الشكل (2-5).

شكل (2-5)

توزيع الطلاب (ذكورا، وإناثا) في احدى الكليات في عامي 1960, 1965

عام 1965

عام 1960



واخيرا فإن عيوب استخدام الدوائر في التمثيل البياني يمكن تحديده فيما يلي :

- تحتاج لحسابات ليست بالبسيطة .
- المقارنة فيما بين قيم الظاهر في الاقسام المختلفة يكون اصعب في حالة الدوائر منه في حالة الاعمدة ابيانية.

إذا كان عدد اقسام الظاهرة كبيرا فإنه من الافضل عدم استخدام الدائرة وذلك لصعوبة تمييز كل قطاع فى هذه الحالة ويفضل هنا استخدام الاعمدة البيانية

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

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

بصفة عامة فإن حالة البيانات المبوبة (الجداول التكرارية) للظاهرة التى تم التعبير عنها بمتغير كمي 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 مساويا لقيمة قيمة المتغير المتقطع .

مثال (11) :

فيما يلى الجدول التالى بين توزيع 20 اسرة حسب حجم الاسرة

جدول (2-23)

توزيع 20 اسرة حسب الحجم

Family size	1	2	3	4	5	6	Σ
Frequency (F _j)	2	3	7	4	3	1	20

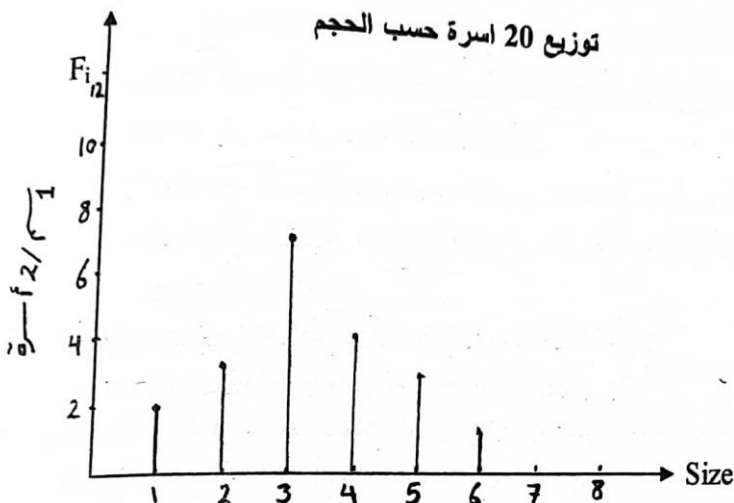
الحل:- يتم تمثيل حجم الاسرة على المحور الافقى والتكرارات على المحور الرأسى بمقياس رسم مناسب (1سم/وحدة تكرار). والشكل (2-6) يوضح التمثيل البيانى للجدول (2-23) :

الشكل (2-6)

توزيع 20 اسرة حسب الحجم

الشكل (2-6)

توزيع 20 أسرة 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 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 :

تكرار الفئة (i)

$$\text{التكرار المعدل للفئة } (MF_i) = \frac{F_i}{L_i}$$

طول الفئة i

اي ان : F_i

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

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

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

(1):المدرج التكرارى Histogram:

والمدرج التكرارى ماهو إلا مجموعة من المستطيلات المتلاصقة كل مستطيل يعبر عن فئة من فئات التوزيع ويتناسب ارتفاعه مع تكرار تلك الفئة.وفى تناولنا لرسم المدرج التكرارى كما سبق وإن ذكرنا يجب بداية ان نفرق فيما بين الجداول التكرارية المنتظمة والجداول التكرارية الغير منتظمة.

(أ):رسم المدرج التكرارى للجداول المنتظمة(متساوية اطوال الفئات):

وتتلخص الخطوات فى الاتى :-

-رسم المحورين الافقى والرأسى. بحيث يتم تمثيل فئات الجدول على المحور الافقى وتكرارات تلك الفئات على المحور الرأسى بمقياس رسم مناسب. هذا ويراعى عند رسم المحور الافقى انه ليس من الضرورى ان يبدأ هذا المحور فى تدريجه من الصفر ولكن يمكن أن يبدأ من فئة سابقة

عن الفئة الاولى فى الجدول وبنفس طول هذه الفئة الاولى . كما يراعى ان يكون هناك مسافة بعد الفئة الاخيرة فى الجدول، اما فيما يتعلق بالمحور الرأسى فأن تدريجه يجب ان يبدأ من الصفر على ان يتم اختيار مقياس الرسم المناسب .

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

والمثال التالى يوضح كيفية رسم المدرج التكرارى لجدول تكرارى منتظم .

مثال(12): فيما يلى لديك التوزيع التكرارى التالى لأجور 50عاملا

جدول(23-2)

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

والمطلوب تمثيل هذا التوزيع بيانيا باستخدام الشكل البيانى الملائم .

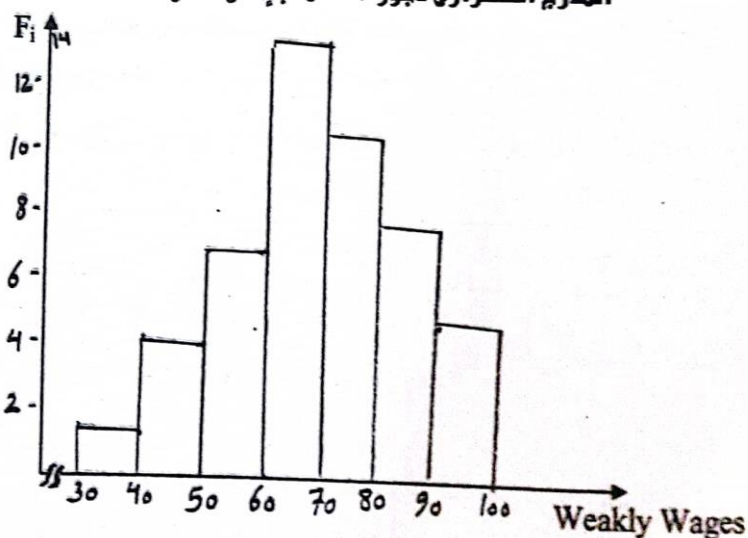
الـحـل:-

بالنظر لطبيعة الظاهرة التى يوضحها التوزيع التكرارى فهى عبارة عن اجور مجموعة من العمال وهى ظاهرة كمية متصلة لذا فإن الشكل البيانى

الملائم هو اما المدرج أو المضلع أو المنحنى التكرارى . فدعنا نقوم بالتمثيل البيانى للجدول السابق من خلال رسم المدرج التكرارى .
 وحيث ان : الجدول التكرارى فئاته متساوية من حيث الطول لذا يتم رسم المدرج التكرارى مباشرة . حيث يتم تمثيل فئات الاجر على المحور الافقى وعدد العمال على المحور الرأسى بمقياس رسم وليكن 1سم/2عامل كما هو موضح فى الشكل (7-2) .

شكل (7-2)

المدرج التكرارى لأجور 50 عاملا بإحدى الشركات



هذا ويمكن أيضا تمثيل تلك البيانات باستخدام التكرارات النسبية بدلا من التكرارات المطلقة. وفى تلك الحالة فإن المحور الرأسى يمثل التكرار النسبى 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:

$$MF_i = F_i / L_i$$

where

$$i = 1, 2, \dots, r$$

وبعد حساب التكرارات المعدلة MF_i يتم اتباع نفس خطوات الرسم
البياني مع فارق وحيد وهو انة 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	-100	Σ
							80	
No.of students	15	70	100	120	200	300	160	965

والمطلوب رسم المدرج التكرارى لتوزيع الدرجات .

الجل :-

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)

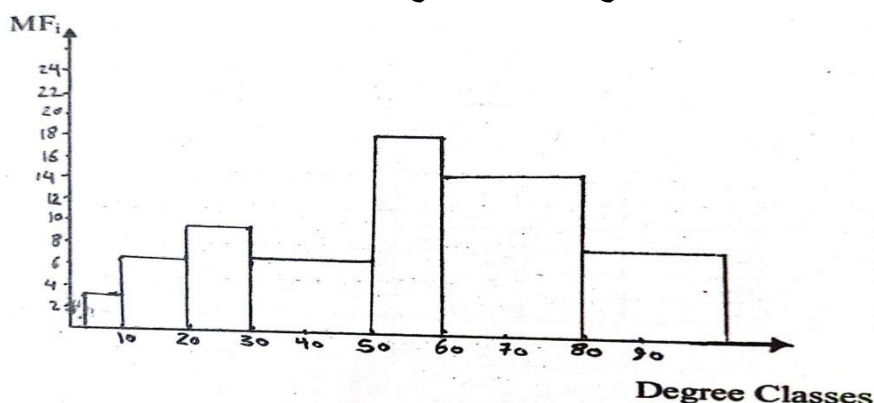
حساب التكرار المعدل (MF_i) لتوزيع الدرجات

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 هو تحقيق التناسب بين مساحات المستطيلات المكونة للمدرج التكرارى وارتفاعاتها . أى تحقيق التناسب ما بين مساحة المستطيل والتكرار الذى يمثله هذا المستطيل. هذا ويرسم الفئات كما سبق على المحور الأفقى بمقياس رسم مناسب وكذا التكرارات المعدلة بمقياس رسم 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) مايلي:-

-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 الاصلى = 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.

(2):المضلع التكرارى : Frequency polygon

وهو وسيلة اخرى لتمثيل بيانات التوزيعات التكرارية لمتغيرات او ظواهر كمية متصلة. وهو عبارة عن خط منكسر يصل ما بين النقاط التى يتم تمثيلها بيانيا من نقاط تلاقى منتصفات أو مراكز الفئات المقابلة للتكرارات الاصلية (فى حالة الجداول التكرارية المنتظمة)أو التكرارات المعدلة(فى حالة الجداول الغير منتظمة) وبصفة عامة فإن مركز الفئة (Mid Point or

X_i) يمكن استنتاجها من خلال احد الزوايا التالية:-

الحد الأدنى للفئة + الحد الأعلى للفئة

مركز الفئة =

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 :

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

Where : $i = 1, 2, 3, \dots, r$

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 أولاً. فإذا كان التوزيع التكرارى منتظما يتم حساب مراكز الفئات والرسم مباشرة حيث يتم مراكز الفئات على المحور الافقى والتكرارات الاصلى 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 نحصل على المضلع التكرارى ويجب فى تلك الحالة إقفال الشكل البيانى، وذلك من خلال افتراض وجود فئة وهمية أو افتراضية تسبق الفئة الاولى وكذلك فئة وهمية تلى الفئة الاخيرة ويتم تحديد مراكز تلك الفئات الوهمية بنقاط على المحور الافقى والمثال التالى يوضح كيفية رسم المضلع التكرارى.

مثال(14): الجدول التالى يوضح توزيع اجور مجموعة من العاملين

الشهرية بإحدى الشركات.

جدول (2-25)

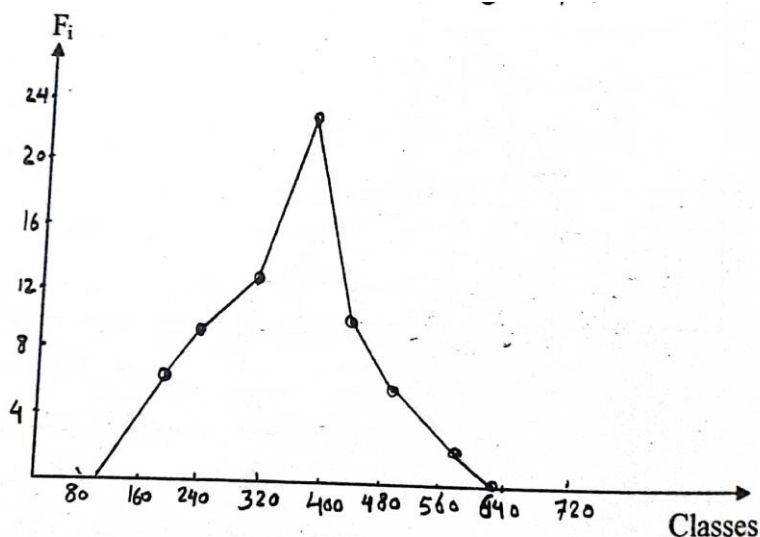
Classes	150-	220-	270-	360-	430-	500-	570- 640
Fi	6	10	18	25	9	6	2

الـحل :

لاحظ ان بيانات التوزيع التكرارى الموضحة فى الجدول (2-23) عبارة عن جدول تكرارى منتظم لذا يجب رسم المدرج مباشرة . حيث يتم تمثيل الفئات على المحور الافقى . كما يتم بيان التكرارات الاصلية على المحور الراسى بمقياس رسم مناسب كما هو موضح بالشكل (2-9).

شكل (2-9)

رسم المضلع التكرارى للأجر الشهري للعاملين

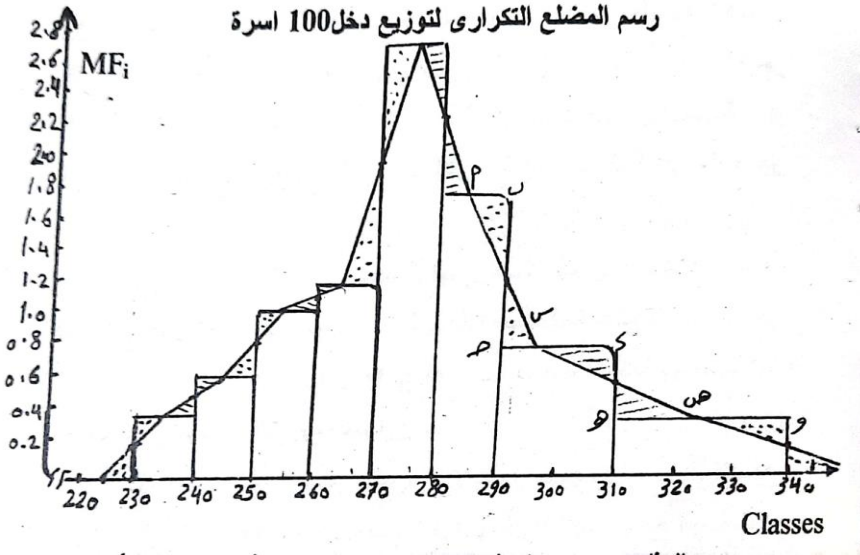


هذا ويمكن الحصول على المضلع التكرارى من المدرج التكرارى كما سبق وان ذكرنا من خلال تنصيف قمم المستطيلات (القواعد العليا للمستطيلات) وتوصيل هذه المنتصفات بخطوط مستقيمة مع إقفال الشكل من البداية

والنهاية وذلك لأن منتصفات قمم المستطيلات ما هي إلا مراكز فئات التوزيع التكرارى. ويمكن توضيح ذلك من خلال الرسم الموضح فى شكل (2-10) حيث تم رسم المدرج التكرارى أولاً ثم تنصيف قمم مستطيلاته وتوصيلها بالخطوط المستقيمة.

شكل (2-10)

رسم المدرج والمضلع التكرارى للاجر الشهرى للعاملين



ومن الشكل (2-10) يمكن ملاحظة مايلى:-

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

- المساحة تحت المضلع 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 لا تتحقق اى لا يتحقق تساوى المساحة تحت المضلع مع المساحة تحت المدرج إلا اذا تم اقفال شكل المضلع التكرارى وذلك من خلال اضافة الفئتين قبل الاولى وبعد الاخيرة ولعل ذلك يبرر اضافة هاتين الفئتين .

وأخيرا ليس بالضرورة رسم المدرج التكرارى قبل المضلع فى حالة الجداول المنتظمة حيث يمكن رسم المضلع مباشرة دون رسم المدرج كما سبق وان اوضحناه فى الشكل (2-9).

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

مثال (15):

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.

الـحـل:-

لاحظ أن التوزيع التكرارى لدخل هؤلاء الاسر غير منتظم. لذا فإنه لرسم المضلع التكرارى يجب الحصول على التكرارات المعدلة اولا ثم رسم المدرج

التكرارى ثانيا ويلى ذلك رسم المضلع التكرارى وذلك على النحو المبين بالجدول (2-27) التالى :

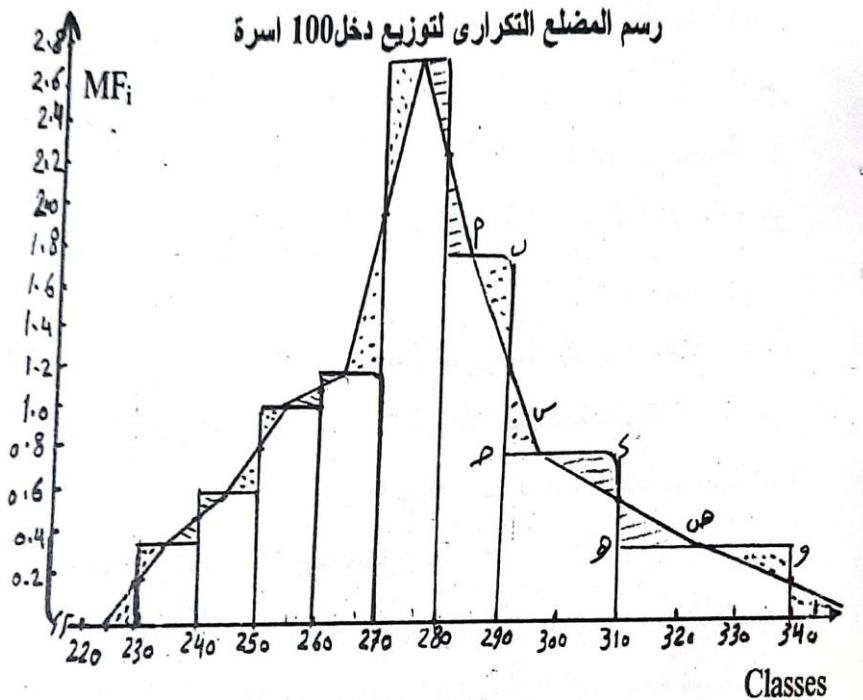
جدول (2-27)

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 باستخدام المدرج التكرارى الموضح فى هذا الشكل فإنه ليست هناك أية تحفظات على توصيل منتصفات قمم المستطيلات الممثلة للفئات الستة الاولى بخطوط مستقيمة حيث ان هذه الفئات متساوية فى الطول (حيث أن طول كل منها 10 جنيهه) وأما فيما يتعلق بالفئة السابعة (290 وحتى اقل من 310 جنيهه) 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 فأنا نرسم الخط المستقيم من النقطة (أ) مارا بمنتصف الارتفاع ب ج (النقص فى الارتفاع) الى أن يقطع قمة المستطيل الممثل للفئة (290 حتى اقل من 310 جنيهه) وليكن فى النقطة (س) وذلك حتى يتحقق أن يكون المثلث المستبعد يساوى فى مساحته المثلث المضاف.

شكل (2-11)

رسم المضلع التكرارى لتوزيع دخل 100 اسرة



وبالمثل يتم رسم الخط المستقيم بدءاً من النقطة (س) الممثلة أخيراً بمنتصف الارتفاع أو الضلع (د هـ) 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 . وأخيراً يتم رسم خط مستقيم بدا من النقطة (ص) وماراً بمنتصف الارتفاع (و ز) الى أن يقطع المحور الأفقى . وبذلك يتم إغلاق الشكل تحت المضلع التكرارى . وفى حقيقة الامر فإن هذا الاجراء يضمن ان تكون المساحة تحت المضلع التكرارى مساوية تماماً للمساحة تحت المدرج التكرارى .

(3): المنحنى التكرارى Frequency curve

إذا تم تمهيد نقاط المضلع التكرارى باليد فإننا نحصل على شكل منحنى يسمى بالمنحنى التكرارى ولذلك فإنه عند رسم المنحنى التكرارى يراعى نفس الشروط المتبعة سابقا عند رسم المدرج أو المضلع التكرارى لتوزيع تكرارى غير منتظم حيث يلزم تعديل التكرارات وتؤخذ مراكز الفئات على المحور الافقى والتكرارات الاصلية فى حالة التوزيعات التكرارية المنتظمة او التكرارات المعدلة فى حالة التوزيعات التكرارية الغير منتظمة على المحور الرأسى . وبصفة عامة إذا كان تمهيد المنحنى جيدا فإن المساحة تحت المنحنى التكرارى سوف تساوى المساحة تحت المضلع التكرارى وكذا تساوى المساحة تحت المدرج التكرارى والتي تعبر هذه المساحات كما بينا فيما قبل عن مجموع تكرارات التوزيع التكرارى .

هذا ونتيجة لأن المنحنى التكرارى لا يمر بجميع نقاط المضلع التكرارى فإن المساحة تحت المنحنى التكرارى قد لا تتساوى مع المساحة تحت المضلع التكرارى . وفى واقع الامر فإنه كلما كانت أطوال فئات التوزيع التكرارى قصيره وفى نفس الوقت زاد حجم البيانات فإن 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 مساحة المضلع التكرارى من مساحة المنحنى التكرارى .

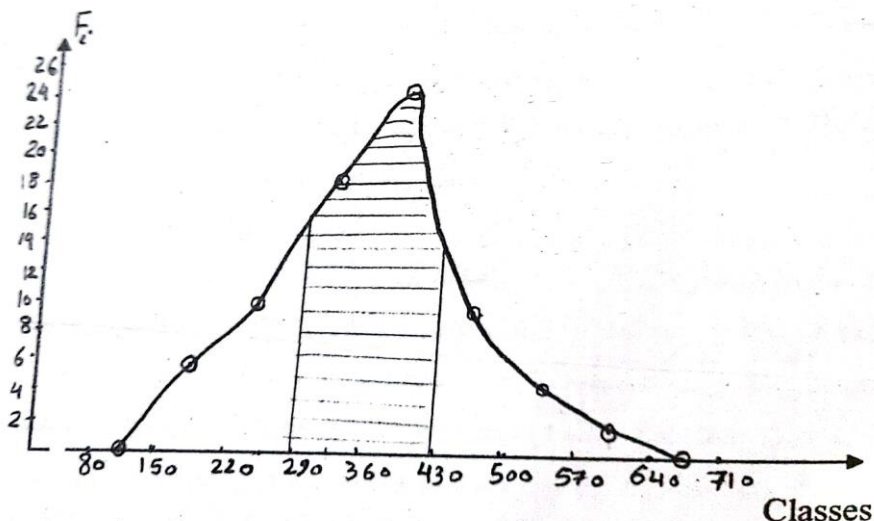
مثال (16) :

ارسم المنحنى التكرارى لتوزيع أجور العاملين بإحدى الشركات الوارد بياناتها بالجدول (25-2) . وهى بيانات لتوزيع تكرارى منتظم . لذا يتم رسم المنحنى التكرارى مباشرة من خلال تحديد نقاط تلاقى مراكز فئات التوزيع ومايقابلها من تكرارات أصلية حيث يتم تحديد الفئات او مراكز الفئات على المحور

الافقى والتكرارات على المحور الرأسى . ويتم تحديد نقاط المنحنى التكرارى
كما هو موضح على الرسم المبين فى شكل (2-12)

شكل (2-12)

المنحنى التكرارى لتوزيع الاجر الشهرى لمجموعة من العاملين بإحدى
الشركات



وفى المنحنى التكرارى لاحظ ان عدد المشاهدات (التكرارات) 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 ما بين 240-290 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.

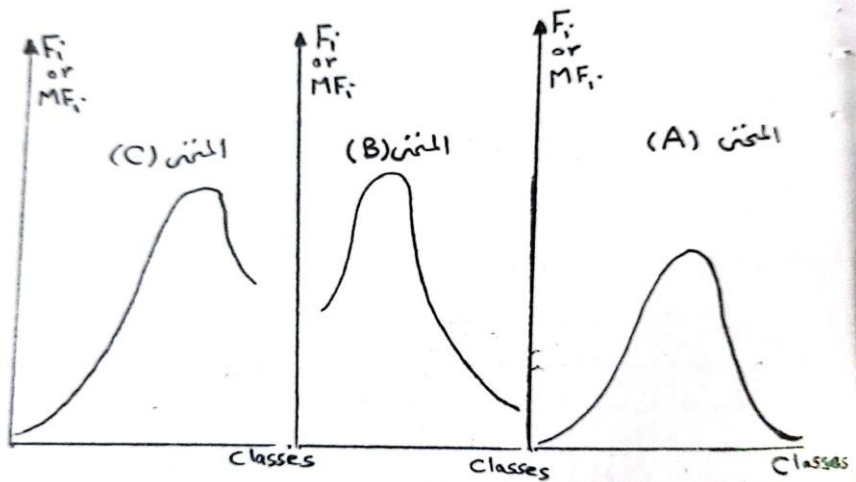
وفيما عدا ذلك فإن المقارنة يجب أن تجرى على أساس التوزيعات التكرارية النسبية . وتعتبر أشكال المنحنيات التكرارية على درجة عالية من الأهمية وذلك لأنها تستخدم في الحكم على مدى التواء التوزيع من عدمه (أي مدى قرب التوزيع من التماثل) كما تستخدم في التعرف على شكل قمة المنحنى (دراسة التفلطح) وهو ماسيرد في أبواب لاحقة بإذن الله.

أشكال المنحنيات التكرارية:

تناولنا فيما سبق كيفية رسم المنحنى التكرارى ورأينا ان شكل المنحنى يتوقف على التوزيع التكرارى الذى يمثله . كما يبرز أهمية دور المنحنيات التكرارية فى استخدامها فى المقارنة فيما بين التوزيعات المختلفة . هذا ومن المتوقع ان نجد عددا كبيرا من الاشكال المختلفة للمنحنيات التكرارية ومع ذلك فإننا سوف نتعرض هنا الى أكثر هذه الاشكال استخداما والموضحة فى شكل (2-13) مرجئين الحديث عن بعض الاشكال الأخرى حتى يم دراستنا لمقاييس الالتواء او التفلطح فيما بعد فى مقررنا الدراسى بمشيئة الله.

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) والتي تناظر على الترتيب المنحنيات التكرارية a,b,c الموضحة فى الشكل(2-13)

جدول(2-28)

توزيعات تكرارية مختلفة

Classes	10 -	22 -	34 -	46 -	58 -	70 -	82 -	Σ
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 تكرارات مختلفة بالشكل الذى يخدم الغرض الذى من أجله نود التفرقة فيما بين أنواع هذه التوزيعات الثلاث . حيث تمكننا التوزيعات الثلاث من تصنيف الاشكال المختلفة للمنحنيات التكرارية 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 النهاية العظمى للمنحنى يقسم المنحنى الى قسمين متطابقين تماما ويسمى المنحنى التكرارى فى هذه الحالة بالمنحنى المتماثل . حيث يكون تزايد او تناقص التكرارات متشابها ومنتظما بطريقة متماثلة على جانبي المحورالرأسى والذى يمر بمركز التوزيع . ويمكن ملاحظة ذلك فى الشكل (2-13) ومن الجدول(2-27) . وبعبارة اخرى فأن المنحنى التكرارى (أ) والموضح فى شكل (2-13) واذى يمثله التوزيع التكرارى (أ) أو A فى الجدول(2-27) .

ويمثل هذا التوزيع صورة للتوزيعات المتماثلة Symmetric Distribution .

(2): التوزيع أو المنحنى (ب) أو (B) :-

وفى هذا التوزيع تتزايد التكرارات أو تتناقص بشكل غير منتظم على جانبي المحور أو العمود المنشأ عند وسط التوزيع . ولذلك فهو توزيع غير متماثل تتزايد فيه التكرارات بسرعة حتى تصل الى القمة وبعدها تتناقص ببطء . ويطلق على هذه التوزيعات والتي يميل فيها التكرارات الكبيرة الى التركيز عند الفئات الدنيا (التى تقع فى بداية التوزيع) بالتوزيعات الموجبة الالتواء Positive Skewness Distributions. والمنحنى التكرارى الذى يمثل مثل هذه التوزيعات يسمى بالمنحنى غير المتماثل حيث يكون طرفه (ذيله) اليمين أطول من طرفه الايسر (انظر الشكل (2-13) المنحنى (ب)) والتوزيع التكرارى (ب) الموضح فى الجدول (2-27) يمثل هذا النوع من المنحنيات.

(3): التوزيع أو المنحنى (ج) أو (C) :-

وفى هذا التوزيع ايضا تتزايد التكرارات او تتناقص بشكل غير منتظم على جانبي المحور المنشأ عند وسط التوزيع وبالتالي فهو توزيع غير متماثل . ووفقا لهذا النوع من التوزيعات فإن التكرارات تتزايد ببطء حتى تصل الى القمة ثم بعدها تأخذ فى التناقص بسرعة. ومثل هذا التوزيع والذي تميل فيه التكرارات الكبيرة الى التركيز عند الفئات العليا (الأخيرة) يسمى بالتوزيع سالب الالتواء (Negative Skewness Distribution). والمنحنى التكرارى الممثل لهذا التوزيع هو منحنى غير متماثل حيث يكون الطرف (الذيل) الايسر فيه أطول من الطرف اليمين (أنظر الشكل (2-13)

والمنحنى (ج). ويمثل هذا النوع من المنحنيات التوزيع التكرارى (ج) الوارد فى جدول (2-27) .

مثال (17):

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

جدول (2-29)

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

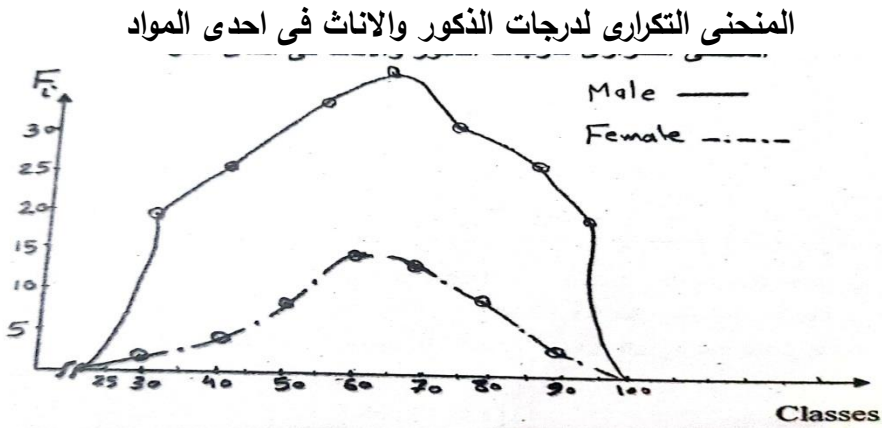
Classes	-25	-35	-45	-55	-65	-75	-85	Σ
No.of Males	20	26	35	43	30	27	19	200
No.of Femals	2	4	7	13	12	9	3	50

والمطلوب:- المقارنة ما بين التوزيعين التكرارين للذكور والاناث وذلك باستخدام الشكل البيانى الملائم.

الجل:- 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 ومنظمة لذا يمكن رسم المدرج أو المنحنى أو المضلع التكرارى مباشرة (لكن فى حالة الجداول الغير منتظمة فلا بد من الحصول على التكرارات المعدلة أولا)والآن دعنا نقوم بالمقارنة من خلال المنحنى التكرارى . والان عزيزى القارئ لاحظ أن مجموع التكرارات فى حالة الذكور مختلفا عن مجموع التكرارات فى حالة الاناث ، فإذا تم إجراء المقارنة فيما بين التوزيعين مباشرة دون الحصول على التكرارات النسبية، سيكون النتيجة

مقارنة غير واضحة وغير موضوعية نظرا لاختلاف مجموع التكرارات .
والشكل (14-2) يوضح رسم المنحنى التكرارى للذكور والاناث مباشرة.

شكل (14-2)



فيتضح من شكل (14-2) ان المقارنة بين التوزيعين غير واضحة إذ أن المساحة تحت كلا من المنحنيين تختلف عن الاخرى نظرا لاختلاف مجموع التكرارات فى كلا من التوزيعين عن الآخر. لذا لى تكون المقارنة موضوعية وعلى اساس إحصائى سليم فأن المساحة تحت المنحنى الممثل لدرجات الذكور يجب أن تتساوى مع المساحة تحت المنحنى الممثل لدرجات الاناث ولى يتحقق ذلك لابد من ايجاد التكرارية 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.

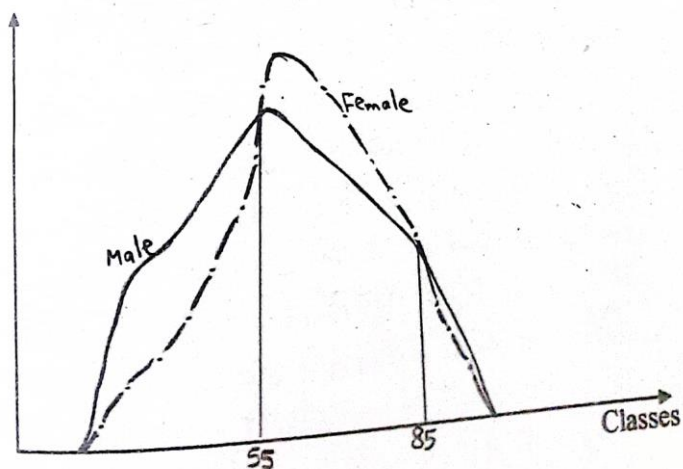
جدول (2-30)

Score Classes	Frequencies		Relative Frequencies %	
	Male	Female	RF _M	RF _F
25 -	20	2	10	4
35 -	26	4	13	8
45 -	35	7	17.5	14
55 -	43	13	21.5	26
65 -	30	12	15	24
75 -	27	9	13.5	18
85 -	19	3	9.5	6
Σ	200	50	100	100

ومن بيانات الجدول (2-30) يمكننا رسم المنحنى الذى يمثل التوزيع التكرارى النسبى لكلا من درجات الذكور والإناث كلا على حده وذلك كما هو موضح بالشكل (2-15) حيث يبين هذا الشكل منحنى التكرارات النسبية لكلا من الذكور والإناث 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 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 باستخدام المدرج أو المضلع أو المنحنى وإن كان يفضل استخدام المضلع أو المنحنى التكرارى لكى تتضح عملية المقارنة سواء باستخدام التكرارات الاصلية (فى حالة تساوى مجموع التكرارات) أو باستخدام التكرارات المعدلة (فى حالة عدم تساوى مجموع التكرارات). حيث يتم تحديد نقاط تقاطع المضلعات أو المنحنيات التكرارية ويتم إسقاط من تلك النقط اعمدة على المحور الأفقى ثم تتم عملية المقارنة فيما بعد ذلك كما سبق فى الشكل (2-15).

المنحنيات التكرارية المتجمعة:- Cumulative Frequency Curves
بصفة عامة تستخدم المنحنيات التكرارية المتجمعة لتمثيل الجداول التكرارية المتجمعة (الصاعدة أو الهابطة) بيانيا وذلك لاستخدامها فى تحديد عدد التكرارات التى تقابل اقل من قيمة ما فى الفئات أو فيما بينها او العكس صحيح.

ولما كان هناك نوعان من الجداول التكرارية المتجمعة (الجداول التكرارية المتجمعة الصاعدة والجداول التكرارية المتجمعة الهابطة) لذا فإنه من الطبيعى ان يكون هناك نوعين للمنحنيات التكرارية المتجمعة وهى:-

*المنحنى التكرارى المتجمع الصاعد Ascending Cumulative Curve

*المنحنى التكرارى المتجمع الهابط Descending Cumulative Curve

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

ولرسم المنحنى التكرارى المتجمع الصاعد فأننا نختار المحور الرأسى لتمثيل التكرارات المتجمعة الصاعدة بينما يتم تمثيل الفئات على المحور الأفقى . حيث يتم تمثيل كل فئة ومايقابلها من تكرار متجمع بنقطة إحداثياها هما الحد الاعلى للفئة والتكرار المتجمع الصاعد لهذه الفئة . ثم نقوم بتوصيل هذه النقاط بخط منحنى ممهد باليد فنحصل بذلك على المنحنى التكرارى المتجمع الصاعد. اما المنحنى التكرارى المتجمع الهابط فيستخدم لرسم الجدول التكرارى المتجمع الهابط بيانيا كما يفيد فى حساب عدد التكرارات(اونسبة التكرارات) التى تبلغ قيمة مافى الفئات (أو بين فئتين)فأكثر او يفيد فى تحديد ماهى القيمة التى يبلغها عددا ما من التكرارات (أونسبة التكرارات)فأكثر .وبأسلوب مماثل يمكن رسم المنحنى التكرارى المتجمع الهابط حيث يتم تمثيل التكرارات المتجمعة الهابطة على المحور الرأسى بينما يمثل المحور الأفقى فئات التوزيع التكرارى المتجمع 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 بذلك على المنحنى التكرارى المتجمع الهابط والمثال التالى يوضح أهمية المنحنيات التكرارية المتجمعة.

مثال(18-2) :- فيما يلى لديك التوزيع التكرارى التالى بين توزيع الاجر الاسبوعى لخمسين عاملا بإحدى الشركات :

جدول(31-2)

Classes	-30	-40	-50	-60	-70	-80	-90	Σ
frequencies	1	4	7	14	11	8	5	50

والمطلوب :-

- 1- رسم المنحنى التكرارى المتجمع الصاعد .
- 2- رسم المنحنى التكرارى المتجمع الهابط .
- 3- باستخدام المنحنى التكرارى التجمع الملائم حدد مايلى:-
 - أ- ماهو قيمة الاجر الاسبوعى الذى يتقاضى اقل من 35 عاملا.
 - ب- اوجد قيمة الاجر الاسبوعى الذى يحصل عليه - كحد أدنى - ثلاثون عاملا .
 - ج- ماهو عدد العمال(وكذا ماهى نسبة العمال) الذين يتقاضون أجورا اسبوعية تقل عن 64 جنيها.
 - د- ماهو قيمة الاجر الاسبوعى الذى يتقاضى أقل منه :
 - 80% من إجمالى عدد العمال .
 - 52% من إجمالى عدد العمال.
 - هـ- 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 .

الجل :-

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

جدول (2-32)

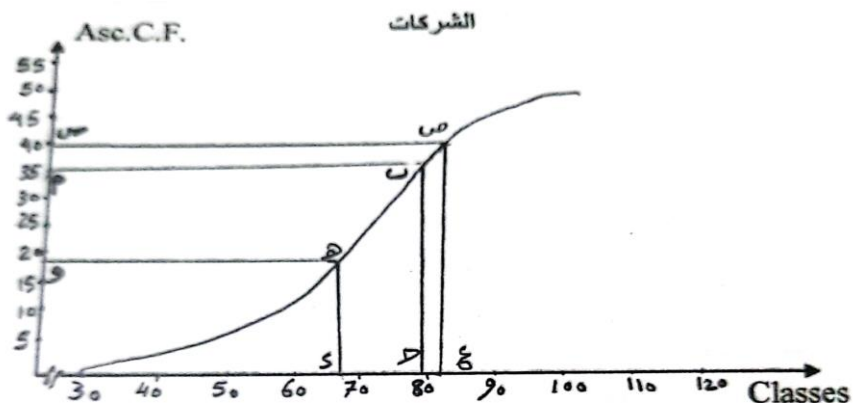
التوزيع التكرارى والتوزيع المتجموع الصاعد لأجور العمال الاسبوعية
فى إحدى الشركات:

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
Σ	50	Less than 100	50

ومن هذا الجدول يتم رسم العمودين الاخيرين لتمثيل المنحنى التكرارى المتجموع الصاعد. والشكل (2-16) يوضح المنحنى التكرارى المتجموع الصاعد للاجور الاسبوعية لمجموعة العاملين المعطاه .

شكل(2-16)

المنحنى التكرارى المتجموع الصاعد لأجور العمال الاسبوعية فى إحدى الشركات



(2): ولرسم المنحنى التكرارى المتجموع الهابط يجب أولاً ايجاد الجدول التكرارى المتجموع الهابط وذلك على النحو المبين فى الجدول التالى:

جدول (2-33)

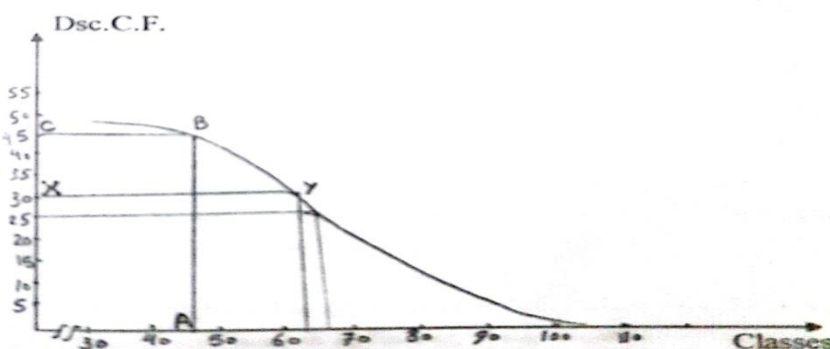
التوزيع التكرارى والتوزيع التكرارى المتجمع الهابط للأجور
الاسبوعية لمجموعة من العاملين بإحدى الشركات

Classes	F_i	More than the Class Lower Limit	Desc.Cum.Freq
30-	1	More than 30	50
40-	4	More than 40	49
50-	7	More than 50	45
60-	14	More than 60	38
70-	11	More than 70	24
80-	8	More than 80	13
90-	5	More than 90	5
Σ	50	More than 100	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 لأجور 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 الإشارة إليها عند دراستنا للجدول التكرارى الصاعدة والهابطه فإن:

أ- لتحديد قيمة الأجر الأسبوعي الذي 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) والتي تمثل عدد العمال فى صورة تكرارات متجمعة صاعدة. ومن هذه النقطة يتم رسم خطا عموديا على المحور الرأسى وموازيا للمحور الأفقى الى أن يقطع هذا الخط المنحنى التكرارى المتجمع الصاعد فى نقطة ولتكن النقطة (ب) فنسقط عمودا من نقطة إلتقاء هذا الخط الموازى للمحور الأفقى بالمنحنى المتجمع الصاعد ليقطع هذا العمود المحور الأفقى (الاجور) فى نقطة ولتكن النقطة (ج) فتكون هى النقطة التي توضح قيمة الأجر الذى يتقاضى أقل منه 35 عاملا. ومن النقطة (ج) على الرسم يتضح لنا أن قيمة الأجر الذى يتقاضى أقل منه 35 عاملا هو الأجر 87 جنيها تقريبا.

ب- لإيجاد قيمة الأجر الذى يتقاضاه كحد أدنى ثلاثون عاملا فيكون ذلك من خلال المنحنى التكرارى المتجمع الهابط لانه كما سبق وأن أوضحنا أن لفظ كحد أدنى يكافئ فى المعنى كلمة فأكثر. وهنا يتم تعيين موضع النقطة 30 عاملا على المحور الرأسى للمنحنى التكرارى المتجمع الهابط كما هو موضح فى شكل (2-17) ولتكن النقطة (X) والتي تمثل عدد العمال فى صورة تكرارات متجمعة هابطة. ومن هذه النقطة يتم رسم خطا عموديا على المحور الرأسى وموازيا للمحور الأفقى الى أن يقطع هذا الخط المنحنى التكرارى المتجمع الهابط فى نقطة ولتكن النقطة (Y) فنسقط عمودا من النقطة (Y) على المحور الأفقى والذي يمثل الحدود الدنيا لفئات التوزيع التكرارى المتجمع الهابط الى ان يقطع هذا العمود المحور الأفقى فى نقطة ولتكن (Z) فتكون النقطة (Z) هى عبارة عن قيمة الأجر الذى يتقاضاه كحد أدنى ثلاثون عاملا. ومن النقطة (Z) على الرسم الموضح فى شكل (2-17) يتضح لنا أن قيمة الأجر الذى يتقاضاه كحد أدنى ثلاثون عاملا هو الأجر 66 جنيها تقريبا.

ج- أما لتحديد عدد العمال (وكذا نسبة العمال) الذين يتقاضون أجرا أسبوعيا تقل عن 64 جنيه . فيتم استنتاج المطلوب باستخدام المنحنى التكرارى المتجمع الصاعد ، حيث يتم تحديد موضع الاجر 64 جنيه بنقطة على المحور الافقى لفئات التوزيع التكرارى المتجمع الصاعد الموضحة فى الشكل (2-16) ولتكن النقطة (د) ومن النقطة (د) يتم إنشاء عمودا على المحور الافقى وموازيا للمحور الرأسى الى أن يقطع هذا العمود المنحنى التكرارى المتجمع الصاعد فى نقطة ولتكن النقطة (هـ) . ومن النقطة (هـ) يتم رسم خطا عموديا موازيا للمحور الافقى ويمتد على استقامته الى أن يقطع المحور الرأسى والذى يعبر عن التكرارات المتجمعة الصاعدة فى نقطة ولتكن النقطة (و) وهى تعتبر بمثابة الاجر 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 ثم فإن :-

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

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

د- ولتحديد قيمة الاجر الذى يتقاضى أقل منه 80% من إجمالى عدد العمال فحيث أن النسبة 80% من إجمالى العمال هى:

$$50 \times (80 / 100) = 40 \text{ عاملا. لذا فإنه يمكن استنتاج قيمة الاجر}$$

الاسبوعى الذى يتقاضى أقل منه 40 عاملا باستخدام المنحنى التكرارى المتجمع الصاعد بنفس آلية المطلوب الاول رقم (أ) من هذا المثال وذلك من خلال الشكل 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 للمحور الأفقى الى أن يقطع المنحنى التكرارى المتجمع الصاعد فى نقطة ولتكن النقطة (ص) والتي يتم منها إسقاط عمودا

على المحور الأفقى وموازيا للمحور الرأسى الى أن يقطع المحور الأفقى فى نقطة ولتكن النقطة (ع) . ومنها فإن قيمة الاجر الذى يتقاضاها اسبوعيا أقل من 40 عاملا (80% من إجمالى العمال) هو الاجر 84 جنيه تقريبا .

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

$$= 50 \times (100 / 52) = 26 \text{ عاملا .}$$

لذا فإنه يمكن استنتاج قيمة الاجر الذى يتقاضى أقل منه 26 عاملا 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 عدد العمال) هو 70 جنيه تقريبا .

ملحوظة:-

لاحظ فى الجدول التكرارى المتجمع الصاعد ان عدد العمال 26 يشكل أحد التكرارات المتجمعة الصاعدة يقابله الفئة 70جنيها كحد أقصى للأجر الذى يتقاضاه عدد 26 عاملا.

ه- لتحديد عدد العمال (وكذا النسبة) الذين يتقاضون أجورا اسبوعية تبلغ 45 جنيها على الأقل (وهذه العبارة تعادل كلمة فأكثر) فإن ذلك يكون من خلال المنحنى التكرارى المتجمع الهابط . حيث يتم تحديد هذا المطلوب بنفس الاسلوب الذى تم به إجابة المطلوب (ب) من نفس المثال . حيث يتم تحديد موضع الاجر 45 جنيها بنقطة على المحور الأفقى للمنحنى التكرارى الهابط الموضح فى الشكل (2-17) ولتكن النقطة (A) ثم يتم إنشاء عمودا

من النقطة (A) على المحور الأفقى وموازيا للمحور الرأسى الى أن يقطع هذا العمود المنحنى التكرارى المتجمع الهابط فى نقطة ولتكن (B) ومن النقطة (B) يتم رسم خطا موازيا للمحور الأفقى فى اتجاه المحور الرأسى الى أن يقطع المحور الرأسى فى نقطة ولتكن النقطة (C) . فتكون النقطة (C) هى بمثابة عدد العمال الذين يتقاضون على الأقل أجرا اسبوعيا يعادل 45 جنيها . ومن الرسم الموضح فى الشكل (2-17) يتضح لنا أن عدد العمال المطلوب هو 47 عاملا تقريبا ومن ثم فإن :-

$$\text{النسبة المطلوبة} = (50 / 47) \times 100 = 94 \% .$$

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

$$50\% \text{ من اجمالى عدد العمال} = 50 \times (100 / 50) = 25 \text{ عاملا .}$$

ومن ثم فإن المطلوب الان هو تحديد قيمة الاجر الذى يتقاضاه كحد أدنى (أى القيمة فأكثر) 25 عاملا . وبنفس الاسلوب الذى تم به اجابة المطلوب (ب) من نفس المثال فإنه من خلال الشكل (2-17) يتضح لنا أن قيمة الاجر الذى يتقاضاه كحد أدنى 50% من عدد العمال (أى 25 عاملا) هو الاجر 69 جنيها تقريبا .

4-مخطط الساق والورقة :- Stem-and-Leaf

يعتبر مخطط الجذع (أو الساق) والورقة من المخططات البيانية المستحدثة فى مجال العرض البيانى للظواهر الكمية المتصلة . وفى هذا الشكل أو هذا المخطط البيانى يتم تقسيم كل مفردة من مفردات الظاهرة أو المتغير محل

الدراسة الى جزئين أساسيين 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 والمنحنى والمضلع التكرارى . والفارق الوحيد بين هذا المخطط البيانى والرسوم السابقة (مدرج- مضلع- منحنى تكرارى) فى أن التكرارات فى حالة المدرج والمضلع والمنحنى التكرارى تمثل بأعمدة بيانية وينتهى دور المفردات الحقيقية فى هذه الحالة ، أما فى حالة مخطط الجذع والورقة فيتم تمثيل القيم الحقيقية للمفردات . وذلك فأن مخطط الجذع والورقة يعكس معلومات عن طبيعة القيم الموجودة أو الفعلية للظاهرة محل الدراسة ذاتها وذلك على عكس المضلع والمنحنى التكرارى والذى يفترض ان تكرارات كل فئة تأخذ قيمة مركز الفئة . والامثلة التالية توضح كيفية إنشاء هذا المخطط البيانى .

مثال(19):-

إذا كانت لديك مجموعة المفردات التالية والتي تعبر عن الدخل اليومي لمجموعة مكونة من 10 أشخاص :-

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

5 ، 7 ، 12

والمطلوب تمثيل مخطط الجذع أو (الساق) والورقة .

الحل:-

لتمثيل مخطط الجذع أو الساق والورقة فإننا نقسم المفردات التى تعبر عن الدخل أو الاجر اليومي لمجموعة المفردات الى جزئين الاول يعبر عن

الجذع أو (الساق) Stem والذي يمثل خانة العشرات والثاني يعبر عن الورق Leaf والذي يمثل خانة الآحاد وكأنه تم تقسيم المتغير الى أقسام أو فئات طول كل منها 10 جنيه فيعطى الشكل البياني التالي (شكل 2-18).

شكل (2-18)

مخطط الجذع(الساق) والاوراق للدخل اليومي 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 في الجانب الايسر . حيث ان هناك مفردتين تبلغ دخولهم اليومية أقل من عشرة جنيهات . وهناك سبعة مفردات تبلغ دخولهم ما بين العشرة جنيهات وأقل من عشرون جنيها . وهناك سبعة مفردات دخولهم ما بين 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 ان هناك مفردتين في 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 جنيهاً 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 تنحصر ما بين عشرة جنيهاً وأقل من عشرون جنيهاً .

- بالنسبة لمفردات الصف الثالث الذى يقع أسفل الجذع والورقة فهناك سبعة مفردات كل منهم يحتوى على عشرون جنيهاً أى اثنين فى خانة العشرات (وهى تعادل $2 \times 10 = 20$) هذا بالاضافة لأرقام الآحاد التى يوضحها الورق . وهذه الأرقام 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 ما بين العشرون جنيهاً وأقل من ثلاثون جنيهاً . 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 ما بين أكبر قيمة وأصغر قيمة فى المفردات .

مثال(20):-

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

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

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 هى 99 سم . فإذا اعتبرنا طول
الفئة 10سم فيكون المخطط الخاص بالجذع (أو الساق) والورقة يبدأ 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 الشكل البيانى التالى:-

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

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

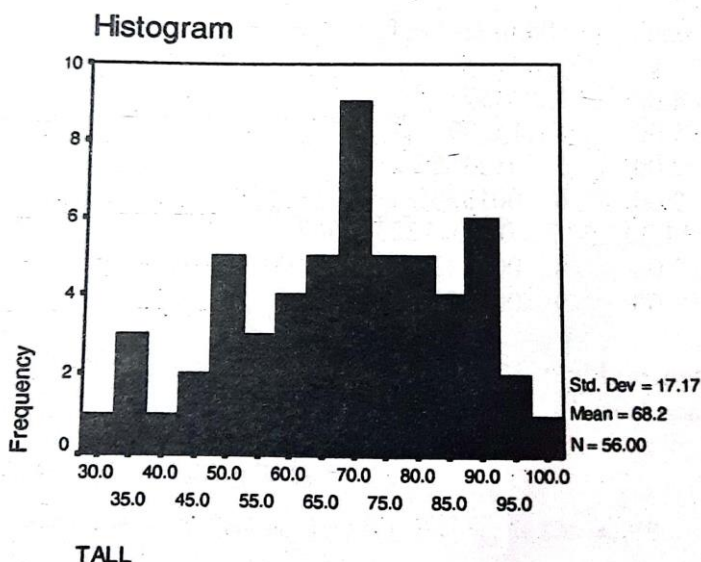
Each Leaf : 1 Case (s)

ومن خلال شكل (2-19) يتضح لنا أن :-

- أصغر قيمة في البيانات هي (32) وأكبر قيمة في البيانات هي (99) .
- طول الفئة في هذا المخطط هو 10 .
- عدد المفردات والتي يمثلها مجموع التكرارات هو 56 مفردة وهو بمثابة حجم العينة المعطى في 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 النباتات 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 هي 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 البياني :- Line Chart

حيث تبين خريطة الخط البياني مدى تطور ظاهرة ما خلال فترات زمنية معينة . حيث يتم تمثيل عنصر الزمن على المحور الأفقى أما قيم الظاهرة محل الدراسة فيتم تمثيلها على المحور الراسى بمقياس رسم يتناسب مع قيم الظاهرة . وبناءا على خريطة الخط البياني يتم تمثيل قيمة الظاهرة محل الدراسة عند كل نقطة زمنية بنقطة إحداثياها هما النقطة الزمنية وقيمة الظاهرة عند تلك النقطة . وبعد الانتهاء من توقيع كافة النقاط التى تمثل

السلسلة الزمنية للظاهرة يتم توصيل كل نقطتين متتاليتين بخط مستقيم أو شبه مضع . فيكون الناتج هو عبارة عن الخط البياني للظاهرة محل الدراسة . والمثال التالي يوضح ذلك .

مثال (21):-

فيما يلي بيان بتطور عدد العاملين (الذكور) بالوظائف العليا بالدولة وذلك خلال الفترة (1980-1984) :

جدول (2-34)

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

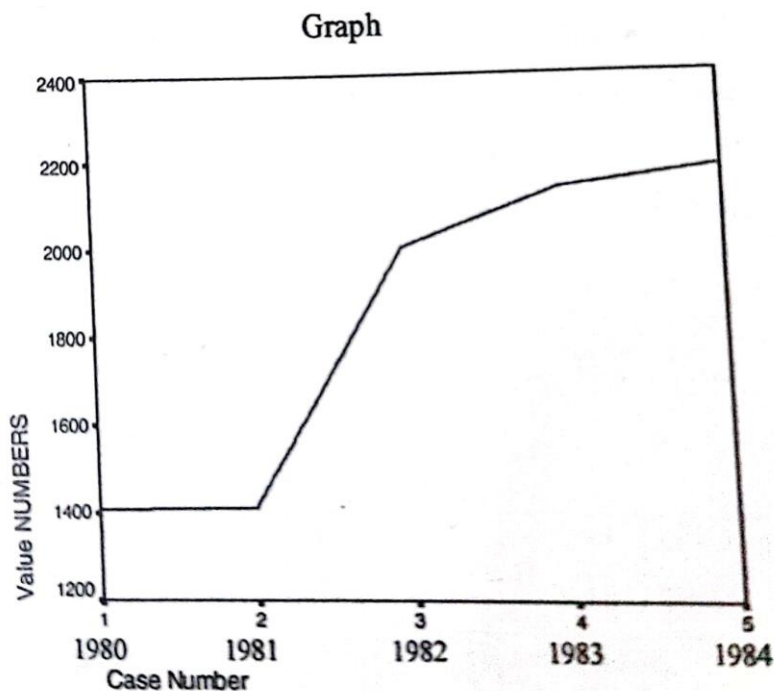
والمطلوب عرض بيانات تلك السلسلة باستخدام خريطة الخط البياني.

الحل :-

يتم تمثيل السنوات المختلفة للسلسلة على المحور الأفقى بينما يمثل عدد العاملين على المحور الرأسى . وحيث أن أصغر قيمة لعدد العاملين هي 1407 عاملا فإنه يكون من الملائم كسر (قطع) المحور الرأسى من أسفل وذلك حتى يمكن اختيار مقياس رسم مناسب يوضح تطور أعداد العاملين دون ترك مساحة كبيرة من ورقة الرسم البياني غير مستخدمة . لذا فإننا سوف نختار مقياس الرسم للمحور الرأسى هو 1سم/100عاملا مع البدء بالقيمة 1400 كما يتضح ذلك فى الشكل (2-21) .

شكل (2-21)

خط بياني يوضح تطور أعداد العاملين (الذكور) بالوظائف العليا في الدولة



ومن خلال هذا الشكل يمكن استنتاج أن أعداد العاملين بالوظائف العليا في الدولة في زيادة مستمرة مع الزمن وذلك بمجرد النظر الى خريطة الخط البياني الموضح في الشكل (2-21) .

2- خريطة الشريط :- Band Chart

وتعتبر خريطة الشريط بمثابة امتدادا لشكل خريطة الخط البياني السابق الاشارة اليها حيث أنها تشمل خطين بيانيين 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 بياني 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 فيما بين القيم المتناظرة للظاهرتين موجبة أحيانا وسالبة أحيانا أخرى وهنا يلزم استخدام لونين مختلفين أو طريقتين مختلفتين فى التظليل وذلك للتفرقة فيما بين الفروق الموجبة والآخرى السالبة بهدف توضيح مدلول كل منهما .

مثال(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
عدد الذكور	120	150	175	210	280	320
عدد الاناث	90	100	120	150	160	175

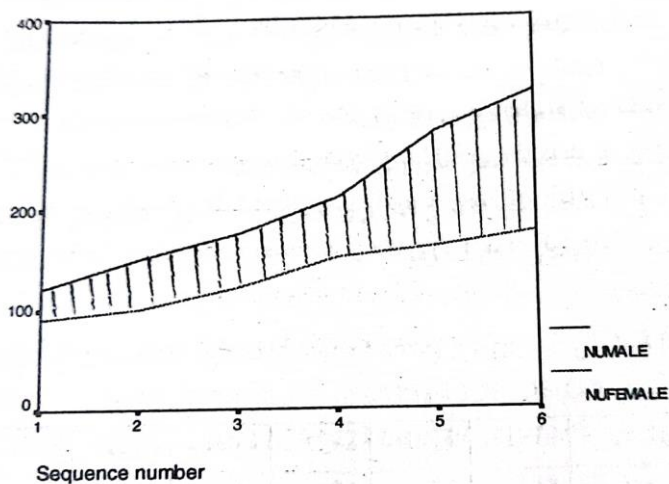
والمطلوب عرض بيانات هذه السلسلة باستخدام خريطة الشريط .

الحل :

الشكل البياني التالى (شكل(2-22)) يوضح خريطة الشريط لبيانات السلسلة الزمنية المعطاه.

شكل (2-22)

خريطة الشريط لأعداد الطلاب (ذكور وإناث)
في إحدى الكليات عن الفترة 1960-1965



ويلاحظ من خريطة الشريط الموضحة في الشكل (2-22) ان عدد الطلاب الذكور يكون دائما أكبر من عدد الطلاب الاناث . كما يتضح من الشكل ايضا ان الفرق بينهما في زيادة مستمرة من سنة لأخرى . وأخيرا فهناك أشكال بيانية أخرى لايسعنا المجال لذكرها حيث أننا نختار منها مايخدم الدراسة . فهناك ما هو يساهم في تمثيل التوزيع أو التركيب العمري للسكان أو مايسمى بالهرم السكاني (Population Pyramid) ومنها ما يستخدم في الدراسات الاقتصادية والذي يعطى صورة عن مدى التوزيع العادل للثروة أو الناتج القومي على جمهور الافراد أو السكان داخل قطر معين وهو مايسمى بمنحنى لورنز (Lorenz Curve) وغير ذلك من الرسوم البيانية الأخرى.

" تمارين "

(1): أخذت عينة من 40 طالب وتم تجميع البيانات الخاصة بالالتحاق بالشعب المختلفة حسب التخصص فكانت كما يلي :

محاسبة	محاسبة	إدارة	محاسبة	محاسبة
محاسبة	إدارة	محاسبة	إدارة	إدارة
إدارة	محاسبة	إقتصاد	محاسبة	محاسبة
إقتصاد	إدارة	محاسبة	إدارة	إقتصاد
محاسبة	محاسبة	إدارة	محاسبة	محاسبة
إحصاء	إقتصاد	إقتصاد	محاسبة	إقتصاد
محاسبة	إدارة	إدارة	إقتصاد	إحصاء
إدارة	إدارة	محاسبة	إحصاء	محاسبة

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 يلي:

جيد	ضعيف	جيد	مقبول	ممتاز
جيد جدا	مقبول	مقبول	ضعيف	ضعيف جدا
جيد	ممتاز	جيد جدا	جيد	جيد
ضعيف	جيد	مقبول	ممتاز	جيد
مقبول	جيد جدا	ضعيف جدا	ضعيف	جيد جدا
جيد	مقبول	ضعيف	جيد جدا	مقبول
ممتاز	مقبول	جيد	جيد	مقبول
ضعيف جدا	مقبول	جيد جدا	مقبول	مقبول
مقبول	مقبول	مقبول	مقبول	مقبول
جيد	مقبول	مقبول	مقبول	مقبول

وضع هذه التقديرات فى جدول تكرارى بسيط حسب التقديرات المختلفة.

(3): فيما يلى تقديرات نجاح 20 طالبا فى مادة الرياضيات :
 جيد جدا، جيد، مقبول، جيد، ممتاز، جيد، جيد جدا، جيد، جيد جدا، مقبول،
 جيد، جيد جدا، جيد، جيد، مقبول، جيد جدا، جيد جدا، مقبول، ممتاز.
 المطلوب :-

- أ- تبويب هذه البيانات فى جدول تكرارى .
 ب- تمثيل البيانات بيانيا وذلك باستخدام شكل الدائرة والاعمدة
 البيانية.

(4): اعتبر الارقام التالية تمثل الدرجات التى حصل عليها 54 طالبا فى
 احدى

الاختبارات (علما بأن الدرجة العظمى = 10 درجات) .

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

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.

المجموع	ممتاز	جيد جدا	جيد	مقبول	ضعيف	ضعيف جدا	الكلية/التقديرات
990	24	78	296	311	203	78	الكلية أ
1987	16	41	291	229	236	274	الكلية ب

والمطلوب: عرض هذه البيانات بالأشكال البيانية المناسبة .

(6): الجدول التكرارى 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 باستخدام الشكل الإحصائى المناسب .

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

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

والمطلوب :

1- وضع هذه البيانات فى جدول تكرارى حسب فئات الدرجات وعدد الطلاب.

2- رسم المدرج التكرارى والمضلع التكرارى لهذا التوزيع .

3- تكوين الجدول التكرارى المتجمع الصاعد والنازل .

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

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

والمطلوب :

أ- ناقش بإيجاز الخطوات التي تستخدم في تبويب هذه البيانات ، ثم استخدم هذه الخطوات في تكوين جدول تكرارى ذى فئات متساوية.

ب- باستخدام الجدول التكرارى الناتج فى (أ) كون الجدول التكرارى المتجمع الصاعد مرة باستخدام التكرارات الاصلية ومرة أخرى باستخدام التكرارات النسبية.

ج- أوجد قيمة العمر الذى يقل عنه 33 شخصا.

د- استخدم شكل المدرج التكرارى للتعبير عن التوزيع الناتج فى (أ) ثم علق على شكل التوزيع .

(9): فيما يلى قيمة الارباح (بالآف الجنيهات) لشركتين متنافستين (س،ص) فى 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

والمطلوب :

أ- قارن بين أرباح الشركتين باستخدام الاعمدة المجزأة والاعمدة المتلاصقة.

ب- استخدام خريطة الخط البيانى لعرض البيانات الخاصة بأرباح الشركة(س).

ج- بين ماهى الشركة الاقدر على المنافسة باستخدام خريطة الشريط .

(10): فيما يلى درجات 30 طالبا فى كل من مادتى الرياضة والإحصاء والمطلوب وضع هذه البيانات فى صورة جدول تكرارى مزدوج.

الرياضة	الإحصاء	الرياضة	الإحصاء	الرياضة	الإحصاء
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 في كل من مادتي الرياضيات (س) والإحصاء (ص):

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

والمطلوب :

- عرض هذه البيانات في شكل جدول تكرارى مزدوج وذلك على النحو

الاتى:

- طول الفئة في كل من س ، ص = 10

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

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

- إستنتج التوزيع التكرارى الهامشى لدرجات الطلاب فى كل من مادتى الرياضيات والاحصاء.
- قارن بين التوزيعين الناتجين فى (ب) من حيث مستوى درجات الطلاب وذلك بإستخدام المنحنى التكرارى.
- بإستخدام المدرج التكرارى ، قارن بين شكلى التوزيعين من حيث التماثل والالتواء .
- عزز نتائجك فى(ج) 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).
- 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 .
- 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 100 شخص موزعين حسب فئات السن المختلفة .

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

والمطلوب :

- أ- رسم المدرج التكرارى والمضلع التكرارى لهذا التوزيع .
 ب- عمل الجدول التكرارى المتجمع الصاعد والنازل ثم أوجد عدد الاشخاص الذين يقل عمرهم عن 45 سنة بالرسم.
 ت- ماهو الحد الادنى لأعمار 40 شخصا بالرسم .

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

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

المطلوب :

- أ- ارسم 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 .
 ج- باستخدام المنحنى التكرارى المتجمع المناسب أحسب نسبة الاشخاص الذين تبلغ أعمارهم 18 سنة على الاقل.

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

المطلوب :

- أ- ارسم المضلع التكرارى والمنحنى التكرارى لبيانات الانفاق .

ب- كون الجدول التكرارى المتجمع الصاعد وذلك باستخدام التكرارات 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 الاسبوعى بالجنيه كالاتى:

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

والمطلوب :

- 1- عرض هذه البيانات باستخدام المدرج التكرارى والمضلع التكرارى.
- 2- رسم المنحنى المتجمع الصاعد ومنه أوجد عدد العمال الذين يتقاضون أجرا أقل من 40 جنيه .
- 3- 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- الحد الادنى الذى يتقاضاه 120 عاملا .

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

كمية الإنتاج	-20	-30	-40	-50	-60	-70	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 كالتالي :

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

والمطلوب :

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

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

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

والمطلوب تمثيل هذه البيانات مستخدما 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 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 فكانت كالتالى:

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 وترفيه	عقيدة عقيدة	رياضة	متعة وترفيه

والمطلوب :

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

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

الإنفاق اليومي	-20	-25	-30	-35	-40	-45	-50	-55
عدد السائحين	18	27	32	51	43	35	24	20

والمطلوب :

1- عرض هذه البيانات باستخدام المدرج التكرارى والمضلع التكرارى.

2- رسم المنحنى التكرارى المتجمع الصاعد ومنه أوجد :

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

ب- الحد الأدنى للإنفاق اليومي لـ 120 سائحا .

3- رسم المنحنى المتجمع النازل ومنه أوجد :

أ- عدد السائحين الذين ينفقون 37.5 جنيه فأكثر يوميا .

ب- الحد الأدنى للإنفاق اليومي لـ 120 سائحا.

ت-

(22): الجدول التالى يوضح التوزيع التكرارى لأجور العمال الاسبوعية

(بالجنيهات) فى ثلاثة مصانع مختلفة (س،ص،ع) :

فئات الاجور	-50	-60	-70	-80	-90	-100	المجموع
عدد العمال فى...	1	3	4	12	12	8	40
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

والمطلوب :

1- استخدام الشكل البيانى المناسب للمقارنة بين توزيع الاجور (من حيث

المستوى) وذلك فى الحالات الاتية :

* المصنعان : س،ع

* المصنعان : س،ص

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

3-باستخدام 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 باستخدام التوزيع التكرارى المناسب .

5- إحسب النسبة المئوية للعمال الذين يتقاضون أجور أقل من 80 جنيها فى كل من المصانع الثلاثة. علق على نتائجك على ضوء ماحصلت عليه فى (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	اقل من	اقل من	اقل من	اقل من
	30	35	45	60	80
التكرار	19	27	48	73	100

والمطلوب :

1-أرسم المدرج التكرارى .

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

3-أذا كانت نسبة الاشخاص الذين تتراوح أعمارهم ما بين 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 وبالتالي 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 اقتصررت فوائدها الإستقرائية على أن يقدم المصنع التكرارى مثلا لعينة من البيانات نقوم بتلخيصها تصورا عن شكل المصنع التكرارى للمجتمع الذى سحبت منه هذه العينة. وفى هذا الباب سنلجأ لتلخيص البيانات فى صورة رقمية وذلك عن طريق حساب بعض المقاييس الاحصائية والتي تعرف بإسم المتوسطات وهى التى تصف لنا كيفية توزيع الظاهرة محل الدراسة فى العينة بطريقة دقيقة ومختصرة وتصلح أيضا للمقارنة بين ظاهرتين أوأكثر .

فضلا عن أنها لا تختلف باختلاف الأشخاص العارضين أو الدارسين للظاهرة. والمتوسط لأية مجموعة من البيانات هو القيمة التي تمثل هذه المجموعة وتعتبر 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 التمرکز حول تلك القيمة الوسطية . فأولئك الذين يتصفون بالنعافة 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 اختبار لقياس درجة الذكاء على مجموعة من الطلاب لوجدنا أن الطلاب المتفوقين الموهوبين قلة والطلاب المبتلين بالبلادة قلة وتنزع درجة الذكاء عند معظم الطلاب الى التمرکز حول الوسط ، ويسمى ذلك الميل الى التجمع حول هذه القيمة الوسطى بالنزعة المركزية ، لذلك تعرف المتوسطات أحيانا (بمقاييس النزعة المركزية) والتي يقصد بها ببساطة ميل البيانات للتركز والتراكم حول هذه القيمة المتوسطة. وفى حياتنا اليومية كثيرا مانستخدم كلمة " فى المتوسط" فنحدث عن الرجل " متوسط الدخل" والشاب " متوسط الثقافة" ، كما نقول أن " إنفاق الاسرة اليومى هو فى المتوسط كذا " الخ.

وهذه الاستخدامات الشائعة لكلمة متوسط تعبر عن شعور داخلى معين يحسه ويفهمه كل منا ولايستطيع ترجمته بدقة ، ومقاييس النزعة المركزية هى محاولة لترجمة هذا الشعور بطريقة دقيقة ومحددة تماما .

وعند حساب احد مقاييس النزعة المركزية يجب مراعاة العوامل الاتية :

(1): ان تكون طريقة حسابه سهلة وأن يعكس التغير فى الظاهرة ولا تتغير قيمته بتغيير 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 وبيان كيفية حسابها فى حالة البيانات غير المبوبة وفى حالة البيانات المبوبة ، هذه المقاييس هى :

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 .

**** الوسط الحسابى:- (Arithmetic Mean (or Avarage)**

يعتبر الوسط الحسابى بحق من أهم مقاييس النزعة المركزية وأكثرها إستخداما فى الحياة العملية وذلك لبساطته وسهولة فهمه ووضوح معناه ، وهو الذى يشار له بإسم المتوسط فى 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 الحسابي من المفردات (بيانات غير مبوبة):

يمكن حساب الوسط الحسابي لمجموعة من البيانات (المفردات) لظاهرة معينة بعدة طرق تختلف فيما بينها في الجهد الحسابي فقط لكنها في النهاية تؤدي للحصول على نفس النتيجة كما سنرى في التحليل البياني التالي :

أولاً : الطريقة الحسابية المطولة (المباشرة) :-

بصفة عامة إذا كان لدينا مجموعة من المفردات ولتكن:

$$x_1, x_2, \dots, x_n$$

حيث (n) تعبر عن عدد المفردات . فإن الوسط الحسابي لهذه المجموعة من هذه المفردات سنرمز له بالرمز \bar{x} وهو عبارة عن النسبة ما بين مجموع القيم وعددها أي أن :

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum x}{n} \quad (1)$$

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

مثال(1) :

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

الحل : حيث أن :

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

لذا فإن :

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

مثال (2):

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

الحل : حيث أن :

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

لذا فإن :

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

ثانياً :- طريقة الانحرافات البسيطة:

وتهدف هذه الطريقة الى إختصار العمليات الحسابية بعض الشئ وبصفة خاصة حينما يكون عدد المفردات وقيم المفردات كبير نسبياً . وهذه الطريقة تقلل الجهد الحسابى المبذول وكذا تقلل من احتمالات الخطأ خاصة حينما لاتتوافر الآلات الحاسبة لإستخدامها فى هذا المجال . وتتلخص هذه الطريقة 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 الحسابي للانحرافات البسيطة الناتجة والتي يمكن معه حساب الوسط الحسابي للمفردات الاصلية وهذه الطريقة تبني على اساس أن من اهم خصائص الوسط الحسابي أنه يتأثر بالعمليات الجبرية الاربعة . ومن ثم فإن خطوات هذه الطريقة تتلخص فيما يلي :

(1):بافتراض أن لديك مجموعة القيم X_1, X_2, \dots, X_n .

(2):يتم اعتبار إحدى القيم بمثابة وسط فرضي ولتكن قيمته (A).ويفضل

اختيار تلك القيمة التي تتكرر بشكل واضح وتكون قريبة من متوسط

مجموعة المفردات محل الدراسة كوسط فرضي . إذ ان هذا سيؤدي الى

تبسيط العملية الحسابية بشكل كبير .

(3):يتم طرح قيمة الوسط الفرضي (A) من جميع المفردات لنحصل على

انحرافات القيم عن هذا الوسط الفرضي ولتكن (di) حيث: $i = 1, 2, \dots, n$

أي أن :

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

او بلغة أخرى فإن :

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

وتسمى قيم (di) بالانحرافات البسيطة .

يتم إيجاد الوسط الحسابي لتلك الانحرافات البسيطة وذلك من خلال قسمة

تلك الانحرافات على عددها أي ان :

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

فيكون الوسط الحسابي لقيم المتغير (X) هو :

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

مثال(3):

أحسب الوسط الحسابي للمفردات الواردة بالمثل (1) مستخدما في ذلك طريقة الانحرافات البسيطة .

الحل : بملاحظة المفردات المعطاه 500 , 600 , 700 , 750 , 800 نجد انه لا توجد مفردة تتكرر أكثر من غيرها لذا دعنا نفترض ان الوسط الفرضى والذي سيتم طرحه من كافة المفردات هو القيمة 700 اى ان $A=700$ فيكون لدينا الجدول التالى :

جدول (1-3)

X_i	$d_i=x_i-700$
500	500-700=-200
600	600-700=-100
700	700-700=0
750	750-700=500
800	800-700=100
Σ	-150

وحيث ان :

$$\begin{aligned} \bar{X} &= A + \frac{\sum d_i}{n} \\ &= 700 + \frac{(-150)}{5} = 700 - 30 = 670 \end{aligned}$$

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

ثالثا : طريقة الانحرافات المختصرة (المعدلة) :-

إمتداد للطريقة السابقة (الانحرافات البسيطة) فإذا كانت قيم الانحرافات البسيطة (d_i) تقبل القسمة على مقدار ثابت وليكن (B) بدون باقى . فإنه

يمكن إختصار العمليات الحسابية أكثر عند حساب الوسط الحسابى وذلك على النحو التالى :

(1): بعد الحصول على الانحرافات البسيطة للمفردات كما سبق فى الطريقة السابقة ولتكن هذه الانحرافات هى (di) لكل قيم $i=1,2,\dots,n$ فإنه يمكن أعتبار عامل مشترك وليكن (B) فيما بين قيم الانحرافات البسيطة وباعتبار هذا العامل المشترك نحصل على الانحرافات الأكثر إختصارا على النحو المبين التالى :

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

(2): لإيجاد الوسط الحسابى للقيم الأصلية xi أى \bar{x} مستخدمين فى

الحساب الانحرافات الاكثر اختصارا (المعدلة) اى (Di) فإنه يجب رد الحسابات بطريقة عكسية بعد حساب متوسط الانحرافات الاكثر إختصارا أى بعد حساب \bar{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 :

$$\bar{X} = A + B \bar{D} = A + B \frac{\sum Di}{n} \quad (3)$$

مثال (4):

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

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

جدول (2-3)

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
Σ		-3

وحيث أن :

$$\bar{X} = A + B \left(\frac{\Sigma D_i}{n} \right)$$

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 . وخالصة 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) .

(2-1-3) خصائص الوسط الحسابى :-

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

أولاً: مجموع انحرافات القيم عن وسطها الحسابى يساوى الصفر : أى 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, \dots, X_2, X_1 ومتوسطه الحسابى هو (x) فتكون

انحرافات القيم (X i) عن متوسطها هى عبارة عن :

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

وبذلك فإن مجموع هذه الانحرافات هو :

$$\begin{aligned} \sum_{i=1}^n (x_i - \bar{x}) &= (x_1 - \bar{x}) + (x_2 - \bar{x}) + \dots + (x_n - \bar{x}) \\ &= (x_1 + x_2 + x_3 + \dots + x_n) - (\bar{x} + \bar{x} + \dots + \bar{x}) \end{aligned}$$

لاحظ ان الحد الثانى عبارة عن مجموع المقدار الثابت (\bar{x}) عددا (n) من

المرات فيكون مجموعه هو : ($n \bar{x}$) فيكون :

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

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 متوسطها الحسابى يساوى دائما أبدا صفرا .
وهذه الخاصية توضح الدور المركزى الذى يلعبه قيمة الوسط الحسابى
للمفردات .

ثانيا : مجموع مربعات إنحرافات القيم عن وسطها الحسابى أصغر ما يمكن
أى اصغر من مجموع مربعات إنحرافات القيم عن أى وسط فرضى آخر أى
أن :

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

حيث (A) مقدار ثابت أختيارى تختلف قيمته عن الوسط الحسابى \bar{X} ،
ولاثبات صحة ذلك فحيث ان :

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

وذلك لأن A، \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 فهي عبارة عن مقدار موجب . كما أن المقدار $(\bar{x} - A)^2$ مقدار مربعاً فهو مقدار موجب أيضاً . وعليه فإن :

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

أى أن :

$$\sum_{i=1}^n (x_i - \bar{x})^2 < \sum_{i=1}^n (x_i - A)^2 \quad (5)$$

أى أن مجموع مربعات إنحرافات القيم عن وسطها الحسابى أقل ما يمكن .

مثال (5):-

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 :

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

وعلق على نتائجك .

الحل :

1- ليجاد مجموع انحرافات القيم عن وسطها الحسابى دعنا نقوم

بحساب الوسط الحسابى للمفردات بأى من الطرق الثلاث التى سبق

الإشارة إليها وليكن بالطريقة المطولة :

فحيث أن:

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

والجدول التالي يبين الحسابات اللازمة :

جدول (3-3)

X_i	$x_i - \bar{x}$	$(\bar{x}_i - x)^2$	$(x_i - 7)$	$(x_i - 7)^2$	$(x_i - 10)$	$(x_i - 10)^2$
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 :

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

وهو ما يؤكد صحة الخاصية الاولى من خصائص الوسط الحسابى .

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

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

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

(3-1-2) حساب الوسط الحسابى للبيانات المبوبة تكراريا :-

سبق وأن ذكرنا فى التمثيل الجدولى للبيانات انه لاختصارا للبيانات 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. فلاحظك 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) فأنا سوف نستخدم الطرق الثلاث السابقة فى حالة البيانات الغير مبوبة (المفردات) وذلك مع إجراء التعديلات اللازمة حتى تلائم هذه الطرق طبيعة عرض البيانات المبوبة فى جداول تكرارية سواء كانت هذه البيانات تمثل متغيرات كمية منفصلة أم متصلة .

أولا : حساب الوسط الحسابى لجدول تكرارى لمتغير كمي منفصل (متقطع):
 فى حالة وجود جدول تكرارى لظاهرة كمية منفصلة (مثل حجم الاسرة أو عدد العاملينالخ) فإن كل قيمة من قيم الظاهرة المعبرة عن هذا المتغير المنفصل تتكرر عددا من المرات . ولذلك فإنه لإيجاد مجموع كافة هذه القيم وهو مايتطلب حساب الوسط الحسابى والذى يساوى مجموع القيم منسوبا الى عددها . فإنه يلزم ضرب كل قيمة فى عدد مرات تكرارها ثم جمع حواصل الضرب ، بعدها يمكن حساب قيمة الوسط الحسابى بقسمة مجموع هذه القيم على عددها . فإذا كانت أقسام الظاهرة التى تظهر فى الجدول التكرارى للمتغير المنفصل هى X_1, X_2, \dots, X_n ذات تكرارات مقابلة لها فى هذا الجدول ولتكن F_1, F_2, \dots, F_n على الترتيب . فإن الوسط الحسابى سيكون عبارة عن :

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

وهنا يمكن ملاحظة أن:

القانون الوارد بالعلاقة (6) وهو الخاص بحساب الوسط الحسابي لبيانات متغير منفصل مبوبا في جدول تكرارى هو نفسه القانون الوارد بالعلاقة (1) الخاص بحساب الوسط الحسابي 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) بقيمة (X_iF_i) فى العلاقة (6). وكذلك استبدال (n) فى العلاقة (1) بـ (ΣF_i) فى العلاقة (6). أى ان التكرارات تؤخذ فى الاعتبار وذلك من خلال ضرب كل قيمة من قيم المتغير المنفصل (X_i) فى عدد مرات تكرارها أى (F_i) .

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

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

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

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

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 .

مثال (6):

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

جدول (3-4)

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

والمطلوب حساب متوسط عدد الغائبين يوميا خلال هذا الشهر فى تلك
الشركة.

الحل: الجدول التالى يوضح الحسابات اللازمة لحساب متوسط عدد
العاملين الغائبين يوميا باستخدام الطرق الثلاث (البسطة والانحرافات
البسيطة والانحرافات 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
Σ	30	120		0		0

* حساب الوسط الحسابي 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 :

$$\bar{X} = \frac{120}{30} = 4 \quad \text{persons}$$

* حساب الوسط الحسابي باستخدام 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 :

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

لاحظ ان المجموع ($d_i F_i$) مساويا للصفر لأن الوسط الفرضى المستخدم فى تلك الطريقة وهو قيمة (A) مساوية للوسط الحسابى . لذا فلا بد أن يكون مجموع هذا العمود ($d_i F_i$) مساويا للصفر وذلك لأن مجموع انحرافات القيم عن وسطها الحسابى هو صفرا (خاصية (1)).

*حساب الوسط الحسابى باستخدام طريقة الانحرافات المعدلة : فحيث ان :

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

ومن الجدول السابق فإن : $B = 2$, $A = 4$

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

لاحظ ان النتيجة واحدة بأى من الطرق الثلاث .

ثانيا : حساب الوسط الحسابى لبيانات متغير متصل (مبوبة تكراريا) :-

عند إيجاد الوسط الحسابى فى حالة التوزيعات التكرارية للمتغيرات المتصلة . فإننا نواجه مشكلة إختفاء بعض المعلومات الاصلية عن الظاهرة محل الدراسة وذلك نتيجة عملية تبويب البيانات تكراريا . فمن الجدول التكرارى محل الدراسة لانعلم القيمة الحقيقية لكل وحدة من وحدات تكرار الجدول . وحيث أننا فى حاجة الى معرفة قيمة المتغير محل الدراسة حيث نتمكن من إيجاد مجموع مفرداته فإننا نلجأ الى افتراض منطقى وهو ماتبنى على أساسه فكرة الجداول التكرارية ألا وهو أن التكرارات تتوزع بانتظام 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 ببياناته تكراريا فإننا 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 تكرارات 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 فإننا نلجأ لحساب مراكز الفئات وهى تقابل قيم الظاهرة فى حالة المتغير المنفصل .

وخلاصة ماسبق هو أن القوانين المستخدمة فى ايجاد الوسط الحسابى فى حالة المتغيرات المتصلة هى نفس القوانين المستخدمة المستخدمة فى حالة المتغيرات المنفصلة مع إجراء التعديل اللازم فى مفهوم قيم (Xi) فقط والتي 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 .

مثال (7):

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	Σ
	-	-	-	-	-	-	-	
							260	
Frequency	2	4	11	38	30	10	5	100

والمطلوب: حساب متوسط الاجر اليومى للعامل فى هذا المصنع باستخدام الطرق الثلاثة .

الحل:

1- باستخدام الطريقة المباشرة (المطولة) :

الجدول التالى يوضح الحسابات اللازمة لحساب متوسط الاجر اليومى :

جدول (7-3)

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
Σ	100		19800

وحيث أن :

$$\bar{X} = \frac{\sum_{i=1}^r x_i F_i}{\sum_{i=1}^r F_i}$$

لذا فإن :

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

2- باستخدام طريقة الانحرافات البسيطة :

باعتبار مركز الفئة التي تتوسط (مركز) التوزيع التكرارى بمثابة وسطا فرضيا يتم طرحه من كافة المراكز . أى بطرح المقدار $A=190$ من كافة مراكز الفئات (X_i) للحصول على الانحرافات (d_i) . ثم ضرب هذه الانحرافات

البيسطة فى التكرارات المقابلة وايجاد المجموع . والجدول التالى يلخص الخطوات اللازمة لحساب المتوسط الحسابى بطريقة الانحرافات 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 :

$$\bar{X} = 190 + \frac{800}{100} = 198 \quad (\text{L.E})$$

وهى نفس الناتج باستخدام الطريقة المباشرة .

3- باستخدام طريقة الانحرافات المختصرة (المعدلة) :-

لاحظ فى الجدول السابق ان عمود الانحرافات البسيطة (d_i) يقبل القسمة على (20) وهى بمثابة طول الفئة فى هذا الجدول المنتظم . لذا يمكن اعتبار طول الفئة بمثابة وسطا فرضيا آخر وحساب الانحرافات المعدلة (D_i)

وذلك من خلال قسمة عمود الانحرافات البسيطة (d_i) على (20) والجدول التالي يوضح الحسابات اللازمة لإيجاد متوسط الأجر اليومي 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 : $A = 190$, $B=20$. لذا فإن :

$$\bar{X} = 190 + 20 \left(\frac{40}{100} \right) = 190 + 8 = 198 \quad (L.E)$$

وهي نفس النتائج المستخلصة من الطريقتين السابقتين لكن أهم ما يميز هذه الطريقة هي تبسيط أرقام الجدول لأقصى درجة ممكنة لدرجة عدم الحاجة إلى استخدام الآلات الحاسبة في 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 النسبية وذلك بعد إجراء التعديل اللازم على صيغ تلك القوانين . وذلك على النحو التالى :-

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

$$\bar{X} = \frac{\sum_{i=1}^r x_i F_i}{\sum_{i=1}^r F_i}$$

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 هى :

$$\bar{x} = \sum_{i=1}^r x_i R_i \quad (9)$$

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

*كذلك وينفس الاسلوب تصبح صيغة ايجاد الوسط الحسابى باستخدام طريقة الانحرافات البسيطة وفى حالة توافر تكرارت نسبية على الصورة .

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

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 فى الجدول التكرارى .
*وأخيرا فأن الصيغة الرياضية لحساب الوسط الحسابى باستخدام طريقة الانحرافات المعدلة (الطريقة الأكثر إختصارا) فى حالة توافر تكرارت نسبية تصبح على الصورة التالية :-

$$\bar{X} = A + B \sum_{i=1}^r D_i R_i$$

حيث (Di) تمثل الانحراف المختصر للفئة (i) كما سبق تعريفه .
والمثال التالى يوضح طريقة حساب الوسط الحسابى باستخدام الطرق الثلاث فى حالة التوزيعات التكرارية النسبية .

مثال(8):-

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 النسبى 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 باستخدام الطريقة المباشرة (المطولة) .

والجدول التالى يبين الحسابات اللازمة لاجاد 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
Σ	1		36

وحيث أن :

$$\bar{x} = \sum_{i=1}^r x_i R_i$$

لذا فإن :

$$\bar{X} = 36 \quad (\text{L.E})$$

2- ايجاد الوسط الحسابى بطريقة الانحرافات البسيطة :-

والجدول (3-12) يوضح الحسابات اللازمة لاجاد متوسط الاجر اليومى فى تلك الشركة:

جدول (12-3)

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
Σ	1			-4

وحيث أن :

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

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

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

وهي نفس النتيجة باستخدام الطريقة المباشرة .

3- ايجاد الوسط الحسابي بطريقة الانحرافات المختصرة (المعدلة) :

والجدول (13-3) يوضح الحسابات اللازمة لاجاد متوسط الاجر اليومي في تلك الشركة:

جدول (13-3)

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
Σ	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:

$$\bar{X} = 40 + 10(-0.4) = 40 - 4 = 36 \quad (\text{L.E})$$

وهي نفس النتيجة التي تم التوصل اليها باستخدام كلا من الطريقة المباشرة وطريقة الانحرافات البسيطة .

(3-1-4) : الوسط الحسابي المرجح (المتوسط العام أو متوسط

المتوسطات) :Weighted Mean

عند ايجاد الوسط الحسابي فقد تم التعامل مع جميع المفردات الداخلة في تكوينه معاملة واحدة . أى أننا اعتبرناها جميعا لها نفس الاهمية .
ويعد ذلك غير صحيحا في غالبية الاحوال. فكثيرا ما تتفاوت القيم في أهميتها النسبية . لذا فإنه عند ايجاد الوسط الحسابي في 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 في تلك الحالة يسمى بالوسط الحسابي المرجح . فإذا اهلنا أوزان الترجيح في تلك الحالة فستكون قيمة الوسط الحسابي مضللة الى حد كبير وبعيدة عن الواقع .

وخلاصة ماسبق هو أنه إذا كانت قيم المتغير (X) ولتكن هي :

$X_1, X_2, X_3, \dots, X_m$ وكانت أوزانها الترجيحية المقابلة هي:

$n_1, n_2, n_3, \dots, n_m$ فإن الوسط الحسابي المرجح يكون عبارة عن :

$$\sum_{i=1}^m n_i \bar{x}_i$$

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

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 هي:

$\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m$ على الترتيب فإن المتوسط العام (او المرجح) ما هو إلا :

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

لاحظ أن كل حد من حدود البسط فى العلاقة الأخيرة ما هو إلا مجموع مفردات العينات المسحوبة من هذا المجتمع. أما المقام والذي يمثل مجموع مفردات العينات المسحوبة من هذا المجتمع . وفى حالة تساوى أحجام تلك العينات فإنه يمكن اعتبار (n_i) مقدار ثابت أى يكون :

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

فتصبح صيغة العلاقة (12) تأخذ الصورة التالية :

$$\bar{X} = \frac{n(\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_m)}{(n)(m)} = \frac{\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_m}{m} = \frac{\sum_{i=1}^m \bar{x}_i}{m} \tag{12}$$

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 الآلات 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 الثالثة 10 وحدات يوميا وتباع منها الوحدة بسعر 40 جنيها .

والمطلوب حساب متوسط سعر بيع الوحدة من إنتاج هذا المصنع عموما .

الحل:-

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

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

وهى نتيجة مضللة لاتعكس حقيقة مستوى السعر الحقيقى لمنتجات الآلات المختلفة الثلاث . اما إذا استخدمنا الكميات المنتجة كأوزان للترجيح فيكون متوسط سعر بيع الوحدة المنتجة للآلات الثلاث هى :

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

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

ملحوظة: من عيوب الوسط الحسابى المرجح عملية تحديد الاوزان الترجيحية. ففى غالبية الاحيان تخضع الاوزان الترجيحية لعملية التقدير الشخصى مما يودى الى اختلاف النتائج من شخص الى آخر .

مثال (10):

يتم تدريس مبادئ الاحصاء الوصفى لثلاث شعب فى كلية ما . وأجرى اختبار 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:

الحل:- معطى لدينا المعلومات التالية :

$$n_1 = 32 , \quad n_2 = 20 , \quad n_3 = 40$$

$$\bar{x}_1 = 70 , \quad \bar{x}_2 = 85 , \quad \bar{x}_3 = 68$$

فيكون المتوسط العام لدرجة الطالب لكافة الشعب عبارة عن :

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

ملحوظة :

لاحظ أن متوسط الدرجات في حالة إهمال عدد المفردات في كل شعبة هو

$$\bar{X} = \frac{70+ 85+ 68}{3} = 74.33 \text{ (degree)}$$

وهذه النتيجة غير صحيحة إحصائياً وتعتبر نتيجة مضللة لاتأخذ كافة البيانات المتاحة حول الظاهرة موضع الدراسة في الاعتبار .

ملاحظات حول الوسط الحسابي (المزيا والعيوب):-

(1): قيمة الوسط الحسابي لا تتغير مهما اختلفت طرق حسابه وأياً كان الوسط الفرضي الذي يستخدم للتسهيل في حسابه . هذا ويفضل إستخدام طريقة الانحرافات المختصرة (المعدلة) لحساب الوسط الحسابي في 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 بواقي . أما طريقة الانحرافات البسيطة فيفضل استخدامها في حالة الجداول غير المنتظمة . أما الطريقة المباشرة (المطولة) فتستخدم غالباً حينما يكون حجم المفردات صغيراً لأنها تحتاج كما رأينا الى عمليات حسابية مطولة واستخدام الآلات الحاسبة نظراً لكبر حجم الأرقام بإستخدام تلك الطريقة.

(2): أهم ما يميز الوسط الحسابي كمقياس من مقاييس النزعة المركزية هو أنه يأخذ في الاعتبار جميع المفردات عند حسابه . لذلك يمكن استخدامه في عمليات احصائية أخرى ، كما أن قيمته وحيدة بمعنى أنه لأي مجموعة من البيانات لابد ان يوجد لها وسط حسابي وحيد (غير متعدد القيم) وهو ما يجعله أهم مقاييس النزعة المركزية ولذلك فإنه الأكثر استخداما في التطبيقات العملية لأساليب التحليل الاحصائي .

(3): أما ما يعاب على الوسط الحسابي أنه لا يفضل حسابه في حالة الجداول التكرارية المفتوحة سواء من أعلى أو من أسفل أو مفتوحة الطرفين وذلك لأن حسابه في الجداول المفتوحة سوف يبنى على افتراضات من أجلها يمكن غلق الجدول المفتوح وهو شرط لازم حتى يمكن تحديد مراكز الفئات والتي تعتبر اساسا في حسابه . وهناك من يرى أنه في ظل الحالات التي يتعذر فيها تحديد حد الفئة المفتوحة (سواء الحد الأدنى للفئة الاولى أو الحد الاعلى للفئة الاخيرة) لا يكون أمامنا سوى إهمال هذه الفئة وذلك إذا كان تكرارها صغير جدا نسبيا بالنسبة لمجموع تكرارات التوزيع التكرارى وهو ما يترتب عليه إهمال جزء من البيانات ومن ثم عدم دقة قيمة الوسط الحسابي في هذه الحالة . وربما من الافضل عدم استخدام الوسط الحسابي في مثل هذه الحالات ويتم استخدام مقياس آخر من مقاييس النزعة المركزية كالوسيط أو المنوال حيث لا يتطلب أى منهم معرفة مراكز الفئات المتطرفة (الاولى أو الاخيرة) كما سنرى عند عرضنا لبقية مقاييس النزعة المركزية في هذا الباب . أما إذا كانت هناك ضرورة ملحة لحساب الوسط الحسابي فإن هناك ثلاثة بدائل يمكن الأخذ بأحداها . وهذه البدائل هي :

(أ): إستبعاد الفئات المفتوحة بتكرارها وحساب الوسط الحسابى لبقية فئات التوزيع. وإن كان هذا البديل يؤدي 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 الأدنى 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 .

(ج): فى حالة التوزيعات التكرارية القريبة من التماثل فإنه يمكن ايجاد قيمة الوسط الحسابى بدلالة الوسيط والمنوال وذلك بموجب العلاقة التالية (لبيرسون) والتي تربط بين مقياس النزعة المركزية الثلاث :

$$3 \times \text{الوسيط} - \text{المنوال}$$

$$\frac{\text{الوسط الحسابى}}{2} =$$

وسوف نتعرض لتلك النقطة بالتفصيل فى نهاية هذا الباب وذلك عند الحديث عن العلاقة بين مقياس المتوسطات الثلاثة (الوسط والوسيط والمنوال) وإستخدامها فى الحكم على مدى تماثل التوزيع من عدمه . كما يؤخذ على الوسط الحسابى أنه لايمكن حسابه بيانيا بدقة إلا اذا كان التوزيع متماثلا تماما كما سنرى فيما بعد حيث يكون الوسط الحسابى متساويا مع كل من الوسيط والمنوال وتكون قيمته ماهى إلا نقطة تلاقى العمود النازل من قمة منحنى التوزيع على المحور الافقى وهو ماسنراه فيما بعد . كما أن من أهم عيوب الوسط الحسابى هو انه يتأثر بدرجة كبيرة جدا بالقيم الشاذة أو المتطرفة من البيانات . حيث يتحيز الوسط الحسابى لها

ويقترَب منها ومبتعد بذلك عن وسط المجموعة الحقيقي لدرجة أنه يصبح في بعض الأحيان غير ممثلاً للبيانات محل الدراسة ولذلك لا ينصح باستخدامه في حالة وجود القيم المتطرفة ، او في ظل وجود منحنيات أو توزيعات تكرارية شديدة الالتواء . فعلى سبيل المثال إذا افترضنا ان لدينا مرتبات خمسة موظفين هي:

185 ، 205 ، 195 ، 180 ، 80 جنيها فإن الوسط الحسابي لمرتباتهم هو 169 جنيها وهو يقل عن مرتب أربعة موظفين من بين الخمسة (أى 80 % منهم) وهذا يعنى أن الوسط الحسابي لا يمثل قيم الظاهرة تمثيلاً صحيحاً من حيث قيمة متوسطه (لاحظ أن القيمة المتطرفة في هذا المثال هي القيمة 80 جنيها).

** الوسيط (Q₂) Median

رأينا حالاً عند دراستنا للوسط الحسابي كأحد مقاييس النزعة المركزية انه إذا كانت هناك إحدى القيم الشاذة أو المتطرفة (سواء كبيرة جداً أم صغيرة جداً) مقارنة مع باقى المفردات للظاهرة محل الدراسة فإن الوسط الحسابي يتأثر بشدة بهذه القيم الشاذة أو المتطرفة لأن تلك القيمة تؤدي الى ميل وإقصاء الوسط الحسابي تجاهها مبتعداً بذلك عن الوسط الحقيقي للمجموعة مما يؤثر على جودة تمثيله لها كما أن هناك بعض البيانات والتي قد تكون في صورة وصفية أو ترتيبية ولا يوجد معنى لكلمة متوسط كما عرفناها في مثل هذه البيانات . فمثلاً إذا كان لدينا تقديرات مجموعة من الطلاب فليس هناك معنى لكلمة متوسط التقدير في تلك الحالة . فيأتى الوسيط هنا كأحد مقاييس النزعة المركزية ويعتبر مناسباً في مثل هذه الحالة حيث يمكن حسابه لكل أنواع البيانات وصفية أو كمية وحتى منها الترتيبية . كما أنه لا يتأثر بوجود القيم الشاذة أو المتطرفة او أن كان هناك أثراً فيكون أثراً طفيفاً .

هذا ويعرف الوسيط بأنه القيمة التي تتوسط البيانات وذلك بعد ترتيبها تصاعديا او تنازليا . أو بمعنى آخر فالوسيط هو عبارة عن قيمة المفردة التي تتوسط التوزيع بحيث يكون عدد المفردات الاقل منها فى القيمة مساويا لعدد المفردات الاكبر منها فى القيمة . وبمعنى ثالث فالوسيط هو القيمة التي 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 البيانات المبوبة تكراريا .

(1-2-3) : إيجاد الوسيط من البيانات الغير مبوبة (المفردات) :

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 رتبته (n/2) والثانية التي تليها ورتبتها هى (n/2)+1 ويكون الوسيط هو عبارة عن الوسط الحسابى لهاتين المفردتين . وفى حالة الظواهر الوصفية كالتقديرات مثلا فإذا كانت أحد المفردتين مثلا جيد والاخرى مقبول فنقول اصطلاحا بأن الوسيط لتك المجموعة هو ما بين الجيد والمقبول .

مثال (11): احسب الوسيط فى الحالات الثلاثة التالية :

(أ): لمجموعة المفردات : 5 ، 18 ، 9 ، 10 ، 15 ، 7 ، 20 .

- (ب): لمجموعة المفردات : 22 ، 27 ، 8 ، 12 ، 30 ، 4 ، 18 ، 25 ،
 (ج): لتقديرات الطلاب فى أحد الاختبارات : جيد - ضعيف - جيد جدا -
 ضعيف جدا - ممتاز - مقبول - جيد - جيد جدا - جيد .

الحل :

(أ): لايجاد الوسيط للمفردات السبعة المعطاه يتم ترتيب المفردات تصاعديا
 أو تنازليا وليكن الترتيب تصاعدى على الصورة :

$$20 ، 18 ، 15 ، 10 ، 9 ، 7 ، 5$$

وحيث أن عدد المفردات (n=7) وهو عددا فرديا لذا فإن ترتيب الوسيط هو
 : $th = \frac{(n+1)}{2} = \frac{(7+1)}{2} = 4$ أى المفردة الرابعة . وعليه فإن قيمة
 الوسيط = $Q_2 = 10$

(ب): بعد ترتيب المفردات تصاعديا تصبح كما يلى :

$$30 ، 27 ، 25 ، 22 ، 18 ، 12 ، 8 ، 4$$

وحيث ان عدد المفردات (=8) عدد زوجيا لذا فإن رتبتي 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 :

$$Q_2 = \frac{18+22}{2} = 20$$

(ج): لتحديد وسيط مجموعة التقديرات المعطاه يتم بداية ترتيب هذه
 المجموعة من التقديرات تصاعديا أو تنازليا . فيكون ترتيب التقديرات
 تصاعديا هو :-

ضعيف جدا - ضعيف - مقبول - جيد - جيد - جيد جدا - جيد جدا - ممتاز.

وحيث أن عدد المفردات هو $n=9$ لذا فإن ترتيب الوسيط هو $(n+1)/2$ أي $(9+1)/2=5$ أي المفردة الخامسة . ومن ثم فيكون الوسيط مساويا للتقدير "جيد".

وكما يتضح مما سبق أن الوسيط لا يتأثر بالقيم الشاذة أو المتطرفة إذ لا تدخل قيمتها بطبيعة الحال في حسابه بينما رأينا أن الوسط الحسابي كان يتحيز أو يميل باتجاهها . ولذلك يعتبر استخدام الوسيط في مثل هذه الحالات أفضل. وهذا لا يعني أن الوسيط مفضل دائما فأحيانا توجد بعض الصعوبات في حسابه كما في حالة المتغيرات المنفصلة والتي لا تأخذ قيما كسرية . فإذا فرضنا ان لدينا ستة أسر عدد افرادها هو 10 ، 8 ، 7 ، 6 ، 4 ، 3 فلما كان عدد المفردات زوجيا فإن الوسيط حسب التعريف السابق وهو الوسط الحسابي للمفردتين الثالثة والرابعة أي : $(6+7) / 2 = 6.5$ أي أن : $Q_2 = 6.5$

وهذه القيمة لا وجود لها ولا معنى لأن تكون قيمة الوسيط لعدد أفراد الاسرة هو 6.5 فردا. هذا بالإضافة الى أنه احيانا يفقد الوسيط دلالاته إذا كان عدد المفردات صغيرا وبينهما قيم كثيرة متكررة . فإذا كان لدينا درجات خمسة طلاب هي : 80 ، 72 ، 70 ، 70 ، 70 فإن الوسيط في هذه الحالة هو الدرجة 70 ولا توجد أي 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 في حساب ترتيب الوسيط ويقدر الوسيط في هذه الحالة على أنه قيمة المفردة ذات الترتيب $n/2$ أي كانت (n) عددا فرديا أم زوجيا وذلك توخيا للاقتصاد في الجهد الحسابي . ولعل السبب في ذلك هو انه في 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 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 الحصول على التوزيع التكرارى المتجمع الصاعد أو الهابط أولا . ثم يتم تحديد رتبة الوسيط والتي تعادل قسمة مجموع التكرارات على (2) ليتحدد بذلك رتبة الوسيط ومن ثم تحدد الفئة التى سيقع فيها قيمة الوسيط اما باستخدام 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 إليها . وعلى ذلك تتلخص خطوات تحديد قيمة الوسيط فى حالة الجداول التكرارية فيما يلى :-

(1): تكوين الجدول التكرارى المتجمع . وليكن التوزيع التكرارى المتجمع الصاعد .

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

(2): تحديد رتبة الوسيط وهى عبارة عن -----

2

حيث (r) عبارة عن عدد فئات التوزيع التكرارى المعطى .

(3): تتحدد الفئة الوسيطة بالفئة التي تقع فيها رتبة الوسيط والتي تم تحديدها في الخطوة السابقة من الجدول المتجمع ومن ثم يتم تحديد الحد الأدنى للفئة الوسيطة (Lower Bound of the Q₂ Class) وكذا التكرار المتجمع الصاعد السابق وليكن F(Pre.Asc) واللاحق وليكن F(Suc.Asc)

(4): يتم إيجاد قيمة الوسيط إما بقانون النسبة والتناسب كما سبق أو باستخدام قانون الوسيط حيث أن الوسيط (Q₂) يمكن تحديده من خلال القانون التالي :

$$Q_2 = L.B.of (Q_2)Class + \frac{F(Suc.Asc) - F(Pre.Asc)}{R.of (Q_2) - F(Pre.Asc)} \times \text{length of}(Q_2) \text{ Class} \quad (14)$$

وفي العلاقة (14) الأخيرة لاحظ أن :

* L.B.of (Q₂)Class تمثل الحد الأدنى للفئة الوسيطة

* R. of (Q₂) تمثل رتبة الوسيط

* F(pre.Asc) تمثل التكرار المتجمع السابق لرتبة الوسيط

* F (suc.Asc) تمثل التكرار المتجمع اللاحق لرتبة الوسيط

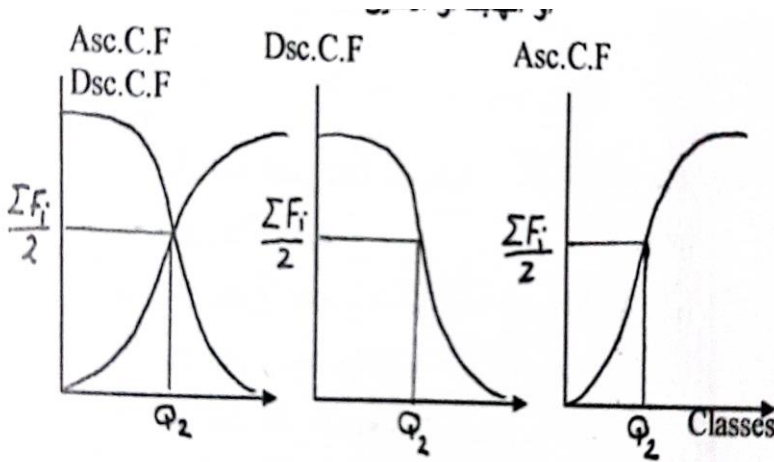
* Length of (Q₂)class تمثل طول الفئة الوسيطة

اما عن إيجاد قيمة الوسيط بيانيا من خلال رسم المنحنى التكرارى المتجمع الصاعد أو الهابط . ثم يتم تحديد رتبة الوسيط (R.of(Q₂)) وهى عبارة عن قسمة مجموع التكرارات (∑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 والتي تحدد قيمة الوسيط . أما إذا تم رسم المنحنى التكرارى المتجمعان الصاعد والهابط معا فى شكلا بيانيا واحدا. فإن إحداثيات نقطة تقاطع المنحنى تعطى قيمة الوسيط على المحور الأفقى وتعطى رتبة الوسيط على المحور الرأسى . أى أنه من نقطة تقاطع المنحنى المتجمعين إذا تم اسقاط عمودا على المحور الأفقى تحدد قيمة الوسيط ، أما إذا 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



مثال (12) : فيما يلى التوزيع التكرارى التالى يوضح الاجر اليومى (بالجنيه) لمجموعة من العاملين بإحدى الشركات :

جدول (14-3)

Classes	8-	18-	28-	38-	48-	58- 68	Σ
Frequency	8	22	25	20	13	12	100

والمطلوب : 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 (15-3) التالى :-

جدول(15-3)

التوزيع التكرارى المتجمع الصاعد للأجر اليومي 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
Σ	100	Less than68	100

ثم يتم تحديد رتبة الوسيط حيث ان :

$$\Sigma^r F_i$$

$$\text{Rank of } (Q_2) = \frac{\Sigma^r F_i}{2} = (100 / 2) = 50$$

ومن ثم فإن الحد الأدنى للفئة الوسيطة هو $L.B = 28$ والتكرار المتجمع السابق لرتبة الوسيط هو (30) والتكرار المتجمع اللاحق لرتبة الوسيط هو (55) وطول الفئة الوسيطة هو (10) جنيه ومن ثم فبالتعويض في العلاقة (14) السابقة فيكون :

$$Q_2 = 28 + \left(\frac{50 - 30}{55 - 30} \right) \times 10$$

$$= 28 + \left(\frac{20}{25} \right) \times 10 = 36 \quad (L.E)$$

أى أن وسيط الأجر اليومي للعامل في هذا التوزيع هو 36 جنيه .
ملحوظة : يمكن ايجاد قيمة الوسيط حسابيا بإستخدام قاعة النسبة والتناسب السابق الإشارة إليها وذلك بعد تحديد رتبة الوسيط فيكون لدينا الرسم الكروكي التالي :-

Less than 28 (L.E) ----- 30 persons
Less than (Q₂) (L.E)----- 50 persons
Less than 38 (L.E) ----- 55 persons

فيكون :

$$\frac{55 - 30}{55 - 50} = \frac{38 - 28}{38 - Q_2}$$

وعليه فإن :

$$10 = \frac{10}{5} \quad , \text{ i.e., } 5 (38 - Q_2) = 10 \quad , \quad 190 - 10 = 5Q_2$$

$$38 - Q_2$$

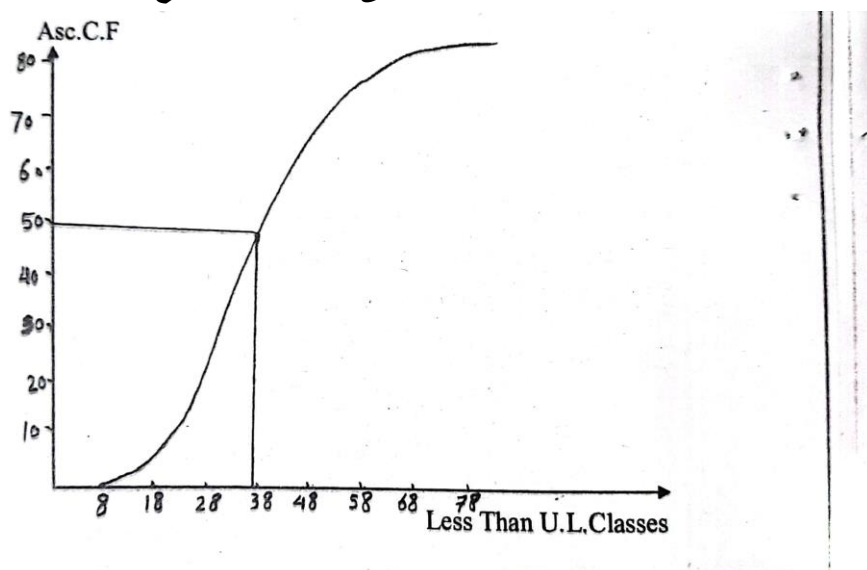
فتكون قيمة الوسيط هي :

$$Q_2 = \frac{180}{5} = 36 \quad (L.E)$$

وهي نفس النتيجة المستخلصة من خلال قانون الوسيط .
 أما عن ايجاد قيمة الوسيط بيانيا فسوف نحدد قيمة الوسيط من خلال رسم
 المنحنى التكرارى المتجمع الصاعد تاركين للقارئ تحديده باستخدام المنحنى
 التكرارى المتجمع الهابط بالإضافة لتحديد الوسيط باستخدام المنحنيين
 المتجمعين الصاعد والهابط معا .

شكل (3-2)

ايجاد قيمة الوسيط من خلال المنحنى التكرارى المتجمع الصاعد



ملاحظات حول الوسيط :-

من اهم الخصائص او المزايا التي يتمتع بها الوسيط عن غيره من مقاييس
 النزعة المركزية ما يلى :

- (1): يتميز الوسيط بأنه لا يتأثر بالقيم الشاذة (Outliers) أو المتطرفة (Extreme Points) لأنه من خلال تعريفنا للوسيط بأنه يعتمد على

المفردة أو القيمة التي تتوسط البيانات أى فى منتصفها . وعلى ذلك فهو يفضل على الوسط الحسابى فى حالة التوزيعات الملتوية أو شديدة الالتواء .

(2): يمكن ايجاد الوسيط فى حالة التوزيعات التكرارية المفتوحة مالم يكن التوزيع شديد الالتواء بحيث لا تقع رتبة الوسيط فى الفئة الاولى او الاخيرة من التوزيع .

(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:

$$\sum_{i=1}^n |x_i - Q_2| < \sum_{i=1}^n (x_i - A)$$

حيث أن (A) أى مقدار ثابت تختلف قيمته عن الوسيط . وهذه النتيجة غير شائعة الاستخدام فى المجالات التطبيقية .

أما عن عيوب الوسيط فتمثل فيما يلى :

(1): من أهم عيوب الوسيط أنه يعتمد على قيمة واحدة تقع فى منتصف المفردات وذلك بعد ترتيبها ولا يدخل فى حسابه باقى المفردات كما هو الحال فى الوسط الحسابى .

(2): كذلك يعاب على الوسيط أنه لا يخضع للعمليات الجبرية خضوع الوسط الحسابى . لذلك ترانا تجاوزنا الدقة التامة عند تحديد رتبة الوسيط فى حالة

التوزيعات التكرارية فأعتبرنا أن رتبة الوسيط هي ناتج خارج قسمة مجموع التكرارات على (2) سواء كان هذا المجموع ($\sum Fi$) عددا زوجيا أو فرديا .

(3):خلافًا لما سبق فإنه يمكن حساب الوسط الحسابي العام لعدة مجموعات جزئية أو عدة عينات باستخدام الأوساط الحسابية لهذه المجموعات أو تلك العينات . فإن هذا الأمر لا يتحقق في حالة الوسيط بمعنى إننا لا نستطيع حساب الوسيط العام لعدة مجموعات جزئية أو عدة عينات إذا علم الوسيط لكل مجموعة جزئية أو كل عينة.

** شبيهات الوسيط :

هناك بعض المقاييس الاحصائية التي تشبه الوسيط في تعريفها وفي طريق حسابها وهي :

• الربيعين **Quartiles**

• العشيرات **Deciles**

• المئنيات **Percentiles**

وسوف نتناول دراسة تلك المقاييس بشئ من التفصيل على النحو التالي :

* الربيعين (الربيع الأدنى والربيع الأعلى):-

ويطلق على الربيعين الأدنى والأعلى أسم الربيع الأول **First Quartile** والربيع الثالث **Third Quartile**.

هذا ويمكن تعريف الربيع الأدنى (أو الأول) أي (Q_1) بأنه القيمة التي يقل عنها ربع المفردات وكذلك القيمة التي يزيد عنها ثلاثة أرباع المفردات . أما الربيع الأعلى (أو الثالث) أي (Q_3) فيمكن تعريفه بأنه القيمة التي يقل عنها ثلاثة أرباع القيم ويزيد عنها ربع القيم . ومن خلال هذا المنطلق فإنه

يتم تعريف الوسيط بأنه بمثابة الربيع الثانى أى (Q_2) حيث ان القيمة التى يقل عنها نصف القيم ويزيد عنها النصف الاخر من تلك القيم . هذا ويتشابه الربيعين الادنى والاعلى مع الوسيط فى طريقة ايجادهما سواء كان ذلك بالطريقة الحسابية (سواء من خلال قانون الوسيط او قاعدة النسبة والتناسب) أو بالطريقة البيانية . ويمكن توضيح ذلك كما يلى :

(أ): ايجاد الربيعين (Q_3, Q_1) حسابيا :

لايجاد الربيعين حسابيا فإنه يمكن استخدام القانون (14) المستخدم فى حساب الوسيط مع اختلاف مسمى المقياس الذى يتم حسابه وذلك على النحو التالى :

$$Q_1 = L.B \text{ of } (Q_1)\text{class} + \left(\frac{R.of(Q_1) - F(\text{pre.Asc})}{F(\text{suc.Asc}) - F(\text{pre.Asc})} \right) \times \text{length of } (Q_1)\text{class} \quad (15)$$

$$Q_3 = L.B \text{ of } (Q_3)\text{class} + \left(\frac{R.of(Q_3) - F(\text{pre.Asc})}{F(\text{suc.Asc}) - F(\text{pre.Asc})} \right) \times \text{length of } (Q_3)\text{class} \quad (16)$$

هذا ويمكن استخدام قاعدة النسبة والتناسب السابق الاشارة اليها فى تحديد قيمة كل من الربيع الادنى (Q_1) والربيع الاعلى (Q_3) .

(ب): ايجاد قيمة الربيعين (Q_3, Q_1) بيانيا :-

يمكن ايجاد قيمة الربيعين (Q_1) ، (Q_3) بأسلوب مشابه تماما للأسلوب المتبع فى حالة الوسيط سواء كان ذلك باستخدام المنحنى التكرارى المتجمع الصاعد أو المتجمع الهابط أو المنحنيين معا . حيث يتم تحديد موقع رتبة الربيعين الادنى والاعلى 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 الأفقى ليقطعانه فى نقطتين الأولى تمثل قيمة الربيع الأدنى (Q1) والثانية تمثل قيمة الربيع الأعلى (Q3) .

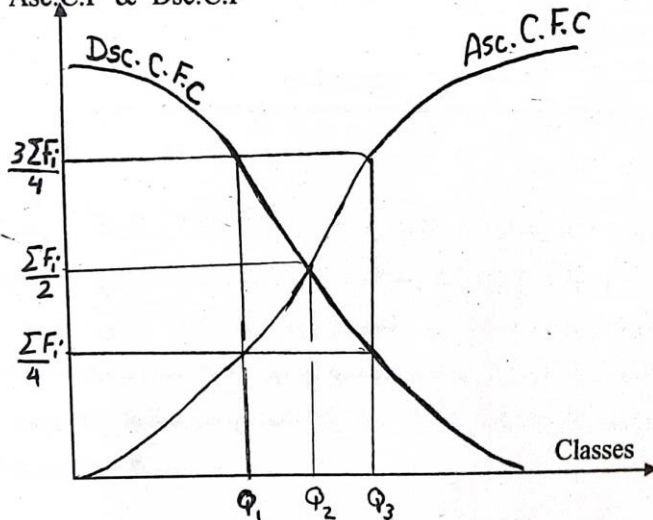
ملحوظة: - فى حالة استخدام المنحنى التكرارى المتجمع الهابط لاحظ أن من خلال دراستنا لمفهوم الربيعيين الأدنى والأعلى فإن :
الربيع الأدنى هو القيمة التى يقل عنها 25% من المفردات ويزيد عنها 75% من المفردات . أى أنه إذ تم رسم المنحنيين التكراريين المتجمعين الصاعد والهابط معا وبتحديد موقع النقطة التى تعبر عن ترتيب (Q1). فإذا تم منها رسم خطا موازيا للمحور الأفقى الى ان يقطع المنحنى التكرارى المتجمع الصاعد والهابط . فيكون العمود النازل من نقطة التقاطع مع المنحنى التكرارى المتجمع الصاعد على المحور الأفقى معبرة عن قيمة الربيع الأدنى . أما العمود النازل من نقطة التقاطع مع المنحنى التكرارى المتجمع الهابط فتعطى قيمة الربيع الثالث (Q3) والشكل البيانى (3-3) يوضح ذلك :

شكل (3-3)

ايجاد الربيعين Q_3, Q_1

من خلال رسم المنحنيين المتجمعيين الصاعد والهابط معا

Asc.C.F & Dsc.C.F



كذلك فإنه من نقطة ترتيب الربيع الاعلى $(\frac{3\sum F_i}{4})$ اذا تم رسم خطاً موازياً للمحور الافقى فإن هذا الخط يقطع أولاً المنحنى التكرارى المتجمع الهابط عند نقطة تقابل تماماً قيمة الربيع الاول Q_1 . أما نقطة تلاقى هذا الخط الموازى مع المنحنى التكرارى المتجمع الصاعد فهي تقابل قيمة الربيع الثالث Q_3 .

وفى نهاية تناولنا للربيعيين الادنى Q_1 والاعلى Q_3 فإنه يمكن القول بأن قيمتى الربيعين تعطيان مؤشراً 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 تصاعدياً مثلاً فإن الشكل (3-4) يوضح التعريف الذى تناولناه سابقاً لمعنى الربيعيين .

شكل (3-4)

نسبة المشاهدات أو المفردات التي تقع ما بين 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 الأدنى Q_1 وان هناك نفس النسبة أي 25% من المفردات تزيد قيمتها عن الربيع الأعلى Q_3 . وكذلك فإنه يمكن القول أن نصف عدد المشاهدات أو المفردات للظاهرة محل الدراسة تتراوح قيمتها فيما بين قيمتي Q_3, Q_1 . لذا فإن الحصول على قيمتي الربيعيين تمكننا من الحصول على معلومات أكثر مما لو استخدمنا قيمة الوسيط بمفرده .

مثال (13):

أوجد قيمة الربيعيين الأدنى والأعلى حسابيا بالطرق المختلفة وبيانيا للتوزيع التكراري المعطى في المثال السابق .

الحل:

أولاً: إيجاد قيم الربيعيين Q_3, Q_1 حسابيا :

- (1): باستخدام قانون الوسيط مع إجراء التعديل اللازم عليه . فبعد الحصول على الجدول التكراري المتجمع الصاعد كما هو مبين في الجدول (3-15) . يتم تحديد رتبة كل ربيع على حده على النحو التالي :

$$\text{Rank of } (Q_1) = \frac{\sum F_i}{4} \times 1 = \frac{100}{4} = 25 ,$$

$$\text{Rank of } (Q_3) = \frac{\sum F_i}{4} \times 3 = \frac{3 \times 100}{4} = 75 ,$$

ومن ثم فيتطبق كل من القانونين (15)،(16) السابق الاشارة اليهما فيكون :

$$Q_1 = L.B \text{ of } (Q_1)\text{class} + \left(\frac{R.of(Q_1) - F(\text{Pre.Asc})}{F(\text{Suc.Asc}) - F(\text{Pre.Asc})} \right) \times \text{length of } (Q_1)\text{class}$$

$$= 18 + \left(\frac{25 - 8}{30 - 8} \right) \times 10 = 18 + \left(\frac{17}{22} \right) \times 10 = 25.727 \text{ L.E}$$

أما عن قيمة Q_3 فالبنظر لعمود التكرارات المتجمعة الصاعدة في الجدول (3-15) يتضح لنا أن رتبة الربيع الاعلى (75) والتي تم تحديدها في الخطوة السابقة ماهى إلا أحد التكرارات المتجمعة الصاعدة وبالتحديد امام الفئة (48) جنيه في 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): ايجاد قيمة الربيعين باستخدام قاعدة النسبة والتناسب :-
وبالتحديد سوف يتم ايجاد قيمة Q_1 فقط باستخدام قاعدة النسبة والتناسب لأن قيمة Q_3 يتم تحديدها مباشرة فى هذا المثال كما سبق وأن أوضحنا .

فإيجاد قيمة Q_1 لاحظ ان الرتبة (25) الخاصة بترتيب Q_1 تقع ما بين التكرارين المتجمعيين الصاعدين 30,8 ومن ثم فئة الربيع الأدنى المقابلة لتلك التكرارت المتجمعة هي من (18) جنيها الى أقل من (28) جنيها.
أى أن :

Less than 18 (L.E) ----- 8 persons

Less than (Q_1) (L.E)----- 25 persons

Less than 28 (L.E) ----- 30 persons

ومن ثم فإن :

$$(28 - 18) / (28 - Q_1) = (30 - 8) / (30 - 25)$$

وعليه فإن :

$$\frac{10}{28 - Q_1} = \frac{22}{5}$$

أى أن :

$$50 = 22(28 - Q_1) , 22Q_1 = 616 - 50$$

فيكون :

$$Q_1 = \frac{566}{22} = 25.272 \quad (\text{L.E})$$

وهي نفس النتيجة السابقة .

ثانيا : إيجاد الربيعين Q_3, Q_1 بيانيا :

يمكن تحديد قيمة الربيعين الأدنى والأعلى بيانيا من خلال إما رسم المنحنى التكرارى المتجمع الصاعد أو رسم المنحنى التكرارى المتجمع الهابط أو رسم المنحنيين التكرارين المتجمعيين الصاعد والهابط معا . وسنترك للطالب

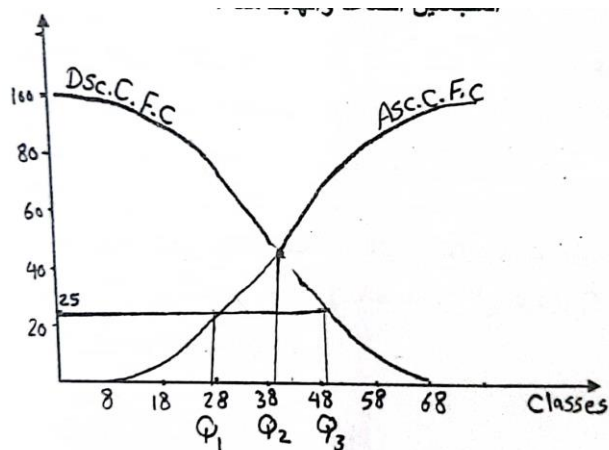
الطريقتين الاول والثانية ونقوم بإيجاد قيمة الربيعيين من خلال رسم

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.



* العشيريات :- Deciles

يمكننا تعريف العشير الاول (D₍₁₎) First Decile بأنه القيمة التى يقل عنها عشر القيم أو المفردات ويزيد عنها تسعة أعشار تلك القيم. وهكذا يمكن تعريف العشيريات الاخرى . هذا ويتم ايجاد العشيريات المختلفة بنفس الاسلوب المتبع فى حالة 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 التكررات ($\sum F_i$) على (10) ثم ضرب ناتج القسمة فى (3) والتى تمثل رتبة العشير الثالث. ويتم استخدام القانون (14) او (15) او (16) مع اجراء التعديلات اللازمة والتى تتلائم مع

مفهوم العشير المطلوب حسابه . هذا ويمكن ايضا ايجاد قيمة العشيريات باستخدام قانون النسبة والتناسب أو الرسم من خلال المنحنى التكرارى المتجمع الصاعد أو الهابط أو المنحنيين معا كما هو الحال فى حالة الوسيط والربيعين الادنى والاعلى .

مثال(14):

باستخدام 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 :

$$\text{Rank of } (D_2) = \frac{\sum F_i}{10} \times 2 = \frac{100}{10} \times 2 = 20 ,$$

كذلك فإن 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 الثاني والعشير الخامس فأمامنا أما الطرق الحسابية (القانون أو قاعدة النسبة والتناسب) بالإضافة للطريقة البيانية من خلال رسم اي منحني تكرارى متجمع . وللدقة دعنا نقوم بحساب العشيرين المطلوبين باستخدام قاعد النسبة والتناسب وذلك على النحو المبين التالى :

(*) فبالنسبة للعشير الثانى فإن الرتبة (20) تقع ما بين التكرارين المتجمعين الصاعدين 8 ، 30 فى جدول (3-15) وعليه فإن فئة العشير الثانى هى من 18 الى 28 جنيه . وعليه يكون لدينا الشكل الكروكى التالى :

Less than 18 (L.E) ----- 8 persons
 Less than (D2) (L.E)----- 20 persons
 Less than 28 (L.E) ----- 30 persons

وعليه فأن :

$$30-8 \quad 28-18$$

$$----- = -----$$

$$30-20 \quad 28 - D_2$$

فيكون :

$$22(28 - D_2) = 100$$

أى ان :

$$516$$

$$D_2 = \frac{516}{22} = 23.45 \quad (L.E)$$

ولاحظ أن قيمة العشير الثانى D_2 تقع ما بين حدى الفئات المتجمعة الصاعدة أى ما بين 18 ، 28 جنيه أى داخل فئة العشير السابق تحديدها . وهذا يعنى أن 20% من العمال تقل أجورهم عن 23.45 جنيه وأن 80% 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 :

$$D_5 = Q_2 = 36 \quad (\text{L.E})$$

وهو مايفيد أن 50% من العمال تقل أجورهم عن 36 جنيه وأن 50% من العمال تزيد أجورهم عن 36 جنيه.

* المئينيات :- Percentiles

بنفس الدلالة السابقة مع اختلاف المفهوم فيمكن تعريف المئين الاول P_1 بان القيمة التي يقل عنها 1% من عدد القيم ويزيد عنها 99% من باقى المفردات وعليه فإن المئين السابع والعشرون مثلا P_{27} وهو عبارة عن القيمة التي يقل عنها 27% من المفردات أو القيمة التي يزيد عنها 73% من باقى المفردات . وهكذا يمكن تعريف العديد من المئينيات بنفس الاسلوب . ولإيجاد قيم المئينيات يمكن ايضا استخدام قانون الوسيط (14) وذلك بعد اجراء التعديلات اللازمة عليه لكي يتفق ومفهوم المئين المطلوب . كما يمكناحساب المئينيات بإستخدام قاعدة النسبة والتناسب وذلك بعد تحديد رتبة المئين فى تلك الحالة . كما يمكن حساب المئينيات ببيانها من خلال رسم المنحنى التكرارى المتجمع الصاعد او الهابط أو من خلال رسم المنحنيين معا

مثال (15): باستخدام التوزيع التكراري التالي للأجر اليومي لـ 50 عاملاً (بالجنيه) بإحدى الشركات المطلوب حساب المئين الثامن والستون

جدول (3 - 16)

Classes	30-	40-	50-	60-	70-	80-	90-	Σ
Frequency	1	4	7	14	11	8	5	50

الحل :-

لحساب المئين الثامن والستون يجب أولاً إيجاد التوزيع التكراري للمتجمع الصاعد أو الهابط ثم تحديد القيمة سواء من خلال قانون الوسيط وشبهياته أو قاعدة النسبة والتناسب أو بيانياً . وسوف نقوم بحسابه من خلال القانون تاركين للطالب التأكد من صحة النتائج باستخدام قاعدة النسبة والتناسب أو الطريقة البيانية .

جدول (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 المئين الثامن والستون هي :

$$\text{Rank of } P_{(68)} = \left(\sum F_i / 100 \right) \times 68 = 34$$

والرتبة (34) تنحصر ما بين التكرارين المتجمعين الصاعدين 26 ، 37 . لذا فإن فئة المئين الثامن والستون هي من (70) حتى أقل من 80 جنييه وعليه فإن :-

$$\text{R.of}(P_{68}) - F(\text{Pre.Asc})$$

$$P_{68} = \text{L.B of } (P_{68})\text{Class} + \left(\frac{\text{F(Suc.Asc)} - \text{F(Pre.Asc)}}{\text{length of Class}} \right) \times \text{length of Class}$$

$$\text{F(Suc.Asc)} - \text{F(Pre.Asc)}$$

$$(34 - 26)$$

$$= 70 + \frac{\quad}{\quad} \times 10$$

$$(37 - 26)$$

$$= 70 + (8 / 11) \times 10 = 77.27 \quad (\text{L.E})$$

وهو ما يعنى أن 68% من إجمالي عدد العمال تقل أجورهم الاسبوعية عن 77.27 جنيه .

إستخدام المننيات :-

لاشك أن مفهوم المننيات 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 .

ففي هذه الحالة فإن قيمة المنين السابع والعشرون أي (27) P هي الدرجة التي لو حصل عليها الطالب 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 المنين الثامنة عشر أي (18) P .

(3): إذا كان الهدف من إجراء أحد أبحاث ميزانية الأسرة هو تحديد قيمة الدخل الشهري الذي يقل عنه 67% (على سبيل المثال) من إجمالي عدد الأسر التي شملها البحث . فإن قيمة هذا الدخل الشهري تتحدد بإيجاد المنين السابع والستون أي P(67) .

(4): المنين 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 .

ملاحظات على الوسيط وشبهاته :

بعد أن استعرضنا طرق إيجاد الوسيط وبعض المقاييس الشبيهة به مثل الربيعين الأدنى والأعلى والعشيرات والمنينيات فإنه تجدر الإشارة لبعض النقاط التالية :

(1): إنه وفقاً 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 تماماً 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 :

$$\frac{\text{الربيع الأدنى } (Q_1) + \text{الربيع الأعلى } (Q_3)}{2} = \text{الوسيط } (Q_2)$$

ومدى الإنحراف (أي القرب أو البعد) عن قيمة الوسيط الحقيقية والقيمة الناتجة من هذه العلاقة يعطي مؤشراً عن مدى تماثل أو 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 دائماً مساوية لمتوسط أي زوج من أزواج العشيريات أو المئينيات وأن كل زوج من العشيريات 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 من العشير الثاني والعشير الثامن يقع على بعد متساوي من الوسيط .
أي أن :-

$$\frac{\text{الوسيط } (Q_2) - \text{العشير الثاني } (D_2)}{\text{العشير الثامن } (D_8) - \text{الوسيط } (Q_2)}$$

ويمكنك عزيزي الطالب التحقق من صحة ذلك من خلال حساب تلك المقاييس 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-	Σ
Frequency	4	8	13	15	13	8	4	65

(3): في حين أن الوسيط لا يساوي الوسط الحسابي للربيعين الأدنى والأعلى أو الوسط الحسابي لأي زوج من أزواج العشيريات أو المننيات إلا في حالة التوزيعات المماثلة فقط إلا أن العلاقات التالية صحيحة دائماً لكافة التوزيعات (المتماثلة أو الملتوية) فيما يتعلق برتبة تلك المقاييس . فمثلاً :
: في أي توزيع تكراري نجد أن الوسط الحسابي لرتبتي أي زوج من تلك المقاييس يكون عبارة ترتيب الوسيط . فمثلاً :

$$\text{Rank of } (Q_1) + \text{Rank of } (Q_3)$$

$$\text{Rank of } (Q_2) = \frac{\text{Rank of } (Q_1) + \text{Rank of } (Q_3)}{2}$$

وذلك سواء كان التوزيع متماثلاً أو غير متماثل . ويمكن اثبات ذلك كما يلي :
: فحيث أن :

$$\begin{aligned} \frac{\text{Rank of } (Q_1) + \text{Rank of } (Q_3)}{2} &= \frac{\sum F_i + 3 \sum F_i}{2} \\ &= \frac{4 \times \sum F_i}{2} \\ &= \frac{\sum F_i}{2} \\ &= \text{Rank of } (Q_2) \end{aligned}$$

(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 والمئين الستون ترتيبه هو 0.6 أو 60 % وهكذا .

** 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 (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 : 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 حسب حالتهم التعليمية وممارستهم للأنشطة الرياضية كما يلي :-

جدول (3 - 19)

أمي	متعلم	الحالة التعليمية
		ممارسة الرياضة
95	110	يمارس الأنشطة الرياضية
100	195	لا يمارس الأنشطة الرياضية

فإن منوال هذا التوزيع التكراري المزدوج هو " متعلم ولا يمارس 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 لأي مجموعة من البيانات قد لا يوجد لها قيمة منوالية . وإن وجد فقد يكون المنوال وحيد القيمة أو ربما تتعدد القيم المنوالية لمجموعة البيانات وذلك بعكس الحال في حالة الوسط الحسابي والوسيط

وشبّهاته حيث نجد أن لأي مجموعة من البيانات لابد وأن يكون لها وسط حسابي واحد أو وسيط واحد .

لذا فإنه جدير بالذكر والملاحظة فإن دور المنوال في علم الاحصاء لا يرقى إلى دور الوسط الحسابي أو الوسيط كمقياس للنزعة المركزية . إذ يهتم بالمنوال عادة أصحاب الأعمال الصناعية والتجارية والدعاية والتسويق . فالقيمة الأكثر تكرار لها مغزى اقتصادي هام بالنسبة لهم . فالمنتج الأكثر رواجاً في صناعة معينة يجذب اهتمام أصحاب هذه الصناعة لزيادة انتاجه وعمل المزيد من الدعاية له . كما يهتم الباحثون في العلوم السلوكية بالمنوال باعتباره قابلاً للحساب في جميع أنواع البيانات .

هذا ويمكن تقدير قيمة المنوال من التوزيعات التكرارية (البيانات المبوبة) شكل 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) تتمثل كما ذكرنا حالاً في مجرد البحث عن تلك المفردة التي تتكرر أكثر من غيرها . ومن ثم فهي تنطوي على قدر كبير من البساطة والوضوح . ولكن الأمر ليس كذلك في حالة البيانات المبوبة في جداول تكرارية . إذ أن الجدول لبيانات ظاهرة ما لا يعطي حقيقة كل مفردة على حدة وبالتالي لا يمكن تحديد تلك القيمة التي تتكرر أكثر من غيرها . وأفضل ما يمكن عمله في هذا الشأن كتحديد قيمة المنوال هو اختيار تلك الفئة التي تحتوي على أكبر تكرار أصلي

(في حالة الجداول المنتظمة أي المتساوية من حيث أطول الفئات) أو أكبر تكرار معدل (في حالة الجداول الغير منتظمة وهي ما تسمى بالفئة المنوالية Model Class . وبعد ذلك يمكن تحديد قيمة المنوال داخل

الفئة المنوالية باستخدام عدة أساليب فتزيد قيمة المنوال عن بداية الفئة المنوالية وتقل عن نهايتها . هذا ويميل أو ينحاز المنوال إلى بداية أو نهاية الفئة المنوالية تبعاً لتكرار الفئتين السابقتين للفئة المنوالية واللاحقة لها . ويتم تحديد قيمة المنوال داخل الفئة المنوالية بإحدى الطرق الثلاث التالية :

أولاً : طريقة مركز الفئة المنوالية :-

فكما سبق وأن أوضحنا يمكن تعريف الفئة المنوالية على أنها الفئة المقابلة لأكبر تكرار أصلي (في حالة التوزيعات التكرارية المنتظمة) أو أكبر تكرار معدل (في حالة التوزيعات التكرارية الغير منتظمة) . ووفقاً لطريقة مركز الفئة المنوالية يتم تحديد الفئة المنوالية ومن ثم يتم تحديد الحد الأدنى والأعلى للفئة المنوالية وبالتالي يمكن تحديد مركز هذه الفئة فتكون هي بمثابة قيمة المنوال . أي أن النوال (Mode) أو (M) هو عبارة عن :

(L.B + U. B) for the Modal Class

$$\text{Mode}(M) = \frac{\quad}{2}$$

2

حيث : L.B هو عبارة عن الحد الأدنى للفئة المنوالية ،
U.B هو عبارة عن الأعلى للفئة المنوالية .

مثال (16) :-

الجدول التالي يبين توزيع أعمار 60 بطارية حسب مدة صلاحية كل منها بالشهور :

جدول (3-20)

Classes	6-	9-	12-	15-	18-21	Σ
Frequency	4	12	20	18	6	60

والمطلوب حدد قيمة المنوال مستخدماً طريقة الفئة المنوالية .

الحل:

بداية وبملاحظة طول الفئة في هذا التوزيع يتبين لنا أن الجدول 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 من (15) شهر ومن ثم فإن المنوال هو عبارة عن مركز تلك الفئة. أى أن :

(L.B + U. B) for the Model Class

$$\text{Mode}(M) = \frac{\quad}{2}$$

$$= \frac{12 + 15}{2} = \frac{27}{2} = 13.5 \text{ (month)}$$

مستخدمين في ذلك طريقة مركز الفئة المنوالية كأحد طرق حساب المنوال .
ملحوظة :- إذا كان التوزيع التكراري غير منتظم فلا بد من حساب التكرارات المعدلة أولاً . وتكون الفئة المنوالية هي الفئة المقابلة لأكبر تكرار معدل .

مثال (17) : الجدول التالي يبين التوزيع التكراري لدرجات مائة طالب في مادة الإحصاء :

جدول (3-21)

Classes	10-	-25	-30	-40	-50	-60	-80	90-100	∑
Frequency	5	7	13	21	19	16	10	9	60

والمطلوب حساب المنوال مستخدماً طريقة مركز الفئة المنوالية .

الحل : حيث أن الجدول (3-21) توزيع تكراري غير متساوي من حيث طول الفئة (توزيع غير منتظم. لذا فإنه يجب حساب التكرارات المعدلة أولاً . وفيما يلي الجدول التالي يوضح التكرارات المعدلة لهذا التوزيع .

جدول (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
Σ	100		

ومن خلال عمود التكرارات المعدلة يتضح لنا أن أكبر تكرار معدل هو (2.1) لذا فالفئة المقابلة له (الفئة المنوالية) هي من (40) حتى أقل من (50) . لذا فإن قيمة المنوال باستخدام طريقة مركز الفئة المنوالية هي :

$$(M) = \frac{(L.B + u. B)}{2} = \frac{(40 + 50)}{2} = 45 \text{ (degree)}$$

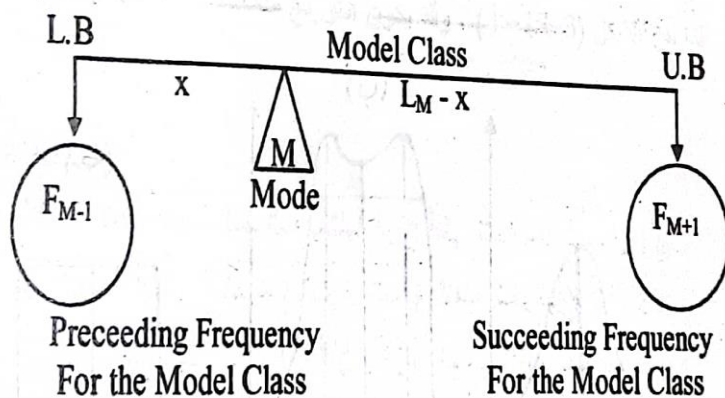
ثانياً: طريقة الرافعة لحساب المنوال :

وهذه الطريقة تعتبر الفئة المنوالية بمثابة ذراع أفقي تتجاذبه قوتان (رافعة من الدرجة الأولى) أحدهما التكرار السابق للفئة المنوالية و الأخرى التكرار اللاحق للفئة المنوالية. فإذا كان التكرار السابق للفئة المنوالية أكبر من التكرار اللاحق للفئة المنوالية فإن قيمة المنوال تكون قريبة من

الحد الأدنى للفئة المنوالية . والعكس صحيح أي أن إذا كان تكرار الفئة اللاحقة للفئة المنوالية أكبر من تكرار الفئة السابقة للفئة المنوالية فإن قيمة المنوال تكون قريبة من الحد الأعلى للفئة المنوالية . أما إذا تساوت القوتان أي تساوي التكرار السابق واللاحق للفئة المنوالية فإن قيمة المنوال تكون عند مركز الفئة المنوالية .

فإذا رمزنا لتكرار الفئة المنوالية بالرمز F_M ولتكرار الفئة السابقة للفئة المنوالية بالرمز $F_{(M-1)}$ ولتكرار الفئة اللاحقة للفئة المنوالية بالرمز $F_{(M+1)}$ وللحد الأدنى للفئة المنوالية بالرمز $(L.B)_M$ ولطول الفئة المنوالية بالرمز L_M . فإن يمكن تصور شكل الرافعة من الدرجة الأولى كما يوضحها الشكل (6-3) . حيث يمثل هذا الشكل خطأً أفقياً يمثل رافعة طولها يساوي 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 قدرها (X) عن بداية الفئة المنوالية فيكون بعد المنوال عن نهاية الفئة المنوالية ما هو إلا طول الفئة المنوالية أي L_M مطروحاً منها البعد (X) كما هو موضح بالشكل الكرويكي (6-3).

شكل (6-3)



هذا وبتطبيق قانون الرافعة والذي يحوي في مؤداه كما نعلم أن:-

$$\text{القوة} \times \text{ذراعها} = \text{المقاومة} \times \text{ذراعها}$$

فنستطيع الحصول على البعد (\times) كآلاتي :-

$$X \cdot (F_{M-1}) = (L_{M-x}) (F_{M+1}) \quad (\text{I})$$

$$X \cdot (F_{M-1}) + F_{M+1} = L_M (F_{M+1}) \quad \text{أي أن}$$

ومنها فإن :

$$X = \frac{F_{M+1}}{(F_{M-1} + F_{M+1})} \times (L_M) \quad (\text{II})$$

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 :

$$M = L \cdot B \text{ for the Model Class} + x$$

أو باستخدام الرموز الواردة في العلاقة II فإن:

$$M = L \cdot B \text{ (Model Class)} + \frac{F_{M+1}}{F_{M-1} + F_{M+1}} \times L_M \quad (\text{III})$$

هذا ويجب أن ننوه بالتأكيد على أنه قبل حساب المنوال بأي من طرق حسابه يجب التحقق من مدى انتظام الجدول التكراري من عدمه . بمعنى إذا كان الجدول منتظماً يتم حساب المنوال مباشرة وإلا فيجب حساب التكرارات المعدلة (في حالة عدم الانتظام) ثم حساب المنوال . وهذه الطريقة وإن كانت تتميز بأنها تأخذ في الاعتبار التكرار السابق والتكرار اللاحق للفئة المنوالية إلا أن من أهم عيوبها أنها تتجاهل تكرار الفئة المنوالية ذاتها .

مثال (18) : احسب قيمة المنوال باستخدام طريقة الرافعة للتوزيع التكراري المعطى في مثال (16) .

الحل :-

بداية حيث أن التوزيع التكراري الوارد في مثال (16) توزيعاً تكرارياً منتظماً لذا يتم حساب المنوال مباشرة . وباستخدام طريقة الرافعة فإنه من خلال الجدول (3-20) يتضح لنا الآتي :

Model Class = (12-15) - الفئة المنوالية

$L_M = 3$ - طول الفئة المنوالية

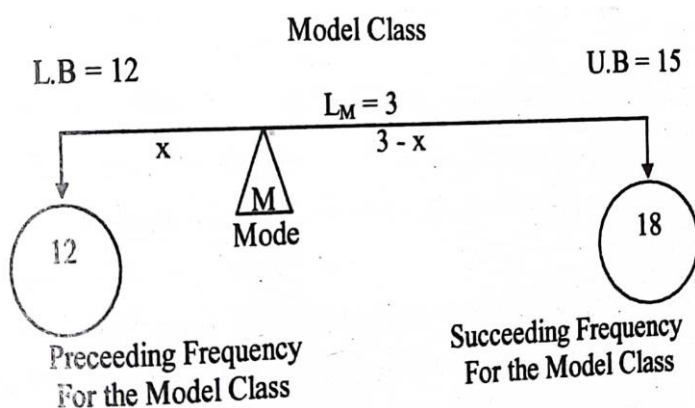
$F_{M-1} = 12$ - التكرار السابق للفئة المنوالية

$F_{M+1} = 18$ - التكرار اللاحق للفئة المنوالية

L . B for the Model Class = 12 - الحد الأدنى للفئة المنوالية

وبافتراض أن المنوال (M) هو نقطة التوازن لرافعة من الدرجة الأولى ويبعد مسافة قدرها (x) عن بداية الفئة المنوالية فنكون بصدد الشكل التالي :

شكل (3 - 7)



فيكون :

$$12 X = 18 (3 - x)$$

$$30 X = 45$$

$$\text{i.e., } X = 54 \div 30 = 1.8 \text{ (month)}$$

فتكون قيمة المنوال .

$$M = L . B \text{ for (Model Class)} + X$$

$$= 12 + 1.8 = 13.8 \quad \text{(month)}$$

مثال (19) :- احسب قيمة المنوال باستخدام طريقة الرافعة للتوزيع التكراري المعطى في مثال (17) .

الحل :-

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 البيانات اللازمة لتحديد قيمة المنوال ومن هذا الجدول فإن :

$$\text{Model Class} = (40 - 50)$$

$$L_M = 10$$

$$F_{(M-1)} = 1.3$$

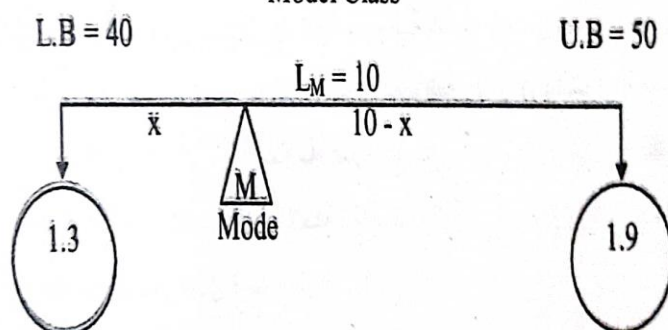
$$F_{(M+1)} = 1.9$$

$$L . B \text{ for the Model Class} = 40$$

وعليه فإنه يمكن تصور صورة الرافعة في الشكل (3-8) . فبافتراض أن المنوال (M) هو نقطة التوازن لرافعة من الدرجة الأولى ويبعد مسافة قدرها (x) عن بداية الفئة المنوالية فنكون بصدد الشكل التالي:

شكل (8 - 3)

Model Class



ولاستنتاج قيمة x فحيث أن

القوة \times ذراعها = المقاومة \times 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 :

$$M = L . B (\text{Model class}) + x$$

$$= 40 + 5.9375$$

$$= 45.9375 = 45.94 (\text{degree})$$

لاحظ أنه في مثالنا هذا وبمجرد النظر إلى بيانات الجدول بعد تعديل التكرارات كما هو في جدول (22-3) فيمكن استنتاج أن قيمة المنوال لابد وأن تزيد عن 45 درجة (وهو بمثابة مركز الفئة المنوالية) وذلك لأن التكرار المعدل اللاحق للفئة المنوالية أكبر من التكرار المعدل السابق للفئة المنوالية .

ثالثاً : طريقة الفروق لبيرسون لحساب المنوال :-

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 .

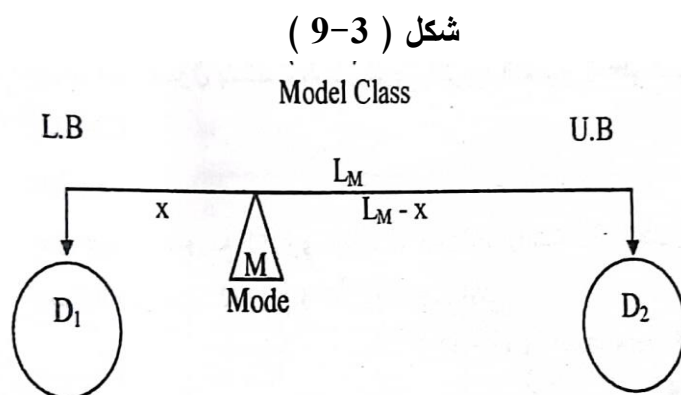
فإذا اعتبرنا أن الفرق الأول D_1 هو عبارة عن القوة أي أن :

$$D_1 = F_M - F_{M-1}$$

وأن الفرق الثاني D_2 هو عبارة عن المقاومة أي أن :-

$$D_2 = F_M - F_{M+1}$$

فتأخذ الرافعة الشكل البياني التالي :-



فيكون :

$$\frac{X}{D_2} = \frac{L_M - x}{D_2}$$

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 :

$$\begin{aligned} M &= L . B \text{ for (Model Class) } + x \\ &= L . B (M) + L_M \times \left(\frac{D_1}{D_1 + D_2} \right) \quad \text{(IV)} \end{aligned}$$

والأمثلة التالية توضح كيفية حساب المنوال باستخدام طريقة الفروق لبيرسون.

مثال (20) : احسب قيمة المنوال باستخدام طريقة الفروق للتوزيع التكراري المعطى في مثال (16) .

الحل :-

بداية حيث أن التوزيع التكراري منتظماً لذا يمكن حساب المنوال مباشرة. فمن خلال الجدول (3-20) يتضح لنا البيانات التالية :

$$\begin{aligned} \text{Model Class} &= (12 - 15) \\ L_M &= 3 \\ D_1 &= 20 - 12 = 8 \\ D_2 &= 20 - 18 = 2 \end{aligned}$$

هذا ومن خلال العلاقة السابقة (IV) فإن قيمة المنوال تكون عبارة عن :

$$\begin{aligned} M &= 12 + 3 \left(\frac{8}{8 + 2} \right) \\ &= 12 + 2.4 = 14.4 \quad \text{(month)} \end{aligned}$$

مثال (21) : احسب قيمة المنوال باستخدام طريقة الفروق للتوزيع التكراري المعطى في مثال 17 .

الحل :-

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 :

$$\text{Model Class} = (40 - 50)$$

$$L_M = 10$$

$$D_1 = F_M - F_{M-1} = 2.1 - 1.3 = 0.8$$

$$D_2 = F_M - F_{M+1} = 2.1 - 1.9 = 0.2$$

وعليه يتم استنتاج المنوال من الصيغة التالية باستخدام طريقة الفروق لبيرسون والتي تأخذ الصورة :

$$D_1$$

$$M = L . B \text{ for (Model Class)} + L_M \times \left(\frac{\quad}{D_1 + D_2} \right)$$

$$0.8$$

$$= 40 + 10 \left(\frac{\quad}{\quad} \right)$$

$$(0.8 + 0.2)$$

$$= 40 + 8 = 48 \quad (\text{degree})$$

ملاحظات :-

(1): بمراجعة نتائجننا في الأمثلة السابقة الخاصة بالمنوال لاحظ أن كل طريقة من الطرق الثلاث (مركز الفئة المنوالية - الرافعة - الفروق) أعطت قيمة مختلفة للمنوال وهذا أمراً طبيعياً ويرجع ذلك إلى أن هذه

التوزيعات التي تم إيجاد المنوال لها توزيعات غير متماثلة فالاختلاف ما بين النتائج يرجع لاختلاف طرق حساب المنوال في كل واحدة عن الأخرى ونتيجة لعدم تساوي تكرار الفئة السابقة للفئة المنوالية مع تكرار الفئة اللاحقة لها أدى ذلك إلى اختلاف النتائج فيما بين الطرق الثلاث . وتتساوى نتائج الطرق الثلاث في حالتي التوزيعات المتماثلة أو التوزيعات التي تتساوى فيها التكرار الأصلية السابقة (في حالة الجداول المنتظمة) واللاحقة للفئة المنوالية أو في 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 على طريقة الرافعة من حيث تقديرها لقيمة المنوال وذلك لأن طريقة الفروق تستخدم معلومات أكثر من تلك التي تستخدمها طريقة الرافعة في تحديد قيمة المنوال .

(3): إذا تساوى التكرار السابق للفئة المنوالية أي (F_{M-1}) مع التكرار اللاحق للفئة المنوالية أي (F_{M+1}) فإن بعد المنوال عن بداية الفئة المنوالية (أي قيمة (x) سوف يأخذ نفس القيمة في الطرق الثلاث . وبالتالي فإن كل من هذه الطرق سيعطي نفس القيمة للمنوال وتكون قيمة x في هذه الحالة عبارة عن خارج قسمة طول الفئة المنوالية أي L_M على (2) أي مركز الفئة المنوالية . وبمعنى أكثر بساطة إذا تساوى التكرار السابق واللاحق للفئة المنوالية فإن قيمة المنوال للطرق الثلاث تكون متساوية وتقدر قيمة المنوال في هذه الحالة بمركز الفئة المنوالية . أي أنه إذا كان :

$$F_{M-1} = F_{M+1}$$

فإن :

(L . B + U.B) for the Model Class

$$\text{Mode} = \left(\frac{\quad}{\quad} \right)$$

2

(3-3-2) إيجاد قيمة المنوال بيانياً من خلال المدرج التكراري :

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 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 :

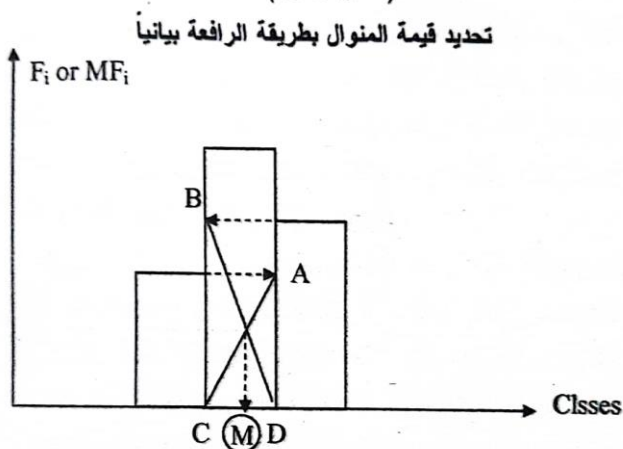
(1): وفقاً لطريقة الرافعة :-

بعد رسم الفئات الثلاث محل الاهتمام في المدرج التكراري (الفئة المنوالية والفئة السابقة والفئة اللاحقة للفئة المنوالية) حسب مقياس الرسم المناسب يتم استكمال قيم المستطيلات الممثلة للفئتين السابقة واللاحقة للفئة المنوالية بخطوط مستقيمة في اتجاه الفئة المنوالية إلى أن يقطعها 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 لتحديد هذه النقطة قيمة المنوال (M) بطريقة الرافعة كما هو موضح في الشكل (10-3) .

(شكل 10-3)

تحديد قيمة المنوال بطريقة الرافعة بيانياً



(ب) وفقاً لطريقة الفروق :-

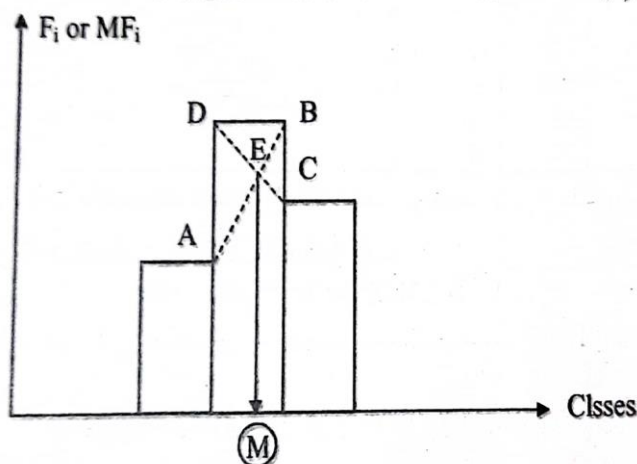
يتم توصيل القمم العليا للمستطيلات الثلاثة المتلاصقة بخطين مستقيمين بطريقة عكسية . حيث يتم توصيل نهاية قمة المستطيل الممثل للفئة المنوالية بخط مستقيم مع نهاية قمة المستطيل الممثل لنهاية الفئة السابقة للفئة المنوالية . ثم يتم توصيل بداية قمة الفئة المنوالية ببداية قمة الفئة اللاحقة للفئة المنوالية بخط 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 (11-3) .

شكل (11-3)

إيجاد قيمة المنوال بطريقة الفروق بيانياً

إيجاد قيمة المنوال بطريقة الفروق بيانياً



وعلى الطالب إيجاد قيمة المنوال بيانياً للمثالين (16) ، (17) باستخدام كل من طريقتي الرافعة والفروق .

مثال (22) :

فيما يلي لديك الجدول التالي يعبر عن توزيع الدخل اليومي (بالجنيه) لمجموعة من العاملين بإحدى الشركات :

جدول (3 - 24)

Daily Income Classes	5 - 10	10 - 20	20 - 40	40 - 50	50 - 60
Frequency	30	150	240	50	20

والمطلوب إيجاد قيمة المنوال حسابياً بكافة الطرق المستخدمة لحسابه.

الحل :- بملاحظة الجدول التكراري المعطى نجد أنه توزيعاً تكرارياً غير منتظم لذا يجب الحصول على التكرارات المعدلة قبل حساب المنوال .
والجدول (25-3) يعطي الحسابات اللازمة لحساب التكرارات المعدلة :

جدول (25-3)

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 (25-3) فإن حساب المنوال باستخدام :-

أولاً : طريقة مركز الفئة المنوالية فإن :

($L . B + U.B$) for the Model Class

$$\text{Mode}(M) = \left(\frac{\quad}{2} \right)$$

وحيث أن أكبر تكرار معدل هو (15) لذا فإن الفئة المنوالية هي :-

$$\text{Model Class} = (10- 20)$$

وعليه فإن المنوال :

$$M = \frac{(10 + 20)}{2} = 15$$

ثانياً : طريقة الرافعة فإن :-

من خلال الجدول السابق يتضح لنا أن :

$$\text{Model Class} = (10- 20)$$

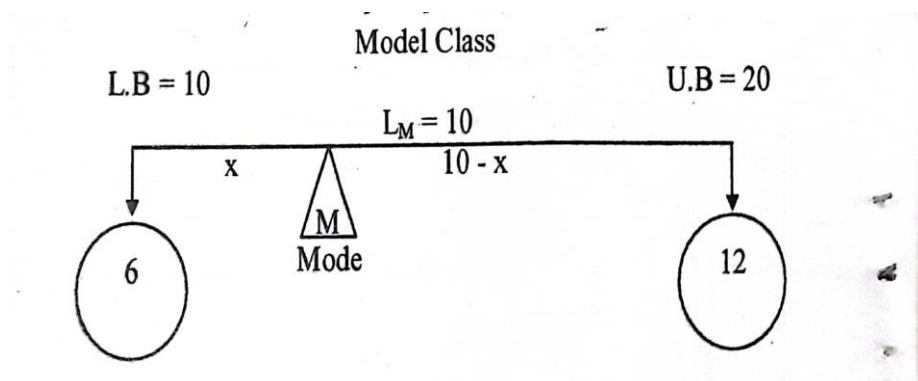
$$L . B \text{ for the Model Class} = 10$$

$$F_{M-1} = 6$$

$$F_M = 15$$

$$F_{M+1} = 12$$

وعليه فإنه يمكن تصور الرسم الكروكي التالي :



وطبقاً لقاعدة الرافعة فإن :

$$6x = 12(10-x)$$

$$\text{i.e., } 18x = 120$$

وعليه فإن قيمة :

$$X = (120 / 18) = 6.667 \quad (\text{L.E})$$

فتكون قيمة المنوال عبارة عن :

$$M = L . B \text{ for the Model Class} + x$$

$$= 10 + 6.667$$

$$= 16.667 \quad (\text{L.E})$$

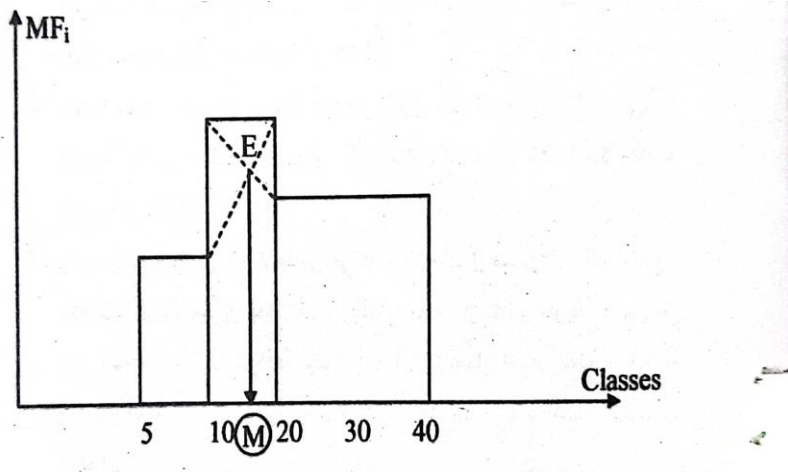
وبيانياً باستخدام طريقة الرافعة يتم رسم المدرج أو بالتحديد الفئات محل

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 $M = 16.667$

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 :-

$$D_1$$

$$M = L . B \text{ for the Model Class} + L_M \left(\frac{D_1}{D_1 + D_2} \right)$$

$$= 10 + 10 \left(\frac{9}{9 + 3} \right)$$

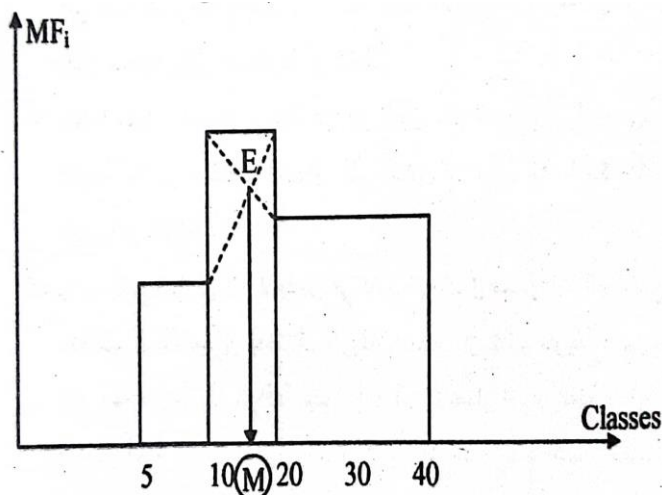
وعليه فإن :

$$= 10 + \frac{90}{12} = 17.5 \quad (L . E)$$

أما عن تقدير المنوال بيانياً باستخدام طريقة الفروق فالشكل (13-3) يوضح إيجاد قيمة المنوال :

شكل (13-3)

تحديد قيمة المنوال باستخدام الفروق بيانياً



ومن الرسم يتضح أن :

$$M = 17.5 \quad L.E$$

ملاحظات حول المنوال :

(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 يفوق عدد من المفردات المحققة لأي صفة أخرى . ولا تعني بالضرورة أنها الصفة التي 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 في التوزيعات التكرارية إذ أننا نحتاج إلى الفئة المنوالية والفئتين السابقتين للفئة المنوالية واللاحقة لها ، ولا يتأثر حسابه بالقيم المتطرفة أو الشاذة .

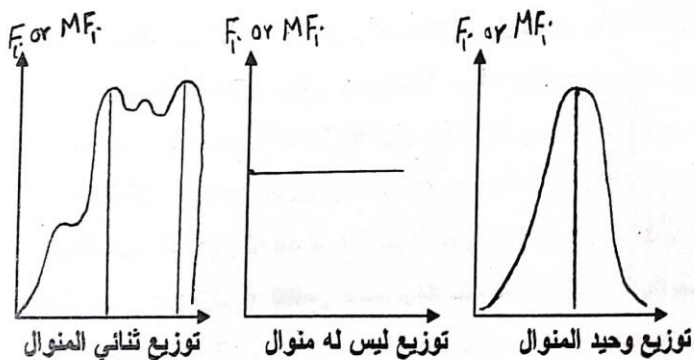
(4): حينما تكون القياسات نوعية كما هو الحال في ظاهرة النوع أو الجنسية أو الديانة فيفضل استخدام المنوال على أساس أنه يمثل الظاهرة الأكثر ظهوراً أو تكراراً .

(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 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 - ج) فهذا يعني وجود منوالين أو أكثر لهذا التوزيع كما هو في الحالتين الأخيرتين .

شكل (3-14)

(لا يوجد منوال أو يوجد أكثر من منوال وهنا لا يصلح
المنوال لأن يكون مقياساً للنزعة المركزية).

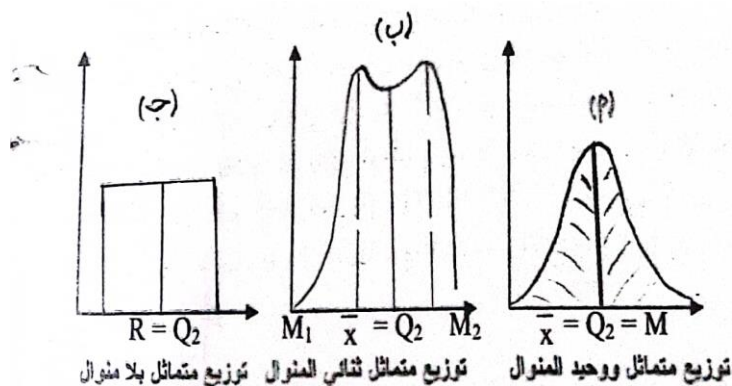


** العلاقة بين الوسط الحسابي والوسيط والمنوال :

رأينا فيما سبق أن الوسط الحسابي يدخل في حسابه كل القيم لذلك فإنه يتأثر تأثراً بالغاً بالقيم الشاذة أو المتطرفة ، أما الوسيط فيتحدد من خلال المواقع النسبية للقيم بعضها من بعض أي أنه يتحدد من خلال رتب هذه القيم ، ولنأخذ على سبيل المثال القيم 10 , 11 , 12 , 13 , 14 فوسطها الحسابي ووسيطها يساوي 12 فإذا أضفنا إليها قيمة واحدة ولتكن 120 نجد أن الوسط الحسابي سيصبح = (6 \ 180) = 30 بينما الوسيط سيصبح 12.5 فالوسط الحسابي قد زاد بمقدار 18 في حين زاد الوسيط بمقدار 0.5 فإضافة المفردة السادسة إلى مجموعة القيم لن تزيد قيمة الوسيط إلا بمقدار نصف مهما كان قيمة تلك المفردة ولكن الزيادة في الوسط الحسابي ستصبح أكثر كلما زادت فيه 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 متمائلاً ووحيد المنوال ويسمى التوزيع فى هذه الحالة بالتوزيع المتماثل **Symmetric Distribution** وهنا يكون المنحنى التكراري للتوزيع ينقسم عند القيمة المتوسطة له إلى قسمين متطابقين ويكون تزايد أو تناقص التكرارات متشابهاً ومنتظماً بطريقة متمائلة على جانبي المحور الرأسي المقام عند القيمة المتوسطة كما يتضح لنا من الشكل (3-15-أ) . وقد يكون التوزيع متمائلاً

شكل (3-15)



وثنائي المنوال ، وفي هذه الحالة فإن الوسط الحسابي يساوي الوسيط بينما يوجد منوالين يختلفان عن قيمة الوسط الحسابي والوسيط كما يتضح من الشكل (3-15-ب) . وقد يكون التوزيع متمائلاً بلا منوال فيظل الوسط الحسابي مساوياً للوسيط كما في الشكل (3-15-ج) .

وفي حالة التوزيعات الملتوية حيث تميل التكرارات الكبيرة إلى التركيز في ناحية المتوسط وتنتشر التكرارات الصغيرة بعيداً على شكل ذيل في الجانب الآخر منه ويكون اتجاه الذيل هو اتجاه الالتواء ، فإذا كان الذيل على

اليسار قلنا أن التوزيع ملتو إلى اليسار أي سالب الالتواء Negative Skewness أما إذا كان الذيل على اليمين قلنا أن التوزيع ملتو إلى اليمين أي موجب الالتواء Positive Skewness وفي التوزيعات الملتوية يقع الوسيط دائماً بين الوسط الحسابي والمنوال ، والوسط الحسابي - كما رأينا - شديد الحساسية للقيم الشاذة لذلك نراه مائلاً إلى اتجاه الذيل بينما يرشح المنوال نفسه عند قمة التوزيع .

فإذا كان التوزيع ملتوياً جهة اليسار (أي سالب الالتواء) حيث تميل التكرارات الكبيرة إلى التركيز عند فئات التوزيع العليا ويكون اتجاه الذيل على اليسار كما يتضح من الشكل (3-16-أ) وهنا تكون العلاقة فيما بين مقاييس النزعة المركزية الثلاثة كالآتي :

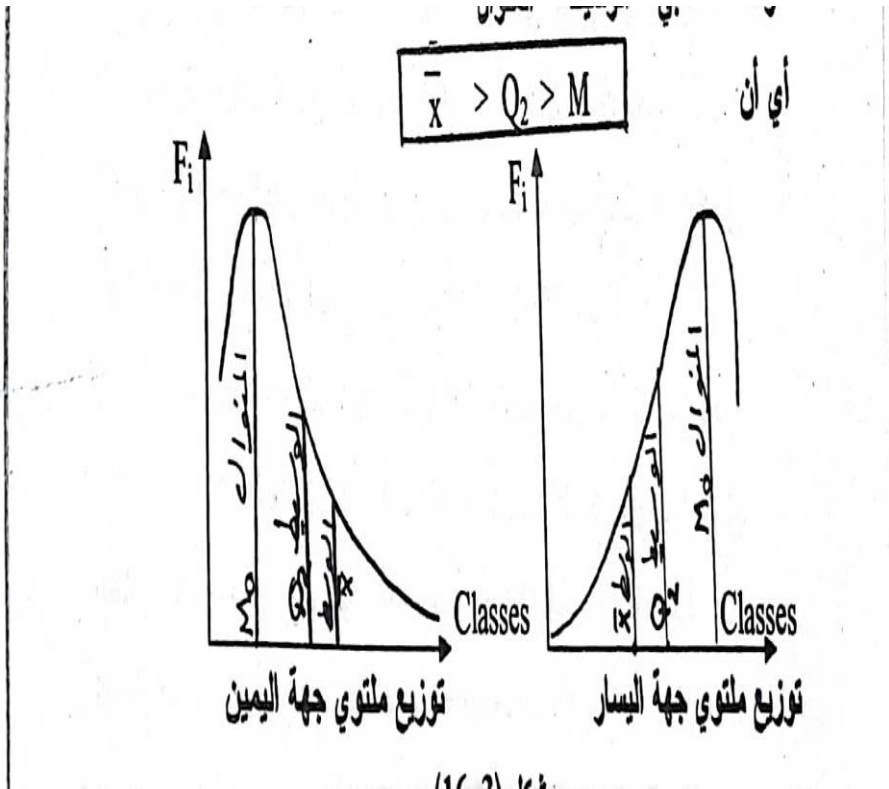
الوسط الحسابي > الوسيط > المنوال . أي أنه إذا كانت :

$$\bar{X} > Q_2 > M \text{ فيكون التوزيع سالب الالتواء.}$$

أما إذا كان التوزيع ملتوياً جهة اليمين (أي موجب الالتواء) حيث تميل التكرارات الكبيرة إلى التركيز عند فئات التوزيع الدنيا ويكون اتجاه الذيل جهة اليمين كما يتضح من الشكل (3-7-ب) فتكون العلاقة بين مقاييس النزعة المركزية الثلاثة هي :

الوسط الحسابي < الوسيط < المنوال . أي أنه إذا كانت :

$$\bar{X} < Q_2 < M \text{ فيكون التوزيع موجب الالتواء.}$$



وعلى ذلك فإن هناك علاقة بين قيمة مقاييس النزعة المركزية الثلاثة وشكل

التوزيع كالتالي : 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

المقاييس الثلاثة السابقة وهي :

الوسط الحسابي - المنوال $= 3(\text{الوسط الحسابي} - \text{الوسيط})$

وإذا استخدمنا الرموز الدالة على تلك المقاييس فإن :

$$\bar{X} - M = 3(\bar{x} - Q_2)$$

وتفيد هذه العلاقة في تقدير قيمة المنوال بطريقة ثالثة بالإضافة إلى طريقتي الرافعة والفروق وبصفة خاصة إذا كانت الفئة المنوالية هي الفئة الأولى أو الأخيرة في التوزيع التكراري . كما تمكن من تقدير قيمة مقربة للوسط الحسابي في حالة التوزيعات التكرارية المفتوحة بعد معرفة قيمة الوسيط والمنوال لهذه التوزيعات بشرط أن تكون هذه التوزيعات قريبة من التماثل .

مثال (20-2) :

إذا علمت أن الجدول الآتي يمثل توزيع تكراري قريب من التماثل :

Classes	Less than 6	6 - 12-	18-	24and more	Σ
F_i	8	20	40	25	7
					100

والمطلوب تقدير قيمة الوسط الحسابي .

الحل :

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 توزيع التكرارات عبر فئات التوزيع. فيمكن إيجاد الوسيط والمنوال ثم نستخدم العلاقة السابقة لإيجاد قيمة الوسط الحسابي .

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
Σ	100		

ولإيجاد الوسيط فإن :

ترتيب الوسيط هو :

$$\text{Ran of } (Q_2) = \frac{\sum F}{2} = \frac{100}{2} = 50$$

ومن ثم تكون قيمة الوسيط هي :

$$Q_2 = 12 + 6 \left(\frac{50 - 28}{68 - 28} \right) = 15.3$$

أما لإيجاد المنوال وليكن بطريقة الفروق فنعلم أن :

$$\begin{aligned} F_M &= 40 \\ \text{Model Class} &= 12 - 18 \\ F_{M-1} &= 20 \\ F_{M+1} &= 25 \\ D_1 &= F_M - F_{M-1} = 40 - 20 = 20 \\ D_2 &= F_M - F_{M+1} = 40 - 25 = 15 \end{aligned}$$

ومن ثم فإن :-

$$\begin{aligned} M &= L \cdot B \text{ for the Model Class} + L_M \left(\frac{D_1}{D_1 + D_2} \right) \\ &= 12 + 6 \left(\frac{20}{20+15} \right) = 15.43 \end{aligned}$$

وحيث أن التوزيع قريب من التماثل لذا يصح تطبيق بيرسون :

$$\begin{aligned} \bar{X} - M &= 3 \left(\bar{x} - Q_2 \right) \\ \text{i.e., } \bar{x} - 15.43 &= 3(\bar{x} - 15.3) \\ \bar{x} &= 15.235 \end{aligned}$$

وبحل المعادلة ينتج أن :

هذا وتجدر الإشارة إلى أن الاختيار بين مقاييس النزعة المركزية الثلاثة (الوسط الحسابي والوسيط والمنوال) للتعبير عن القيمة المتوسطة للتوزيع قد يتوقف أحيانا على الغرض الذي نبتغيه من المقياس : ولبيان ذلك نفرض أن شركة تدفع لعينة من موظفيها رواتب شهرية (بالجنيه) كما يلي :

400 , 5600 , 400 , 400, 7200, 5000, 400 , 1400

فلهذه القيم نجد أن :

$$\bar{X} = 2600 \text{ جنيهاً} , Q_2 = 900 \text{ جنيهاً} , M = 400 \text{ جنيهاً}$$

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 جنيهاً مقياسًا للنزعة المركزية ، في حين أن مدير الشركة سيختار الوسط الحسابي وهو 2600 جنيهاً مقياسًا للنزعة المركزية ليثبت أن رواتب الشركة مرتفعة ، ولو أن باحثًا محايدًا أراد أن يعبر بموضوعية عن حقيقة الرواتب بالشركة فسيختار الوسيط وهو 900 جنيهاً مقياسًا للنزعة المركزية . هذا ويظل الوسط الحسابي في أغلب الأحوال مقياسًا هامًا للنزعة المركزية يتمتع بخصائص جيدة تؤهله لأن يستخدم على نطاق واسع في علم الإحصاء كما سيتضح لنا فيما بعد . ولو أخذنا عينات مختلفة ذات أحجام متساوية من مجتمع الدراسة وحسبنا لكل عينة الوسط الحسابي والوسيط والمنوال لوجدنا أن التغيير من عينة إلى أخرى هو أقل في قيم الوسط الحسابي منه في قيم كل من الوسيط والمنوال ، مما يعني أن الوسط الحسابي أكثر استقرارًا من كل من الوسيط والمنوال عبر عينات نسحبها من مجتمع الدراسة .

**** الوسط الهندسي Geometric Mean (G.M) يمثل**

الوسط الهندسي أحد مقاييس النزعة المركزية والذي يفضل استخدامه عندما تكون قيم الظاهرة في صورة نسب أو معدلات وكذلك عند تقدير عدد السكان بين سنوات التعداد .

هذا ويعرف الوسط الهندسي لمجموعة من القيم عدده وليكن
(ن) من القيم بأنه عبارة عن الجذر النوني لحاصل ضرب هذه القيم .
(1-4-3): الوسط الهندسي في حالة البيانات غير المبوية :

إذا كان لدينا المتغير (X_i) يأخذ القيم X_1, X_2, \dots, X_n الموجبة والغير صفرية فإن الوسط الهندسي لهذه القيم وسنرمز له بالرمز

$$(G.M) \text{ يأخذ الصورة : } = \sqrt[n]{X_1 \cdot X_2 \cdot X_3 \dots X_n}$$

$$= \sqrt[n]{\pi^n X_i}$$

$$\text{Log (G.M)} = \frac{1}{n} \sum^n \text{Log } X_i$$

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 (G.M) Log الذي
نحصل عليه من جدول الأعداد المقابلة للوغاريتمات .

وبصفة عامة فإن الوسط الهندسي لمجموعة من القيم دائماً أقل من الوسط
الحسابي ويمكن إثبات ذلك رياضياً على النحو التالي :

$$\bar{X} = \frac{X_2 + X_1}{2} = \text{الوسط الحسابي}$$

$$G.M = \sqrt{X_1 X_2} \text{ فهو :}$$

وحيث أن :

$$(\sqrt{X_1} - \sqrt{X_2})^2 > 0$$

أي أن

$$X_1 - 2\sqrt{X_1 X_2} + X_2 > 0$$

$$2\sqrt{X_1 X_2} < X_1 + X_2$$

i.e.,

$$\sqrt{X_1 X_2} < \frac{X_1 + X_2}{2}$$

$$G.M < \bar{X}$$

أي أن:

وبذلك يكون الوسط الهندسي أصغر من الوسط الحسابي لنفس المجموعة من القيم .

مثال (3-21) :

احسب الوسط الهندسي للنسب التالية :

6% , 7% , 12% , 43%

الحل :

$$G.M = \sqrt[4]{6 \times 7 \times 12 \times 43}$$

$$\begin{aligned} \text{Log (G.M)} &= \frac{1}{4} (\text{Log } 6 + \text{Log } 7 + \text{Log } 12 + \text{Log } 43) \\ &= \frac{1}{4} (0.7782 + 0.8451 + 1.0792 + 1.6335) \\ &= \frac{1}{4} (4.336) = 1.084 \end{aligned}$$

بالكشف في جدول الأعداد المقابلة للوغاريتمات نجد

$$(G.M) = 12.13\%$$

أي أن :

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 بازياد عدد السكان فإن معدل النمو السكاني يتغير بمعدل متزايد لذلك فإنه يفضل استخدام الوسط الهندسي لتقدير عدد السكان حيث نجد أن الوسط الهندسي هو :

$$G.M = \sqrt{1500000 \times 2000000} = 1732050 \text{ person}$$

والمثال التالي يوضح أن الوسط الهندسي يكون أكثر دقة من الوسط الحسابي عندما تكون البيانات في صورة معدلات أو نسب .

(3-4-2) الوسط الهندسي في حالة البيانات الميوية:

إذا كان لدينا توزيع تكراري وكانت (x_i) تمثل مراكز فئات التوزيع ، F_i هي التكرارات المناظرة حيث ($i = 1, 2, \dots, r$) ، r هي عدد فئات التوزيع ، فإن الوسط الهندسي لهذا التوزيع هو :

$$G.M = \sqrt[\sum]{(x_1)^{F_1} \cdot (x_2)^{F_2} \cdot \dots \cdot (x_r)^{F_r}}$$

$$\text{Log } (G.M) = \frac{1}{\sum F_i} (F_1 \text{ Log } x_1 + F_2 \text{ Log } x_2 + \dots + F_r \text{ Log } x_r)$$

$$= 1/\sum F_i (\sum F_i \text{Log } x_i)$$

مثال (24-3) :

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	$\text{Log } x_i$	$F_i \log x_i$
55 -	10	60	1.7782	17.782
65 -	30	70	1.8451	55.353
75 -	40	80	1.9031	76.124
85-95	20	90	1.9542	39.084
\sum	100			188.343

$$\text{Log(G.M)} = \frac{\sum F_i \log x_i}{\sum_1 F_i} = \frac{188.343}{100} = 1.88343$$

ومن جدول اللوغاريتمات نجد أن :

$$\text{G.M} = 76.38$$

**** الوسط التوافقي : (H . M) Harmonic Mean**

الوسط التوافقي يمثل أحد مقاييس النزعة المركزية ويفضل استخدامه في حالة ما إذا كان المتغير أو الظاهرة محل الدراسة في صورة معدلات زمنية أو في صورة معدل تغيير أو معدل السرعة بالنسبة للزمن ، ويعرف الوسط التوافقي لمجموعة من القيم بأنه مقلوب الوسط الحسابي لمقلوبات القيم .

(1-5-3): الوسط التوافقي في حالة البيانات غير المبوية :

إذا كان لدينا مجموعة القيم x_1, x_2, \dots, x_n الغير صفرية فإن الوسط التوافقي ونرمز له بالرمز (H.M) يحسب كالآتي :

$$(H.M) = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} = \frac{n}{\sum \frac{1}{x_i}}$$

مثال (25-3) :

احسب الوسط التوافقي والوسط الهندسي للقيم التالية :
5, 25, 5, 1 وماذا تستنتج ؟

الحل :

الوسط التوافقي هو :

$$(H.M) = \frac{4}{\frac{1}{1} + \frac{1}{5} + \frac{1}{25} + \frac{1}{5}} = \frac{4}{\frac{36}{25}} = 2.78$$

$$(G.M) = \sqrt[4]{1 \times 5 \times 25 \times 5} = \sqrt[4]{625} = 5$$

ونستنتج أن الوسط التوافقي أصغر من الوسط الهندسي .

مثال(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 (ج) هي
60 كم قطعها القطار بسرعة 90 كم/ساعة من المدينة (ج) إلى المدينة (د)
هي 30 كم قطعها القطار بسرعة 120 كم/ساعة . احسب متوسط سرعة
القطار خلال الرحلة .

الحل :

إذا أردنا تقدير متوسط سرعة القطار في الساعة باستخدام الوسط الحسابي
والوسط التوافقي لمعرفة أيهما أكثر ملائمة فإن :

أ - باستخدام الوسط الحسابي :

متوسط سرعة القطار في الساعة باستخدام مفهوم الوسط الحسابي هو :

$$80+90+120$$

$$\bar{X} = \frac{\quad}{3} = 96.66 \quad \text{K.M/hour}$$

3

ب- باستخدام الوسط التوافقي :

متوسط سرعة القطار في الساعة باستخدام الوسط التوافقي هو :

$$\bar{X} = \frac{3}{\frac{1}{80} + \frac{1}{90} + \frac{1}{120}} = 96.66 \quad \text{K.M/hour}$$

لمعرفة أي المقياسين أكثر ملائمة نجد أن :

القطار قطع مسافة من المدينة (أ) إلى المدينة (ب) في زمن قدره:

(ج) $80 / 40 = 0.5$ ساعة ، وقطع المسافة من المدينة (ب) إلى المدينة (ج) في زمن قدره : $90 / 60 = 0.67$ ساعة ، وقطع المسافة من المدينة (ج) إلى المدينة (د) في زمن قدره : $120 / 30 = 0.25$ ساعة وبالتالي 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
 ويتضح من ذلك 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 $(1 / xi) \cdot Fi$.
- 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 (27-3) :

احسب الوسط التوافقي للتوزيع التكراري التالي :

Classes	5 -	15 -	25 -	35-45	Σ
Frequency	20	50	90	40	200

الحل :

لحساب الوسط التوافقي يتم تكوين الجدول التالي :

Classes	F_i	x_i	$1/x_i$	F_i / x_i
-5	20	10	1.0	2
-15	50	20	05.0	2.5
-25	90	30	0.033	3
35-45	40	40	0.025	1
Σ	200			8.5

فيكون الوسط التوافقي عبارة عن :

$$H.M = \frac{\sum_1^r F_i}{\sum_1^r (F_i/x_i)} = \frac{200}{8.5} = 23.53$$

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

$$\text{الوسط الحسابي} \leq \text{الوسط الهندسي} \leq \text{الوسط التوافقي}$$

وإذا استخدمنا الرموز فإن :

$$\bar{X} \geq G.M \geq H.M$$

(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 -	70-80
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 تكوين 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
Σ	80				-25		126.6198	2.3361

$$\Sigma D_i F_i \quad -25$$

$$\bar{X} = A + B \left(\frac{\Sigma D_i F_i}{\Sigma F_i} \right) = 45 + 10 \left(\frac{-25}{80} \right) = 41.875$$

$$\Sigma F_i \quad 80$$

وبفرض أن الوسط الهندسي هو G.M فإن :

$$\Sigma F_i \log x_i \quad 126.6198$$

$$\log(G.M) = \frac{\Sigma F_i \log x_i}{\Sigma F_i} = \frac{126.6198}{80}$$

$$\Sigma F_i \quad 80$$

ومن جدول الأعداد المقابلة للوغاريتمات نجد أن :

$$G.M = 38.25 \quad \text{ميكروجرام/م}^3$$

أما عن الوسط التوافقي فإن :

$$H.M = \frac{\sum F_i X_i}{\sum F_i} = \frac{2.3361}{80} = 34.2451 \quad \text{ميكروجرام/م}^3$$

ويتضح لنا من النتائج الثلاث السابقة أن الوسط الحسابي أكبر من الوسط الهندسي والوسط الهندسي أكبر من الوسط التوافقي .

(2): لحساب الوسيط والرابعين الأدنى والأعلى يلزم تكوين الجدول التكراري المتجمع الصاعد :

Classes	F _i	Less than the Class U.L	Asc.C.F
10 –	9	Less than 20	9
20 –	11	Less than 30	20
30 -	16	Less than 40	36
40 -	20	Less than 50	56
50 -	12	Less than 60	68
60 -	8	Less than 70	76
80-70	4	Less than 80	80
Σ	80		

ولحساب الوسيط أي Q₂ فإن :

$$\text{Rank of } (Q_2) = \frac{\sum F_i}{2} = \frac{80}{2} = 40$$

$$Q_2 = 40 + 10 (40 - 36) / (56 - 36) = 42 \quad \text{ميكروجرام/م}^3$$

لحساب الربع الأدنى (Q₁) فإن :

$$\text{Rank of } (Q_1) = \frac{\sum F_i}{4} = \frac{80}{4} = 20$$

وحيث أن الترتيب 20 يوجد مباشرة في عمود التكرار المتجمع الصاعد فيكون الربع الأدنى هو القيمة التي تناظر هذا الترتيب من الجدول، أي أن :

$$Q_1 = 30 \text{ ميكروجرام/م}^3$$

ولحساب الربع الأعلى أي Q_3 فإن :

$$\text{Rank of } (Q_3) = \frac{3 \sum F}{4} = 3 \times \frac{80}{4} = 60$$

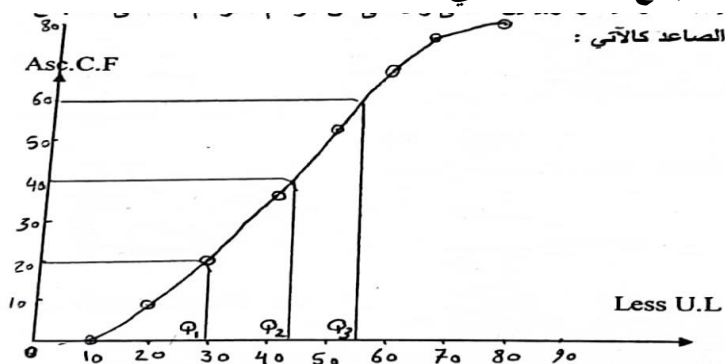
وتكون قيمة Q_3 هي : -

60-56

$$Q_3 = 50 + 10 \left(\frac{60-56}{68-56} \right) = 53.33 \text{ ميكروجرام/م}^3$$

68-56

ولإيجاد الوسيط والربعين الأدنى والأعلى من الرسم ، يتم رسم المنحنى التكراري المتجمع الصاعد كالآتي :



وحيث أن الجدول التكراري منتظم فيتم حساب المنوال مباشرة ، ويمكن حساب المنوال بطريقة الرافعة أو بطريقة الفروق ، فإذا استخدمنا نجد أن :

$$F_M = 20$$

$$F_{M-1} = 16$$

$$F_{M+1} = 12$$

وفئة المنوال هي من 40 إلى أقل من 50 أي أن :

$$\text{Model Class} = (40-50)$$

الفرق الأول هو :

$$D_1 = F_M - F_{(M-1)} = 20 - 16 = 4$$

الفرق الثاني هو

$$D_2 = F_M - F_{(M+1)} = 20 - 12 = 8$$

فتكون قيمة المنوال هي :

$$D_1$$

$$\text{Mode} = L . B \text{ for the Model Class} + L_M \left(- \frac{\quad}{\quad} \right)$$

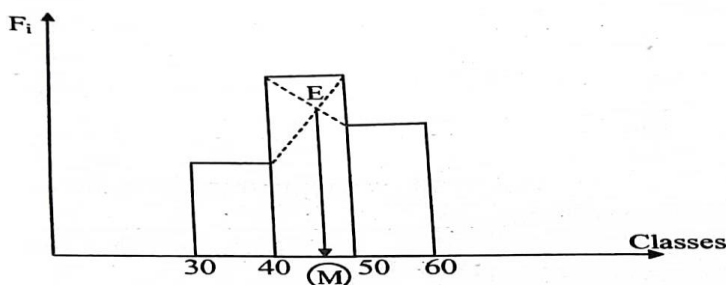
$$D_1 + D_2$$

$$4$$

$$= 40 + 10 \left(- \frac{\quad}{\quad} \right) = 43.33 \quad \text{ميكروجرام}$$

$$4+8$$

أما لإيجاد المنوال بيانياً فإن يتم رسم ثلاثة مستطيلات متلاصقة تمثل الفئة المنوالية والفئتين السابقتين عليها والفئة اللاحقة لها كالاتي :



ومن الرسم نستنتج أن : المنوال = 43.33 ميكروجرام / م³

(3): لإيجاد نسبة المدن التي تزيد فيها درجة التلوث عن 45 ميكروجرام / م³ من المنحنى التكراري المتجمع الصاعد والسابق رسمه وذلك من خلال رسم خط مستقيم على المحور الأفقي وذلك عند النقطة 45 حيث يتم رسم خطأ رأسياً يوازي المحور الرأسي إلى أن يقابل المنحنى المتجمع الصاعد في

نقطة نرسم منها خطأ أفقياً إلى أن يقابل المحور الرأسى عند النقطة 47 مدينة ، وبالتالي فإن :

عدد المدن التي تقل فيها درجة التلوث عن 45 ميكروجرام /م³ = 47 جنيه .
وعليه فإن :

عدد المدن التي تزيد فيها درجة التلوث عن 45 ميكروجرام /م³

$$= 80 - 47 = 33 \text{ مدينة .}$$

نسبة المدن التي تزيد فيها درجة التلوث عن 45 ميكروجرام /م³ =

$$(33/80) \% = 41.25 \%$$

(4): لإيجاد درجة التلوث التي تكون عندها 30% من المدن التي في

عينة الدراسة أكثر نظافة، فحيث أن نسبة ال 30 % من المدن عبارة عن:

نسبة 30% من المدن في عينة الدراسة = $80 \times (30/100) = 24$ مدينة

لذا فإنه من المنحنى المتجمع الصاعد نرسم على المحور الرأسى وبالتحديد

عند النقطة 24 خطأ أفقياً إلى أن يقابل المنحنى الصاعد في نقطة نرسم

منها خطأ رأسياً موازياً 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 /م³ .

تمارين الفصل الثالث

(1):- الجدول التالي يبين توزيع الطلاب حسب أيام الغياب خلال العام

الماضي:

Σ	-40	-35	-30	-25	-20	-15	-10	- 5	صفر-	عدد أيام الغياب
655	2	3	6	10	52	117	241	195	29	عدد الطلبة

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 .

ج - 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 طريقة الحساب .
- 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 (بالملي متر) للبيانات التالية :

202	201	200	199	198	الطول بالملي متر
1	2	17	4	1	التكرار

(6):- فيما يلي توزيع درجات الحرارة :

22	21	20	19	18	17	16	درجات الحرارة
1	1	9	8	6	4	1	التكرار

والمطلوب حساب الوسط الحسابي لدرجة الحرارة .

(7):-الجدول التالي يمثل توزيع مستويات أسعار 100 سلعة .

فئات أسعارالغلة	- 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 .
- ج- الوسيط .

د - القيمة الأكثر شيوعاً للسعر حسابياً وبيانياً .

(10):- الجدول التالي يبين توزيع العمال حسب الدخل :

فئات الدخل	أقل من 20	- 20	- 30	- 40	-50	60 فأكثر	المجموع
عدد العمال	5	14	20	15	10	6	70

المطلوب إيجاد :

أ - مقياس مناسب للنزعة المركزية موضعاً 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 هي 10 جنيهاً وأكبر قيمة هي 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 ، ج 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 :

-350 400	- 300	- 250	- 150	- 100	فئات التكاليف
5	10	12	20	8	عدد أفراد عينة الفندق الأول
2	7	12	15	6	عدد أفراد عينة الفندق الثاني

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 لـ 300 أسرة حسب دخلها اليومي (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 ما بين الوسط الحسابي والوسيط والمنوال ومرة أخرى باستخدام الربيعين الأدنى والأعلى والوسيط ، قارن بين النتيجة في الحالتين .
- (د): احسب قيمة الدخل الذي يقل عنه 60% من إجمالي عدد الأسر .

- (هـ): احسب النسبة المئوية لعدد الأسر التي يقل دخلها عن 43 جنيهاً .
 (و): حساب العشير السابع والمئتين الخامس والثمانين . اذكر بعض النتائج الأكثر تفصيلاً التي يمكن الحصول عليها باستخدام قيمة الربيعين الأدنى والأعلى والوسيط ، العشير السابع ، المئتين 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 .

112	87	76	95	90	105	86	99	84	103
101	94	104	87	100	67	109	75	118	88
122	72	112	92	65	115	80	111	74	112
90	101	92	87	108	90	97	93	104	98
120	86	103	77	98	84	117	69	123	94

والمطلوب :

- (أ): حساب الوسط الحسابي والوسيط لهذه البيانات .
 (ب): عرض هذه البيانات في صورة جدول تكراري متساوي الفئات طول كل فئة 10 مبتدئاً 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 .

(هـ) احسب نسبة الشقق التي يتراوح إيجارها الشهري ما بين 70، 90 جنيهاً.

(17):- الجدول التالي يوضح التوزيع التكراري لدرجات الطلاب بإحدى كليات التجارة حسب نظام الدراسة (انتظام- انتساب) وذلك في دور مايو 1987 .

Σ	- 60 100	- 50	- 40	- 30	- 20	- 0	فئات الدرجات
							نظام الدراسة
702	15	36	167	170	148	166	انتظام
285	1	5	24	59	88	108	انتساب

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-	Σ
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 لـ 40 أسرة حسب الحجم :

حجم الأسرة	1	2	3	4	5	6	المجموع
عدد الأسر	4	5	8	12	8	3	40

والمطلوب : 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 كل من الكليتين وقارن بينهما (الحد الأدنى لتقدير جيد هو 65 درجة).
- ج - 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

$$\begin{aligned} \text{أ - } n_1 = 10, \quad \bar{x}_1 = 25, \quad n_2 = 12, \quad \bar{x}_2 = 15 \\ \text{ب - } n_1 = 30, \quad \bar{s}_1 = 8, \quad n_2 = 8, \quad \bar{s}_2 = 12 \end{aligned}$$

(22):- إذا كان :

$$n_1 = 10, \quad \bar{s}_1 = 10, \quad n_2 = ? , \quad \bar{s}_2 = 20$$

فإذا كان الوسط الحسابي العام للعينتين هو 16، احسب حجم العينة الثانية .

(23):- الجدول التالي يوضح توزيع 80 أسرة حسب إنفاقها 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 .
 ج - 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 ما بين ما تحصل عليه في (ب) بقيمة الوسيط . علق على ما تحصل عليه من نتائج .

(25):- التوزيع التكراري التالي يوضح الدخل الاسبوعية (بالجنيهات)

لعينة حجمها 50 أسرة :

فئات الدخل	40 - 50 - 60 - 70 - 80 - 90	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 هذا التوزيع .

(26):- فيما يلي التوزيع التكراري للأجور اليومية لمائة عامل في إحدى الشركات :

فئات الأجور	35-45	45-55	55-65	65-75	75-85	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	35-45	45-55	55-65	65-75	75-85	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 الرابع

مقاييس التشتت

Measures of Dispersion

فى الباب السابق لاحظنا أن مقاييس النزعة المركزية أو المتوسطات (الوسط الحسابى- الوسيط - المنوال - الوسط الهندسى - الوسط التوافقى) تمثل قيمة متوسطة تتجمع حولها معظم البيانات الخاصة بظاهرة ما وتعبر عنها بصورة عامة . والآن فالسؤال الذى يطرح نفسه فى هذا المجال هو : هل يكفى معرفة مقياس من مقاييس النزعة المركزية ليعطى مؤشراً حقيقياً للبيانات محل الدراسة ؟ وهل يمكن من خلال مقاييس النزعة المركزية مقارنة هذه البيانات بمجموعة أخرى من البيانات؟

فى واقع الأمر أن هذا النوع وحدة من تلك المقاييس - مقاييس النزعة المركزية- لا يكفى لإعطاء صورة متكاملة عن توزيع المتغير محل الدراسة. إذ يتطلب ذلك نوعاً آخر من أنواع المقاييس الإحصائية ليعطى فكرة واضحة ودقيقة عن درجة أو مدى تجانس توزيع الظاهرة أو بمعنى آخر مدى تركيز 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 مضملة. فقد تشترك مجموعتين أو أكثر فى المتوسط الحسابى ولكنها قد تختلف فى درجة تجانسها. ولبيان ذلك دعنا نفترض أن لدينا مجتمعين أ ، ب وكان الدخل الأسبوعى (بالجنية) لخمسة أفراد من كل من المجتمعين على النحو المبين التالى:

الدخل الأسبوعي فى المجتمع (أ) هو 5 ، 50 ، 100 ، 150 ، 195 جنيهه
 وأن الدخل الاسبوعى فى المجتمع (ب) هو 98 ، 99 ، 100 ، 101 ، 102
 جنيهه . فنجد أن الوسط الحسابى للدخل الاسبوعى لكل من المجتمعين (أ ، ب)
 هو مساوياً 100 جنيهه . كما أن وسيط الدخل الأسبوعى لكل من المجتمعين
 (أ ، ب) أيضاً مساوياً 100 جنيهه . أى أنه وفقاً لمقاييس النزعة المركزية فإننا
 نجد أن البيانات فى المجتمعين (أ ، ب) لا يختلفان من حيث متوسط الدخل
 الاسبوعى . ولكن هذا مخالفاً لواقع البيانات الفعلية فى المجتمعين (أ ، ب)
 حيث يوجد هناك تفاوتاً واضحاً فى توزيع مفردات الدخل داخل كل مجتمع عن
 الآخر. فالبيانات فى المجتمع الأول (أ) يتراوح فيها الدخل الأسبوعى فيما بين
 195 جنيهه و 5 جنيهات أى أن المدى هو :-

Range = Maximum value – Minimum value.

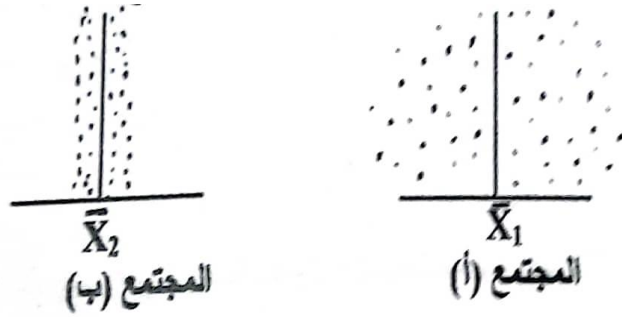
$$= 195 - 5 = 190 \quad (\text{L.E})$$

بينما نجد أن فى المجتمع الثانى (ب) يتراوح فيه الدخل الأسبوعى فيما بين
 102 ، 98 جنيهه. أى أن المدى هو :

$$\text{Range} = 102 - 94 = 4 \quad (\text{L.E})$$

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 يوحي 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 فإن التشتت لأي مجموعة من المفردات يقصد به مدى التباعد أو الاختلاف فيما بين قيم تلك المفردات فهذا التشتت يكون صغيراً إذا كان الاختلاف بين قيم مفردات الظاهرة محل الدراسة قليلاً أى إذا كانت مفردات الظاهرة تحتشد وتتجمع حول متوسطاتها أو قريبة منه (فإذا تساوت جميع القيم للظاهرة محل الدراسة فإن التشتت يندم أى يكون مساوياً للصفر) فى حين يكون التشتت كبيراً إذا ما كان الاختلاف فيما بين قيم المفردات كبيراً أى إذا كانت المفردات تنتشر بعيداً على جانبي الوسط الحسابى (أو مقياس النزعة المركزية) ولذلك فكما يعيننا 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). وفى الشكل (1-4) نجد أن مفردات المجتمع (أ) أكثر تشتتاً (أو أقل تجانساً) من مفردات المجتمع (ب) :



شكل (1-4)

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 وهي تختلف فيما بينها من حيث الدقة وطريقة الحساب. فكما سيتضح لنا فيما بعد إنه لأختيار المقياس الأفضل يجب ان نبحث عن توافر بعض الخصائص التي سبق أن ذكرناها عند الكلام عن المتوسطات من حيث سهولة فهم المقياس وأن يأخذ في الاعتبار جميع المفردات عند حسابه وإمكانية وضعه في صورة جبرية مما يسهل معه أن يخضع للعمليات الحسابية وقلة تأثرة بإختيار العينة والقيم الشاذة أو المتطرفة في قيم المفردات وسوف ندرس لهذا الغرض نوعين أساسيين من مقاييس التشتت هما:-

النوع الأول : وهي ما تسمى بمقاييس التشتت المطلقة وتشمل المدى ونصف المدى الربيعي والانحراف المتوسط والانحراف المعياري.

النوع الثانى: وهو ما يسمى بمقاييس التشتت النسبية وهى تشمل ما يسمى بمعاملات الاختلاف سواء النسبى أو الربيعى والتى سيرد دراستنا لها فى هذا الباب.

هذا وسوف نتعرض لمختلف هذه المقاييس بنفس الطريقة التى أتبعناها فى الباب السابق حيث يتم التعرف على كيفية حساب تلك المقاييس فى حالة توافر بيانات مفردة (غير مبوبة) ثم فى حالة البيانات المبوبة تكرارياً .

أولاً :- مقاييس التشتت المطلقة :-

Measures of Absolute Dispersion

وتستخدم هذه المقاييس فى دراسة مقارنة تشتت مجموعتين (أو أكثر) متحدتين فى وحدة القياس : بمعنى تستخدم هذه المقاييس المطلقة للتشتت مثلاً فى دراسة مقارنة تشتت درجات مجموعتين مختلفين من الطلاب أو الطالبات فى مادة معينة.

وبالإضافة إلى ان مقاييس التشتت المطلقة تقيس تشتت البيانات أو بعثرتها فأهم ما يميزها أنها تأخذ فى النهاية وحدة قياس هى نفس وحدة قياس القيم الأصلية كما هو الحال فى حالة مقاييس النزعة المركزية. فإذا كانت القيم الأصلية مثلاً بالكيلومتر فستكون النتيجة النهائية لقيمة هذه المقاييس بالكيلومتر أيضاً.

وتتمثل مقاييس التشتت المطلقة فيما يلى:-

1-المدى : Range: ويعرف المدى لمجموعة من المفردات كما سبق بأنه الفرق ما بين أكبر القيم وأصغرها أو ما يعرف بالقيمتين الصغرى والكبرى معاً ويستخدم فى بيان مدى الإتساع الذى تنتشر عليه المفردات. أى أن المدى :-

أ- في حالة المفردات (البيانات الغير مبوبة) : هو :

Range = Maximum value – Minimum value.

فإذا كان لدينا متغيراً يأخذ مجموعة القيم X_1, X_2, \dots, X_n وتم ترتيبها تصاعدياً أو تنازلياً وليكن $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ حيث يتم التعبير عن أصغر القيم بالقيمة $X_{(1)}$ وأكبر قيمة بالقيمة $X_{(n)}$ فإن المدى يكون عبارة عن:

Range = Maximum value – Minimum value.

$$= X_{(n)} - X_{(1)}$$

وبلغه أخرى نقول أن المدى للمتغير (X) هو من $X_{(1)}$ حتى $X_{(n)}$.

فمثلاً

إذا كان لدينا مجموعة القيم : 20 ، 15 ، 38 ، 12 ، 30 فإن المدى لتلك المجموعة هو :

$$\text{Range} = 38 - 12 = 26$$

أو نقول بأن المدى من 12 حتى 38 وإن كان هذا التعبير الأخير هو الأفضل حيث أنه يعطى معلومات أكثر من التعبير الأول (القانون) .

ب- المدى في حالة البيانات المبوبة تكرارياً :-

إذا كان لدينا توزيعاً تكرارياً فإن هناك مجموعة من الصيغ التي يمكن من خلالها حساب المدى . الأولى والتي تفيد بأن المدى هو عبارة عن الفرق ما بين الحد الأعلى للفئة الأخيرة والحد الأدنى للفئة الأولى للتوزيع التكرارى المعطى أى أن :-

Rang = U.L for the Latter Class – L.L for the First Class.

حيث :

* U.L for the Latter Class تمثل الحد الاعلى للفئة الاخيرة

* L.L for the First Class تمثل الحد الادنى للفئة الاولى

أما الصيغة الثانية والتي يمكن من خلالها حساب المدى ايضاً وذلك من خلال ايجاد الفرق بين مركزى الفئة الاخيرة والفئة الاولى للتوزيع التكرارى .ووفقاً لهذا التعريف فان المدى يمكن كتابته على الصورة :

Range=Midpoint for the Latter Class –Midpoint for the First Class.

ويعتبر المدى بذلك من أسهل مقاييس التشتت المطلق وأسرعها فى الحساب وهو مفيد أكثر فى العمليات الصناعية وخصوصاً فى مجال إحصاءات مراقبة جودة الإنتاج حيث تؤخذ عينات صغيرة ومتساوية من حيث الحجم (ومن ثم فإن حجم العينة لن يؤثر على المدى وهو ما يعيبه كمقياس لدرجة التشتت) وذلك على فترات متقاربة ونحتاج الى حساب سريع لتشتت كل عينة من تلك العينات المسحوبة من قطاع الإنتاج. كذلك نستخدم المدى فى دراسة التغيرات فى درجات الحرارة اليومية والتغيرات التى تنشأ فى الأجور أو الأسعار خلال فترة زمنية معينة. والمدى بهذا التعريف كمقياس من مقاييس التشتت يعاب عليه ما يلى :-

(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): نظراً لإعتماد المدى فى حسابه على المفردتين المتطرفتين فى القيمة فقط Extreme points فإنه يتأثر جداً بالقيم المتطرفة أو الشاذة إذ أن وجود قيمة

شأذه فى مجموعة قيم الظاهرة محل الدراسة قد يسبب زيادة كبيرة فى المدى فىستدل منه على أن مفردات الظاهرة متشتتة جداً بينما قد تكون مفرداتها كلها تقريباً – أى فىأ عدا القيمة الشأذه أو المتطرفة – متقاربة. فمثلاً إذا كان لدينا مجموعة القيم 15، 18، 20، 13، 19 فإن المدى لهذه القيم مساوياً :

$$\text{Range} = 20 - 13 = 7$$

فإذا ما أضفنا قيمة واحده ولتكن المفردة 100 مثلاً لمجموعة المفردات السابقة فإن المدى سيصبح :-

$$\text{Range} = 100 - 13 = 87$$

أى أن المدى تغيرت قيمته لتصبح 87 بدلاً من 7 نتيجة إضافة المفردة التى تأخذ القيمة الشأذه أو المتطرفة عن باقى المفردات (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 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 في مجال واسع في مجال دراسات ضبط جودة الإنتاج Quality control. والسبب في ذلك يرجع إلى أنه كما ذكرنا أنه في مثل هذه الدراسات (دراسات ضبط الجودة) فإننا نحتاج إلى حساب مقياس للتشتت بسرعة. والمدى بطبيعة الحال هو أسرع مقياس يمكن إيجاده ويعطى دلالة على التشتت. هذا وتجدر الإشارة إلى أنه في مثل تلك الحالات فإنه المدى يتم حسابه أكثر من مره وذلك باستخدام عدد كبير من العينات وذلك للتقليل ما أمكن من فرص الحصول على قيم شاذه أو متطرفة. وبصفة عامه فإنه المدى عادة ما يستخدم كمقياس للتشتت في حالة الحاجة إلى حساب مقياس سريع لتشتت البيانات دون الإهتمام بالدقة في القياس وكذلك عندما يكون لقيم المفردات المتطرفة أهمية خاصة في البحث يراد إبرازها من خلال الدراسة أو البحث.

2-شبهيات المدى :- Quasi - Range

للتغلب على مشكلة القيم المتطرفة أو الشاذه والتي يتأثر بها المدى في حسابه بشكل كبير فقد أتجه الإحصائيون إلى مقاييس بديلة شبيهه بالمدى ولكنها تهمل القيم المتطرفة. ووفقاً لهذه البدائل فإن قيم المفردات يتم ترتيبها تصاعدياً ثم يتم حذف عدداً متساوي من القيم عند طرفي مجموعة المفردات ثم يتم حساب المدى لبقية القيم. وتسمى هذه المقاييس بشبهيات المدى ويمكن تعريفها على النحو التالي:

$$\text{First Range} = X_{(n-1)} - X_{(2)}$$

$$\text{Second Range} = X_{(n-2)} - X_{(3)}$$

وهكذا بالنسبة لبقية مقاييس شبيهات المدى حيث يتم التعريف المدى الرائي أو ذو الرتبة (r) بأنه على الصورة:

$$R^{\text{th}} \sim \text{Range} = X_{(n-r)} - X_{(r+1)}$$

فمثلاً

إذا كان لدينا مجموعة البيانات التالية :

$$16, 84, 18, 15, 17, 14, 13, 10, 11, 12$$

فأنه يمكن إيجاد المدى وشبيهاته وذلك على النحو التالي :-

فبعد ترتيب المفردات تنازلياً أو تصاعدياً مثلاً فإن البيانات تأخذ الصورة التالية

$$10, 11, 12, 13, 14, 15, 16, 17, 18, 84$$

فيكون المدى هو :

$$\text{Range} = X_{(n)} - X_{(1)} = X_{(10)} - X_{(1)} = 84 - 10 = 74$$

أما شبيهات المدى فهي عبارة عن :

$$\text{First Range} = X_{(n-1)} - X_{(2)} \quad \text{فالمدى الأول} :$$

$$\text{or 1st Range} = X_{(9)} - X_{(2)} = 18 - 11 = 7$$

والمدى الثاني هو :

$$\text{Second range} = X_{(n-2)} - X_{(3)}$$

$$\text{or 2}^{\text{nd}} \text{ Range} = X_{(8)} - X_{(3)}$$

$$= 17 - 12 = 5$$

والمدى الثالث هو :

~

$$\text{Third Range} = X_{(n-3)} - X_4$$

$$3^{\text{rd}} \text{ Range} = X_{(7)} - X_4$$

$$= 16 - 13 = 3$$

لاحظ مدى الأثر على شبيهات المدى من حيث عدم تأثر مثل هذه المقاييس بالقيم الشاذة أو المتطرفة أى أنه بإستخدام شبيهات المدى فإننا نكون قد تغلبنا على مشكلة القيم الشاذة أو المتطرفة.

هذا وتجدر الإشارة إلى أن مقاييس شبيهات المدى يمكن إستخدامها سواء كانت عدد المفردات أو القيم صغيراً أو كبيراً. ففي 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 كبيراً كافياً فإنه يمكن إيجاد المدى بين المئين الخامس والتسعين والمئين الخامس أو بين المئين السادس والتسعون والمئين الرابع على سبيل المثال. ومثل هذه المقاييس كثيراً ما تستخدم فى مجالات بحوث التربية وعلم النفس على وجه الخصوص. ولعل أشهر مقاييس شبيهات المدى وأكثرها إستخداماً هو المدى بين الربع الثالث (Q_3) والربع الأول (Q_1) وهو ما يسمى بالمدى الربيعى . وسوف نتعرض لهذا المقياس بشئ من التفصيل فيما يلى .

* نصف المدى الربيعى : Semi – Interquartile Range لتلافى

عيوب المدى المطلق يوجد مقياس آخر من مقاييس التشتت يحاول معالجة حساسية المدى للقيم الشاذة 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 الشاذة أو المتطرفة سواء فى بداية أو نهاية البيانات. هذا المقياس يسمى بالمدى الربيعى أو الانحراف الربيعى والذى إذا ما تم قسمته على (2) نحصل على مايسمى بنصف المدى الربيعى. أى أن :

$$\text{Semi. Int. qu. Range} = \frac{Q_3 - Q_1}{2}$$

ويتميز نصف المدى الربيعى بأنه سهل الحساب والفهم والتطبيق كما أنه لا يتأثر نوعاً ما بالقيم الشاذة 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 مهماً 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 البىانات حول الربيعين فإنه يصبح نصف المدى الربيعى غير ملائماً لقياس
تشتت الظاهرة محل الدراسة.

مثال :- فيما يلى لديك التوزيع التكرارى لدرجات مجموعة من الطلاب فى
اختبار مادة الإحصاء.

جدول (1-4)

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 البىانات المعطاه لتوزيعاً تكرارياً لذا فإن :

1- المدى هو :

Range = Midpoint for the Latter Class - Midpoint for the
First Class

$$= 45 - 5 = 40 \quad (\text{degree})$$

or

Range = U.L. for the Latter Class – L.L for the First Class

وعليه فإن قيمة المدى هي:

$$R = 50 - 0 = 50 \quad (\text{degree})$$

لاحظ أن إختلاف القيم ليس إلا نتيجة أختلاف مفهوم الصيغة الرياضية التي 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 . أي أن المدى الأول هو أما أن يكون :

$$1^{\text{st}} \text{ Range} = 35 - 15 = 20 \quad (\text{degree})$$

$$\text{or} = 40 - 10 = 30 \quad (\text{degree})$$

أما لحساب نصف المدى الربيعي فيجب في البداية إيجاد الجدول التكراري المجتمع الصاعد او الهابط وذلك على النحو التالي :

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
Σ	60		

ثم تحديد رتبة كل ربيع ومن ثم تطبيق القانون في تحديد كل قيمة من قيم الربيعين اى ان :

$$\text{Rank of } Q_1 = \frac{\sum F}{4} = \frac{60}{4} = 15$$

وبملاحظة عمود التكرار المجتمع المساعد 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:

$$Q_1 = 20 \text{ degree}$$

اما عن الربيع الثالث فان رتبة الربيع الثالث هي :

$$\text{Rank of } (Q_3) = \frac{\sum F}{4} \times 3 = \frac{60}{4} \times 3 = 45$$

وبملاحظة رتبة الربيع الثالث (45) تقع ما بين التكرارين المجتمعين 51,39 ومن ثم فان بداية فئة الربيع الثالث هي 30 درجة وبتطبيق الصيغة الحسابية للربيع الثالث فان :

$$Q_3 = \text{L.L for } (Q_3) \text{ class} + \frac{\frac{3\sum F}{4} - \text{Pre. Asc. C.F. for the rank of } (Q_3)}{\text{Suc. Asc. C.F. for} - \text{Pres. Asc. C.F. for}} \times LQ_3$$

اى ان :

$$Q_3 = 30 + 10 \left(\frac{45-39}{51-39} \right) = 30 + 10 \left(\frac{6}{12} \right)$$

i.e.,

$$Q_3 = 35 \quad (\text{degree})$$

وعليه فان نصف المدى الربيعى يكون :

$$\text{Semi - Int. Qu. Range} = \frac{Q_3 - Q_1}{2}$$

$$= \frac{35-20}{2} = 7.5 \quad (\text{degree})$$

واخيرا فبالإضافة الى ان مقياس نصف المدى الربيعي كمقياسا للتشتت المطلق يتغلب على مشكلة القيم المتطرفة في البيانات كما انه يمكن ايجاده من الجداول التكرارية المفتوحة فان هذا المقياس - نصف المدى الربيعي- يكون ملائما في حالة التوزيعات التكرارية المتلوية (اللهم الا اذا كان التوزيع ملتو التواء شديدا بحيث تتركز ربع القيم في الفئة الاولى او الفئة الاخيرة مما يصعب معه حساب الربعين الادنى والاعلى) حينئذ يصعب حساب نصف المدى الربيعي ومع ذلك فان نصف المدى الربيعي يؤخذ عليه انه غير حساس للقيم الاقل من الربع الادنى (Q_1) والقيم الاكبر من الربع الاعلى وكذلك القيم التي تنحصر بينهما وهو ما يعنى ان هذا المقياس - مثل المدى - يقتصر في حسابة على قيمتين فقط هما Q_1 , Q_3 في حالة نصف المدى الربيعي وعلى قيمتى $X_{(1)}$ ، $X_{(n)}$ كما هو في حالة المدى .

وفيما يلى نتعرض بالدراسة الى مقياسين آخرين من مقياس التشتت المطلقة والذين لا يعتمدان في 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 المعياري .

3- الانحراف المتوسط : Mean Deviation (MD)

او متوسط الانحرافات المطلقة : Mean of Absolute Deviations

رأينا فيما سبق ان كل من المدى ونصف المدى الربيعي 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}) هو عبارة عن الفرق بين تلك القيمة وبين الوسط الحسابى مع اهمال اشارة الفرق اى ان :

$$d_i = | X_i - \bar{X} | \quad \text{where } i = 1, 2, \dots, n$$

وقد تم اعتبار الفرق المطلق اى تم اهمال 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 :

$$\text{Mean Deviation} = \frac{\sum_{i=1}^n | X_i - \bar{X} |}{n}$$

(MD)

ويتضح من تعريفنا للانحراف المتوسط ان قيمته تعتمد على مدى تقارب او تباعد المفردات عن الوسط الحسابي (او الوسيط) .

* الانحراف المتوسط في حالة البيانات الغير مبوبة (المفردات) :

سبق وان ذكرنا حالاً انه اذا كان لدينا متغيراً يأخذ مجموعة القيم :

X_1, X_2, \dots, X_n فان الانحراف المتوسط عن الوسط الحسابي يعرف كالآتي :

$$\text{Mean Deviation about Mean (MD)} = \frac{\sum_{i=1}^n |X_i - \bar{X}|}{n}$$

كما ان الانحراف المتوسط عن الوسيط يعرف على الصورة التالية :

$$\text{Mean Deviation about Median} = \frac{\sum_{i=1}^n |X_i - Q_2|}{n}$$

حيث تغني $\sum |X_i - \bar{X}|$ ، $\sum |X_i - Q_2|$ مجموع الانحراف المطلق

(اي بعد إهمال اشارة كل انحراف) عن الوسط او عن الوسيط على الترتيب .

هذا ويمكن حساب الانحراف المتوسط عن اي قيمة متوسطة اخرى كالمنوال او

الوسط الهندسي أو التوافقي الا ان هذا نادر الحدوث ويعتبر المقياس الاكثر

استخداماً هو الانحراف المتوسط عن الوسط الحسابي او الوسيط . واذا ذكر

الانحراف المتوسط بدون تخصيص بانه حول اي من قيم المتوسطات

المختلفة فيكون المقصود به هو الانحراف المتوسط عن الوسط الحسابي .

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 :

$$\begin{aligned}\bar{X} &= \frac{\sum X}{n} = \frac{133 + 137 + \dots + 131 + 150}{7} \\ &= \frac{959}{7} = 137\end{aligned}$$

والجدول التالى يعطى الفرق فيما بين مجموعة المفردات والوسط الحسابى

ومن ثم الانحرافات المطلقة لتلك الفروق :

جدول (3-4)

X_i	$d_i = X_i - 137$	$d_i = X_i - \bar{X} $
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 :

$$\text{Mean Deviation} = \frac{\sum |X_i - \bar{X}|}{n} = \frac{\sum |d_i|}{n} \quad \text{i.e.,}$$

$$MD_{\bar{X}} = \frac{30}{7} = 4.28$$

أما لايجاد الانحراف المتوسط عن الوسيط فيجب ترتيب المفردات وليكن تصاعديا لتحديد قيمة الوسيط اولا فتكون مجموعة القيم على الصورة :

131 , 133 , 134 , 135 , 137 , 139 , 150

وحيث ان عدد المفردات $(n) = 7$ وهو عددا فرديا لذا فان رتبة الوسيط هى:

$$\text{Rank of } Q_2 = \frac{n+1}{2} = \frac{7+1}{2} = 4$$

اى ان الوسيط هو عبارة عن المفردة الرابعة للبيانات بعد ترتيبها اى ان قيمة الوسيط هى 135 اى ان: $Q_2=135$. وعليه فانه لحساب قيمة الانحراف المتوسط عن الوسيط يتم تكوين الجدول التالى لحساب الفروق المطلقة للقيم عن الوسيط أى قيمة $|X_i - 135|$: فإن الجدول (4-4) التالى يوضح الحسابات اللازمة لذلك :

جدول (4-4)

X_i	$d_i = X_i - Q_2$ $= X_i - 135$	$ d_i = x_i - 135 $
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

$$\text{i. e., } MDQ_2 = \frac{28}{7} = 4$$

*الانحراف المتوسط في حالة البيانات المبوبة :

في حالة التوزيعات التكرارية فانه لايجاد الانحراف المتوسط يتم حساب الانحرافات المطلقة وذلك من خلال ايجاد الفرق فيما بين كل قيمة من قيم مراكز فئات التوزيع وبين الوسط الحسابي او الوسيط حسب المطلوب آخذين في الاعتبار تكرارات فئات التوزيع اي 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 الفئة (i) على الترتيب حيث $i = 1, 2, \dots, n$ تعبر عن عدد فئات التوزيع . حيث ان :

$$\bar{X} = \frac{\sum X_i F_i}{\sum F_i}$$

اما عن الانحراف المتوسط عن الوسيط فهو عبارة عن :

$$\text{Mean Deviation} = \frac{\sum |X_i - Q_2| F_i}{\sum F_i}$$

about Median

مثال :

احسب الانحراف المتوسط عن الوسط الحسابي للتوزيع التكرارى التالى والذى يمثل توزيع اعمار مجموعة معينة من موظفى احد الادارات بالجامعة (بالسنة):

جدول (5-4)

Classes	20-	28-	36-	44-	52-60	Σ
Frequency	4	12	17	4	3	40

الحل :

لحساب الانحراف المتوسط عن الوسط الحسابي يتم حساب الوسط الحسابي اولاً (وفى حالة الجداول المنتظمة يفضل حسابه بطريقة الانحرافات المعدلة)

ومن ثم نكون بصدد الجدول التالي (6-4) والذي يوضح الحسابات اللازمة
لحساب الوسط الحسابي ومن ثم الانحراف المتوسط عن قيمة الوسط الحسابي

جدول (6-4)

Classes	F_i	X_i	$d_i = X_i - 40$	$D_i = \frac{D_i}{8}$	$D_i F_i$	$ X_i - \bar{X} $	$ X_i - \bar{X} 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
Σ	40				-10		256

فيكون :

$$\bar{X} = A + B \frac{\sum D_i F_i}{\sum F_i}$$

$$= 40 + 8 \left(\frac{-10}{40} \right) = 40 - 2 = 38 \quad (\text{year})$$

وعليه فان الانحراف المتوسط على الوسط :

$$MD_{\bar{X}} = \frac{\sum |X_i - \bar{X}| F_i}{\sum F_i} = \frac{256}{40} = 64 \quad (\text{year})$$

مثال: احسب الانحراف المتوسط عن الوسيط للتوزيع التكراري التالي:

جدول (7-4)

Classes	Less than 40	40-	48-	54-	60-80	Σ
Frequency	10	15	50	22	3	100

الحل: لحساب الانحراف المتوسط عن الوسيط فيجب حساب الوسيط اولاً ثم يتم حساب الانحراف المتوسط عن الوسيط والجدول (8-4) يوضح الحسابات اللازمة لذلك :

جدول (8-4)

Classes	F_i	X_i	Less than the Class Upper Limit	ASC. C.F	$ X_i - Q_2 $ = $ X_i - 51 $	$ X_i - Q_2 ^{F_i}$
30-	10	35	Less than 40	10	16	160
40-	15	44	Less than 48	25	7	105
48-	50	51	Less than 54	75	0	0
54-	22	57	Less than 60	97	6	132
60-80	3	70	Less than 80	100	19	57
Σ	100				48	454

$$\text{Rank of } Q_2 = \frac{\Sigma F}{2} = \frac{100}{2} = 50$$

وحيث ان ترتيب الوسيط (50) يقع في نقطة المنتصف ما بين التكرارين المتجمعين الصاعدين 25، 75 لذا فان قيمة الوسيط تنصف البعد فيما بين

الحدين الأدنى والاعلى للفئة الوسيطة اى تنصف البعد ما بين 48 , 54 اى ان:

$$Q_2 = \frac{48-54}{2} = 51 \quad (\text{year})$$

(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 :

$$\text{Mean Deviation} = \frac{\sum |x_i - Q_2| F_i}{\sum F_i}$$

About Median

ومن خلال الجدول السابق فان :

$$\text{M.D. about } Q_2 = \frac{454}{100} = 4.45 \quad (\text{year})$$

وأخيرا فان الانحراف المتوسط كقياس للتشتت المطلق يعتبر غير شائع الاستخدام وذلك لعدة اسباب هى :

1- عدم سهولة العمليات الحسابية اللازمة لاجاده .

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 متفاوتة وارادنا تقدير الانحراف المتوسط للمجتمع بدلالة الانحرافات المتوسطة المحسوبة لهذه العينات لما امكنا ذلك . وفى تلك الحالة فاننا نكون مضطرين الى ادماج هذه العينات جميعا فى عينة كلية ثم حساب الانحراف المتوسط لهذه العينة الكلية .

هذا وتجدر الإشارة 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 في وحده القياس . اما اذا كان الهدف هو ايجاد مقياس للتشتت لاستخدامه فيما بعد في عمليات التحليل الاحصائي فاننا يجب ان نستخدم مقياسا غيره يكون اكثر ملائمة لهذا الغرض.

ولقد كانت الفكرة الاساسية وراء هذا المقياس هي اننا اخذنا الانحرافات المطلقة عن الوسط لتكون كلها موجبة بغض النظر عن اشارتها وذلك تجنباً لمشكلة ان مجموع انحرافات القيم عن وسطها الحسابي يساوى صفر دائماً ابداً . مما يعكس بدوره على ان الانحراف المتوسط سوف ياخذ في كل الحالات القيمة صفراً وذلك اذا تم التعامل مع الانحرافات باشارتها الفعلية كما هي 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 دائماً ذا قيمة موجبة . ومن ثم فبايجاد متوسط هذه المربعات فاننا نحصل على ما يسمى بالتباين وهو ما سوف نتعرض له الان بالتفصيل وكذا الانحراف المعياري .

4 - التباين والانحراف المعياري :

Variance and Standard Deviation

سبق وأن رأينا ان انحرافات القيم عن وسطها الحسابي تعطي صورة جيدة لدرجة تباين مفردات الظاهرة موضع الدراسة. كما بينا ايضا ان مجموع انحرافات القيم عن وسطها الحسابي يساوى الصفر دائما. وللتغلب على ذلك كانت وسيلتنا فى الانحراف المتوسط هي اهمال اشارات الانحراف واعتبارها جميعا موجبة اى الاعتماد على الانحرافات المطلقة وهذا الاهمال للاشارات لا يستند كما سبق ان اوضحنا الى اى منطق رياضى 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 الانحرافات موجبة ثم نوجد حاصل جمع هذه المربعات لتلك الانحرافات ونوجد متوسطها لنحصل على ما يسمى بالتباين وهو ما يرمز له بالرمز (σ^2) .

فالتباين يعرف بانه عبارة عن متوسط مربعات انحرافات القيم عن وسطها الحسابي اما الانحرافات المعيارى والذى يرمز بالرمز (σ) والذى يعتبر من اهم مقاييس التشتت المطلقة فما هو الا الجذر التربيعى (الموجب) للتباين اى ان الانحراف المعيارى ما هو الا الجذر التربيعى (الموجب) للوسط الحسابي لمربعات انحرافات القيم عن وسطها الحسابي وعادة ما يستخدم الانحراف المعيارى (σ) كمقياس للتشتت عندما لا يكون قياس تشتت الظاهرة هو نهاية المطاف بل بداية لعملية التحليل الاحصائى الاكثر اهمية والمتعلقة بالاستنتاج الاحصائى لمقاييس المجتمع محل الدراسة. وترجع تلك الاهمية الى خضوعه للطرق الرياضية وسهولة التعامل معه جبريا فهو يستخدم على نطاق واسع

للغاية فى نظرية التقدير الاحصائى واختبارات الفروض الاحصائية كأحد المهام الرئيسية لفرع الاحصاء الاستدلالى - الاحصاء التطبيقى - وفيما يلى نوضح كيفية حساب الانحراف المعياري ومن ثم التباين فى حالة المفردات وكذا فى حالة البيانات المبوبة تكراريا .

أولاً: الانحراف المعياري للبيانات غير المبوبة (المفردات) :

إذا كان لدينا متغيراً وليكن X يأخذ مجموعة القيم X_n, \dots, X_2, X_1 ظاهرة معينة وكان متوسطها الحسابى هو (\bar{X}) فان انحرافات هذه القيم عن وسطها الحسابى هي :

$$(X_1 - \bar{X}), (X_2 - \bar{X}), (X_3 - \bar{X}), \dots, (X_n - \bar{X})$$

ومن ثم يكون التباين عبارة عن متوسط مربعات انحرافات القيم عن وسطها الحسابى اى ان :

$$\sigma^2 = \frac{1}{n} [(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2]$$

اما عن الانحراف المعياري (σ) فما هو الا الجذر التربيعى للتباين والعكس صحيح اى ان التباين ما هو الا مربع الانحراف المعياري. اى ان :

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}}$$

والسبب فى اخذ الجذر التربيعى للتباين σ^2 للحصول على الانحراف المعياري 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 الانحرافات 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 المعياري فهو مقياس التشتت المطلق واهمها نظرا لاهميته في مجال التحليل الاحصائي .
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

مثال :

احسب الانحراف المعياري لمجموعه القيم التالية :

$$3 ، 12 ، 9 ، 8 ، 15 ، 6$$

الحل :-

لاستخدام الصيغه الرياضية السابقه في حساب الانحراف المعياري يتم حساب الوسط الحسابي لتلك المفردات بداية : فحيث ان:

$$\bar{X} = \frac{\sum X}{N} = \frac{6 + 10 + 8 + 9 + 12 + 3}{6} = \frac{48}{6} = 8$$

ثم دعنا نقوم بأجراء الحسابات اللازمة لصيغة الانحراف المعياري (σ) من خلال الجدول التالي :

جدول (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 فإن

$$\sigma = \sqrt{\frac{\sum(X_i - \bar{X})^2}{n}} = \sqrt{\frac{50}{6}}$$

$$= \sqrt{8.333} \quad \simeq \quad 2.887$$

اي ان الانحراف المعياري لتلك القيم هو تقريبا 2.887 اما التباين فهو عبارة عن الناتج تحت الجذر التربيعي . اي ان : $\sigma^2 = 8.333$ وجدير بالذكر انه 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 بالصيغة السابقة فإنه غالبا ما نجد ان قيمة الوسط الحسابي تأخذ قيمة كسرية مما يصعب معه تعقيد الحسابات اللازمه لحساب الانحراف المعياري او التباین بتلك الصيغة هذا بالاضافة الي التقريب المستخدم غالبا والذي قد يؤدي الي عدم دقة في النتائج لذا ينصح بالتغلب علي ذلك من خلال ما يلي :

فحيث ان :

$$\sigma = \sqrt{\frac{\sum(X-\bar{X})^2}{n}}$$

$$\begin{aligned}\sum(X_i - \bar{X})^2 &= \sum_{i=1}^n (X_i^2 - 2\bar{X}X_i + \bar{X}^2) \\ &= \sum X_i^2 - 2\bar{X}\sum X_i + n\bar{X}^2 \\ &= \sum X_i^2 - 2n\bar{X}^2 + n\bar{X}^2 \\ &= \sum X_i^2 - n\bar{X}^2 \\ &= \sum X_i^2 - \frac{(\sum X_i)^2}{n}\end{aligned}$$

وعليه فإنه :-

$$\begin{aligned}\sigma &= \sqrt{\frac{1}{n} \left[\sum X_i^2 - \frac{(\sum X_i)^2}{n} \right]} \\ &= \sqrt{\frac{\sum X_i^2}{n} - \left(\frac{\sum X_i}{n}\right)^2} \quad (2)\end{aligned}$$

ومما لا شك فيه ان هذه الصيغة 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 وفقا لتلك الصيغة مجرد حساب مجموع القيم اي $\sum x_i$ ومجموع مربعات هذه القيم اي $(\sum x_i^2)$. 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 باتباع الصيغة الاولى امرا اكثر تعقيدا . ولكن وكما سبق وأن ذكرنا فأن الصيغة 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 الانحرافات عن الوسط الحسابي

ب - للقيم الاصلية.

وماذا نستنتج بين النتيجتين . علق علي نتائجك .

الحل:

لحساب الانحراف المعياري بالطريقتين سواء بالانحرافات عن الوسط الحسابي أو باستخدام القيم الصلية دعنا اولاً نقوم بحساب الوسط الحسابي للمفردات . فحيث ان :

$$\bar{X} = \frac{\sum X}{n} = \frac{3 + 1 + 9 + \dots + 6}{7} = \frac{33}{7} = 4.71$$

والآن 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 = -3.71	13.7641
9	81	4.29	18.4041
8	64	3.29	10.8241
2	4	- 2.71	7.3441
4	16	- 0.71	0.5041
6	36	1029	106641
33	211	0.03 \simeq 0	55.4287

وعلي ذلك فان :

قيمة الانحراف المعياري باستخدام :-

ا - طريقة الانحراف عن الوسط هي :

$$\sigma = \sqrt{\frac{\sum(X-\bar{X})^2}{n}} = \sqrt{\frac{55.4287}{7}} = \sqrt{7.918385714} = 2.813962636$$

إما باستخدام القيم الأصلية فإن

$$\begin{aligned} \sigma &= \sqrt{\frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2} = \sqrt{\frac{211}{7} - \left(\frac{33}{7}\right)^2} \\ &= \sqrt{7.918367347} \simeq 2.813959372 \end{aligned}$$

لاحظ ان الاختلاف فيما بين قيمة الانحراف المعياري يأتي في الرقم الخامس من بعد العلامة 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 . والخالصه ان الوسط الحسابي باستخدام القيم الاصلية يعتبر اكثر دقة من طريقة الانحرافات حيث لا تتضمن هذه الطريقة اي 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 قيمة كسرية) فسوف يتساوي قيمة الانحراف المعياري باستخدام ايا من الاسلوبين سواء باستخدام طريقة الانحرافات حول الوسط او باستخدام القيم الاصلية لانه لن يكون هناك اي عمليات تقريب وإن كان من المفضل استخدام الطريقة الخاصة 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 النتيجة اكثر دقة .

خصائص الانحراف المعياري :

- 1 – لا يتاثر الانحراف المعياري بعملياتي الجمع او الطرح : بمعنى أنه إذا كان لدينا قيم المتغير x هي مجموعه القيم x_1, x_2, \dots, 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 الناتجة حول هذا المقدار الثابت هي :

$$d_1 = X_1 - a, d_2 = X_2 - a, \dots, d_n = X_n - a$$

ومن ثم فان :

الوسط الحسابي للقيم d_i حيث $i = 1, 2, \dots, n$ هو :

$$\begin{aligned} \bar{d} &= \frac{\sum d_i}{n} \\ &= \frac{\sum (X_i - a)}{n} = \frac{\sum X_i - na}{n} \\ &= \frac{\sum X}{n} - a = \bar{X} - a \end{aligned}$$

كما ان الانحراف المعياري لمجموعة القيم d_i هو :

$$\sigma_{d_i} = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n}}$$

وبالتعويض في المعادله الاخيره عن ($\bar{d} = \bar{X} - a$ ، $d_i = x_i - a$) 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 . اي ان الانحراف المعياري لم يتأثر بعملية الطرح . وكذا يمكن إثبات عدم تأثر الانحراف المعياري بالعملية العسكية للطرح (اي الجمع) . ومن ثم فإن الانحراف المعياري لا يتأثر بعملية الجمع او الطرح . ولعل السبب في ذلك يرجع الي ان قيمة الانحراف المعياري 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- الانحراف المعياري يتأثر بعمليتي الضرب والقسمه :

ويمكن اثبات ذلك رياضيا بنفس الالية السابقة المستخدمه في الخاصية الاولي وذلك علي النحو التالي :

فنفرض ان لدينا مجموعة القيم X_1, X_2, \dots, X_n للمتغير (X) فان :
الوسط الحسابي لمجموعة تلك المفردات هو :

$$\bar{X} = \frac{\sum x}{n}$$

وكذلك فإن الانحراف المعياري 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 أنه قيمة (a) لا تساوي الصفر اي (a ≠ 0) فان مجموع الصيغ الجديده ما

$$D_1 = ax_1, D_2 = ax_2, \dots, D_n = ax_n \quad : \text{هي الا}$$

فيكون الوسط الحسابى لها هو :

$$\bar{D} = \frac{\sum D_i}{n} = \frac{\sum(ax_i)}{n} = a \frac{\sum X}{n}$$

$$\bar{D} = a\bar{X} \quad : \text{أى أن}$$

أما الانحراف المعياري لتلك القيم فهو عبارة عن :-

$$\sigma_D = \sqrt{\frac{\sum(D_i - \bar{D})^2}{n}} = \sqrt{\frac{\sum(ax_i - a\bar{X})^2}{n}}$$

أى أن :

$$\begin{aligned} \sigma_D &= \sqrt{\frac{a^2 \sum(X_i - \bar{X})^2}{n}} \\ &= a \sqrt{\frac{\sum(X_i - \bar{X})^2}{n}} = a \sigma_X \end{aligned}$$

وهو ما يفيد أن الانحراف المعياري

للقيم الجديدة (اي بعد ضرب قيم المتغير الاصلى في المقدار الثابت) ما هو الا حاصل ضرب الانحراف المعياري للقيم الاصلية (اي قبل عملية الضرب) فى المقدار الثابت والذي تم ضرب قيم المتغير الاصلى فيه . وعليه فإن :

$$\sigma_x = \frac{\sigma_D}{a}$$

اي ان الانحراف المعياري للقيم الاصلية (x_i) = الانحراف المعياري للقيم الناتجة بعد اجراء عملية ضرب المفردات الاصلية في المقدار القابت (a) .

كذلك يتم التعامل مع العملية العكسية لعملية الضرب (اي عملية القسمة) نفس معاملة عملية الضرب ولكن بطريه عكس ية. بمعنى انه اذا تم قسمه قيم المتغير (x_i) وليكن (a) حيث $(a \neq 0)$

فانه الانحراف المعياري للقيم الاصلية اي σ_x سوف يكون مساويا للانحراف المعياري للقيم الجديدة (اي d_x) مضروبا في هذا المقدار الثابت (a) . هذا

ويمكن بيان مدي صحة هاتين الخاصيتين السابقتين من خلال المثال البسيط التالي :

افرض ان لديك المفردات الخمس التاليه 100,200,300,400,400 فان :
*الوسط الحسابي لتلك المفردات الاصلية هو:

$$\begin{aligned}\bar{X} &= \frac{\sum x}{n} \\ &= \frac{100+200+300+400+500}{5} = \frac{1500}{5} = 300\end{aligned}$$

كما ان الانحراف المعياري لتلك المفردات الاصلية هو :

$$\sigma_x = \sqrt{\frac{\sum(x - \bar{x})^2}{2}}$$

$$= \sqrt{\frac{(100-300)^2 + (200-300)^2 + \dots + (500-300)^2}{5}} = 141.2413562$$

والآن دعنا نفرض أنه تم طرح مقدار ثابت وليكن 100 من المفردات الخمسة أي أن ($a=100$). فإن مجموعة الانحرافات للقيم الأصلية عن الوسط الفرض 100 هي عبارة عن :

$$d_i = 0, 100, 200, 300, 400$$

فيكون الوسط الحسابي لتلك الانحرافات هو

$$\bar{d} = \frac{0 + 100 + 200 + 300 + 400}{5} = 200$$

كما أن الانحراف المعياري لتلك الانحرافات هو :

$$\sigma_{di} = \sqrt{\frac{(0-200)^2 + (100-200)^2 + \dots + (400-200)^2}{5}} = 141.4213562$$

اي ان اثر التغير فقط وقع علي قيمة الوسط الحسابي اما الانحراف المعياري فلا يتاثر بالطرح او الجمع لكن يتاثر فقط بالضرب او القسمة وللتأكد من تاثره بعملية الضرب والقسمة : دعنا نقوم بقسمة المفردات محل هذا المثال علي مقدار الثابت (100) فتكون نواتج قسمة المفردات الاصلية علي 100 هي :

$$D_i = 1, 2, 3, 4, 5$$

فيكون الوسط الحسابي لتلك القيم الجديدة 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 عبارته عن :

$$\sigma_D = \sqrt{\frac{\sum(D_i - \bar{D})^2}{n}}$$

$$= \sqrt{\frac{(1-3)^2 + 2-3^2 + \dots + (5-3)^2}{5}} = 1.414213562$$

أى أن :-

$$\bar{X} = 300 = 100 (3) = a \cdot \bar{D}$$

وكذلك فإن :-

$$\sigma_x = 141.4213562 = 100 (1.414213562) = a \cdot \sigma_D$$

اي ان الوسط الحسابي وكذا الانحراف المعياري للقيم الاصلية هو عبارة عن حاصل ضرب المقدار (a) في الوسط الحسابي والانحراف المعياري للقيم الجديدة علي الترتيب .

ومن خلال دراستنا لهاتين الخاصيتين وللمثال المبسط السابق يمكننا الان دراسة طريقتي الانحراف البسيطة والانحرافات المعدلة والتتين 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 مناسباً من مجموعة القيم الاصلية بحيث نحصل علي اصغر اعداد ممكنه تسمى 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 الاولى من خصائص الانحراف المعياري) كما سبق ان اوضحنا ذلك .

فاذا اعتبرنا ان قيم الانحرافات هي $d_i = x_i - a$ ، حيث $i = 1, 2, \dots, n$ تمثل الانحرافات البسيطة حول الوسط الفرضي (a) فان :

$$\sigma_x = \sigma_d = \sqrt{\frac{\sum di^2}{n} - \left(\frac{\sum di}{n}\right)^2}$$

والوسط الفرضي (a) يتم اختياره بالشكل الذي يؤدي الي تبسيط العمليات الحسابية . ويفضل استخدام قيمة هذا الوسط الفرضي بمثابة قيمة المفردة التي تتكرر أكثر من غيرها في باقى المفردات وذلك لتبسيط الحسابات .

طريقة الانحرافات المعدلة (الاكثر اختصارا) :-

بعد الحصول علي الانحرافات البسيطة $d_i = x_i - a$ ، فاذا كانت جميع الانحرافات البسيطة d_i تقبل القسمة علي وسط فرضي آخر و بدون بواقي . فإنه يتم قسمة قيم الانحرافات البسيطة (d_i) على الوسط الفرضي وليكن (b) فنحصل علي الانحرافات المعدلة D_i حيث ان :

$$D_i = \frac{x_i - a}{b} = \frac{d_i}{b} \quad \text{where } i = 1, 2, \dots, n$$

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 المعياري للانحراف المعدلة (D_i) وذلك بعد ضربها في قيمة

الوسط الفرضي (b) حيث ان الانحراف المعياري يتأثر بعملية الضرب والقسمة كما سبق وان اسلفنا (الخاصية الثانية من خصائص الانحرافات المعيارية) اي ان :

$$\sigma_x = b \sigma_d = b \sqrt{\frac{\sum di^2}{n} - \left(\frac{\sum di}{n}\right)^2}$$

مثال : احسب الانحراف المعياري لمجموعة القيم :-

625 , 700 , 650 , 725

وذلك باستخدام :

أ-القيم الاصلية .

ب- طريقة الانحرافات البسيطة .

ج- طريقة الانحرافات المعدلة .

الحل :-

أحساب النحراف المعياري باستخدام القيم الاصلية :

الجدول التالي يوضح الحسابات اللازمة لحساب الانحراف المعياري من واقع

استخدامنا للقيم الاصلية :

X_i	X_i^2
625	390625
700	490000
650	422500
725	525625
2700	1828750

وحيث أن :

$$\sigma = \sqrt{\frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2}$$

فإنه :

$$\sigma = \sqrt{\frac{1828750}{4} - \left(\frac{2700}{4}\right)^2} = 39.53$$

ب- حساب الانحراف المعياري باستخدام طريقة الانحرافات البسيطة :

يتم اختيار اى من قيم المفردات بمثابة وسط فرض ولتكن $a=650$ والجدول التالى يوضح الحسابات اللازمة لحساب الانحراف المعياري باستخدام طريقة الانحراف البسيطة :

وحيث أن :

X_i	$d_i = x_i - 365$	d_i^2
625	625-650= -25	625
700	50	2500
650	0	0
725	75	5625
	100	8750

$$\sigma = \sqrt{\frac{\sum d_i^2}{n} - \left(\frac{\sum d_i}{n}\right)^2}$$

فإنه :

$$\sigma = \sqrt{\frac{8750}{4} - \left(\frac{100}{4}\right)^2} = 39.53$$

وهى نفس النتيجة التى تم الحصول عليها فى (أ) لاحظ اثر تبسيط الحسابات فى هذه الطريقة عن السابقة .

ج - حساب الانحراف المعياري باستخدام طريقة الانحرافات المعدلة :

بملاحظة عمود قيم المتغير (di) فى الطريقة السابقة يلاحظ انه يمكن قسمة

قيم الانحرافات

البسيطة على

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 :

$$\sigma_x = b \sigma_p = b \sqrt{\frac{\sum Di^2}{n} - \left(\frac{\sum Di}{n}\right)^2}$$

$$= \sqrt{\frac{14}{4} - \left(\frac{4}{4}\right)^2} = 39.53$$

وهى نفس النتيجة التى حصلنا عليها فى (أ) ، (ب) مع تبسيط اكثر فى العمليات الحسابية وجهد التجميع .

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 بأيا من الطرق الثلاث السابقة (الطريقة المباشرة) او القيم الصلية) - طريقة الانحرافات البسيطة - طريقة الانحرافات المعدلة) وذلك مع اعتبار ان مراكز

فئات التوزيع هي قيم (X_i) ويراعى ترجيحها بالتكرارات المناظرة لتلك المراكز او تلك الفئات اي (F_i) .

1 – الطريقة المباشرة لحساب الانحراف المعياري في حالة البيانات المبوبة :

اذا كانت لدينا مراكز فئات توزيع تكراري هي X_1, X_2, \dots, X_n وما يناظرها من تكرارات F_1, F_2, \dots, F_n فمن المعروف ان كل مركز فئة يعتبر مكررا عدد من المرات مساويا لتكرار تلك الفئة المناظرة وفي هذه الحالة فانه لحساب الانحراف المعياري باستخدام الطريقة المباشرة يتم ضرب مركز الفئات (X_i) في التكرارات المناظرة لها (F_i) ثم ايجاد مجموع النواتج لحاصل الضرب وقسمة هذا المجموع على مجموع التكرارات لتعطي بذلك الوسط الحسابي للتوزيع التكراري اي ان :

$$\bar{X} = \frac{\sum X_i F_i}{\sum F_i}$$

وبعد حساب قيمة الوسط الحسابي للتوزيع يتم ايجاد انحرافات قيم مراكز فئات التوزيع (X_i) عن الوسط الحسابي للتوزيع (\bar{X}) اي نوجد $(X_i - \bar{X})$ لكل فئة من فئات التوزيع ثم يتم ايجاد مربعات انحرافات المراكز عن الوسط الحسابي اي $(X_i - \bar{X})^2$ ثم ضرب نواتج مربعات انحرافات المراكز عن وسطها الحسابي في التكرارات المقابلة لها ثم ايجاد المجموع اي :

$$\sum (X_i - \bar{X})^2 F_i$$

فيكون الانحراف المعياري للتوزيع التكراري عبارة عن :

$$\sigma = \sqrt{\frac{\sum (X_i - \bar{X})^2 F_i}{\sum F_i}}$$

ويمكن تبسيط هذه الصورة الاخيرة كما سبق لتأخذ الصورة التالية :

$$\sigma = \sqrt{\frac{\sum X_i^2 F_i}{\sum F_i} - \frac{(\sum X_i F_i)^2}{\sum F_i}}$$

ويفضل استخدام هذه الصيغة عن السابقة لها لحساب الانحراف المعياري خصوصا اذا كانت قيمة الوسط الحسابي تحتوى على قيما كسرية كما سبق وان بينا ذلك .

2- طريقة الانحراف البسيطة :

وهنا يتم اخذ احد مراكز الفئات بمثابة وسط فرضي وليكن (a) ويفضل اختيار قيمة الوسط الفرضي (a) على ان يكون مركز الفئة التي تتوسط فئات التوزيع التكراري او مركز الفئة الذي يقابل اكبر تكرار وذلك لتبسيط العمليات الحسابية بشكل ملحوظ . فاذا افترضنا ان انحرافات مراكز فئات التوزيع التكراري عن الوسط الفرض بمثابة القيم (d_i) حيث :

$$d_i = x_i - a \quad i=1,2,\dots,n$$

وتعبر (n) عن عدد فئات التوزيع التكراري ثم يتم ايجاد حاصل ضرب الانحرافات (d_i) في التكرارات المقابلة ومن ثم المجموع $\sum d_i F_i$ ومرة اخرى يتم ضرب $d_i F_i$ في عمود الانحرافات (d_i) ثم الجمع لنحصل على المجموع $\sum d_i^2 F_i$ فيكون الانحراف المعياري عبارة عن :

$$\sigma_x = \sigma_d = \sqrt{\frac{\sum d_i^2 F_i}{\sum F_i} - \left(\frac{\sum d_i F_i}{\sum F_i}\right)^2}$$

3 – طريقة الانحرافات المعدلة (الاكثر اختصارا) :

اذا كانت قيم الانحراف البسيطة (d_i) المحسوبة فى الطريقة السابقة تقبل القسمة على مقدار اوسط فرضى آخر وليكن (b) بدون بواقي (وغالبا ما تكون قيمة (b) بمثابة طول الفئة فى حالة الجدول التكرارية المنتظمة) فانه يفضل قسمة الانحرافات البسيطة (d_i) على الوسط الفرضى (b) لينتج لنا الانحرافات المختصرة ولتكن D_i حيث :

$$D_i = \frac{X_i - a}{b} = \frac{d_i}{b} \quad \text{where : } i = 1, 2, \dots, n$$

فانه يتم ضرب قيم الانحرافات المختصرة (D_i) فى التكرار المقابل ثم ايجاد المجموع اى ($\sum D_i F_i$) ومرة اخرى يتم ضرب عمود ($D_i F_i$) فى عمود الانحرافات المختصرة (D_i) ثم الجمع لينتج $\sum D_i^2 F_i$ ثم يتم حساب الانحراف المعياري من الصيغة التالية :

$$\sigma = b \sigma_d = b \sqrt{\frac{\sum D_i^2 F_i}{\sum F_i} - \left(\frac{\sum D_i F_i}{\sum F_i}\right)^2}$$

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 (11-4)

No. of defective units	0	1	2	3	4	5
Frequency	13	15	18	25	20	19

المطلوب اوجد الانحراف المعياري 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 (X_i) ومن ثم فالجدول التالي يبين الحسابات اللازمة لايجاد الانحراف المعياري 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
Σ	100	301	1107

فيكون الانحراف المعياري لعدد الوحدات المعيبة في الصندوق هو :

$$\sigma = \sqrt{\frac{\sum X_i^2 F_i}{\sum F_i} - \left(\frac{\sum X_i F_i}{\sum F_i} \right)^2}$$

اي ان الانحراف المعياري لعدد الوحدات المعيبة هو :

$$\sigma \sqrt{\frac{1107}{100} - \left(\frac{301}{100}\right)^2} = 1.42 \text{ unite}$$

مثال :

احسب الوسط الحسابي والانحراف المعياري للتوزيع التكراري التالي :

Classes	12-	16-	30-	40-	50-70	Σ
Frequency	8	13	15	10	4	50

الحل :

لحساب الوسط الحسابي والانحراف المعياري لهذا التوزيع يتم تكوين الجدول

التالي :

جدول (4-15)

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
Σ	50			-124	8900

وعليه فان الوسط الحسابي باستخدام طريقة الانحرافات البسيطة هو :

$$\bar{X} = a + \frac{\sum d_i F_i}{\sum F_i}$$

$$= 35 + \left(\frac{-124}{50} \right) = 35 - 2.48 = 32.52$$

اما الانحراف المعياري باستخدام طريقة الانحرافات البسيطة فهو :

$$\begin{aligned} \sigma_x = \sigma_d &= \sqrt{\frac{\sum di^2 F_i}{\sum F_i} - \left(\frac{\sum di F_i}{\sum F_i} \right)^2} \\ &= \sqrt{\frac{8900}{50} - \left(\frac{-124}{50} \right)^2} \end{aligned}$$

مثال :

فيما يلي التوزيع التكراري التالي لاطوال عينة من طلاب كلية التجارة بالسنتيمتر .

جدول (4 - 16)

Classes	140-	150-	160-	170-	180-190	Σ
Frequency	6	10	22	7	5	50

والمطلوب احسب الوسط الحسابي والانحراف المعياري لطول الطالب بالعينة وكذا تباين هذا التوزيع :

الحل :

حيث ان التوزيع التكراري المعطى توزيعا تكراريا منتظما (اي متساوي من حيث اطوال فئاته) لذا يفضل استخدام طريقة الانحرافات المعدلة . هذا والجدول (4-17) يبين الحسابات اللازمة لاجاد كلا من الوسط الحسابي

والانحراف المعياري لطول الطالب في العينة باستخدام طريقة الانحرافات المعدلة :

جدول (4 - 17)

Classes	F_i	X_i	$d_i=X_i-165$	$D_i = \frac{d_i}{10}$	$D_i f_i$	$D_i^2 F_i$
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
Σ	50				-5	61

وحيث ان صيغة القانون المستخدم لحساب الوسط الحسابي باستخدام طريقة الانحرافات المعدلة هو :

$$\bar{X} = a + b \frac{\sum D_i F_i}{\sum F_i}$$

لذا فان :

$$\bar{X} = 165 + 10 \left(\frac{61}{50} \right) = 177.2 \quad (C.M)$$

اما عن صيغة الانحراف المعياري باستخدام طريقة الانحرافات المعدلة فهي :

$$\sigma_x = b \sigma_D = b \sqrt{\frac{\sum d_i^2 F_i}{\sum F_i} - \left(\frac{\sum d_i F_i}{\sum F_i} \right)^2}$$

$$= 10 \sqrt{\frac{61}{50} - \left(\frac{-5}{50}\right)^2} = 11 \text{ (C.M)}$$

اما عن التباين (σ^2) فهو عبارة عن مربع الانحراف المعياري

$$\sigma^2 = (11)^2 = 121 \quad \text{(C.M)} \quad \text{اي ان}$$

*الانحراف المعياري المصحح لشبيرد (تصحيح الانحراف المعياري للبيانات

المبوبة)

سبق وان راينا انه قد تؤدي عملية تبويب البيانات في الجداول التكرارية الى فقدان بعض المعلومات عن الظاهرة محل الدراسة حيث لا يكون معروفا كل مفردة 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 في مركزها او بمعنى اخر ان الجداول التكرارية منشأة كما ذكرنا سابقا تحت افتراض ان قيم المفردات تتوزع بانتظام (اي متجانسة الانتشار) على مدار طول كل فئة من فئات الجدول وهو ما يدعونا الى استخدام مركز كل فئة كقيمة تأخذها كل مفردة من المفردات داخل هذه الفئة. وهذا الافتراض وإن كان لا يشوبه الدقة الا إنه يعتبر افتراضا تقريبا الغرض منه تسهيل العمليات الحسابية وهو ما يؤدي بطبيعة الحال الى اخطاء في قيم المقاييس الاحصائية وبالتحديد تلك المقاييس الاحصائية (كالوسط والانحراف المعياري) المحسوبة من البيانات المبوبة سوف لا تنطبق أو لا تتساوى تماما مع قيمة المقاييس المحسوبة من القيم الاصلية قبل تبويبها) وهو ما يسمى 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 جميع مربعات الفروق بين قيم المفردات ووسطها الحسابى يميل الى التراكم وليس الى التلاشى (حيث ان مربعات الفروق السالبة تكون موجبة) وبالتالي فان قيمة الانحراف المعيارى المحسوبة من بيانات مبوبة فى جدول تكرارى تميل الى ان تختلف عن قيمة الانحراف المعيارى فى حالة حسابه من البيانات الاصلية مباشرة هذا ويزداد هذا الاختلاف كلما زاد طول الفئات للتوزيع التكرارى .

وتحاشيا لمسئلة الاختلاف فيما بين قيمة الانحراف المعيارى للمفردات الاصلية عن الانحراف المعيارى لتلك المفردات بعد تبويبها فقد قام شبرد Shepperd بدراسة ذلك على التوزيعات المتصلة التى تلامس طرفا كل منها المحور الافقى واوجد تصحيحا Correction للانحراف المعيارى وذلك من خلال طرح خارج قسمة مربع طول الفئة فى هذا التوزيع على (12) اى طرح المقدار

$\left(\frac{L^2}{12}\right)$ من صيغة قانون الانحراف المعياري المعروفة اي ان صيغة الانحراف المعياري المصحح لشبرد تأخذ الصورة التالية :

$$\sigma_{cor.} = \sqrt{\frac{\sum X^2 F}{\sum F} - \left(\frac{\sum XF}{\sum F}\right)^2 - \frac{L^2}{12}}$$

أى أن :

$$\sigma_{cor.} = \sqrt{\sigma^2 - \frac{L^2}{12}}$$

حيث ان σ_{cor} نقصد بها الانحراف المعياري 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 المعياري على مستوى عال من الدقة . ويمكنك عزيزى القارئ التحقق من انه كلما قصر طول الفئة فى التوزيع التكرارى كلما كان الفارق ما بين الانحراف المعياري المصحح والانحراف المعياري الغير مصحح قليلا ومن جهة اخرى فان قيمة التصحيح عادة ما تكون طفيفة الاهمية اذا كان مجموع التكرارات قليل نسبيا (اقل من 1000 مفردة 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

المعروف بالاسى السالب او الاسى الموجب) او فى حالة التوزيعات المستطيلة.

مثال :

احسب الانحراف المعيارى المصحح للتوزيع الوارد فى المثال السابق جدول (16-4) .

الحل :

من خلال نتائجنا فى المثال السابق اتضح لنا ان التباين لهذا التوزيع قيمته (121) اى ان $\sigma^2 = 121$ وحيث ان طول الفئة فى هذا التوزيع هو $L=10$ لذا فان الانحراف المعيارى المصحح لهذا التوزيع هو :

$$\begin{aligned}\sigma_{cor.} &= \sqrt{\sigma^2 - \frac{L^2}{12}} = \sqrt{121 - \frac{(10)^2}{12}} \\ &= \sqrt{112.6667} = \sqrt{10.614} \text{ (C.M)}\end{aligned}$$

*الانحراف المعيارى فى حالة التوزيعات التكرارية النسبية:- 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 فان :-

$$\sigma = \sqrt{\frac{\sum X_i^2 F_i}{\sum F_i} - \left(\frac{\sum X_i F_i}{\sum F_i}\right)^2}$$

$$= \sqrt{\sum X_i^2 \frac{F_i}{\sum F_i} - \left(\sum X_i \frac{F_i}{\sum F_i}\right)^2} \quad (1)$$

وحيث ان التكرار النسبي ما هو الا تكرار الفئة منسوبا الى مجموع التكرارات
اي ان :

$$R_i = \frac{F_i}{\sum F_i} \quad \text{where } i = 1, 2, 3, \dots, n$$

حيث (n) تمثل عدد فئات التوزيع

لذا فإنه بالتعويض عن التكرارات النسبية في القانون (1) ينتج ان :

$$\sigma = \sqrt{\sum X_i^2 R_i - \left(\sum X_i R_i\right)^2}$$

وبالمثل فإنه وفقا لطريقة الانحرافات البسيطة تصبح صيغة قانون الانحراف
المعياري في حالة الجداول التكرارية النسبية هي :

$$\sigma = \sqrt{\sum d_i^2 R_i - \left(\sum d_i R_i\right)^2}$$

اما في حالة طريقة الانحرافات المعدلة فان الانحراف المعياري للتوزيع
التكراري النسبي هو :

$$\sigma = b \sqrt{\sum D_i^2 R_i - \left(\sum D_i R_i\right)^2}$$

مثال :

التوزيع التالي 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 .

الحل :

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	$D_i = \frac{X_i - 30}{5}$	D _i R _i	D _i R _i ²
17.5-	0.12	20	-2	-0.24	0.48
22.5-	0.2	25	-1	-0.2	0.2
27.5-	0.28	30	0	0	0
32.5-	0.25	35	1	0.25	0.25
37.5-42.5	0.12	40	2	0.3	0.6
Σ	1			0.11	1.53

وعليه فإن :

الوسط الحسابي لدرجة الحرارة المتوقعة في هذا اليوم هو :

$$\bar{x} = a + b \sum_{i=1}^n D_i R_i = 30 + 5 (0.11)$$

$$= 30.55 \quad (\text{degree})$$

أما الانحراف المعياري فهو :

$$\sigma = b \sqrt{\sum D_i^2 R_i - \left(\sum D_i R_i\right)^2}$$

$$= 5\sqrt{1.53 - (0.11)^2} = 6.16 \quad (\text{degree})$$

ملاحظات حول مقاييس التشتت المطلقة :-

- 1- يعتبر الانحراف المعياري من اهم مقاييس التشتت المطلقة واكثرها استخداما في الاحصاء وذلك نظرا لانه ياخذ في الاعتبار جميع المفردات في الحساب

فضلا على انه يخضع للمعالجة الجبرية ويستخدم فى معالجات اخرى لذلك يفضل استخدام الانحراف المعيارى حينما لا يكون قياس تشتت الظاهرة هو نهاية المطاف او نهاية التحليل الاحصائى بل انه بداية لعمليات احصائية اخرى اكثر اهمية كعملية الاستدلال الاحصائى او مراقبة جودة الانتاج **Quality Control** .

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 قيمة لا تختلف كثيرا من عينة لاخرى مسحوبة من نفس المجتمع محل الدراسة .

3- رغم ان الانحراف المعيارى ياخذ فى الاعتبار جميع 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 الوسط الحسابى فى ايجاد .

4- يعاب ايضا على الانحراف المعياري انه لا يمكن استخدامه في التوزيعات التكرارية المفتوحة كما لا يمكن حسابه بيانيا وفي هذه الحالة يفضل استخدام نصف المدى الربيعي لدراسة تشتت التوزيع .

5- اختلاف قيم مقاييس التشتت المطلق للتوزيع التكراري الواحد امرا طبيعيا ومتوقعا وذلك نظرا لاختلاف تلك المقاييس من حيث المفهوم او التعريف فكل مقياس له قانونه الخاص به . وهذا في حد ذاته يعطى مؤشرا الى انه في حالة مقارنة تشتت ظاهرة واحدة في عدة مجتمعات او مقارنة عدة ظواهر لها نفس وحدات القياس فإنه يلزم استخدام مقياس واحد فقط من هذه المقاييس وذلك بما يتناسب مع طبيعة البيانات وذلك حتى لا تكون النتائج المتحصل عليها مضللة في عملية المقارنة .

6- في 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 المعياري

- نصف المدى الربيعي = $\frac{2}{3}$ الانحراف المعياري.

ومن هاتين العلاقتين يمكن التوصل لعلاقة تقريبية تربط ما بين الانحراف المتوسط ونصف المدى الربيعي وهي :

$$\text{الانحراف المتوسط} = 1.2 \text{ نصف المدى الربيعي}$$

اخيرا فمن اهم خصائص التوزيع المعتاد الطبيعي **Normal Distribution** ما يلي :

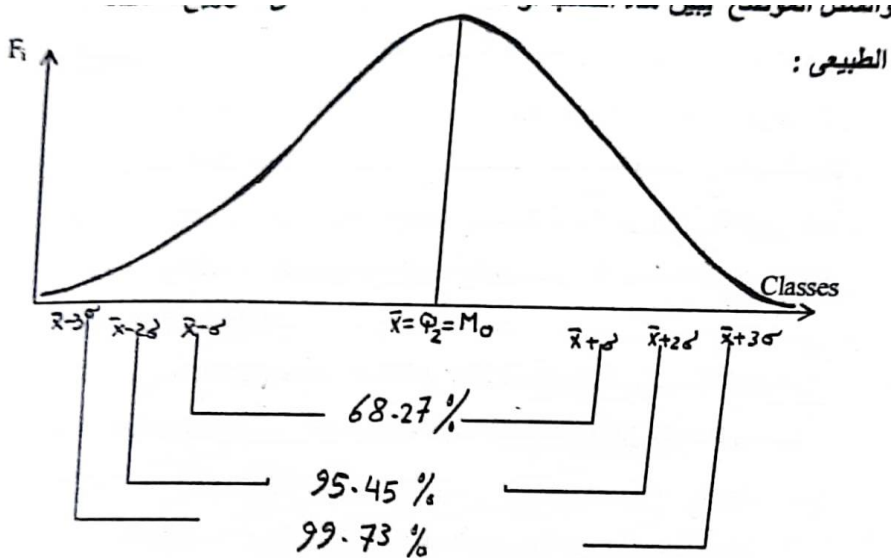
* ان 68.27% من جميع المفردات تقع فى المدى ما بين $(\bar{X} \pm \sigma)$.

* ان 95.45% من جميع المفردات تقع فى المدى ما بين $(\bar{X} \pm 2\sigma)$.

* وان 99.73 من جميع المفردات تقع فى المدى ما بين $(\bar{X} \pm 3\sigma)$.

والشكل الموضح يبين هذه النسب او المساحات تحت منحنى التوزيع المعتاد

الطبيعى :



وجدير بالذكر فى هذا الخصوص ان كثيرا من الظواهر الطبيعية او المتغيرات

تفترب او توزيعها يقترب من التوزيع المعتاد الطبيعى وذلك اذا كان حجم او

عدد المشاهدات كبيرة نسبيا (نظرية الحد المركزية Central Limit

. (Theorem

ثانيا : مقاييس التشتت النسبية (معاملات الختلاف) :

Coefficients of Variation .

راينا فيما سبق ان مقاييس التشتت المطلقة والتي عرضناها سابقا تاخذ نفس تمييز الوحدات الاصلية للظاهرة محل الدراسة ومن ثم فإن مقاييس التشتت المطلقة باختلاف مسمياتها تستخدم فى المقارنة بين مجموعتين (او اكثر) متحدين فى وحدة القياس. اى ان هذه المقاييس لا تصلح للمقارنة بين تشتت مجموعتين (او اكثر) مختلفين فى وحدات القياس كالأوزان (بالكيلوجرام مثلا) والاعمار (بالسنوات مثلا) فليس من المعقول اجراء مقارنة بين الكيلوجرامات والسنوات ونقول ان الكيلوجرامات اقل او اكثر تشتتا من السنوات . اصف الى ذلك انه قد تتحد ظاهرتين او اكثر من وحدات القياس لكن طبيعة الظاهرتين قد تختلف اختلافا جوهريا فيما بين المجموعتين المطلوب المقارنة بينهما . فمثلا بافتراض ان هناك مجموعة من القيم تعبر عن الدخل الشهرى لمجموعة من المستثمرين او رجال الاعمال (دخول مرتفعة جدا) وهناك مجموعة اخرى من القيم تعبر عن الدخل الشهرى لمجموعة من العمال (دخول بسيطة) فإن هذا يودى بدوره الى انه بالطبع يختلف متوسط الدخل الشهرى للمجموعتين اختلافا كبيرا . وبالتالي فإن استخدام مقاييس التشتت المطلقة لمقارنة تشتت هاتين المجموعتين فى هذه الحالة سوف يودى بطبيعة الحال الى نتائج مضللة واستنتاجات خاطئة بالمرّة .

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. وفي مثل تلك الحالات فإنه 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):-تخلص المقاييس النسبية من اثر اختلاف قيمة المتوسطات الذى ربما قد يوجد ما بين المجموعات.

(2):- حيث أن المقاييس النسبية تاتى على شكل نسبة مئوية خالية من التمييز وبالتالي يمكن استخدامها لمقارنة تشتت مجموعتين متحدين او مختلفين فى وحدات قياسها .

وبصفة عامة فإن الصورة العامة لمعامل الاختلاف ياخذ الصورة التالية :

$$\text{معامل الاختلاف} = 100 \times \frac{\text{مقياس للتشتت المطلق}}{\text{مقياس من مقاييس المتوسطات}}$$

على ان يكون مقياس المتوسطات المستخدمة فى المقام له نفس طبيعة (أو طريقة حساب) مقياس التشتت المستخدمة فى البسط : لذا فإن هناك مقياسين اساسيين لدراسة التشتت النسبى وهما :

(1):معامل الاختلاف (بدلاله الوسط الحسابى والانحراف المعيارى) :

وهذا المعامل يعتبر اكثر معاملات الاختلاف انتشارا وهو ما يسمى بمعامل الاختلاف النسبى وهو عبارة عن النسبة ما بين الانحراف المعيارى والوسط الحسابى للظاهرة محل الدراسة اى ان :

$$\text{معامل الاختلاف} = \frac{\text{الاختلاف المعيارى}}{\text{الوسط الحسابى}} \times 100$$

أى أن :

Coefficient of Variation هو عبارة عن :

$$C_V = \frac{\sigma}{\bar{X}} \times 100$$

هذا ويفضل استخدام هذا المعامل فى حالة التوزيعات التكرارية المتماثلة او القريبة من التماثل. لكن ما يعيب هذا المقياس هو انه لا يمكن حسابه من الجداول التكرارية المفتوحة اذا انه يعتمد فى حسابه على ايجاد كل من الوسط الحسابى والانحرافى المعيارى وكلاهما لا يمكن حسابه من الجداول التكرارية المفتوحة . كما ان هذا المقياس يعطى نتائج مضلله اذا كان التوزيع شديد الالتواء وذلك نظرا لتحيز كل من الوسط الحسابى والانحراف المعيارى للقيم الشاذة 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 فانه

يفضل حساب معامل الاختلاف الربيعى وهو عبارة عن النسبة ما بين نصف المدى الربيعى والوسيط اى ان :-

$$\text{معامل الاختلاف الربيعى} = 100 \times \frac{\text{نصف المدى الربيعى}}{\text{الوسيط}}$$

$$Q.C_V = \frac{Q_3 - Q_1}{2 Q_2} \times 100$$

هذا وفى حالة ما اذا كان التوزيع محل الدراسة قريبا من التماثل فأن قيمة الوسيط تساوى تقريبا الوسيط 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 السابقة يكون:

$$Q.C_V = \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$$

ويفضل استخدام اى من هاتين الصيغتين فى حالة الجداول التكرارية المفتوحة وهناك صيغ اخرى لمعاملات الاختلاف تقيس التشتت النسبى للظاهرة موضع الدراسة ولكنها غير شائعة الاستخدام وهى :-

*معامل الاختلاف باستخدام الانحراف المتوسط عن الوسط وياخذ الصورة التالية:-

فى حالة المفردات :

$$C_V.M_D = \frac{\sum |X - \bar{X}|}{\bar{X}} \times 100$$

أما في حالة البيانات المبوبة فإن صيغة هذا المعامل تأخذ الصورة التالية :

$$= \frac{\frac{\sum |X_i - \bar{X}|}{\sum F_i}}{\bar{X}} \times 100$$

*معامل الاختلاف بدلالة الانحراف المتوسط عن الوسيط وياخذ الصورة التالية:

$$C_{V.M_Q} = \frac{\frac{\sum |X_i - Q_2|}{n}}{Q_2} \times 100$$

ويتم تعديل الصيغة في حالة الجداول التكرارية كما هو في الصيغة الخاصة بالمعامل بدلالة الانحراف المتوسط عن الوسط من خلال ترجيح الفروق المطلقة بتكرارات الفئات المقابلة .

وفي نهاية هذا الجزء يجب ان ننوه الى ان قيمة معامل الاختلاف سوف تختلف باختلاف المعادلة المستخدمة في الحساب وهذا امرا طبيعيا نظرا لاختلاف الاساس الرياضى لكل معادلة من المعادلات السابقة ولذلك يجب استخدام صيغة واحدة لاحد تلك المعاملات بما يتناسب مع طبيعة البيانات وذلك عند دراسة مقارنة التشتت النسبى ما بين المجموعات المختلفة .

مثال :

اجريت دراسة لمجموعة مكونة من 80 عاملا باحد المصانع شملت العمر والاجر الاسبوعى وحصلنا على البيانات التالية :

بالنسبة للعمر فإن : الوسط الحسابى = 35 ، التباين = 49

اما بالنسبة للاجر الاسبوعى فكان موزعا على النحو التالى :

Classes	65-	69-	73-	77-	81- 85	Σ
Frequency	6	14	42	10	8	80

والمطلوب تحديد اى من الظاهرتين (العمر ام الاجر الاسبوعى) الاكثر تشتتا ؟

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 : فإن 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 المعياري لحساب معامل الإختلاف النسبى للدخل الأسبوعى فيتم تكوين الجدول التالى :

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 فإن الوسط الحسابي والانحراف المعياري للدخل الاسبوعي هما :

$$\bar{X} = a + b \frac{\sum D_i F_i}{\sum F_i} = 75 + 4 \left(\frac{0}{80} \right) = 75 \text{ (L.E)}$$

كما أن الإنحراف المعياري هو :

$$\sigma = b \sqrt{\frac{\sum D_i^2 F_i}{\sum F_i} - \left(\frac{\sum D_i F_i}{\sum F_i} \right)^2}$$

$$= 4 \sqrt{\frac{80}{80} - \left(\frac{0}{80} \right)^2} = 4 \text{ (L.E)}$$

ومن ثم فإن معامل الاختلاف النسبي للدخل الاسبوعي هو:

$$C_V = \frac{\sigma}{\bar{X}} \times 100 = \frac{4}{75} \times 100 = 5.3 \%$$

وبمقارنة معامل الاختلاف للعمر (20%) والدخل الاسبوعي (5.2%) نجد

ان توزيع العمر اكثر تشتتا من توزيع الاجر او الدخل الاسبوعي .

مثال: فيما يلى لديك توزيع درجات مجموعة من الطلاب فى مادة الإحصاء:

Classes	Less than 50	50-	60-	70-	80 and more
Frequencies	5	15	55	20	5

والمطلوب حساب معامل الاختلاف المناسب لهذا التوزيع .

الحل:

حيث ان الجدول التكرارى الذى يعبر عن توزيع الدرجات جدولاً تكرارياً مفتوح الطرفين . لذا يتعين حساب معامل الاختلاف الربيعى كذلك فإنه من شكل توزيع التكرارات يتضح من الجدول ان التوزيع قريباً من التماثل . لذا يمكن حساب الربيعين الأدنى والأعلى فقط دون الحاجة لحساب الوسيط اى يتم استخدام صيغة معامل الاختلاف الربيعى المستخدمة للتوزيعات القريبة من التماثل وهى:

$$Q.C_V = \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$$

والجدول التالى يبين التوزيع التكرارى المجتمع الصاعد اللازم لاجاد قيمة كل

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
Σ	100		

وحيث ان :

$$\text{Rank of } Q_1 = \frac{\Sigma F}{4} \times 1 = \frac{100}{4} = 25$$

$$\text{Rank of } Q_3 = \frac{\Sigma F}{4} \times 3 = \frac{100}{4} \times 3 = 75$$

وعليه فان :

$$\begin{aligned} Q_1 &= 60 + \left(\frac{25 - 20}{75 - 20} \right) \times 10 \\ &= 60 + \left(\frac{5}{55} \right) \times 10 = 60.9 \quad (\text{degree}) \end{aligned}$$

$$Q_3 = 75 \quad (\text{degree})$$

لاحظ ان ترتيب Q_3 كان بمثابة احد التكرارات المتجمعة الصاعدة لذا فان قيمة Q_3 هي القيمة المقابلة لهذا التكرار وهي القيمة 75 درجة وعليه فان معامل الاختلاف الربيعي هو :

$$Q.C_V = \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$$

$$= \frac{75 - 60.9}{75 + 60.9} \times 100 = 6.59 \%$$

مثال :-

فى دراسة لمستوى الاجور الشهرية للعمال فى قطاع صناعة الغزل والنسيج وقطاع الصناعات الدوائية تم اخذ عينة عشوائية من عمال كل قطاع من القطاعين وحصلنا على النتائج التالية لكل منهما :

العينة الثانية (من قطاع الصناعات الدوائية)	العينة الاولى (من قطاع الغزل والنسيج)
$\sum_{i=1}^{50} X_{2i} = 22500$ <p>ومجموع مربعات المفردات هو :</p> $\sum_{i=1}^{50} X_{2i}^2 = 10141200$	$\sum_{i=1}^{30} X_{1i} = 12000$ <p>ومجموع مربعات المفردات هو :</p> $\sum_{i=1}^{30} X_{1i}^2 = 4817280$

والمطلوب :

- (1): احسب الوسط الحسابى والانحراف المعياري لكل عينة .
- (2): قارن بين العينات من حيث التشتت موضحا ايها الاكثر تشتتاً .

(3): إذا تم دماج العينين فى عينة واحدة فاحسب متوسط العينة الجديدة وكذا معامل الاختلاف لها .

الحل :-

(1): حساب الوسط الحسابى والانحراف المعياري لكل عينة على حده.

فأنه بالنسبة للعينة الاولى: فإن الوسط الحسابى هو :

$$\bar{X}_1 = \frac{\sum_{i=1}^{30} X_{1i}}{30} = \frac{12000}{30} = 400 \text{ (L.E)}$$

اما بالنسبة للانحراف المعياري للعينة الاولى فهو:

$$\sigma_1 = \sqrt{\frac{\sum X_{1i}^2}{n} - \left(\frac{\sum X_{1i}}{n}\right)^2}$$

أما بالنسبة للعينة الثانية: فإن الوسط الحسابى والانحراف المعياري هما :

$$\bar{X}_1 = \frac{\sum_{i=1}^{50} X_{2i}}{50} = \frac{22500}{50} = 450 \text{ (L.E)}$$

$$\begin{aligned} \sigma_1 &= \sqrt{\frac{\sum x_{2i}^2}{n} - \left(\frac{\sum x_{2i}^2}{n}\right)^2} = \sqrt{\frac{10141200}{50} - \left(\frac{22500}{50}\right)^2} \\ &= 18 \text{ (L.E)} \end{aligned}$$

ب- ولمقارنة تشتت المجموعتين لتحديد ايهما الاكثر تشتتاً لاحظ ان هناك اختلاف فى معيارين يحولاً دون استخدام مقاييس التشتت المطلق فى عملية المقارنة وهما اختلاف حجم العينتين اولاً اى اختلاف قيمة (n) هذا بالاضافة

الى الاختلاف الجوهرى فيما بين مستوى الاجور فى القطاعين والذي تعكسه
المجاميع المعطاه لذا يفضل ان تتم المقارنة من خلال احد معاملات الاختلاف
لمقاييس تدرس التشتت النسبى وعليه فان :

$$C_V \text{ for the first sample} = \frac{\sigma_1}{\bar{X}_1} \times 100$$

$$= \frac{24}{400} \times 100 = 6 \%$$

اما عن معامل الاختلاف النسبى للعينة الثانية فهو :

$$C_V \text{ for the second sample} = \frac{\sigma_2}{\bar{X}_2} \times 100$$

$$= \frac{18}{45} \times 100 = 4 \%$$

ومن خلال مقارنة معاملى الاختلاف النسبى يتضح لنا ان اجور العينة الاولى
تعتبر الاكثر تشتتا (اى الاقل تجانسا) من اجور العينة الثانية مستخدمين فى
عملية المقارنة معامل الاختلاف النسبى كاحد مقاييس التشتت النسبى .

ج - اذا تم دمج العينتين فى عينة واحدة فان :

حجم العينة الجديدة (الكلية) هو :

$$n = n_1 + n_2 = 30 + 50 = 80$$

والوسط الحسابى للعينة الجديدة هو :

$$\bar{X} = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2}{n_1 + n_2}$$

اما عن تباين العينة الجديدة (الكلية) فأنا سوف نورد صيغة التباين المشترك للعينتين (دون اثبات) فهو :-

$$\sigma^2 = \frac{n_1 (\sigma_1^2 + D_1^2) + n_2 (\sigma_2^2 + D_2^2)}{n_1 + n_2}$$

حيث D_1, D_2 هي عبارة عن انحراف وسط العينة العينة عن الوسط الحسابى العام او الفرق ما بين متوسط العينة والمتوسط العام (للعينتين) اى ان :

$$D_1 = \bar{X}_1 - \bar{X} = 400 - 431.25 = 31.25 \quad (L.E)$$

ومن ثم فإن :

$$D_1^2 = (-31.25)^2 = 976.5625$$

كذلك فإن انحراف العينة الثانية عن المتوسط العام هو :

$$D_2 = \bar{X}_2 - \bar{X} = 450 - 431.25 = 18.75 \quad (L.E)$$

وعليه فإن مربع هذا الانحراف هو :

$$D_2^2 = (18.75)^2 = 351.5625$$

ومن ثم فبالتعويض فى صيغة التباين المشترك فيكون :

$$\sigma^2 = \frac{n_1 (\sigma_1^2 + D_1^2) + n_2 (\sigma_2^2 + D_2^2)}{n_1 + n_2}$$

$$= \frac{30(576 + 976.57625) + 50(324 + 351.5625)}{30 + 50}$$

$$= 1004.4375$$

وعليه فإن الانحراف المعياري للعينة الكلية الجديدة هو :

$$\sigma = \sqrt{1004.4375} = 31.693 \quad (L.E)$$

ومن ثم فإن معامل الاختلاف النسبي للعينة الكلية الجديدة هو:

$$C_v = \frac{\sigma}{\bar{X}} \times 100 = \frac{31.693}{431.25} = 7.35 \%$$

الدرجات المعيارية : Standard Scores

تناولنا فيما سبق كيفية المقارنة ما بين تشتت مجموعتين (او اكثر) باستخدام مقاييس التشتت المطلقة والنسبية الا انه في بعض الحالات قد يكون الباحث في حاجة لاجراء عملية مقارنه فيما بين مفردتين او قيمتين في مجموعتين مختلفتين من حيث مقاييس المتوسطات والتشتت وفي هذه الحالة فأننا نكون في حاجة الى مقياس احصائي ياخذ في الاعتبار الترتيب النسبي لكل من هاتين المفردتين داخل المجموعة التي تنتمي اليها . فعلى سبيل المثال لمقارنة مستوى اداء طالب واحد في مادتين او اكثر في مجموعة تلك المواد فلا تصح مقارنه درجة الاختبار في مادة باخرى الا من خلال قياس مستوى اداء هذا الطالب في تلك المواد المختلفة مقارنة باداء مجموعة زملائه في نفس المواد فعلى سبيل المثال اذا كانت درجة طالب معين في مادة الاحصاء هي 75 درجة بينما كانت درجته في مادة الاقتصاد هي 80 درجة . فهل هذا يعنى ان مستوى هذا الطالب في مادة الاقتصاد اعلى من مستواه في مادة الاحصاء. فلو اخذنا

الامور بظاھرھا 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 75 درجة . 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 معينة 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 معه ان هذا الطالب يعتبر اقل الطلاب فى ادائه فى هذه المادة . ومن وجهة اخرى فقد تكون درجة الطالب فى مادة الاحصاء رغم انها الاقل ظاهريا الا انه قد يكون الامتحان الخاص بمادة الاحصاء على درجة عالية من الصعوبة بحيث ان يكون بقية زملائه حصلوا على درجات منخفضة فى تلك المادة وانه نسبيا يعتبر متفوقا فى الاحصاء اذا ما قورنت درجته بدرجات بقية زملائه . وعلى ذلك فأنه لمقارنة مستوى اداء هذا الطالب فى هاتين المادتين فأن يلزمنا استخدام مقياس احصائى ياخذ فى الاعتبار الاداء النسبى للطلاب 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 الشأن 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 المعياري لدرجات الطلاب فى مادة الاقتصاد هما 71 , 6 درجات على الترتيب . فلقياس مستوى الاداء النسبى لهذا الطالب فى هاتين المادتين فأنا 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 المعياري لدرجات هذه المادة ايضا . ويتم ذلك من خلال طرح قيمة الوسط الحسابى لكل توزيع من قيمة الدرجة الطبيعية الخاصة بتلك المادة ثم قسمة الناتج على الانحراف المعياري لهذا التوزيع اى ان الدرجة المعيارية هى :

$$\text{Standard Score} = Z = \frac{X - \bar{X}}{\sigma}$$

حيث تمثل (X) الدرجة الطبيعية للطالب اما (\bar{X} ، σ) فهما عبارة عن الوسط الحسابى والانحراف المعياري لتوزيع درجات الطلاب فى تلك المادة . والهدف من القسمة على الانحراف المعياري هو تجنب الاختلاف فى وحدات القياس حيث ان الاكتفاء بحساب القيمة ($\bar{X} - X$) فى عمليات المقارنة قد تنطوى على نتائج مضلله ومن ثم فبايجاد الدرجة المعيارية لهذا الطالب فى كل مادة من المادتين نجد ان :

درجة الطالب المعيارية فى مادة الاحصاء هى:

$$Z_1 = S.Sc. \text{ for stat. degree} = \frac{X_1 - \bar{X}_1}{\sigma_1} = \frac{75 - 67}{4} = 2$$

اما درجة الطالب المعيارية فى مادة الاقتصاد فهى :

$$Z_2 = S.Sc. \text{ for stat. degree} = \frac{X_2 - \bar{X}_2}{\sigma_2} = \frac{80 - 71}{6} = 1.5$$

ومن خلال مقارنة الدرجتين المعياريتين نجد ان مستوى الاداء النسبي لهذا الطالب فى مادة الاحصاء يفوق المتوسط بما يعادل ضعف الانحراف المعياري اما عن ادائه النسبي فى مادة الاقتصاد فقد كان فوق المتوسط ايضا ولكن بما يعادل مره ونصف فقط من قيمة الانحراف المعياري وهو ما يدل بدوره على ان مستوى الاداء النسبي لهذا الطالب (اى مقارنة بالنسبة لاداء بقية زملائه) فى مادة الاحصاء كان افضل من ادائه النسبي فى مادة الاقتصاد . اى انه يمكن استنتاج ان مستوى هذا الطالب فى مادة الاحصاء افضل منه فى مادة الاقتصاد وهو عكس ما يوحي اليه درجتى الطالب الطبيعية فى هاتين المادتين اى عكس الاستنتاج المبني على ظواهر الامور . وفى ختام هذا الجزء يجب ان نؤكد على ان القيم المعيارية عبارة عن قيم بدلاله وحدات من الانحراف المعياري لذلك فهي تكون مجردة من وحدات القياس وتستخدم بذلك لمقارنة قيم مختلفة من مجموعات مختلفة وللقيم المعيارية عدة خصائص هي :

1- الوسط الحسابي للقيم المعيارية يساوى الصفر دائما بغض النظر عن الوسط

الحسابي للقيم الاصلية ويمكن اثبات ذلك على النحو التالي :

فحيث ان الدرجة المعيارية للدرجة الطبيعية (x_i) هي (z_i) حيث :

$$\bar{z} = \frac{X_i - \bar{X}}{\sigma}$$

فيكون الوسط الحسابي للدرجات المعيارية اي (\bar{z}) هو :

$$\bar{z} = \frac{\sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)}{n} = \frac{1}{n\sigma} \sum_{i=1}^n (X_i - \bar{X})$$

وحيث ان مجموع انحرافات القيم عن وسطها الحسابي يساوي صفرا . لذا فإن

$$\bar{Z} = \frac{1}{n} \sum_{i=1}^n (0) = 0$$

2

- التباين او الانحراف المعياري للقيم المعيارية يساوي الواحد الصحيح ويمكن

اثبات ذلك علي النحو التالي :

$$\sigma_z^2 = \frac{\sum_{i=1}^n (Z_i - \bar{Z})^2}{n} = \frac{\sum_{i=1}^n Z_i^2}{n}$$

وذلك لان \bar{Z} تساوي الصفر (من الخاصية السابقة) . وحيث ان الدرجة المعيارية (z)

$$\bar{Z} = \frac{X_i - \bar{X}}{\sigma}$$

هي :

$$\begin{aligned} \sigma_z^2 &= \frac{\sum_{i=1}^n Z_i^2}{n} = \frac{\sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2}{n} = \frac{1}{\sigma^2} \cdot \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} \\ &= \frac{1}{\sigma^2} (\sigma^2) = 1 \end{aligned}$$

ومن خلال الخاصيتين السابقتين يتضح لنا انه لمعايرة مجموعتين من المفردات فان هذا يعني اننا نردها الي مجموعتين لها نفس المتوسط الحسابي (وهو الصفر) ونفس الانحراف المعياري (وهو الواحد الصحيح) .

مثال :

في إمتحان الشهادة الاعدادية بمحافظة قنا وجد احد الطلاب ان درجاته في الرياضيات واللغة العربية واللغة الانجليزية كانت علي الترتيب هي 45 ، 40 ،

42 درجة . فاذا علمت ان متوسط درجات الطالب بالمحافظة في المواد الثلاثة السابقة هو علي الترتيب 42 ، 35 ، 40 درجة . وان الانحراف المعياري للدرجات في تلك المواد الثلاثة بالمحافظة هو علي الترتيب 6 ، 4 ، 5 فما هو الحكم الموضوعي علي مستوي هذا الطالب في المواد الثلاثة ؟

الحل :

لكي يكون الحكم صحيحا علي مستوي اداء الطالب في المواد الثلاث يجب حساب الدرجات المعيارية للمواد الثلاث . حيث ان :

الدرجة المعيارية للطالب في مادة الرياضيات ولتكن (Z_1) هي :

$$\bar{Z}_1 = \frac{X_1 - \bar{X}_1}{\sigma_1} = \frac{45 - 42}{6} = 0.5$$

والدرجة المعيارية للطالب في مادة اللغة العربية ولتكن (Z_2) هي :

$$\bar{Z}_2 = \frac{X_2 - \bar{X}_2}{\sigma_2} = \frac{40 - 35}{4} = 1.25$$

اما عن الدرجة المعيارية للطالب في مادة اللغة 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 التي تعطى بياناً موضوعياً عن مستوى الطالب في المواد الثلاث نجد أن مستوى الطالب في مادة اللغة العربية أفضل من مستواه في مادة الرياضيات وكذا أفضل من مستواه في مادة اللغة الإنجليزية. وهذا الحكم هو الحكم الصحيح على مستوى أداء الطالب في المواد الثلاث وهو عكس ما توحى به الأرقام المجردة لدرجات الطالب في 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 المجموعة الاولى هي :

17, 22 , 10, 18 , 27 , 15 , 13 , 30

واعمار المجموعة الثانية هي :

38 , 41 , 33 , 39 , 41 , 73, 36 , 35

والمطلوب :-

أ – قارن بين التشتت في المجموعتين وذلك باستخدام المدس وشبهات 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 نتايجك 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 والنسبي 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 المدى 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-	Σ
No. of employees	10	15	30	50	25	20	10	160

والمطلوب حساب :

- 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 التالي 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 مادة 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 الاختلاف الملائم .

(10): اوجد المدي والانحراف المتوسط والانحراف المعياري للقيم التالية :

(أ) : 5 ، 3 ، 8 ، 4 ، 7 ، 6 ، 12 ، 4 ، 3

(ب) : 6 ، 4 ، 9 ، 4 ، 8 ، 6 ، 10 ، 6 ، 6 ، 5 ، 8 ، 8

(11): اوجد الانحراف المعياري للبيانات التالية :

أ- 5 ، 7 ، 1 ، 2 ، 6 ، 3

ب- 4 ، 4 ، 5 ، 2 ، 2 ، 8 ، 4 ، 6 ، 3 ، 2

ج- 1 ، 1 ، 1 ، 0 ، 0 ، 0 ، 0 ، 0

(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

والمطلوب حساب :

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 الخليلى حسب السعر فكانت كالتالى :

Pro. Price 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 والرابع 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 درجة 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 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 والتي 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 المعياري ولكنهما يختلفان من حيث التماثل او الاعتدال لذا فإنه قد يلزم ايجاد مقاييس احصائية اخرى تقيس مدى تماثل التوزيع او التواءه وكذلك تقيس درجة تفرطح (شكل قمة منحني) 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 .

اولا : العزوم : Moments

ان كلمة عزوم بدايتها اساساً فى علم الاستاتيكا فى تحليل القوى التي تؤثر فى جسم او كتلة معينة حيث يقاس عزم القوة التي تؤثر فى جسم ما بحاصل ضرب مقدار القوة في مركز العزم ويتم حساب عزوم مجموعة من القوى بانه عبارة مجموعة حواصل ضرب كل قوة فى ذراع عزمها . والعزم فى الاحصاء ناتج عن تشبيه قيمة الظاهرة بزراع العزم اما تكرار تلك القيمة فهو بمثابة القوة

وان كان العزم يقاس فى الاحصاء بطريقة مختلفة إلى حد ما عن طريقة قياسه فى علم الاستاتيكا .

وتعتبر العزوم من اهم معالم التوزيعات الاحتمالية للظواهر المختلفة حيث يمكن معرفة اى توزيع احتمالى من خلال عزوم هذا التوزيع . وهناك انواع مختلفة من العزوم وهى :

(1):العزوم حول الصفر.

(2):العزوم حول الوسط الحسابى.

(3):العزوم حول وسط فرض معين.

(4):العزوم المعيارية .

وبصفة عامة اذا كان لدينا متغيرا وليكن (X) ياخذ مجموعة القيم

X_1, X_2, \dots, X_n فان العزوم ذات الرتبة (r) حول وسط فرضى وليكن (α)

هو عبارة عن متوسط انحرافات القيم (X_i) عن هذا الوسط الفرضى (α)

مرفوعه للقوه (r) . وهذه القوه (r) تفيد فى تحديد قيم العزوم المتتالية الرتبة

اى العزوم الاولى والعزم الثانى و وهكذا لاي رتبة موجبة . ومن ثم فانه

يمكن وضع الصيغة العامة للعزم ذو الرتبة (r) حول الوسط الفرضى وليكن

(α) على انه:

$$M_r^{(\alpha)} = \frac{\sum_{i=1}^n (X_i - \alpha)^r}{n} \quad (1)$$

وهذه الصيغة (1) تستخدم فى حساب العزوم المتتالية فى حالة البيانات

المفردة اما فى حالة الجدول او التوزيعات التكرارية فانه يتم ترجيع الانحرافات

المحسوبة فى البسط بقيم التكرارات المقابلة لمراكز فئات التوزيع والتي تاخذ

مجموعة القيم المعبرة عن مراكز الفئات أي (X_i) ومن ثم فان الصيغة العامة

للعزم ذو الرتبة (r) حول الوسط الفرضى (α) فى حالة البيانات المبوبة هو
عبارة عن :

$$M_r^{(\alpha)} = \frac{\sum_{i=1}^n (X_i - \alpha)^r F_i}{\sum_{i=1}^n F_i} \quad (2)$$

فإذا تم التعويض عن $r = 1$ فى اى من المعادلتين (1) او (2) نحصل على العزم الاول حول الوسط الفرضى (α) واذا تم التعويض عن $r = 2$ فى هاتين المعادلتين نحصل على العزم الثانى حول الوسط الفرضى (α)..... وهكذا .
وطبقا للانواع المختلفة من العزوم السابق الاشارة اليها فإنه اذا تم التعويض عن الوسط الفرضى (α) فى المعادلتين (1) ، (2) بالقيمة صفر فاننا نحصل على العزوم المتتالية حول الصفر . اما عن العزوم حول الوسط الحسابى فأنها تنتج حينما يتم التعويض فى هاتين المعادلتين عن قيمة الوسط الفرضى على انها تساوى الوسط الحسابى للتوزيع اى (\bar{X}) واخيرا فالعزوم حول وسط فرضى معين وليكن (a) فأنها تنتج من خلال التعويض المباشر فى المعادلتين (1) او (2) عن قيمة $\alpha = a$ اما العزوم المعيارية ففيها يتم ايجاد البسط فى المعادلتين (1) ، (2) من خلال حساب متوسط الدرجات المعيارية وهو ما سيرد فيما بعد الا انه يمكن القول ان العزوم حول الصفر والعزوم حول الوسط تعتبر الاكثر استخداما فى مجال التحليل الاحصائى وفيما يلى نتناول الانواع المختلفة للعزوم وسوف نكتفى من كل نوع بالاشارة الى كيفية ايجاد العزوم الاربعة الاولى ويمكن بالطبع الحصول على العزوم من قوة اكبر اى بدءا من العزم الخامس والسادس و وهكذا .

1- العزوم حول الصفر (العزوم اللامركزية) :-

من خلال المعادلتين (1) ، (2) السابقتين فإنه عند التعويض عن قيمة $\alpha =$ صفرا تصبح صيغة العزم ذو الرتبة (r) حول الصفر في حالة البيانات الغير مبوبة على الصورة :

$$M_r = M_r^{(0)} = \frac{\sum_{i=1}^n (X_i - 0)^r}{n} = \frac{\sum_{i=1}^n X_i^r}{n} \quad (3)$$

حيث $r = 1, 2, \dots, n$

اما صيغة العزم ذو الرتبة (r) حول الصفر في حالة البيانات المبوبة تكراريا فتأخذ الصورة التالية :

$$M_r = M_r^{(0)} = \frac{\sum_{i=1}^n (X_i - 0)^r F_i}{\sum F_i} = \frac{\sum_{i=1}^n X_i^r F_i}{\sum_{i=1}^n F_i} \quad (4)$$

حيث (X_i) تعبر عن مركز الفئة (i) حيث $i=1, 2, \dots, n$ كما ان (n) تعبر عن عدد فئات التوزيع في هذه الحالة .

ومن خلال المعادلات (3) ، (4) يمكن بيان العزوم الاربعة الاولى حول الصفر وذلك في حالتى البيانات الغير مبوبة والبيانات المبوبة والجدول (1-5) يبين هذه العزوم الاربعة حول الصفر :

جدول (1-5)
العزوم الأربعة الأولى حول الصفر

Case r^{th} Moment	Values (القيم) (المفردة)	Frequency Distributions
$M_1^{(0)} = M_1$	$\frac{\sum_{i=1}^n X}{n} = \bar{X}$	$\frac{\sum_{i=1}^n X_i F_i}{\sum F_i} = \bar{X}$
$M_2^{(0)} = M_2$	$\frac{\sum_{i=1}^n X_i^2}{n}$	$\frac{\sum_{i=1}^n X_i^2 F_i}{\sum F_i}$
$M_3^{(0)} = M_3$	$\frac{\sum X_i^3}{n}$	$\frac{\sum_{i=1}^n X_i^3 F_i}{\sum F_i}$
$M_4^{(0)} = M_4$	$\frac{\sum X_i^4}{n}$	$\frac{\sum_{i=1}^n X_i^4 F_i}{\sum F_i}$

لاحظ انه من التعريف السابق للعزوم حول نقطة الاصل اى حول الصفر سواء
فى حالة القيم (المفردات) او البيانات المبوبة تكراريا فان العزم الاول حول
الصفر ما هو الا الوسط الحسابى اى ان :

$$M_1^{(0)} = \bar{X}$$

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 $\alpha = \bar{X}$ في المعادلتين (1) ، (2) السابقتين ومن ثم فان صيغة العزم ذو الرتبة (r) حول الوسط الحسابي (المركزية) في حالة البيانات المفردة تأخذ الصورة التالية :

$$\hat{M}_r = M_r(\bar{X}) = \frac{\sum_{i=1}^n (X_i - \bar{X})^r}{n} \quad (5)$$

Where $r = 1, 2, 3, \dots$

اما عن صيغة العزم ذو الرتبة (r) حول الوسط الحسابي في حالة البيانات المبوبة تكراريا فهي تأخذ الصورة التالية :

$$\hat{M}_r = M_r(\bar{X}) = \frac{\sum_{i=1}^n (X_i - \bar{X})^r F_i}{\sum F_i} \quad (6)$$

Where $r = 1, 2, 3, \dots$

حيث (X_i) تعبر عن مركز الفئة (i) ، (F_i) تعبر عن تكرار الفئة (i) ، $i=1,2,\dots,n$ ، (n) تعبر عن عدد فئات التوزيع التكرارى .

والجدول (2-5) يلخص صيغة العزوم الاربعة الاولى حول الوسط (المركز) وذلك في حالة البيانات المفردة (الغير مبوبة) والبيانات المبوبة .

جدول (5 - 2)

Case r^{th} Moment	Values (القيم المفردة)	Frequency distributions
$\hat{M}_1 = M_1^{(\bar{x})}$	$\frac{\sum (X_i - \bar{X})}{n} = 0$	$\frac{\sum (X_i - \bar{X}) F_i}{\sum F_i}$
$\hat{M}_2 = M_2^{(\bar{x})}$	$\frac{\sum (X_i - \bar{X})^2}{n} = \sigma^2$	$\frac{\sum (X_i - \bar{X})^2 F_i}{\sum F_i} = \sigma^2$
$\hat{M}_3 = M_3^{(\bar{x})}$	$\frac{\sum (X_i - \bar{X})^3}{n}$	$\frac{\sum (X_i - \bar{X})^3 F_i}{\sum F_i}$
$\hat{M}_4 = M_4^{(\bar{x})}$	$\frac{\sum (X_i - \bar{X})^4}{n}$	$\frac{\sum (X_i - \bar{X})^4 F_i}{\sum F_i}$

لاحظ ايضا انه من خلال التعريف السابق للعزوم حول الوسط الحسابي 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 وليكن (a) :

إذا اعتبرنا ان $\alpha = a$ حيث (a) عبارة عن مقدار ثابت معين فإنه بالتعويض فى المعادلتين (1) ، (2) السابق الاشارة لهما نحصل على العزوم المختلفة

حول هذا الوسط الفرضى (a) ومن ثم فان صيغة العزم ذو الرتبة (r) حول الوسط الفرضى (a) فى حالة المفردات تأخذ الصورة التالية :

$$M_r^{(a)} = \frac{\sum (X_i - a)^r}{n} \quad \text{where } r = 1, 2, 3, \dots \quad (7)$$

اما عن صيغة العزم ذو الرتبة (r) حول الوسط الفرضى (a) فى حالة البيانات المبوبة تكراريا فتأخذ الصورة التالية :

$$M_r^{(a)} = \frac{\sum (X_i - a)^r F_i}{\sum F_i} \quad \text{where } r = 1, 2, 3, \dots \quad (8)$$

والجدول (3-5) يلخص صورة العزوم الاربعة الاولى حول الوسط الفرضى (a) دول (3-5)

Case r^{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 :

$$M_r^{(a)} = \frac{\sum d_i^r}{n} \quad \text{في حالة المفردات}$$

$$= \frac{\sum d_i^r F_i}{\sum F_i} \quad \text{في حالة البيانات المبوبة}$$

وذلك باعتبار ان d_i تمثل انحرافات القيم (المفردات في حالة البيانات الغير مبوبة او مراكز فئات التوزيع (في حالة البيانات المبوبة) عن الوسط الفرضي (a) اي ان :

$$d_i = X_i - a$$

سواء كانت X_i تعبر عن المفرده رقم (i) في حالة المفردات او تعبر عن مركز الفئة (i) في حالة البيانات المبوبة .

4- العزوم المعيارية : Standard Moments

بصفة عامة تأخذ صورة العزم المعيارى ذو الرتبة (r) الصيغة التالية في حالة المفردات :

$$M_r^{(z)} = \frac{\sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^r}{n} = \frac{\sum Z^r}{n} \quad \text{where } r = 1, 2, \dots \dots \dots (9)$$

اما عن صيغة العزم المعيارى ذو الرتبة (r) في حالة البيانات المبوبة فانه يأخذ الصورة التالية:

$$M_r^{(z)} = \frac{\sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^r F_i}{\sum F_i} = \frac{\sum Z^r F_i}{\sum F_i} \quad (10)$$

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 التوالي .

والعزم الاول المعيارى دائما ابدا قيمته صفرية والعزم الثانى المعيارى دائما ابدا قيمته مساوية للواحد الصحيح وذلك لان العزم الثانى المعيارى ما هو الا تباين القيم المعيارية والتي سبق وان اوضحنا ان الوسط الحسابى للقيم المعيارية هو صفرا اما التباين لها فهو الواحد الصحيح.

هذا ويمكن من خلال صيغ العزوم المختلفة فى حالة الجداول التكرارية تحويلها لى تتناسب مع التوزيعات التكرارية النسبية . وسوف اترك للقارئ هذه الجزئية على ان يتعامل معها كما تم فى المقاييس السابقة لهذا الباب باختلاف انواعها .

مثال :

اذا كان لديك مجموعة القيم التالية : 4 , 5 , 3 , 6 , 7

فالمطلوب حساب :

- 1- العزوم الاربعة الاولى حول الصفر.
- 2- العزوم الاربعة الاولى حول الوسط الحسابى .
- 3- العزوم الاربعة الاولى حول المعيارية .

الحل :

1- العزوم الاربعة الاولى حول الصفر تتطلب حساب المجاميع التالية :

$\sum x$, $\sum x^2$, $\sum x^3$, $\sum x^4$ 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

$$M_r^{(0)} = \frac{\sum X^r}{n} \quad \text{في حالة المفردات هي :}$$

لذا فان العزوم الاربعة الاولى هي :

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 الصفر هو :

$$M_3^{(0)} = \frac{\sum X^3}{n} = \frac{575}{5} = 115$$

واخيرا فان العزم الرابع حول الصفر هو :

$$M_4^{(0)} = \frac{\sum X^4}{n} = \frac{4609}{5} = 921.8$$

-2 العزوم الاربعة الاولى حول الوسط الحسابي (العزوم المركزية) :

من خلال العزم الاول حول الصفر والذي يعتبر بمثابة الوسط الحسابي للقيم فإنه صيغة العزوم الاربعة الاولى حول الوسط الحسابي تأخذ الصورة التالية باعتبار ان $\bar{X} = 5$ هي :

$$\hat{M}_r = M_r^{(\bar{X})} = \frac{\sum (X_i - 5)^r}{n}$$

والجدول التالي يوضح حسابات العزوم الاربعة الاولى لتلك المفردات
جدول (5-5)

المجاميع اللازمة لحساب العزوم الاربعة الاولى حول الوسط الحسابي

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 فإن العزوم الاربعة الاولى حول الوسط الحسابي هي :

$$\hat{M}_1 = M_1^{(\bar{X})} = \frac{\sum (X - 5)}{n} = \frac{0}{5} = 0$$

$$\hat{M}_2 = M_2^{(\bar{X})} = \frac{\sum(X - 5)^2}{n} = \frac{10}{5} = 2$$

،

$$\hat{M}_3 = M_3^{(\bar{X})} = \frac{\sum(X - 5)^3}{n} = \frac{0}{5} = 0$$

،

$$\hat{M}_4 = M_4^{(\bar{X})} = \frac{\sum(X - 5)^4}{n} = \frac{34}{5} = 6.8$$

3- العزوم الاربعة الاولى المعيارية:

لحساب العزوم الاربعة الاولى المعيارية يتم حساب القيم المعيارية للمفردات من خلال التحويل المعيارية التالية :

$$Z_i = \frac{X_i - \bar{X}}{\sigma}$$

وحيث ان العزم الاول حول الصفر ما هو إلا الوسط الحسابي (\bar{X}) وهو في مثالنا يساوى القيمة (5) والعزم الثانى حول الوسط الحسابى هو بمثابة تباين المفردات المعطاه اى ان $\sigma^2 = 2$

لذا فان الانحراف المعيارى هو $\sigma = \sqrt{2} \cong 1.414$

ومن ثم يمكن حساب العزوم الاربعة الاولى المعيارية 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
6	0.707	1.999396	-2.827145944	3.997584365
7	1.414	0.499849	0.353393243	0.249849022
	0	1.999396	2.827145944	3.997584365
	0	4.99849	0	8.494866775

وعليه فان العزوم الاربعة الاولى المعيارية هي :

$$M_1^{(z)} = \frac{\sum \left(\frac{X_i - \bar{X}}{\sigma} \right)}{n} = \frac{0}{5} = 0$$

$$M_2^{(z)} = \frac{\sum \left(\frac{X_i - \bar{X}}{\sigma} \right)^2}{n} = \frac{4.99849}{5} = 0.999698 \simeq 1$$

$$M_3^{(z)} = \frac{\sum \left(\frac{X_i - \bar{X}}{\sigma} \right)^3}{n} = \frac{8.494866775}{5} = 1.698973355$$

لاحظ ان العزم الاول المعيارى ما هو الا الوسط الحسابى للقيم المعيارية وهو يساوى الصفر دائما . كذلك فان العزم الثانى المعيارى 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 .

مثال :-

احسب العزوم الاربعة الاولى حول الصفر وكذا حول الوسط الحسابي للتوزيع التكراري التالي :

Classes	1-	5-	9-	13-17
Frequency	2	3	4	1

الحل :-

لحساب العزوم الابعه الاولى حول الصفر نكون الجدول التالي :-

Classes	F_i	x_i	$X_i F_i$	$X_i^2 F_i$	$X_i^3 F_i$	$X_i^4 F_i$
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 الاربعة الاولى 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 الاولى حول الوسط الحسابي

Classes	F _i	X _i	(X _i - 8.6)	(X _i - 8.6)F _i	(X _i - 8.6) ² F _i	(X _i - 8.6) ³ F _i	(X _i - 8.6) ⁴ F _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 الاولى حول الوسط الحسابي هي :

$$M_1^{(\bar{X})} = \frac{\sum (X_i - \bar{X}) F_i}{\sum F_i} = \frac{0}{10} = 0 \quad (\text{بالتعريف})$$

$$M_2^{(\bar{X})} = \frac{\sum (X_i - \bar{X})^2 F_i}{\sum F_i} = \frac{134.4}{10} = 13.44 = \sigma^2$$

$$M_3^{(\bar{X})} = \frac{\sum (X_i - \bar{X})^3 F_i}{\sum F_i} = \frac{-46.08}{10} = -4.608$$

$$M_4^{(\bar{X})} = \frac{\sum (X_i - \bar{X})^4 F_i}{\sum F_i} = \frac{3796.992}{10} = 379.6992$$

• 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 راينا ان العزم الاول حول نقطة الاصل (الصفري) ما هو الا عباره عن الوسط الحسابي . كما تمكنا العزوم ايضا من معرفة اهم مقياس من مقاييس التشتت المطلق وهو الانحراف المعياري ومن ثم التباين حيث راينا ان العزم الثاني حول الوسط الحسابي (العزم الثاني المركزي) ما هو الا التباين . والاهم من ذلك فأن العزوم حول الوسط اي العزوم المركزية تمكنا من دراسة مدي تماثل التوزيع من عدمه . الا ان طريقة حساب العزوم المختلفة حول الوسط تحمل في فحواها حسابات صعبة ومعقدة وخصوصا عندما يزداد عدد المفردات والقيم وبصفه خاصة اذا احتوت قيمة الوسط الحسابي (\bar{X}) علي قيم كسرية فنضطر الي تقريبيها 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 العلاقة ما بين العزوم حول الوسط الحسابي والعزوم حول نقطة الاصل وسوف نورد هذه العلاقات دون برهان للتبسيط علي القارئ .

وهذه العلاقة هي :

$$M_3^{(\bar{X})} = M_3^{(0)} - 3 M_2^{(0)} M_1^{(0)} + 2 (M_1^{(0)})^2$$

$$M_4^{(\bar{X})} = M_4^{(0)} - 4 M_3^{(0)} M_1^{(0)} + 6 M_2^{(0)} (M_1^{(0)})^2 - 3 (M_1^{(0)})^4$$

• العزوم المصححة لشبيرد :

سبق وان ذكرنا انه عند تبويب المفردات او البيانات في صورة جدول تكراري فاننا نعلم 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 بانتظام علي مدار طول الفئة وهذا الافتراض وأن كان يشوبه عدم الدقه لذا قام شبيرد بعمل تصحيح للعزوم المركزية حيث اوجد تصحيحا للعزم الثاني حول الوسط بحيث باخذ الصورة التالية :

$$\text{correction of } M_2^{(\bar{X})} = M_2^{(\bar{X})} - \frac{L_i^2}{12}$$

حيث (L_i) تعبر عن طول الفئة في التوزيع التكراري كما قام شبيرد بتصحيح العزم الرابع حول الوسط الحسابي ليأخذ الصورة المصححة التالية :

$$\text{correction of } M_4^{(\bar{X})} = M_4^{(\bar{X})} - \frac{L_i^2}{12} M_2^{(\bar{X})} + \frac{7}{240} L_i^4$$

ويشترط لاجراء هذه التصحيحات ان يكون منحنى التوزيع التكراري متصلا وأن يقابل طرفي المنحنى المحور الافقي وان تكون فئات التوزيع متساوية الطول . وعلي المستوي الواقع العملي فانه نادرا ما تستخدم العزوم المصححة ولن تقوم بحسابها الا اذا نص المطلوب صراحة علي حسابها .

ثانياً :- الالتواء : skewness

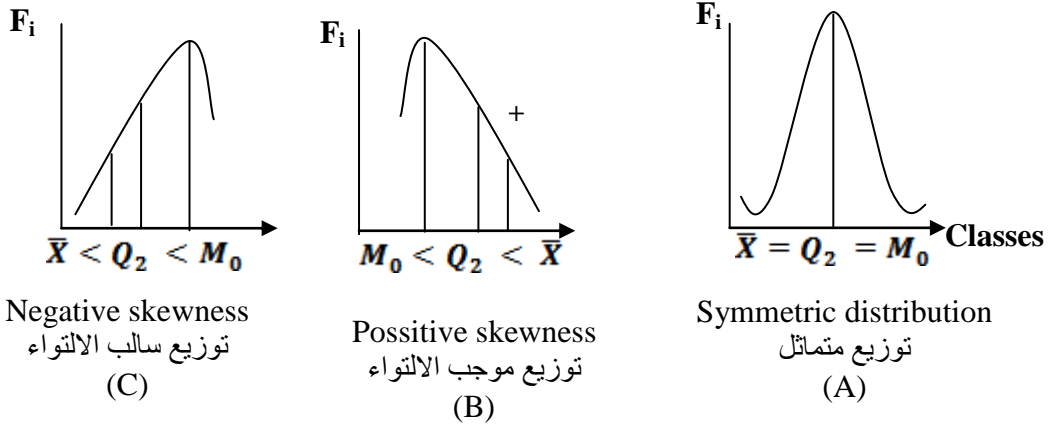
بعد ان تعرضنا في دراستنا السابقة الي مقاييس النزعة المركزية ومقاييس التشتت بنوعها المطلقة والنسبية والعزوم الرياضية للبيانات الغير مبوبة او المبوبة تكراريا لظاهرة ما تأتي الان دراسة ما اذا كان التوزيع الخاص بظاهرة ما متماثلا او ملتويا وما هو شكل ودرجة التواء التوزيع دون ان نقوم برسم منحنى التوزيع . حيث يفيد ذلك في وصف وتحليل توزيع الظاهرة وذلك لان التوزيع المتماثل يعني ان المنحنى التكراري للظاهرة موضع الدراسة تنقسم عند القيمة المتوسطة لها إلى قسمين متطابقين تماما ويكون تزايد او تناقص التكرارات منتظما او متشابهها او بطريقة متماثلة علي جانبي المحور الراسي المقام عند القيمة المتوسطة للتوزيع .

اما التواء التوزيع فنقصد به بعد المنحنى عن التماثل اي تركيز 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 الاولى) للتوزيع اما التكرارات الصغيرة فنتشر عند الفئات العليا من التوزيع ويكون ذيل التوزيع جهة اليمين اطول من جهة اليسار والعكس صحيح حيث يكون التوزيع سالب الالتواء Negative skewness (اي ملتوي جهة اليسار) نتيجة لتركيز عدد كبير من التكرارات عند الفئات العليا للتوزيع بينما تنتشر قليل من التكرارات عند الحدود الدنيا للتوزيع ويكون حينئذ التوزيع ذو ذيل اطول جهة اليسار (توزيع ذو ذيل ايسر) .

وخلاصة ما سبق هو ان التوزيع التكراري من حيث التماثل او الالتواء يكون واحدا من الصور الثلاث : فاما ان يكون التوزيع متماثلا او ملتوي جهة اليمين او ملتوي جهة اليسار .

وفي دراستنا لمقاييس الالتواء سوف نميز بين نوعين من المقاييس حيث يعتمد النوع الاول في ايجاد علي مقاييس النزعة المركزية ومقاييس التشتت التي تعرضنا لها سابقا . اما النوع الثاني من مقاييس الالتواء فيعتمد في حسابه علي العزوم .

هذا ويمكن من خلال مقارنة مقاييس المتوسطات (الوسط – الوسيط – والنوال) دراسة ماهية تماثل التوزيع من عدمه كما يوضحه شكل (1-5)



شكل (1-5)

* فاذا كان التوزيع متماثلا فان الوسط الحسابي يتساوى مع الوسيط يتساوى مع النوال اي ان :

$$\bar{X} = Q_2 = M_0$$

كما هو واضح في شكل (A-1-5)

اما اذا ابتعدت التوزيعات عن التماثل فان قيم مقاييس المتوسطات ستختلف عن بعضها . وحيث ان الوسط الحسابي يتاثر بالقيم المتطرفة او الشاذة لذلك نراه دائما مانلا في اتجاه ذيل التوزيع بينما تكون قيمة النوال 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 (c - 1 - 5) هذا ويزداد قرب او بعد التوزيع عن التماثل بمدى قرب او بعد مقاييس المتوسطات الثلاثة عن بعضها البعض. إلا أن هذه الدراسة تعطى اشارة لنوع الالتواء فقط دون قياس لدرجة الالتواء. لكن لتحديد درجة الالتواء بالإضافة لنوعه أيضا فيجب حسب اى من نوعى معاملات الالتواء السابق الاشارة لها والتي سترد حالا فى هذا القسم .

1- معاملات الالتواء بدلاله مقاييس المتوسطات و التشتت : - ذكرنا حالا أنه كلما بعدت مقاييس المتوسطات عن بعضها البعض كلما بعد التوزيع عن التماثل ومن ثم فإنه يمكن استخدام الفروق بين قيم هذه المتوسطات فى ايجاد مقاييس الالتواء . لكن فى الحقيقة فان الفروق فيما بين مقاييس المتوسطات وحدها فى حد ذاتها لا تقيس درجة او حدة الالتواء بشكل صحيح اذا انه من الممكن ان يكون الالتواء كبيرا 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 في 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 في حالة التوزيعات المتماثلة.
 - * ان تكون قيمة معامل الالتواء مجردة من وحدات القياس .
- وفيما يلي اكثر معاملات الالتواء شيوعا :

$$B_1 = \frac{\bar{X} - M_0}{\sigma}$$

أ - معامل الالتواء الاول لبيرسون (B_1) وهو عبارة عن :

اى أن معامل الالتواء الأول لبيرسون هو عبارة عن نسبة الفرق ما بين الوسط الحسابى والمنوال الى الانحراف المعيارى للتوزيع .

ب - معامل الالتواء الثانى لبيرسون (B_2) وهو عبارة عن :

$$B_2 = \frac{3(\bar{X} - Q_2)}{\sigma}$$

اى النسبة بين ثلثه امثال الفرق بين الوسط والوسيط الى الانحراف المعيارى للتوزيع.

وتتراوح قيمتى معاملى الالتواء الاول والثانى لبيرسون (B_1 , B_2) فما بين $(-3$, $3+)$ إلا أن ما يعاب على هذين المعاملين ان كلاهما لا يمكن حسابه من الجداول التكرارية المفتوحة حيث يصعب حساب كل من الوسط الحسابى والانحراف المعيارى كما يتاثر هذين المعاملين بالقيم الشاذة او المتطرفة .

وهذين المعاملين تتساوى قيمتهم بالصفر فى حالة التوزيعات المتماثلة اما اذا كان الوسط الحسابى اكبر من كل من المنوال والوسيط فان التوزيع يكون موجب الالتواء والعكس صحيح .

وللتغلب على عيوب معاملى بيرسون للالتواء B_1 , B_2 من حيث عدم امكانية حسابهم فى حالة التوزيعات التكرارية المفتوحة وكذلك تاثرهم بالقيم الشاذة فقد امكن استنتاج معامل ثالث للالتواء وهو ما يسمى بمعامل بولى Bowley للالتواء او ما يسمى بمعامل الالتواء الربيعى والذى ياخذ الصورة التالية :

$$B_3 = \frac{3(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1}$$

وأهم ما يميز هذا المعامل بالاضافة لعدم تأثره بالقيم الشاذة او المتطرفه فهو المعامل الوحيد الذى يمكن ايجاده من خلال الرسم البيانى حيث يمكن من خلال رسم المنحنى التكرارى المجتمع الصاعد او الهابط ايجاد كل من الربيعين الادنى والاعلى والوسيط ثم التعويض فى صيغة المعامل.

هذا وتتساوى قيمة معامل بولى (B_3) بالصفر فى حالة التوزيعات المتماثلة وذلك لان الربيعين الادنى (Q_1) والاعلى (Q_3) يقعان على بعدين متساويين من الوسيط (Q_2) فى هذه الحالة وتتراوح قيمة معامل بولى للالتواء ما بين (-1 , +1) لكن يعاب على هذا المعامل انه لا يأخذ فى الاعتبار مجموعة القيم التى تسبق الربيع الادنى او التالية للربيع الاعلى ولذلك يعتبر اقل دقة من معاملى بيرسون للالتواء حيث لا ياخذ كل القيم فى الحساب .

د - هناك معامل رابع يقيس نوع ودرجة الالتواء وان كان اقل استخداما من معاملات السابقة وهو ما يسمى بمعامل الالتواء الميننى نظرا لاعتماده على حساب الميننيات وهو ياخذ الصورة التالية :

$$B_4 = \frac{(P_{90} - Q_2) - (Q_2 - P_{10})}{P_{90} - P_{10}}$$

حيث P_{10} , P_{90} تمثل المنين العاشر والمنيين التسعون على الترتيب.

2- قياس الالتواء باستخدام العزوم المركزية (حول الوسط) :

قياس الالتواء باستخدام العزوم يعتمد على قيمة العزم الثالث حول الوسط الحسابي اى $(M_3^{(\bar{X})})$ فاذا كانت قيمة العزم الثالث حول الوسط مساوية للصفر يكون التوزيع متماثلا حيث تكون الانحرافات السالبة مساوية للانحرافات الموجبة ويكون المجموع الجبرى لمكعبات تلك الانحرافات مساويا للصفر اما اذا كانت قيمة العزم الثالث حول الوسط تختلف عن الصفر فيكون التوزيع متلوياً وله نفس اتجاه اشارة هذا العزم بمعنى اذا كانت :

$$M_3^{(\bar{X})} > 0$$

فيكون التوزيع موجب الالتواء اما اذا كان :

$$M_3^{(\bar{X})} < 0$$

فيكون التوزيع سالب الالتواء اما عن قياس درجة الالتواء فتكون من خلال النسبة ما بين العزم الثالث حول الوسط والجذر التربيعى لمكعب العزم الثانى حول الوسط اى ان معامل الالتواء بدلاله العزوم (B_5) ياخذ الصورة التالية :

$$B_5 = \frac{M_3^{(\bar{X})}}{\sqrt{(M_2^{(\bar{X})})^3}} = \frac{M_3^{(\bar{X})}}{(M_2^{(\bar{X})})^{3/2}}$$

وبصفه عامة فان معاملات الالتواء الخمسة السابقة تشترك فى دراسة درجة

تماثل او التواء التوزيع فيما يلى :

- (1): فإذا كانت قيمة معامل الالتواء تساوى الصفر يكون التوزيع متماثلاً .
 (2): وإذا كانت قيمة معامل الالتواء موجبا يكون التوزيع ملتويا جهة اليمين
 أى أن له التواء موجب.
 (3): وإذا 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 اى ان الالتواء وحده غير كاف لتحديد معالم التوزيع التكرارى او لاجراء المقارنات فيما بين التوزيعات التكرارية المختلفة . فقد تكون احداها ذات قمة عريضة (مفرطح) او قد تكون اكثر تدببا لذلك فانه لا بد من وجود مقياس احصائى يدرس شكل قمة التوزيع وهو ما يسمى بمعامل التفرطح والذى يقىس درجة تفرطح منحنى التوزيع اى شكل قمته وهناك معاملين اساسيين يمكن استخدامهم فى دراسة شكل قمة المنحنيات التكرارية يسميان بمعاملات التفرطح وهما :

أ – معامل التفرطح بدلاله العزم :

وهذا المعامل هو خارج قسمة العزم الرابع المعيارى حول الوسط على مربع العزم الثانى حول الوسط ويرمز له بالرمز (α_1) حيث ان :

$$\alpha_1 = \frac{M_4(\bar{X})}{(M_2(\bar{X}))^2}$$

وتمت القسمة على مربع العزم الثاني حول الوسط (مربع التباين) وذلك للتخلص من وحدات القياس ونحصل منه على مقياس نسبي .
وقد وُجد ان قيمة هذا المعامل (α_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 الاخرى. لذا فإنه بعد حساب معامل التفرطح بدلالة العزوم (α_1) فتتم مقارنة معامل تفرطح التوزيع بمعامل التفرطح الخاص بالتوزيع المعتدل (اي بالقيمة 3). وهنا نكون بصدد احد الحالات الثلاث التالية :

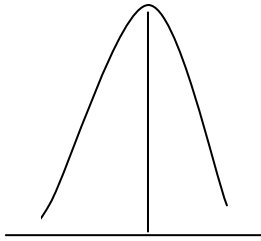
* اذا كانت قيمة $\alpha_1 = 3$: فان هذا يعني ان المنحني طبيعي او معتدل اي متوسط التفرطح Mesokurtic كما موضح في شكل (A-2-5). * واذا كانت α_1 للمنحني اقل من (3) : فان هذا يعني ان منحني التوزيع مفرطح Platykurtic اي ذو قمة عريضة (اي يكون اكثر اتساعا من وسطه وتنخفض قمته عن قمة المنحني المعتدل) كما هو موضح في شكل (B-2-5).
اما اذا كانت قيمة (α_1) اكبر من (3) : فان هذا يعني ان منحني التوزيع مدبب Leptokurtic اي يكون اكثر ضيقا من وسطه وترتفع قمته عن المنحني المعتدل .

ب – معامل التفرطح بدلالة الربيعيين والمئينيات :

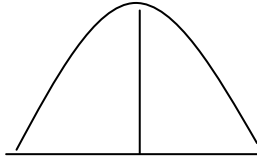
وهذا المعامل عبارة عن النسبة ما بين نصف المدى الربيعي وبين الفرق ما بين المئين التسعين والمئين العاشر ويرمز له بالرمز (α_2) اي ان

$$\alpha_2 = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}$$

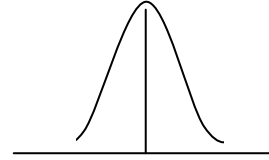
وتبلغ قيمة هذا المعامل 0.263 في حالة التوزيع الطبيعي في حين يكون منحنى التوزيع مفرطاً إذا كانت قيمة معامل التفرطح (α_4) أكبر من (0.263) ويكون منحنى التوزيع مدبباً إذا كانت قيمة معامل التفرطح (α_4) أقل من القيمة 0.263 هذا ويفضل استخدام هذا المعامل في حالة الجداول التكرارية المفتوحة حيث يصعب إيجاد المعامل الأول للتفرطح والمعتمد في قياسه علي العزوم . كما يفضل استخدامه في حالة وجود قيم شاذة او متطرفه في البيانات.



Leptokurtic Dist.
توزيع مدبب



Platykurtic Dist.
توزيع مفرطح



Mesokurtic Distributio.
توزيع معتدل (طبيعي)

شكل (2-5)

ملاحظات حول الالتواء والتفرطح :-

(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 وهذا يعني ضرورة استخدام المعامل نفسه عند مقارنة درجة الالتواء لتوزيعين او اكثر .

(3): تتوقف اشارة ايأ من معاملات الالتواء علي اشارة البسط وذلك لان اشارة المقام دائما ابدا موجب باستثناء حالة وحيدة وهي حينما تكون قيمه (σ) او

الفرق بين الربيعيين الاعلى والادني مساوية للصفر وهي تمثل حالة التجانس الكامل للبيانات (ظاهرة تأخذ قيمة ثابتة دون تغير او اختلاف)

مثال :

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 . فكانت قيمة الوسط الحسابي كما يلي :

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

$$\bar{X} = \frac{\sum X_i F_i}{\sum F_i} = \frac{1310}{50} = 26.2$$

وعليه فإن العزوم الأربعة الأولى حول الوسط هي :

$$M_1^{(\bar{X})} = \frac{\sum (X_i - \bar{X}) F_i}{\sum F_i} = \frac{\sum (X_i - 26.2) F_i}{\sum F_i} = \frac{0}{50} = 0$$

$$M_2^{(\bar{X})} = \frac{\sum (X_i - \bar{X})^2 F_i}{\sum F_i} = \frac{\sum (X_i - 26.2)^2 F_i}{\sum F_i} = \frac{336}{50} = 6.72$$

$$M_3^{(\bar{X})} = \frac{\sum (X_i - \bar{X})^3 F_i}{\sum F_i} = \frac{\sum (X_i - 26.2)^3 F_i}{\sum F_i} = \frac{-96}{50} = -1.92$$

$$M_4^{(\bar{X})} = \frac{\sum (X_i - \bar{X})^4 F_i}{\sum F_i} = \frac{\sum (X_i - 26.2)^4 F_i}{\sum F_i} = \frac{5614.08}{50} = 112.2816$$

وعليه فإن معامل الألتواء بدلالة العزوم هو :

$$B_4 = \frac{M_3^{(\bar{X})}}{\sqrt{(M_2^{(\bar{X})})^3}} = \frac{-1.92}{\sqrt{(6.72)^3}} = \frac{-1.92}{17.420231} = -0.1102$$

وهو ما يعني أن هذا التوزيع وإن كان يقترب من التماثل إلا أنه يميل للالتواء السالب بدرجة ضعيفة.

أما عن معامل التفرطح بدلالة العزوم (α_1) فهو عبارة عن :

$$\alpha_1 = \frac{M_4^{(\bar{X})}}{(M_2^{(\bar{X})})^2} = \frac{112.2816}{(6.72)^2} = 2.4864$$

وحيث أن قيمه معامل التفرطح أقل من (3) فهذا يفيد أن التوزيع مفرطح وإن كانت القيمة تقترب من القيمة (3) أي أن التوزيع يقترب من الاعتدال .

مثال:

إذا كان العزم الاول حول الرقم (3) هو (8) فما هي قيمة الوسط الحسابي للتوزيع .

الحل :

حيث أن صيغة العزم ذو الرتبة (r) حول الوسط الفرضى (a) يأخذ الصورة التالية (في 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) فإن :

$$M_1^{(3)} = \frac{\sum(X_i - 3)^1}{n} = 8$$

ومن ثم فإن :

$$\sum_{i=1}^n X_i - \sum_{i=1}^n 3 = 8n$$

$$\sum X_i - 3n = 8n$$

$$\sum X_i = 8n + 3n = 11n$$

أى أن :

$$\frac{\sum X_i}{n} = 11$$

$$i. e., \bar{X} = 11$$

مثال :

إذا كان لديك التوزيع التكرارى المتجمع التالى لتوزيع منتظم لأجر عينة من 50 عاملا (بالجنيه) فى احد المصانع فى اليوم :

Less than 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).

ج - حدد نسبة العمال الذين تزيد اجورهم عن 90 جنيها اسبوعيا .

الحل :-

معطى فى المثال أن التوزيع التكرارى منتظما لذا فيمكن معرفة الحد الادنى للفئة 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 متجمعا لجدول اصلى مغلق من اعلى وحيث أن العينة حجمها 50 عاملا وأن التكرار المتجمع الصاعد للفئة الاخيرة هو 50 عاملا ايضا لذا فإن 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 . والخاصة أن هذا التوزيع المتجمع مشتق من توزيع تكرارى مغلق الطرفين لذا فإن معامل الاختلاف المناسب هو معامل الاختلاف النسبى . ولكى يتم حساب هذا المعامل يجب رد التوزيع المتجمع الصاعد لصورة التوزيع التكرارى الاصلى حتى يتم حساب كل من الوسط الحسابى 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:

Classes	F_i	X_i	$D_i = \frac{X_i - 80}{10}$	$D_i F_i$	$D_i^2 F_i$
45-	2-0=2	50	-3	-6	18
55-	8-2=6	60	-2	-12	24
65-	15-8=7	70	-1	-7	7
75-	35-15=20	80	0	0	0
85-	43-35=8	90	1	8	8
95-	47-4=3	100	2	8	16
105-115	50-47=3	110	3	7	27
Σ	50			0	100

وعليه فإن :

$$\bar{X} = a + b \frac{\sum D_i F_i}{\sum F_i} = 80 + 10 \left(\frac{0}{50} \right) = 80 \quad (L.E)$$

،

$$\sigma = b \sqrt{\frac{\sum D_i^2 F_i}{\sum F_i} - \left(\frac{\sum D_i F_i}{\sum F_i} \right)^2}$$

$$= 10 \sqrt{\frac{100}{50} - \left(\frac{0}{50}\right)^2} = 10\sqrt{2} = 14.14 \quad (L.E)$$

ومن ثم فإن معامل الاختلاف النسبي هو :

$$C_v = \frac{\sigma}{\bar{X}} \times 100 = \frac{14.14}{80} \times 100 = 17.675 \%$$

ب- لإيجاد معامل الالتواء الثانى لبيرسون أى :

$$B_2 = \frac{3(\bar{X} - Q_2)}{\sigma}$$

وهذا المعامل يتطلب حساب الوسيط اما الوسط والانحراف المعياري فتم حسابهم فى المطلوب السابق فحيث أن:

$$\text{Rank of } Q_2 = \frac{\sum F}{2} = \frac{50}{2} = 25$$

وعليه فإنه من خلال الجدول التكرارى المتجمع الصاعد فان قيمة الوسيط هي:

$$Q_2 = 75 + 10 = \left(\frac{25 - 15}{35 - 15}\right) = 80 \quad (L.E)$$

ومن ثم فإن معامل الالتواء الثانى (B_2) عبارة عن:

$$B_2 = \frac{3(80 - 80)}{14.14} = 0$$

وهو ما يعنى أن التوزيع متماثل.

ج- لإيجاد نسبة العمال الذين تزيد أجورهم عن 90 جنيهاً فإن يمكن من الجدول التكرارى المتجمع الصاعد معرفة نسبة الاسر الذين تقل أجورهم عن 90 جنيهاً ثم بطرح النسبة الناتجة من 100% لينتج المطلوب مباشرة بدلا من ايجاد التوزيع التكرارى المتجمع الهابط واستنتاج النسبة المطلوبة .

(على الطالب إستنتاج المطلوب والاجابة 22%)

تمارين على الفصل الخامس

- 1- فيمايلي التوزيع العمري لافراد عينة حجمها 60 من العاملين في كل من 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 :

أولاً: احسب المقاييس التالية لكل من التوزيعين :

الوسط الحسابي – الوسيط – المنوال – الانحراف المعياري – الانحراف الربيعي .

ثانياً: قارن بين درجة تماثل كل من التوزيعين وذلك باستخدام الطرق التالية :
(أ): إيجاد معاملات الالتواء المختلفة وذلك باستخدام المقاييس المحسوبة في (أولاً).

(ب): تطبيق العلاقات ما بين الانحراف المعياري وبين كل من الانحراف المتوسط ونصف المدى الربيعي والمتعلقة بالتوزيع المعتاد مع ابداء التعليق المناسب .

(ج): حساب نسبة الموظفين الذين تتراوح مرتباتهم ما بين :

$$\bar{X} \pm \sigma, \bar{X} \pm 2\sigma, \bar{X} \pm 3\sigma$$

حيث: \bar{X} ، σ هما على الترتيب الوسط الحسابي والانحراف المعياري للمرتبات.
قارن بين قيم هذه النسب بما يجب ان تكون عليه في حالة التوزيع المعتاد وهي 68.27 %، 95.45 %، 99.73 %، مبدئياً ما تراه مناسب من تعلق .

2- فيما يلي التوزيع التكرارى للدخل الأسبوعى لمائة أسرة:

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 الأدنى والأعلى ومن ثم حساب معامل الاختلاف الربيعى .
 (ب): قياس الالتواء والتفرطح لهذا التوزيع.

3- استخدام بيانات أجور العمال الأسبوعية فى المصنعين (أ، ب) لصناعة

المنسوجات القطنية والوادة فى التمرين الأول من تمارين هذا الباب 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 للالتواء.

(ب):معاملا الالتواء الربيعى.

(ج):معاملا الالتواء المثبنى .

ثالثا : قارن بين التوزيعين من حيث الاختلاف النسبى . علق ما تحصل عليه من نتائج .

رابعا: قياس التفرطح فى كل من التوزيعين .

4- باستخدام بيانات درجات الطلاب في كل من كليتي التجارة والتربية والواردة في تمارين الباب الثالث .

المطلوب :

(أ):قارن بين تشتت التوزيعين باستخدام المقاييسين التاليين:

- نصف المدى الربيعي - الانحراف المعياري

ب- باستخدام العزوم ، قارن بين تفرطح التوزيعين .

ج- قدر التواء كل توزيع واختلافه النسبي.

5- البيانات التالية تمثل درجات 80 طالبا في مادة الاقتصاد:

Degree Classes	40-	50-	60-	70-	80-	90-
No.Of stndents	5	12	32	21	7	3

والمطلوب :

أ - حساب المقاييس التالية :

- الربع الادنى والربع الاعلى ، الوسيط

- المئين العاشر والمئين التسعون.

ب - من نتائجك في (أ) احسب المقاييس التالية :

- معادل الاختلاف - معامل الالتواء - معامل التفرطح

ج- احسب نسبة الطلاب الذين تتراوح درجاتهم ما بين $\bar{X} \pm \sigma$

حيث \bar{X} ، σ هما على الترتيب الوسط الحسابي والانحراف المعياري لهذا

التوزيع .ثم بين كيف يمكن استخدام قيمة هذه النسبة كمؤشر على تماثل هذا

التوزيع من عدمه.

6- فيما يلي التوزيع التكرارى لدرجات طلاب كليات التجارة موزعين حسب نظام الدراسة، وذلك فى احد الامتحانات .

المجموع	انتساب موجه	انتظام	نظام الدراسة فئات الدرجات
78	40	38	-0
203	77	126	-14
311	107	204	-20
296	57	239	-26
78	5	73	-32
24	صفر	24	40-36
990	286	704	المجموع

والمطلوب :

(أ): المقارنة بين تشتت درجات الطلاب فى كل من نظامى الدراسة مستخدما الانحراف المعياري .

(ب): احسب معامل الالتواء لكل من التوزيعين التكراريين ، ثم قارن بينهما .

ج- قارن بين نسبة الطلاب الذين تتراوح درجاتهم ما بين 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 students	4	98	169	192	163	33	702

والمطلوب :

حساب مايلى :

(أ):مقاييس التشتت التالية :

- الانحراف المعياري. - نصف المدى الربيعي.

ب - نسبة رسوب الطلاب (باعتبار أن الطالب الحاصل على درجة أقل من 25 درجة يعتبر راسبا).

ج- عدد الطلاب الذين تتراوح درجاتهم ما بين 35،52 درجة .

د - درجة ونوع التواء توزيع الدرجات .

هـ - اذا كانت الدرجة النهائية لامتحان الاحصاء فى هذا الدورى 60 درجة ،فما هو 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 اقتصرنا 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 لدراسة العلاقة بين ادمان المخدرات بمرض السرطان او الايدز ...الى غير ذلك من الامثلة العديدة فى جميع فروع العلوم والمعرفة لمختلف الانشطة مما يودى بدوه الى اهتمام متزايد بدراسة الارتباط والانحدار .

هذا ويطلق على عملية دراسة العلاقة بين متغيرين او اكثر بعلاقة الارتباط حيث يهتم الباحث بدراسة ما اذا كانت هذه المتغيرات ترتبط بعلاقة خطية كما يهتم بعملية تحديد اتجاه تلك العلاقة اى تحديد ما اذا كانت علاقة الارتباط

طردية ام عكسية ، اما عن تعيين شكل الصورة الرياضية التى تربط متغيرات الدراسة ببعضها البعض لاستخدامها فى التنبؤ بقيم متغير فى المستقبل فى ضوء مجموعة من القيم التى يأخذها متغير او مجموعة متغيرات اخرى تؤثر بشكل كبير على المتغير الاصلى او الظاهرة موضع الدراسة والتحليل فهو ما يعرف بتحليل الانحدار والذى سوف نقوم بدراسته فى الباب التالى . وتجدر الاشارة الى ان الارتباط والانحدار يختلفان فى الفرض القائم وقت الدراسة. فالارتباط هو عبارة عن دراسة للتغيير المشترك فيما بين متغيرين (او اكثر) كل منها يأخذ قيما مختلفة دون تدخل من الباحث. اى ان ازواج المشاهدات المتناظرة من المتغيرين هى بمثابة قيما عشوائية لا يعيننا عند دراسة الارتباط التمييز بين المتغيرين اى منها يعتبر متغيرا مستقلا وايهما يعتبر متغيرا تابعا. اما الانحدار فيفترض وجود متغيراً واحداً تابعا يتاثر بواحد 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 يأخذها المتغير التابع عند تلك المستويات المحددة للمتغير او المتغيرات المستقلة والمحددة سلفا . ومن ازواج القيم المتناظرة فى الحالة الاولى يمكن تقدير معامل الارتباط والذى يُستدل منه على اتجاه 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 الانحدار وهو ما سيرد الفصل التالى لدراستنا .

وسوف نهتم فى هذا الباب بدراسة الارتباط الخطى البسيط بين متغيرين والارتباط ما هو الاميل متغيرين (او اكثر) الى التغيير معا فهو يبين درجة قوة العلاقة بين متغيرين كما يبين اتجاه تلك العلاقة (طردية كانت او عكسية) ايضا. فاذا كان لدينا متغيرين وليكن x, y وكان التغيير فى احد المتغيرين يتبعه تغير فى المتغير الاخر فى نفس الاتجاه بمعنى ان زيادة احدهما يودى الى زيادة فى قيمة المتغير الاخر او ان نقص احدهما يودى الى نقص فى قيمة المتغير الاخر فيقال فى تلك الحالة ان الارتباط بين المتغيرين طرديا (اى موجب) والعكس صحيح اى انه اذا كان التغيير فى احد المتغيرين يتبعه تغير فى قيمة المتغير الاخر فى الاتجاه المضاد بمعنى ان زيادة احدهما يودى الى نقص فى قيمة المتغير الاخر او ان نقص احدهما يودى الى زيادة فى قيمة المتغير الاخر فيقال فى هذه الحالة ان الارتباط بين المتغيرين ارتباطا 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 الانتشار Scatter

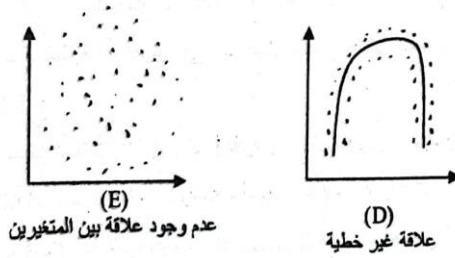
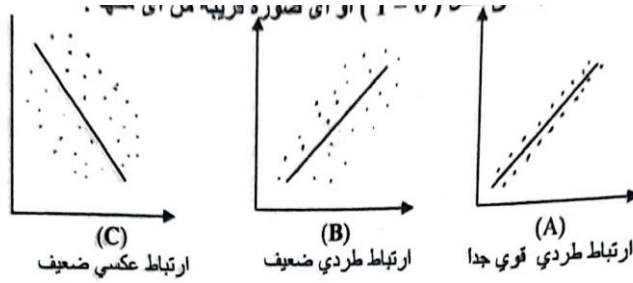
. Diagram

فاذا كان لدينا المتغيرين x, y وبفرض ان المتغيرين يرتبطان من خلال علاقة . فانه اذا سحبنا عينة عشوائية حجمها (n) من ازواج المشاهدات لقيم (X_i) وقيم (Y_i) المناظرة حيث $i=1,2,\dots,n$ اى يصبح لدينا ازواج المشاهدات التالية :

$$(X_1, Y_1), (X_2, Y_2), (X_3, Y_3), \dots, (X_n, Y_n)$$

فاذا تم تمثيل هذه الازواج من المشاهدات بيانيا فانا نحصل على شكل الانتشار والذى يعطى للباحث صورة مبدئية واضحة عن طبيعة العلاقة ما بين المتغيرين محل الدراسة ويساعده على اختيار النموذج القياسى الذى يناسب طبيعة

البيانات موضع الدراسة. وهناك صوراً مختلفة لاشكال الانتشار نرصد منها الاشكال التالية والمبينة في شكل (6 - 1) او اي صورة قريبة من اي منها .



صورة مختلفة لاشكال الانتشار

(1-6)

حيث يتضح لنا من خلال اشكال الانتشار الوارده في الشكل (1-6) ما يلي :

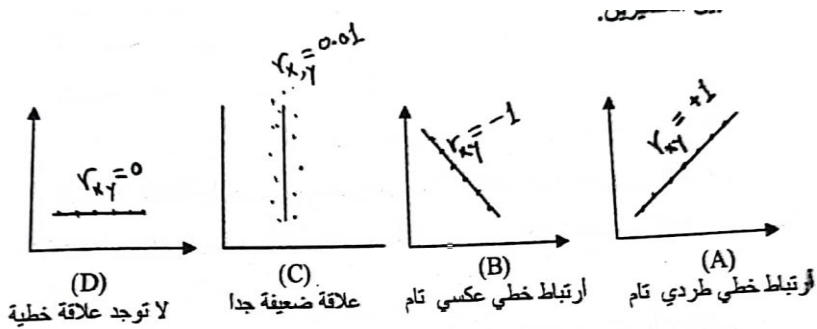
- بالنسبة للشكل (A) في شكل (1-6) فانه يوضح لنا وجود علاقة طردية وقوية جدا ما بين المتغيرين Y, X بمعنى ان اي تغير في احد المتغيرين وليكن مثلا (X) يصاحبه تغير في قيمة المتغير (Y) وفي نفس الاتجاه . وكلما كانت نقاط شكل الانتشار التي تعبر عن ازواج المشاهدات (Y_i , X_i) قريبة من

الخط المستقيم والذي يتوسط هذه النقاط فكلما دل ذلك على قوة الارتباط بين المتغيرين . والعكس صحيح فكلما ابتعدت النقاط عن الخط المستقيم الذي يتوسط تلك النقاط فان هذا يعنى ان الارتباط ضعيف كما هو فى الشكل (B) والذي يوضح وجود علاقة طردية ضعيفة بين المتغيرين. اما الشكل (C) فهو يوضح ان التغير فى اى من قيم المتغيرين يصاحبه تغير فى قيمة المتغير الاخر ولكن فى الاتجاه المضاد وهو ما يفيد وجود علاقة عكسية ضعيفة نظرا لبعدها عن 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 ارتباطا خطيا حيث يمكن توفيق خط مستقيم يمر باكبر عدد ممكن من نقاط شكل الانتشار ويتوسط الباقي 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 منظم) 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 التى تنشأ فى 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 (A-2-6) فى معنى 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 (B-2-6) فان هذا يدل على ان الارتباط بين المتغيرين هو ارتباطا خطيا عكسيا تاماً بين المتغيرين.



صور اخرى لاشكال الانتشار

شكل (2-6)

اما اذا تركزت نقاط شكل الانتشار حول خط مستقيم كما هو فى شكل (C-2-6) فان هذا يعنى وجود علاقة ضعيفة جدا بين المتغيرين .

وفى حالة وقوع جميع النقاط على خط مستقيم يوازى احد المحورين كما هو الحال فى شكل (D-2-6) فان هذا معناه انه لا توجد علاقة خطية بين المتغيرين اى ان معامل الارتباط الخطى بين المتغيرين فى هذه الحالة يكون مساويا للصفر .

واخيرا فانه بالرغم من ان اشكال الانتشار توضح ما اذا كانت هناك علاقة ما بين المتغيرين محل الدراسة واما 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 Coefficient** 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 موجبا او سالبا فان هذا يعنى قوة العلاقة بين المتغيرين . والعكس صحيح فكلما اقتربت قيمة معامل الارتباط من الصفر كلما دل ذلك على ضعف العلاقة الخطية فيما بين المتغيرين . واذا تساوت قيمة معامل الارتباط بالصفر معنى ذلك عدم وجود علاقة ارتباط او نزعة او تأثير بين قيم احد المتغيرات وقيم المتغير الاخر . اما عن اشارة معامل الارتباط فهى تبين اتجاه العلاقة بين المتغيرين حيث تكون الإشارة موجبة إذا كانت العلاقة طردية والعكس صحيح تكون الإشارة سالبة اذا كانت العلاقة عكسية بين المتغيرين . هذا وعند دراستنا للارتباط فى هذا الباب فاننا سوف نقتصر فى هذه الدراسة على الارتباط الخطى البسيط بين متغيرين فقط .

(1-6) العلاقة بين الظواهر (المتغيرات) :

تأخذ العلاقة بين المتغيرات صوراً مختلفة ويتوقف ذلك على طبيعة او نوعية البيانات وكذا عدد المتغيرات المطلوب دراسة العلاقة فيما بينها وتتلخص هذه الصور فيما يلى :

1- دراسة علاقة الارتباط :

وهى العلاقة بين المتغيرات لبيانات يمكن قياسها رقمياً (كمية) وينقسم الارتباط فى هذه الحالة الى عدة اقسام يتوقف كل منها على عدد المتغيرات المطلوب دراسة علاقة الارتباط فيما بينها .
وتتلخص أقسام علاقة الارتباط فيما يلى :-

(أ): الارتباط الخطى البسيط : وهو الذى يستخدم فى قياس العلاقة بين متغيرين كميين فقط .

(ب): الارتباط المتعددة : وهو يتناول العلاقة بين اكثر من متغيرين كميين .

(ج): الارتباط الجزئى : 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 كمياً فان قياس العلاقة بين الظاهرتين يكون من خلال ما يلى :

أ- علاقة الإقتران : وقياس الإقتران العلاقة بين ظاهرتين (متغيرين) احدهما على الاقل وصفياً على ان يتكون كل من المتغيرين من قسمين فقط .

ب- علاقة التوافق : ويقاس التوافق العلاقة بين ظاهرتين (متغيرين) احدهما على الأقل وصفيا وتحتوى احدى هاتين الظاهرتين 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 Coefficient** 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 أى 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 أى ان

$$-1 \leq r_{xy} \leq +1$$

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): معامل الارتباط لايتأثر باى من العمليات الجبرية الاربعة (الطرح او الجمع

او الضرب او القسمة) اى ان :

$$r_{x,y} = r_{dx dy} = r_{Dx , Dy}$$

حيث :-

طرحه من كافة المفردات سواء X أو Y كما سبق وان عرفنا الانحرافات المباشرة فى دراستنا للوسط الحسابى والانحراف المعياري اما:

$$D_{yi} = \frac{y_i - a}{b} , D_{xi} = \frac{x_i - a}{b}$$

السابق الاشارة اليها .

هذا وسوف نقوم بدراسة معاملات الارتباط التالية :

- 1- معامل الارتباط الخطى البسيط لبيرسون.
- 2- معامل ارتباط الرتب لسبيرمان.
- 3- معامل ارتباط كندال.
- 4- معامل ارتباط فهنر.

1- معامل الإرتباط الخطى البسيط لبيرسون :

Simple Linear Correlation Coefficient

وينسب هذا المعامل الى كارل بيرسون K. Pearson وهناك اكثر من صورة رياضية لهذا المعامل وهو يقيس ماهية العلاقة الخطية ما بين متغيرين كميين سواء كانت البيانات مبوبة او غير مبوبة وذلك على النحو التالى :

أولاً: في حالة البيانات الغير مبوية (المفردات) :

فإذا كان لدينا عدد (n) من ازواج المشاهدات للمتغيرين X , Y اى لدينا ازواج المشاهدات :

$$(X_1, Y_1) , (X_2, Y_2) , \dots \dots \dots (X_n, Y_n)$$

وكان الوسط الحسابى للمتغير X هو \bar{X} والوسط الحسابى للمتغير Y هو

\bar{Y} وكان الانحراف المعياري للمتغيرين X , Y هما σ_x ، σ_y على الترتيب

فان معامل الارتباط الخطى البسيط لبيرسون ($r_{X,y}$) هو عبارة عن متوسط

حاصل ضرب القيم المعياريه للمتغيرين y.x اى أن :

$$r_{x,y} = \frac{1}{n} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma_x} \right) \left(\frac{y_i - \bar{y}}{\sigma_y} \right)$$

$$\begin{aligned} &= \frac{\sum(X_i - \bar{X})(y_i - \bar{y})}{n \sigma_x \sigma_y} \\ &= \frac{\sum(X_i - \bar{X})(y_i - \bar{y})}{n \sqrt{\frac{\sum(X_i - \bar{X})^2}{n} \cdot \frac{\sum(y_i - \bar{y})^2}{n}}} \quad (1) \end{aligned}$$

$$= \frac{\sum(X_i - \bar{X})(y_i - \bar{y})}{\sqrt{\sum(X_i - \bar{X})^2 \cdot \sum(y_i - \bar{y})^2}} \quad (2)$$

والمقدار $\frac{\sum(X_i - \bar{X})(y_i - \bar{y})}{n}$ يسمى بالتغاير Covariance بين

المتغيرين x ، y . ومن ثم فإذا كان معامل الارتباط الخطى البسيط هو عبارة عن خارج قسمة التغيرات بين المتغيرين على حاصل ضرب الإنحرافات المعياريه للمتغير. هذا وعادة ما تكون قيم الاوساط الحسابية للمتغيرين قيما كسرية مما يؤدي الى صعوبة استخدام صيغة المعادلة (1) فى حساب قيمة معامل الارتباط بين المتغيرين x ، y وللتغلب على هذه المشكلة فإن يمكن استنتاج (دون الخوض فى الإثبات) الصيغة التاليه والتي يسهل استخدامها لحساب معامل الارتباط الخطى البسيط لبيرسون :

$$r_{x,y} = \frac{\frac{\sum xy}{n} - \frac{\sum X}{n} \cdot \frac{\sum Y}{n}}{\sqrt{\left[\frac{\sum X^2}{n} - \left(\frac{\sum X}{n} \right)^2 \right] \left[\frac{\sum Y^2}{n} - \left(\frac{\sum Y}{n} \right)^2 \right]}} \quad (3)$$

ووفقا لصيغة المعادلة (3) فان يلزم لحساب معامل الارتباط الخطى البسيط المجاميع التالية:

$$\sum X, \sum y, \sum X^2, \sum y^2, \sum Xy$$

وإنطلاقا من الخاصية الثالثة لخصائص معامل الارتباط والتي تفيد بان معامل الارتباط لايتاثر بأيا من العمليات الجبريه الاربعة. فإنه يمكن بالاضافة للطريقة المباشرة والمستخدم فيها المعادلة (3)السابقة فانه يمكن حساب معامل الارتباط 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) هي :-

$$r_{x,y} = \frac{\frac{\sum d_x d_y}{n} - \frac{\sum d_x}{n} \cdot \frac{\sum d_y}{n}}{\sqrt{\left[\frac{\sum d_x^2}{n} - \left(\frac{\sum d_x}{n} \right)^2 \right] \left[\frac{\sum d_y^2}{n} - \left(\frac{\sum d_y}{n} \right)^2 \right]}}$$

حيث d_x . d_y تعبر عن إنحرافات قيم أزواج المشاهدات عن وسط فرض معين سواء كان هذا الوسط الفرضي متساويا للمتغيرين او مختلفا.

• صيغة معامل الارتباط الخطى البسيط لبيرسون فى حالة استخدام طريقة

الإنحرافات المعدلة:

$$r_{x,y} = r_{D_x, D_y} = \frac{\frac{\sum D_x D_y}{n} - \frac{\sum D_x}{n} \cdot \frac{\sum D_y}{n}}{\sqrt{\left[\frac{\sum D_x^2}{n} - \left(\frac{\sum D_x}{n} \right)^2 \right] \left[\frac{\sum D_y^2}{n} - \left(\frac{\sum D_y}{n} \right)^2 \right]}}$$

مثال:

فيما يلى بيان باسعار الفائدة 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

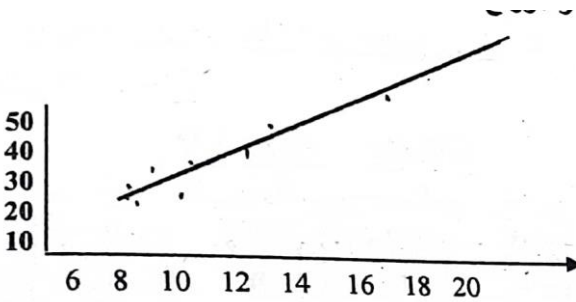
والمطلوب :-

1- ارسم شكل الانتشار وعلق على نتائجك .

2- احسب معامل الارتباط الخطى البسيط لبيرسون وذلك باستخدام الانحرافات عن الوسط الحسابى ثم باستخدام القيم الاصلية وذلك باستخدام الطريقة المباشرة وطريقة الانحرافات البسيطة حول وسط فرض خلاف قيمة الوسط الحسابى للمتغيرين .

الحل :

1- رسم شكل الانتشار لأزواج المشاهدات وذلك على النحو التالى:-



ومن خلال هذا الشكل يتضح لنا ان نقاط ازواج المشاهدات تتركز بدرجة كبيرة حول خط مستقيم ذو ميل موجب اى ان العلاقة طردية وقوية جدا بين المتغيرين (x) ، (y) .

2- حساب معامل الارتباط الخطى البسيط :

(أ) : باستخدام انحرافات قيم المتغيرين عن اوساطها الحسابية اى باستخدام الصيغة التالية :

$$r_{x,y} = \frac{\sum(X_i - \bar{X})(y_i - \bar{y})}{\sqrt{\sum(X_i - \bar{X})^2 \cdot (y_i - \bar{y})^2}}$$

لذا يلزم بداية حساب قيمة كل من \bar{x} ، \bar{y} فحيث ان :

$$\bar{X} = \frac{\sum X}{n} = \frac{8 + 9 + 8 + 11 + \dots + 18}{7} = \frac{77}{7} = 11$$

$$\bar{Y} = \frac{\sum Y}{n} = \frac{30 + 33 + \dots + 57}{7} = \frac{252}{7} = 36$$

ثم يتم تكوين الجدول التالي لاجراء الحسابات اللازمة لحساب معامل الارتباط في هذه 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	348

وعليه فان معامل الارتباط الخطى البسيط هو :

$$r_{x,y} = \frac{212}{\sqrt{(76)(648)}} \simeq 0.955$$

وهو ما يفيد ان العلاقة طردية قوية جدا بين x ، y .

(ب) حساب معامل الارتباط الخطى البسيط باستخدام القيم الاصلية :-

1- باستخدام الطريقة المباشره :-

والجدول التالي يبين الحسابات اللازمة لحساب معامل الارتباط الخطى البسيط

ليبرسون باستخدام الطريقة المباشره :-

X_i	Y_i	$X_i Y_i$	X_i^2	Y_i^2
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 :

$$r_{x,y} = \frac{\frac{2984}{7} - \frac{77}{7} \cdot \frac{252}{7}}{\sqrt{\left[\frac{923}{7} - \left(\frac{77}{7} \right)^2 \right] \left[\frac{9720}{7} - \left(\frac{252}{7} \right)^2 \right]}}$$

$$= \frac{30.2857}{\sqrt{[10.8714][92.5714]}} = \frac{30.2857}{31.7235} \simeq 0.955$$

وهي نفس النتيجة السابقة .

2- باستخدام طريقة الانحرافات البسيطة :

دعنا نقوم بحساب انحرافات قيم المتغير (x) عن الوسط الفرضي وليكن القيمة (10) ونقوم بحساب انحرافات قيم المتغير (y) عن الوسط الفرضي وليكن القيمة (30) . أى أن :

$$d_x = x_i - 10 \quad \text{where } i=1,2,\dots,7$$

$$d_y = y_i - 30 \quad \text{where } i=1,2,\dots,7$$

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 :

$$r_{x,y} = r_{D_x, D_y} = \frac{\frac{\sum D_x D_y}{n} - \frac{\sum D_x}{n} \cdot \frac{\sum D_y}{n}}{\sqrt{\left[\frac{\sum D_x^2}{n} - \left(\frac{\sum D_x}{n} \right)^2 \right] \left[\frac{\sum D_y^2}{n} - \left(\frac{\sum D_y}{n} \right)^2 \right]}}$$

لذا فإن

$$r_{x,y} = \frac{\frac{254}{7} - \frac{7 \cdot 42}{7 \cdot 7}}{\sqrt{\left[\frac{83}{7} - \left(\frac{7}{7} \right)^2 \right] \left[\frac{900}{7} - \left(\frac{42}{7} \right)^2 \right]}}$$

$$= \frac{27.42857}{\sqrt{(10.85714)(92.57143)}} = \frac{30.28571}{31.7027} \approx 0.955$$

وهو نفس ناتج الطريقتين السابقتين.

مثال:

فيإلى بيان عن الدخل الاسبوعى (x) والانفاق الاستهلاكي الاسبوعى (y) بالجنيه لعينه من خمسة أسر:-

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 :-

$$D_x = \frac{Xi - 80}{10}, D_y = \frac{yi - 40}{10}$$

ومن خلال المجاميع المستنتجة فى الجدول السابق فان قيمة معامل الارتباط الخطى البسيط لبيرسون فيما بين الدخل والأنفاق الأسبوعى للأسر هو :

$$r_{x,y} = r_{D_x, D_y} = \frac{\frac{\sum D_x D_y}{n} - \frac{\sum D_x}{n} \cdot \frac{\sum D_y}{n}}{\sqrt{\left[\frac{\sum D_x^2}{n} - \left(\frac{\sum D_x}{n} \right)^2 \right] \left[\frac{\sum D_y^2}{n} - \left(\frac{\sum D_y}{n} \right)^2 \right]}}$$

$$= \frac{\frac{22}{7} - \frac{1}{7} \cdot \frac{3}{7}}{\sqrt{\left[\frac{33}{7} - \left(\frac{1}{7}\right)^2\right] \left[\frac{21}{7} - \left(\frac{3}{7}\right)^2\right]}} = \frac{3.081633}{\sqrt{(4.693878)(2.816326)}} \approx 0.848$$

وهذه النتيجة تفيد أن هناك علاقة طردية قوية بين المتغيرين x , y .

ملاحظات على معامل الارتباط الخطى البسيط لبيرسون :

(1):معامل الارتباط الخطى البسيط لبيرسون يستخدم القيم الفعلية لبيانات المتغيرين فى الحساب ولا تختلف قيمته باختلاف الطريقة المستخدمة فى حسابة . فكما اتضح لنا من الامثلة السابقة فإن معامل الارتباط لا يتأثر بالعمليات الجبرية الاربعة .

(2):معامل الارتباط الخطى البسيط يقيس قوة العلاقة الخطية البسيطة بين المتغيرين وهو على هذا الاساس يحدد نسبة التغير فى احد المتغيرين والتي يمكن تفسيرها على اساس وجود هذه العلاقة الخطية .ونسبة التغير هذه يتم قياسها عن طريق ما تسمى بمعامل التحديد **Coefficient of Determination** وهو عبارة عن مربع معامل الارتباط الخطى البسيط لبيرسون اى (r^2) ففى مثالنا السابق مثلا كانت قيمة معامل الارتباط فيايبين الدخل الاسبوعى والانفاق الاستهلاكى الاسبوعى هى 0.848 وهنا تبلغ قيمة معامل التحديد (r^2) القيمة 0.719104 (لاحظ ان $(0.8480)^2=0.719104$) وهو ما يعنى م ان 71.9104 % من التغيرات التى تنشأ فى قيمة الانفاق الاستهلاكى الاسبوعى يرجع فى أساسه للتغيرات التى تنشأ على قيمة الدخل

الاسبوعى وذلك بموجب العلاقة الخطية التي تربط بينهما وان 28.0896 % فقط من هذه التغيرات انما يرجع لعوامل اخرى متعلقة بالعلاقة الخطية فيما بين الدخل الاسبوعى و الانفاق الاستهلاكى الاسبوعى. وبعبارة اخرى يمكن القول أن التغير فى قيمة الانفاق الاستهلاكى الاسبوعى يرجع 71.91 % منه الى التغير فى الدخل الاسبوعى وذلك بموجب العلاقة الخطية التي تربط ما بين الدخل والانفاق الاستهلاكى الاسبوعى وان النسبة الباقية (28.09 %) ترجع الى عوامل اخرى تؤثر على الانفاق الاستهلاكى الاسبوعى بخلاف الدخل الاسبوعى للأسرة.

(3): تتوقف اشارة معامل الارتباط (اى اتجاه العلاقة) على قيمة البسط فى صيغة هذا المعامل وذلك لان المقام الذى يعبر عنه بالانحرافات المعيارية للمتغيرين موجب دائما .

ثانيا: حساب معامل الارتباط الخطى لبيسون فى حالة البيانات المبويه :
اذا كانت قيم ازواج المشاهدات الخاصة بالمتغيرين او الظاهرتين محل الدراسة كبيره بحيث يصعب معها اجراء العمليات الحسابية اللازمه لتحليلها احصائياً فإنه يجب فى هذه الحالة عرض البيانات الخاصة بالمتغيرين فى صورة جدول تكرارى مزدوج وهو ما سبق فى مطلع دراستنا . والجدول التكرارى المزدوج ما هو إلا جدول تكرارى فى إتجاهين حيث يتم فيه تخصيص فئات احد المتغيرين فى الصفوف وفئات المتغير الثانى فى الاعمدة وتمثل القيم الموجودة بباقي خلايا الجدول التكرارات المشتركة المناظرة لفئات احد المتغيرين فى مقابل فئات المتغير الاخر . وانتشار التكرارات المشتركة داخل الجدول

التكرارى المزدوج يعطى فكرة مبدئية عن طبيعة العلاقة بين المتغيرين (x ، y) من حيث كونها طردية ام عكسية ، فتركيز التكرارات حول أحد قطرى الجدول التكرارى المزدوج يفيد بقوة العلاقة فيما بين المتغيرين . اما عن صيغة معامل الارتباط الخطى البسيط فى حالة البيانات المبوبة فهى نفس صيغ المعامل فى حالة البيانات المفردة مع ترجيح مراكز فئات كل توزيع من التوزيعين بالتكرارات المزدوجة.

وبصفة عامه فإنه لحساب معامل الإرتباط من البيانات المبوبة يتم اشتقاق التوزيعين التكراريين الهامشيين من الجدول التكراري المزدوج. ثم حساب المجاميع اللازمة لحساب الانحراف المعياري لكل توزيع هامشي ثم من خلال الجدول التكراري المزدوج واستبدال فئاته بمراكز الفئات او الانحرافات البسيطة او المعدلة عن اوساط فرضيه معينه يتم ايجاد المجموع المشترك ثم يتم التعويض في احد الصيغ الثلاثة التالية لاجاد معامل الارتباط الخطى البسيط لبيرسون :

$$r_{x,y} = \frac{\sum \sum x_i y_j F_{x_i y_j} - \left(\frac{\sum X_i F_x}{\sum F_x} \right) \left(\frac{\sum y_j F_y}{\sum F_y} \right)}{\sqrt{\left[\frac{\sum x_x^2 F_x}{\sum F_x} - \left(\frac{\sum x_i f_i}{\sum x} \right)^2 \right] \left[\frac{\sum y_j F_y}{\sum F_y} - \left(\frac{\sum y_j F_y}{\sum F_y} \right)^2 \right]}}$$

حيث تمثل x_i, y_i مراكز الفئتين (i),(j) للمتغيرين Y،X علي الترتيب. أما F_y, F_x فهي تعبر عن التكرارات الهامشية لكل من y,x أما $F_{x_i y_j}$ فتعبر عن التكرارات المزدوجة لخلايا الجدول المزدوج والنتيجة عن تقاطع الفئتين j,i للمتغير y,x علي الترتيب .

وتمثل الصيغة السابقة استخدام القيم الاصلية دون طرح او قسم اي صيغه الطريقة المباشرة لحساب معامل الارتباط . اما عن صيغة المعامل باستخدام طريقة الانحرافات البسيطة فهي نفس الصيغة السابقة مع الفارق الوحيد هو استبدال قيم المراكز (X_i) بالانحرافات (d_{xi}) وكذلك استبدال قيم المراكز y_j بالانحرافات البسيطة d_{yj} . وكذلك صيغة طريقة الانحرافات المعدلة هي نفس الصيغة لكن مع استبدال المراكز x_i , y_j بالانحرافات المعدلة D_x , D_y علي الترتيب .

مثال :

الجدول التالي يبين عدد سنوات الخبرة (x) والمرتب الشهري (y) لعينه من مائة عامل تم اختيارهم بطريقة عشوائية من احدي الشركات:

Y \ X	150-	180-	210-	240-	270 – 300	Σ
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
Σ	10	15	50	18	7	100

والمطلوب حساب معامل الارتباط الخطي البسيط وفسر ما تراه مناسباً باستخدام معامل التحديد .

الحل :

إذا نظرنا للجدول التكراري المزدوج نجد ان التكرارات المزدوجة تتركز حول القطر الرئيسي للجدول (من اعلي جهة اليسار حتي اسفل جهة اليسار) مما يدل بشكل مبني علي وجود علاقة طردية بين المتغيرين x , y وحيث ان أطوال الفئات متساوية لكل من المتغيرين (اي ان التوزيعات الهامشية منتظمة) 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 الخبرة (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	D _y = $\frac{d_y}{30}$	D _y F _{yj}	D ² _y F _{yj}
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
Σ	100				-3	101

ومن هذا الجدول فاننا نجد ان : $\sum d_y^2 F_y = 101$ ، $\sum d_y F_y = -3$.

*اما لايجاد المجموع المشترك $\sum D_x D_y F_{xy}$: فيتم إستبدال فئات المتغيرين في الجدول التكراري المزدوج بالانحرافات المعدلة لهذه المتغيرات ثم حساب إيجاد حواصل ضرب انحراف الصف (i) أي D_{xi} في تكرار الخلية المزدوجة F_{xy} في العمود (j) أي D_{yj} ثم الجمع لكل هذه الحواصل وهو ما يوضحه الجدول التالي:

Dy \ Dx	-2	-1	0	1	2	Σ
-2	8 <u>32</u>	7 <u>14</u>				<u>46</u>
-1	2 <u>4</u>	7 <u>7</u>	10 <u>0</u>	6 <u>-6</u>		<u>5</u>
0		1 <u>0</u>	19 <u>0</u>	10 <u>0</u>		<u>0</u>
1			13 <u>0</u>	2 <u>2</u>		<u>2</u>
2			8 <u>0</u>		2 <u>8</u>	<u>8</u>
3					5 <u>30</u>	<u>30</u>
Σ	10 <u>36</u>	15 <u>21</u>	50 <u>0</u>	18 <u>-4</u>	7 <u>38</u>	100 <u>91</u>

ومن الجدول فإن المجموع المشترك $\sum D_x D_y F_{xy}$ هو (91) . ومن ثم يكون معامل الارتباط الخطي البسيط لبيرسون هو:-

$$r_{x,y} = r_{D_x D_y} = \frac{\sum D_x D_y F_{xy} - \left(\frac{\sum D_x F_x}{\sum F} \right) \cdot \left(\frac{\sum D_y F_y}{\sum F} \right)}{\sqrt{\left[\frac{\sum D_x^2 F_x}{\sum F_x} - \left(\frac{\sum D_x F_x}{\sum F_x} \right)^2 \right] \left[\frac{\sum D_y^2 F_y}{\sum F_y} - \left(\frac{\sum D_y F_y}{\sum F_y} \right)^2 \right]}}$$

$$= \frac{\frac{91}{100} - \left(\frac{-5}{100} \right) \cdot \left(\frac{-3}{100} \right)}{\sqrt{\left[\frac{185}{100} - \left(\frac{-5}{100} \right)^2 \right] \left[\frac{101}{100} - \left(\frac{-3}{100} \right)^2 \right]}}$$

$$= \frac{0.9085}{\sqrt{[1.8475] [1.0091]}} = 0.665$$

اي ان العلاقة طردية متوسطة الي حد ما فيما بين عدد سنوات الخبرة والراتب الشهري وهو ما يؤكد صحة التنبؤ باتجاهه من خلال شكل تركيز التكرارات المزدوجة علي جانبي القطر الرئيسي . اما عن معامل التحديد

0.4422=0.665) وهو ما يفيد ان 44.22 % من التغيرات التي تنشأ في اي من المتغيرين يرجع لتغير الاخر والنسبة الباقية ترجع لعوامل اخري .

ملاحظات على معامل الارتباط البسيط:

1- عند حساب معامل الارتباط الخطي البسيط لبيرسون في التطبيقات العملية يجب ان يكون عدد القيم المستخدمة علي الاقل (30) مفردة . لانه في هذه الحالة فأن وجود اي قيمة شاذة (او حتي اكثر من قيمة) او غير دقيقة سيكون لها تأثير كبير علي قيمة معامل الارتباط الخطي البسيط لبيرسون . لذا ينصح بأنه اذا كان عدد القيم المستخدمة في حساب معامل الارتباط اقل من ثلاثون مفردة (والمفرده هي زوج من المشاهدات (y, x) فأن وجود قيمة شاذة سوف يكون لها تأثير كبيرا علي دقة معامل بيرسون للارتباط وفي هذه الحالة يفضل البحث عن مقياس اخر للارتباط يكون اكثر دقه لان معامل الارتباط الخطي البسيط لبيرسون سوف يعطي نتائج مضللة في هذه الحالة نتيجة لتأثره بالقيم الشاذة او المتطرفه . فعلى سبيل المثال لنوضح تلك النقطة دعنا نفرض ان لدينا البيانات التالية عن المتغيرين x ، y :

x:	1	2	3	4	5
y:	1	2	3	4	5

2- يستخدم معامل الارتباط الخطي البسيط لقياس درجة قوة العلاقة الخطية بين متغيرين ، اي درجة تركيز النقاط حول خط الانحدار **Regression Line** فكلما زادت القيمة المطلقة لمعامل الارتباط ($|r_{xy}|$) كلما زادت قوة العلاقة بين المتغيرين وبالتالي زادت درجة تركيز النقاط حول خط الانحدار . فعلى سبيل

المثال ، تكون قيمة ($|r_{xy}|$) مساوية الواحد الصحيح في الشكلين 2-6 / B ، A ، بينما هي في الشكل (A / 1-6) اكبر منها في الشكل (C,B / 1-5) .

3- تتوقف اشارة معامل الارتباط علي اشارة البسط والذي يمثل التغير بين المتغيرين ، وذلك لان المقام دائما موجب حيث انه حاصل ضرب الانحرافات المعيارية σ_x ، σ_y في وكلاهما مقدار موجب .

4- عند ايجاد معامل الارتباط لبيانات جدول تكراري مزدوج فإنه ليست هناك اية قيود او اشتراطات فيما يتعلق بطول الفئة او عدد الفئات في المتغيرين. فعلي سبيل المثال يمكن ان يكون عدد الفئات لأحد المتغيرين مختلفا عنه للمتغير الاخر ، كما يمكن ان تكون الفئات متساوية الطول لأحد المتغيرين وغير متساوية الطول للمتغير الاخر .

5- أن إقتراب قيمة معامل الارتباط 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 فإن 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 .

6- عند حساب معامل الارتباط للبيانات المفردة (غير المبوبة) فإنه اذا كان الوسط الحسابي لكل من المتغيرين x ، y عددين صحيحين وكانت القيم كبيرة. حينئذ يفضل اتباع طريقة الانحرافات البسيطة باستخدام الوسطين الحسابيين

كوسطين فرضين . وهذا يؤدي الى تبسيط العمليات الحسابية ،بالإضافة الى انه يمكننا التأكد 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 استخدام الطريقة المباشرة وذلك اختصارا للعمليات الحسابية .

7- عند حساب معامل الارتباط لبيانات مبوبة فى جدول تكرارى مزدوج فانه اذا كانت فئات كل من المتغيرين متساوية الطول يفضل استخدام طريقة الانحرافات المعدلة .

8- فى حالة حساب معامل الارتباط لبيانات المتغيرين x ، y والمبوية فى جدول تكرارى مزدوج ، فانه قبل اجراء أية عملية حسابية يمكن الاستدلال على طبيعة العلاقة بين المتغيرين وذلك على النحو التالى :

*اذا كانت التكرارات المشتركة داخل خلايا الجدول تتركزة حول قطر الجدول الممتد من اعلى ركن من جهة اليسار الى اسفل الجدول جهة اليمين فأن هذا يعنى ان هناك علاقة طردية بين المتغيرين وتزداد قوة هذه العلاقة بزيادة تركز التكرارات حول هذا القطر ، والعكس صحيح أى أنه اذا كانت التكرارات المشتركة داخل خلايا الجدول تتركزة حول قطر الجدول الممتد من اعلى اليمين الى ادنى اليسار فأن هذا يعنى ان هناك علاقة عكسية بين المتغيرين وتتوقف درجة قوة او ضعف هذه العلاقة على مدى تركز هذه التكرارات حول هذا القطر.

9- تلزم الإشارة الى نقطة هامة وهى : انه فى دراستنا للارتباط ، يجب علينا ان نفرق بين العلاقة الاحصائية – متمثلة فى معامل الارتباط بين المتغيرين – وبين العلاقة السببية . فحين نجد ان هناك ارتباطا قويا بين متغيرين (اي قيمة $|r_{xy}|$ تقترب من الواحد الصحيح) . فان ذلك لايعنى بالضرورة ان العلاقة بينهما هى بالضرورة علاقة سببية بمعنى ان التغير فى احد المتغيرين هو السبب فى التغير فى قيمة الاخر . فعلى سبيل المثال : لو كنا بصدد ايجاد معامل الارتباط بين درجات مجموعة من الطلاب فى مادة الاحصاء وبين درجاتهم فى مادة القانون ، ووجدنا ان معامل الارتباط يساوى 0.95 فان هذا يعنى ان هناك ارتباطا قويا بين درجات الطلاب فى مادتي الاحصاء والقانون ، ولكن هل يعنى ذلك ان ارتفاع (او انخفاض) درجات الطلاب فى مادة القانون انما يرجع الى ارتفاع (او انخفاض) درجاتهم فى مادة الاحصاء او العكس ؟ الاجابة بكل تأكيد هى : لا . وهنا يمكن القبول ان هناك علاقة احصائية قوية بين المتغيرين رغم انعدام العلاقة السببية بينها . ومن جهة اخرى ، قد تنعدم العلاقة الاحصائية – ممثلة فى معامل الارتباط 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) ، 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 امر يخرج عن نطاق علم الاحصاء والذى يهتم فقط بمجرد قياس العلاقة بين المتغيرات.

10- أنه أحياناً ما نصل إلى نتائج غير منطقية بمعامل الارتباط الخطي البسيط لقياس ماهية العلاقة الخطية بين المتغيرين. بمعنى أنه قد يكون من المعروف أن العلاقة الخطية بين متغيرين هي علاقة عكسية ، ولكن قيمة هذا المعامل أظهرت لنا أنها علاقة طردية أو العكس. وفي هذه الحالة يمكن تفسير هذا بأحد أمرين :-

أولهما : قد يكون هناك متغير ثالث يؤثر على المتغيرين (x ، y) موضع الاعتبار وذلك بالشكل الذي يؤدي إلى النتيجة غير المنطقية ولنضرب مثلاً 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 نستطيع القول حينئذ بأن هذه القيمة تفيد بأن هناك علاقة طردية ما بين كمية الإنتاج وعدد ساعات تعطل الآلات ؟ إن قلنا ذلك فإنه يعتبر منطقاً

معكوساً ومخالفاً للواقع. لأنه إن صح أمر كهذا فما علينا إلا أن نوقف جميع الآلات حتى نستطيع أن نحقق أكبر كمية ممكنة من الإنتاج. وهذا يبدو أمراً لا يقبله منطق أو يُقرّه 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 في 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 في 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 في كل من كمية الإنتاج وعدد ساعات الأعطال، مما يؤدي في النهاية إلى أن معامل الارتباط 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.

ثانيهما : أن العلاقة بين المتغيرين (y, x) موضع الإعتبار هي علاقة ضعيفة، بمعنى أن معامل الارتباط 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 الارتباط 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.

11- أنه في حالة الحصول على علاقة خطية ضعيفة بين متغيرين (y . x) فإنه يجب عدم التسليم بذلك حيث يتم التأكد مما يلي:

- أنه لا يوجد قيم شاذة 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 احتمال أن تكون هناك علاقة قوية (بل قوية جداً) أو ربما تامة بين المتغيرين ولكنها 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 أخرى تعبر عن وجود علاقة غير خطية قوية بين المتغيرين.

2 – معامل فهنر للارتباط:

يبني هذا المعامل على اساس ان العلاقة ما بين متغيرين كميّين لا تعتمد على القيم الاصلية للمتغيرين وانما تعتمد اساساً على انحرافات قيم كل متغير عن وسطه الحسابي على حده .

فاذا كانت لدينا عدد (n) من ازواج المشاهدات لمتغيرين كميّين Y,X ولتكن هذه الازواج من القيم هي ، $(y_1 - x_1)$ ، $(y_2 - x_2)$ ، ... ، $(y_n - x_n)$ فإنه لحساب معامل فهنر للارتباط يتم أتباع الخطوات التالية :-

- حساب الوسط الحسابى لكل متغير اى \bar{x} ، \bar{y} .
- حساب انحرافات قيم كل متغير عن وسطه الحسابى اى يتم حساب قيم:

$$(x_i - \bar{x}) , (y_i - \bar{y}) \quad \text{For } i = 1, 2, \dots, n$$

- يتم حصر عدد الحالات التى تكون فيها ازواج الانحرافات للمتغيرين ذات اشارات متشابهة (موجبة كانت ام سالبة) ويمكن اجراء ذلك من خلال اعطاء رمز وليكن (S) لكل حالة من حالات تشابه تلك الانحرافات فى الاشارة فيكون مجموع حالات التشابه هو $\sum S$.

- يتم حصر عدد الحالات التى تكون فيها ازواج الانحرافات للمتغيرين ذات اشارات مختلفة. ونرمز لكل حالة من تلك الحالات بالرمز (d). ومن ثم يكون اجمالي مجموع تلك الحالات وليكن المقدار $\sum d$.

يتم حساب معامل فهنر للارتباط (r_f) والذى يأخذ الصورة التالية :

$$r_f = \frac{\sum s - \sum d}{n} \quad \text{where } n = \sum s + \sum d$$

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 شئى وانما يهمننا 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 الحسابي لهذا المتغير والمثال التالي يوضح كيفية حساب معامل فهنر للارتباط .

مثال:

البيانات التالية تمثل درجات عشرة طلاب في مادتي الاحصاء والمحاسبة:

Statistic	82	98	67	40	75	61	31	72	53	83
Degree(x)										
Accounting	80	83	52	63	96	71	71	75	35	85
Degree(y)										

والمطلوب حساب معامل فهنر للارتباط .

الحل:

لحساب معامل فهنر للارتباط (r_f) فانه يجب في البداية حساب الوسط الحسابي لكل من المتغير Y,X حيث نجد ان :

$$\bar{x} = \frac{\sum x}{n} = \frac{82+98+\dots+83}{10} = \frac{662}{10} = 66.2 \quad \text{degree}$$

$$\bar{y} = \frac{\sum y}{n} = \frac{80 + 83 + \dots + 58}{10} = \frac{711}{10} = 71.1$$

ثم يتم بعد ذلك تحديد عدد الانحرافات المتشابهة وكذا المختلفة عن الوسط الحسابي لكل متغير كما يوضحها الجدول التالي :

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 فإن :

$$\sum S = 9 , \sum d = 1 , n = 10$$

لذا فإن معامل فهنر للارتباط هو :

$$r_f = \frac{\sum s - \sum d}{n} = \frac{9 - 1}{10} = 0.8$$

وحيث أن قيمة معامل فهنر تقترب من الواحد الصحيح. لذا فإن 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 ناد . وفى احيان اخرى قد يتعدى قياس احدى الظاهرتين على الاقل كما هو الحال فى الظواهر الوصفية او قد تكون الدقة العالية فى النتائج غير مطلوبة وفى مثل هذه الحالات يستعاض عن القيم العددية الحقيقية لكل من المتغيرين بالرتب التى يأخذها كل منها على اساس موقعه فى المجموعة التى ينتمى اليها .

فاذا اتفقت رتب قيم المتغير (x) مع رتب قيم المتغير (y) اتفقا تماما اى لكافة قيم ازواج المشاهدات (بمعنى ان اصغر قيمة فى قيم (x) تقابلها اصغر قيمة لـ (y) نكون بصدده حالة ارتباط طردى تام . والعكس صحيح فاذا كانت رتب (x) ، رتب (y) رتب عكسية تماما (بمعنى ان اصغر قيمة لـ (x) تقابلها اكبر قيمة لـ (y) والقيمة البعد الصغرى لـ (x) تقابلها القيمة قبل الكبرى لـ (y) وهكذا الى ان نصل بان تكون اكبر قيمة من قيم (x) تقابلها اصغر قيمة فى قيم (y) فيكون الارتباط عكسيا وتاما. الا ان معظم التوزيعات تقع بين هذا وذاك اى قد لا يكون لا هو ارتباطا طرديا تماما ولا هو عكسيا تماما ولقياس ارتباط الرتب فهناك معامل ارتباط الرتب لسبيرمان ومعامل كندال وسوف نعرض هاذين المعاملين على النحو المبين التالى :

أ- معامل ارتباط الرتب لسبيرمان :

يعتمد معامل ارتباط الرتب الذى وضعه سبيرمان على ايجاد فروق الرتب للقيم المتناظرة 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 .

أولاً: حساب معامل ارتباط الرتب لسبيرمان فى حالة البيانات الغير مبوبة :

اذا كان لدينا عدد (n) من ازواج المشاهدات للمتغيرين X , Y اى لدينا مجموعة القيم : $(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n)$ فإنه لحساب معامل الارتباط لسبيرمان يتم الاتي :

(1):ضع رتب تصاعديّة او تنازليّة لقيم كل متغير على حده على اساس موقع القيمة بالنسبة لجميع قيم المتغير الذى ينتمى له . فاذا تم وضع رتب تصاعديّة 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) .

- يتم حساب فروق الرتب المتناظرة للمتغيرين ولتكن هذه الفروق هي مجموعة القيم D_i حيث $i = 1, 2, \dots, n$ حيث ان :

$$D_i = r_i - r_y$$

ويلاحظ ان بعض قيم (d_i) ستكون موجبة والبعض الاخر ستكون سالبة وقد تأخذ قيم صفرية . الا ان معيار صحة الحسابات هو ان يكون المجموع الجبرى لهذه الفروق مساوية 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 .
- 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 ارتباط الرتب لسبيرمان بين -1 , +1 وعند حسابنا لمعامل ارتباط الرتب لسبيرمان بين متغيرين y , x نكون بصدد احد الصور الثلاثة التالية :

الاولى : عندما تكون الرتب المتناظرة لقيم المتغيرين متساوية تماما . وهو ما يودى الى تطابق الرتب ومن ثم تكون فروق الرتب صفرية 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 .
الثانية : اذا كانت الرتب المتناظرة لازواج المشاهدات عكسية تماما حيث تكون رتب قيم المتغير (x) هى : 1,2,.....,n,(n-1),n تقابلها تماما رتب المتغير (y) والتي تأخذ الصورة 1,2,.....,(n-1),n وهنا تكون العلاقة بين المتغيرين عكسية تامة .

الثالثة : فى حالة عدم وجود علاقة ارتباط الرتب بين المتغيرين x , y فان قيمة معامل ارتباط الرتب لسبيرمان تكون صفرية .

ملحوظة : ان المؤشر الحقيقى لتمام العلاقة الخطية (سواء طردية تامة او عكسية تامة) هو قيمة معامل الارتباط الخطى البسيط لبيرسون .

بمعنى اذا كانت هناك علاقة خطية بين متغيرين x , y فانه يمكن القول بانه اذا كانت العلاقة بينهما علاقة تامة. فان معامل الارتباط الخطى البسيط سيكون مساويا +1 , -1 وفى هذه الحالة فانه معامل ارتباط الرتب لسبيرمان لا بد 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 ارتباط الرتب لسبيرمان مساويا +1 او -1 فان ذلك لا يعنى بالضرورة ان تكون العلاقة الخطية بينها تامة اى ليس بالضرورة ان تكون قيمة معامل الارتباط الخطى البسيط لسبيرمان مساوية ايضا ± 1 .

هذا بالاضافة الى ان الاختلاف فى قيم المعاملين (بيرسون وسبيرمان) يرجع اصلا لاختلاف طريقة الحساب حيث يعتمد معامل الارتباط الخطى البسيط لبيرسون فى حساباته على قيم ازواج المشاهدات فى حين يعتمد معامل ارتباط

الرتب لسبيرمان على رتب ازواج المشاهدات وليس قيمها فاختلف طرق الحساب من شأنه ان يؤدي الى اختلاف النتائج .

مثال:

أجرى احد اختبارات الذكاء لمجموعة مكون من ثمانية افراد في اعمار مختلفة فكانت نتيجة الاختبار على النحو التالي :

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 المعامل دعنا نفترض ان العمر هو المتغير (x) والتقدير هو المتغير (y) ولحساب معامل سبيرمان يتم تصوير جدول الحسابات التالى:

Age (x)	Grade(Y)	R_X	R_Y	$D_i = R_X - R_Y$	D_i^2
20	Very Good	6	7	6-7 =-1	1
17	Good	3.5	5	3.5-5 = -1.5	2.25
25	Excellent	8	8	0	0
18	Good	5	5	0	0
12	Weak	1	2	-1	1
15	Pass	2	3	-1	1
17	Very Weak	3.5	1	2.5	6.25
22	Good	7	5	2	4
Σ				0	15.5

لاحظ ان هناك قيم مكرره فى قيم المتغيرين x , y فبالنسبة للمتغير (x) الذى يعبر عن العمر نجد ان القيمة 17 مكرره مرتان فلو لم تكن هاتين القيمتان مكررتان لكانت رتبهم هى 3 , 4 لذا تم اعتبار الوسط الحسابى للرتبتين المتتاليين اى $3.5 = \frac{3+4}{2}$ وكذلك بالنسبة للمتغير (y) الذى يعبر عن التقدير فان التقدير good مكرر ثلاث مرات . فلولا هذا التكرار لكانت رتب هذه المفردات هى 4 , 5 , 6 ومن ثم يتم اعتبار الوسط الحسابى لتلك الرتب الثلاث اى:

$$\frac{4+5+6}{3} = 5$$

لذا يتم وضع الرتبة 5 امام كل تقدير "Good" كما بالجدول. وعليه فان:

$$r = 1 - \frac{6 \sum D_i^2}{n(n^2 - 1)}$$

ومن خلال حسابات الجدول السابق فإن 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 .

ثانياً: حساب معامل ارتباط الرتب لسبيرمان في حالة البيانات المبوبة :

يتم حساب معامل ارتباط الرتب في حالة البيانات المبوبة في جدول تكرارى مزدوج للمتغيرين x , y سواء كانت البيانات كمية أو وصفية حيث يتم إعطاء الفئات رتباً تصاعدياً (أو تنازلياً) ويشترط أن يكون الجدولين التكرارين منتظمين لكل من X ، Y بمعنى ان كون كل فئات x متساوية الطول وكذلك كل فئات y متساوية الطول . ويتعذر استخدام هذه الطريقة في حالة عدم تساوى اطوال الفئات لاحد المتغيرين او كليهما ويمتاز هذا المعامل بانه يصلح لقياس الارتباط في حالة الجداول التكرارية المفتوحة .

وجدير بالذكر ان حساب معامل ارتباط الرتب في حالة البيانات المبوبة بالجدول التكرارى المزدوج يعطى نفس النتيجة التى تم التوصل اليها باستخدام معامل ارتباط بيرسون بالرغم من اننا نستخدم فى المعامل الاول رتب x , y بينما فى المعامل الثانى قيم x , y .

ويتم حساب معامل ارتباط الرتب لسبيرمان في حالة البيانات المبوبة أى لجدول تكرارى مزدوج باتباع الخطوات الاتية :

(1): نعطي فئات المتغير x رتبا تصاعدياً تبدأ بالوحدة اى 1 , 2 , ..., n وكذلك نعطي فئات المتغير y رتبا تصاعدياً 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 x , y لهذا الاقطار المتوازية حيث يكون هذا الفرق مساويا للصفر بالنسبة للقطر الرئيسى ويأخذ قيما مختلفة عن الصفر بالنسبة لباقي الاقطار ويلاحظ ان هذا الفرق بين رتبتى x , y يكون متساويا على جميع اجزاء كل قطر من هذه الاقطار لذلك تسمى هذه الطريقة احيانا 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 تكراراته 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 $\sigma_d^2, \sigma_y^2, \sigma_x^2$ هو تباين رتب x ، تباين رتب y وتباين فروق الرتب

بين y, x علي الترتيب .

هذا ولتوضيح خطوات حساب معامل ارتباط الرتب لسبيرمان في حالة البيانات

المبوبة دعنا نقدم المثال التالي :

مثال :

الجدول التالي يمثل الانفاق الاستهلاكي (x) بالجنيهات والادخار (y) بالجنيهات

في احد الشهور لعينة مكونة من 200 اسره موزعين علي النحو التالي :

Saving Classes (y)	120-	125-	130-	135-	140- 145	Σ
Exp. Classes (x)						
280-			5	25	10	40
290-			40	20		60
300-			25	5	20	50
310-		30				30
220-330	10	10				20
Σ	10	40	70	50	30	200

والمطلوب حساب معامل ارتباط الرتب لسبيرمان بين الانفاق الاستهلاكي وبين

الادخار للاسرة ؟

الحل :

بالنظر الى شكل توزيع التكرارات بالجدول التكرارى المزدوج نجد ان التكرارات تتركز حول القطر العكس (من اعلى جه اليمين الى ادنى اليسار) مما يكشف بصورة مبدئية عن وجود ارتباط عكسى بين المتغيرين محل الدراسة ولحساب معامل ارتباط الرتب لسبيرمان يتم حساب تباين الرتب لكل من توزيع المتغيرين كلا حده اولا : وذلك على النحو التالي:

فبالنسبة لجدول المتغير (x) فان :

Rank of (x) Classes = R _x	F _x	R _x F _x	R _x ² F _x
1	40	40	40
2	60	120	240
3	50	150	450
4	30	120	480
5	20	100	500
Σ	200	530	1710

فيكون تباين فئات المتغير (x) هو :

$$\begin{aligned}\sigma_x^2 &= \frac{\sum R_x^2 F_x}{\sum F} - \left(\frac{\sum R_x F_x}{\sum F} \right)^2 \\ &= \frac{1710}{200} - \left(\frac{530}{200} \right)^2 = 1.5275 \quad (L.E)\end{aligned}$$

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
Σ	200	650	2350

وعليه فان تباين رتب فئات المتغير (y) هو :

$$\sigma_y^2 = \frac{\sum R_y^2 F_y}{\sum F} - \left(\frac{\sum R_y F_y}{\sum F} \right)^2$$

$$= \frac{2350}{200} - \left(\frac{650}{200} \right)^2 = 1.1875 \quad (L.E)$$

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 :

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 المبيئية فان تبأين فروق رتب فئات التوزيع التكرارى
المزدوج هو :

$$\sigma_d^2 = \frac{\sum d_i^2 F_d}{\sum F_i} - \left(\frac{\sum d_i F_d}{\sum F_i} \right)^2$$

$$= \frac{980}{200} - \left(\frac{-120}{200} \right)^2 = 4.54 \quad (L.E)$$

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 الاتفاق الاستهلاكى (x) والادخار (y) للاسرة فى تلك العينة.

ملاحظات حول معامل ارتباط الرتب لسبيرمان :

(1): يستخدم معامل ارتباط الرتب لسبيرمان لقياس العلاقة بين متغيرين سواء كانت القياسات ترتيبية اصلا او كمية تم تحويلها الى قياسات ترتيبية بينما معامل ارتباط 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 رتبين ثم تربيع هذا الفرق . لذلك لا يفضل استخدام هذا المعامل الا فى الحالات التى لا تتطلب دقة عالية فى النتائج أى فى حالة تعذر حساب معامل بيرسون نظرا لان البيانات وصفية ويتعذر قياسها كميًا بل يمكن ترتيبها فيما بينها ووفقا لخاصية معينة .

(3): عند حساب معامل ارتباط الرتب لسبيرمان فانه يلزم ترتيب المتغيرين x , y بنفس الطريقة بمعنى اذا رتبنا المتغير (x) ترتيبا تنازليا مثلا فانه يلزم حينئذ ترتيب المتغير (y) تنازليا ايضا والعكس صحيح فلا يصح ترتيب احد المتغيرين تنازليا مثلا وترتيب المتغير الاخر تصاعديا فى نفس الوقت .

ب – معامل كندال للارتباط :

راينا ان معامل ارتباط الرتب لسبيرمان يعاب عليه عدم وجود معنى طبيعى لحساب الفرق بين رتبتي المتغيرين ثم يتم ايجاد مربع هذه الفروق . لذا يوجد معامل لقياس ارتباط الرتب لكن يعتبر افضل من المعامل السابق يعرف بمعامل كندال لارتباط الرتب ومن ثم فان معامل كندال شأنه شأن معامل سبيرمان يعتمد على رتب قيم المتغير وليس على قيم المتغيرين الاصلية . ويحسب معامل كندال لارتباط الرتب اذا ما توافرت لدينا ازواج من المشاهدات ولتكن :

$$(X_1, Y_1) , (X_2, Y_2) , \dots, (X_n, Y_n)$$

فانه يتم اتباع الخطوات التالية لحساب هذا المعامل :

- (1): ضع رتبا تصاعديا 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 , , 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 ($t = 1 , 2, \dots, n$) 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) والذي يأخذ الصورة الرياضية التالية :

$$r_k = \frac{2C}{n(n-1)}$$

هذا وتتراوح قيمة هذا المعامل بين +1 , -1 اي ان :

$$-1 \leq r_k \leq 1$$

والمثال: التالي يوضح كيفية حساب معامل كندال لارتباط الرتب .

مثال:

احسب معامل كندال للارتباط وذلك لبيانات درجات الطلاب العشرة فى مادتي الاحصاء والمحاسبة والذي تم له حساب معامل الارتباط لفهرس سابقا .

الحل:

الجدول التالي يوضح الحسابات اللازمة لاجاد معامل كندال للارتباط :

Stat . Degree(x)	Accounting Degree(y)	R _x	R _y	Normal Ranking for(R _x) and its corresponding (R _y)		N _t	M _t	C=N _t - M _t
				R _x	R _y			
82	80	8	7	1	4.5	5	3	5-3=2
98	83	10	8	2	3	6	2	4
67	52	5	2	3	1	7	0	7
40	63	2	3	4	4.5	5	1	4
75	96	7	10	5	2	5	0	5
61	71	4	4.5	6	6	4	0	4
31	71	1	4.5	7	10	0	3	-3
72	75	6	6	8	7	2	0	2
53	35	3	1	9	9	0	1	-1
83	85	9	9	10	8	0	0	0
Σ		55	55	55	55			24

وعليه فان معامل كندال للارتباط هو

$$r_k = \frac{2C}{n(n-1)} = \frac{2(24)}{10(10-1)} = \frac{48}{90} = 0.533$$

اي ان هناك علاقة طردية بين المتغيرين ولكنها ليست قوية .

4- الارتباط بين الصفات (المبوبة تكراريا في جداول تكرارية مزدوجة):

درسنا في الاجزاء السابقة من هذا الباب كيفية قياس العلاقة بين الظواهر الكمية والتي 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) ففي هذه الحالة لا نستطيع استخدام معامل الارتباط لقياس

العلاقة بين الظاهرتين وسندرس اثنين من هذه المقاييس هما معامل الاقتران ومعامل التوافق .

(أ): معامل الاقتران : **Coefficient of Association**

قدم يول Yule مقياسا لدراسة العلاقة بين ظاهرتين احدهما على الاقل وصفية و كليهما يتكون من قسمين فقط يسمى بمعامل الاقتران ليول . وهذا المعامل يقيس العلاقة بين ظاهرتين احدهما 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 $F_{11}, F_{12}, F_{21}, F_{22}$ تشير الى تكرارات خلايا الجدول التكرارى المزدوج حيث $F_{i,j}$ تشير الى تكرار الخلية التى تقع فى الصف (i) والعمود (j) حيث $J=1,2$ ، $i=1,2$

فى هذه الحالة فان افضل مقياس لقياس الارتباط بين الظاهرتين x, y هو 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 x , y اقترانا تاما . اما اذا كانت قيمة هذا المعامل قيمة سالبة فانه يجب اهمال الاشارة السالبة لانه لا يصح ان يقال ان هناك علاقة طردية او عكسية بين ظاهرتين وصفيتين وعندما تنقدم قيمة هذا المعامل اى 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) والتي شملت 200 فردا تم اختيارهم عشوائيا تم الحصول على البيانات لتالية :

X \ Y	Black	Coloured	Σ
Black	110	40	150
Yellow	20	30	50
Σ	130	70	200

والمطلوب دراسة ماهية وجود علاقة بين لون الشعر (x) ولون العينين (y) وذلك باستخدام المقياس الاحصائى الملائم .

الحل :

حيث أن الجدول التكرارى المزدوج للمتغيرين x , y مكون من صفين وعمودين اي يضم اربعة خلايا فقط لظاهرتين احدهما على الاقل وصفية لذا فان انسب معامل لقياس درجة العلاقة بين الظاهرتين هو معامل يول للاقتران (C_{ASS}) وهو :

$$C_{ASS} = \frac{F_{11} F_{22} - F_{12} F_{21}}{F_{11} F_{22} + F_{12} F_{21}}$$

$$= \frac{110 (30) - 20(40)}{110 (30) + 20(40)} = 0.61$$

وهو ما يدل على ان هناك علاقة طردية وقوية الى حد ما (متوسطة) فيما بين لون شعر الفرد ولون عينيه .

هذا ومن أهم خصائص معامل يول للاقتران ما يلى :

* اذا كانت احد التكرارات الاربعة F_{11} او F_{12} او F_{21} او F_{22} على الاقل مساوية لصفر فان معامل يول للاقتران يكون حينئذ مساويا $+1$ وتكون علاقة الاقتران تامة. اما اذا كانت كل من التكرارات الاربعة اكبر من الصفر فان معامل الاقتران تتراوح قيمته بين $+1$, -1 وكلما اقتربت قيمة هذا المعامل من الصفر دل ذلك على ضعف العلاقة بين الظاهرتين والعكس صحيح .

* إذا كان : $F_{12} F_{21} = F_{11} F_{22}$ فإن 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 عن 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 2x2) لذا فانه لقياس العلاقة بين الظاهرتين يكون ذلك من خلال معامل يول للاقتران وهو :

$$C_{ASS.} = \frac{F_{11} F_{22} - F_{12} F_{21}}{F_{11} F_{22} + F_{12} F_{21}} = \frac{50 (20) - 15 (35)}{50 (20) + 15 (35)} = 0.311$$

وهذا يدل على ان هناك علاقة ضعيفة ما بين التدخين والاصابة بمرض القلب ومن بيانات الجدول المعطى نستطيع القول ان التدخين يؤدي الى حد ما الى الاصابة بمرض القلب حيث نجد ان نسبة المصابين بالمرض من بين المدخنين تبلغ حوالى 59% بينما تبلغ تلك النسبة حوالى 43% بين غير المدخنين.

مثال:

الجدول التالى يوضح البيانات الخاصة بتوزيع 250 طفلا حسب الاصابة بمرض الحصبة (x) وعدم التحصين ضد هذا المرض (y).

Vaccination(y) Illness disease(X)	Not Vaccinate	Vaccinate
	Illness	40
Not Illness	12	180

والمطلوب: حساب معامل الاقتران بين الاصابة بالمرض وعدم التحصين ضده مع التعليق على نتائجك .

الحل:

حيث ان معامل الاقتران :

$$C_{ASS.} = \frac{F_{11} F_{22} - F_{12} F_{21}}{F_{11} F_{22} + F_{12} F_{21}} = \frac{40 (180) - 18 (12)}{40 (180) + 15 (12)} = 0.94$$

اي ان هناك علاقة قوية ما بين الاصابة بمرض الحصبة وعدم التحصين ضده ومن بيانات الجدول المعطى نستطيع ان نستنتج ان الاطفال الذين لم يحصنوا ضد هذا 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 في حين ان هذه النسبة تبلغ %77 من بين الذين لم يحصنوا ضد هذا المرض .

هذا ويجب التنوية عن ما سبق تناوله من نقطة هامة وهى انه يؤخذ دائما بالقيمة الموجبة لمعامل الاقتران مهملين الاشارة السالبة للنتائج السالبة ان وُجدت . إذ أنه لا يصح ان يقال ان هناك علاقة طردية او عكسية بين ظاهرتين وصفتين (او احدهما على الاقل وصفية) وذلك لان الاشارة تنتج عن جدولة البيانات فى الجدول التكرارى المزدوج . بمعنى انه اذا اعدنا 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 .

(ب): معامل التوافق: Contingency Coefficient

درسنا فى المقياس السابق قياس علاقة الاقتران ما بين ظاهرتين احدهما على الاقل وصفية وكل منها يتكون من قسمين فقط . اما عن علاقة التوافق التى نحن بصددنا الان فهى تبحث ايضا عن العلاقة بين ظاهرتين ايضا احدهما على الاقل وصفية ولكن تتكون احدهما ايضا على الاقل من قسمين . فقد قدم لنا

كارل بيرسون K.Pearson مقياسا احصائيا يقيس العلاقة 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	Σ
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.}
Σ	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 $i = 1, 2, \dots, n$ ،

$j = 1, 2, \dots, m$

(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 :

$$C = \sum_{i=1}^n \sum_{j=1}^m (F_{ij}^2 / F_i \cdot F_j)$$

هذا ويأخذ معامل التوافق دائما ابدا قيمة موجبة (ان لم تكن قيمة مساوية للصفر) كما ان الحد الاقصى (اكبر قيمة يمكن ان يأخذها) لمعامل التوافق يمكن تحديده من خلال العلاقة التالية :

$$\text{Maximum Value of } C_{cont.} = \sqrt{\frac{(n-1)(m-1)}{n \cdot m}}$$

حيث (n) , (m) هي 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 هي :

$$\text{Maximum Value of } C_{cont.} = \sqrt{\frac{n-1}{n}}$$

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 التالي يوضح توزيع 80 أسرة حسب درجة تعليم الام وعدد الاطفال فى الاسرة وذلك بعد مدة زواج قدرها ستة سنوات :

No. children (y)	Less than 3 children	3 children and More	Σ
Mother Education(x)			
Not Qualified	2	10	12
Preparatory	3	6	9
Secondary	20	3	23
University Qualification and more	35	1	36
Σ	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 :

$$\begin{aligned}
 C &= \sum_i \sum_j (F_{ij}^2 / F_i \cdot F_j) \\
 &= \frac{(2)^2}{12 \times 60} + \frac{(10)^2}{12 \times 20} + \frac{(3)^2}{9 \times 60} + \dots \\
 &\dots + \frac{(1)^2}{36 \times (20)} = 1.519
 \end{aligned}$$

$$C_{Cont.} = \sqrt{\frac{c-1}{c}} = \sqrt{\frac{1.519}{1-1.519}} = 0.58$$

وحيث ان الحد الاقصى لمعامل التوافق في هذا المثال هو :

$$\begin{aligned} \text{maximum Value of } C_{cont.} &= \sqrt{\frac{(n-1)(m-1)}{n.m}} \\ &= \sqrt{\frac{(2-1)(4-1)}{2(4)}} \end{aligned}$$

ولتحديد ماهية قوة العلاقة يتم حساب النسبة ما بين قيمة معامل التوافق والحد الاقصى له وهي :

$$\frac{0.58}{0.78} = 0.74$$

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): إذا 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 يلي الدرجات التي حصل عليها 10 طلاب في مادتي 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 يلي الدرجات التي حصل عليها 10 طلاب في امتحان مادتي 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): في أحد الدراسات الاحصائية عن رأى مجموعتين من السائحين في ستة فنادق بهدف معرفه الافضليه من حيث حسن التعامل مع السائح فكانت نتجة الدراسة كما يلي :

الفندق	هيلتون	شيراتون	سونستا	موفنبيك	ميريديان	ماريوت
المجموعة الاولى	2	5	1	4	3	6
المجموعة الثانية	5	4	3	2	1	6

والمطلوب تحديد قيمة معامل الارتباط (معامل الرتب) بين افضلية المجموعتان .

(9): فيما يلي اسعار السلع الجلدية والخزف (للوحدة الواحدة) خلال عشرة ايام فى منطقة خان الخليلي :

السلع الجلدية 5 7 8 4 9 3 2 5 4 3

السلع الجلدية 2 4 5 5 6 4 5 4 3 2

والمطلوب ايجاد معامل بيرسون وايضا معامل ارتباط سبيرمان (الرتب) ومقارنة قيمة معاملى الارتباط وبيان سبب الاختلاف فى قيمتها أن وجد .

(10): يوضح الجدول التالى توزيع عينة بحسب فئات دخولهم الشهرية 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 : Y ، X .

(11): إذا كان لدينا البيانات الآتية عن كلا من الظاهرة Y ، X .

$$\sum_{XF_X} = 39000 \quad , \quad \sum_{X^2F_X} = 1605000$$

$$\sum_{yF_y} = 29950 \quad , \quad \sum_{y^2F_y} = 919750$$

$$\sum_{XYF_{XY}} = 1197250 \quad , \quad \sum_{F_X} = \sum_{F_Y} = \sum_{F_{XY}}$$

والمطلوب : ايجاد :

1-الوسط الحسابي والانحراف المعياري ومعامل الاختلاف للظاهرتين

. و قارن تشتت الظاهرتين Y ، X .

2-حساب معامل الارتباط بين Y ، X .

(12): فيما يلي بيانات عن عينة من نزلاء فندق سونستا بالقاهرة حسب

أعمارهم (X) وعدد الليالي السياحية التي قضاها بالقاهرة y

$Y \backslash X$	2	3	5	6	Σ
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
Σ	5	19	18	7	50

والمطلوب :

ايجاد قوة العلاقة ونوعها بين اعمار نزلاء الفندق و عدد الليالي السياحية

وعلق على النتيجة .

(13): ارسم الشكل الانتشارى ثم أوجد معامل الارتباط الخطى البسيط بين

المتغيرين x, y وذلك من واقع البيانات التالية :

X: 71 ، 62 ، 56 ، 45 ، 32 ، 28 ، 24 ، 21 ، 18 ، 12

Y: 65 ، 60 ، 50 ، 42 ، 35 ، 30 ، 23 ، 22 ، 13 ، 11

(14): فيما يلى الدخل الاسبوعى (x) والانفاق الاسبوعى (y) لعينة مكون من

8 أسر :

X: 76 ، 98 ، 72 ، 81 ، 62 ، 55 ، 58 ، 43

Y: 62 ، 50 ، 45 ، 79 ، 61 ، 53 ، 55 ، 35

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 تلقته مجموعة من

العاملين باحدى الشركات ونتيجة تقييم أدائهم فى نهاية الملائم (y) .

X: 65 60 40 35 50 42 20

Y: مقبول ممتاز جيد جدا 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 10 أسر :

X : 100 ، 95 ، 91 ، 76 ، 65 ، 82 ، 75 ، 130 ، 110 ، 80

Y : 94 ، 80 ، 82 ، 63 ، 50 ، 55 ، 60 ، 102 ، 93 ، 75

Z : 6 ، 5 ، 2 ، 3 ، 2 ، 3 ، 3 ، 4 ، 6 ، 5

والمطلوب :

- أ- ايجاد معاملات الارتباط الخطى بين أزواج المتغيرات المختلفة.
 ب- احسب معاملات التحديد المختلفة مع توضيح مدلول كل منهما.
 ج- على ضوء نتائجك فى أ ، ب علق بما تراه مناسباً .

(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 :

المجموع	-85	-75	-65	-55	-45	-35	-25	X Y
11					3	5	3	-40
13				1	1	6	5	-50
19			1	10	7	1		-60
25		3	7	9	6			-70
18	1	5	8	3	1			-80
14	9	5						-90
100	10	13	16	23	18	12	8	المجموع

والمطلوب :

أ- إيجاد معامل الارتباط الخطى البسيط بين Y ، X .

ب- باستخدام معامل التحديد فسر ما تحصل عليه من نتائج.

(20): فيما يلى التوزيع التكرارى المزدوج لدرجات 50 طالبا فى كل من مادتي

الاحصاء (x) والرياضيات (y) :

المجموع	-90	-80	-70	-60	-50	(x)	(y)
11		2	3	4	2		-60
26		5	9	8	4		-65
22	2	3	11	4	2		-70
30	4	9	15	2			-75
7	2	2	3				-80
3	1	2					-85
1	1						-90
50	10	23	41	18	8		المجموع

والمطلوب :

أ - حساب معامل الارتباط الخطى بين x , y . ثم تفسير ما تحصل عليه من

نتائج باستخدام معامل التحديد .

ب- قارن بين الاختلاف النسبى لكل من x , y .

ج - 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 :

المجموع	غير ذلك	سوداء	X
	Other	Black	Y
23	8	25	سوداء Black
27	15	12	غير ذلك Other
60	23	37	المجموع

والمطلوب : دراسة اثر العوامل الوراثية فى لون العينين وذلك باستخدام المقياس الاحصائى المناسب .

(22): فيما يلى توزيع عينة من 100 شخص حسب درجة تعليمهم (x) وارئهم تجاه قضية عمل المرأة (y) :

المجموع	Prep	Second	Univ.	x
	And Less			y
42	9	28	5	مويد
58	6	2	50	معارض
100	15	30	55	المجموع

والمطلوب وضع الى اى حد تعتمد وجهة النظر تجاه عمل المرأة على درجة التعليم مستخدما فى ذلك المقياس الاحصائى الملائم .

(23): فيما يلى نتيجة امتحان رياضيات الاستثمار (X) فى احدى كليات التجارة حسب حالة الطلاب (منتظمين و منتسبين) (y) وذلك فى دور يناير 1987 :

المجموع	Pass	Failed	X Y
908	716	192	إنتظام
352	208	144	إنتساب
1260	924	336	المجموع

والمطلوب : وضح الى اى حد تتوقف نتيجة الامتحان على نظام الدراسة مستخدماً في ذلك المقاييس الاحصائى الملائم ؟

(24): الجدول التالى يوضح توزيع الطلاب المنتظمين و المنتسبين فى احدى كليات التجارة وذلك حسب تقديراتهم فى نتيجة امتحان مادة الاحصاء لدور مايو 1987 :

المجموع	ممتاز	جيد جداً	جيد	مقبول	ضعيف	ضعيف جداً	التقدير نظام الدراسة
702	15	36	167	170	148	166	إنتظام
285	1	5	24	59	88	108	إنتساب
987	16	41	191	229	236	274	المجموع

والمطلوب : باستخدام المقاييس الاحصائى الملائم ،حدد درجة قوة وطبيعة العلاقة ما بين نظام الدراسة وتقديرات الطلاب فى مادة الاحصاء.

(25):البيانات التالية تمثل درجات عشرة طلاب فى كل من مادتي الفلسفة وعلم نفس :

19	14	16	18	15	12	13	17	10	16	X
15	12	16	17	13	16	14	15	12	14	Y

والمطلوب: حساب معامل الارتباط الخطى البسيط . ثم علق على النتيجة التي تحصل عليها باستخدام معامل التحديد .

(26): فيما يلي التوزيع التكرارى المزدوج للمتغيرين :-

المجموع	60-50	-40	-30	-20	-10	X Y
4			1	1	2	-25
7			2	4	1	-30
13		3	5	3	2	-35
14		6	6	2		-40
12	3	4	5			50-45
50	3	13	19	10	5	المجموع

والمطلوب

أ- حساب معامل الارتباط الخطى ثم فسر النتيجة التي تحصل عليها وذلك باستخدام معامل التحديد.

ب- قارن بين التشتت النسبى لكل من المتغيرين .

ج- باستخدام معامل بيرسون الاول قارن بين الالتواء فى كل من التوزيعين.

(27): في دراسة لتحديد العلاقة ما بين درجة تعليم الرجل ووجهة نظره تجاه عمل المرأة تم استطلاع آراء مائة رجل وكانت آراؤهم على النحو المبين بالجدول التالي :

اعدادى فأقل	ثانوى	جامعى فما فوق	درجة التعليم وجهة نظر تجاه عمل المرأة
5	22	28	مويد
30	10	5	معارض

المطلوب :

- 1- تحديد طبيعة ودرجة قوة العلاقة ما بين درجة التعليم للرجل ووجهة النظر تجاه عمل المرأة وذلك باستخدام المقياس الاحصائى الملانم .
- 2- لو إفترضنا ان هناك شكوكا حول صحة آراء الرجال الحاصلين على مؤهل اعدادى فأقل وانه تم استبعادهم نهائيا من الدراسة. والمطلوب هنا قياس درجة قوة العلاقة ما بين درجة التعليم وتأيد عمل المرأة فى هذه الحالة مستخدما المقياس الاحصائى الملانم .
- 3- قارن بين النتيجين فى كل من (أ) ، (ب) مع إبداء التعليق المناسب .

الفصل السابع

الانحدار الخطي البسيط

Simple linear Regression

رأينا في الباب السابق ان الارتباط 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 بين متغيرين كميين احدهما مقيد عند مستويات معينة 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 **Dependent Variable (y)** ولقد عُرف الانحدار عام 1885 علي يد فرانسيس جالتون حينما اهتم بدراسة الصفات الوراثية للابناء والاباء. فقد لاحظ ان الابناء لآباء طوال القامة يميلون في المتوسط الي ان يكونوا قصارا عن ابائهم في حين ان الابناء لآباء قصار القامة يميلون في المتوسط الي ان يكونوا طوالا عن ابائهم فاستخدم جالتون لفظ انحدار للتعبير عن هذا الاتجاه نحو المتوسط او الدرجة المتوسطة في هذه العلاقة . ومنذ ذلك الحين ظل هذا اللفظ متداولاً في سائر العلوم الادارية والاقتصادية والبيولوجية ...
والخ

ويقصد بتحليل الانحدار عرض شكل العلاقة بين المتغيرين (او المتغيرات) في صورة معادلات رياضية بهدف التنبؤ بقيمة المتغير التابع (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 والسياحة والتعليم والصحة والقوي العامله و... الخ . فتخصيص الميزانية لحظة خمسية قادمة يتطلب التنبؤ بعدد السكان المتوقع خلال سنوات الخطة وتوزيعهم العمري والنوعي والتنبؤ باعداد الطلاب المتوقع في المراحل التعليمية المختلفة واعداد الخريجين وقوة العمل وحجم المباني والالات والاجهزة المطلوبة لتنفيذ المستهدف من الخطة . وتلعب معادلات الانحدار الدور الرئيسي في الوصول الي مثل هذه التنبؤات كما سنري فيما بعد .

وكمثال مبسط لتعريف فكرة المتغير التابع والمتغير المستقل في علاقة الانحدار نشير الي انه من الطبيعي ان تكون هناك علاقة بين الدخل الشهري للأسرة 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 فإن الدخل يعتبر بمثابة المتغير المستقل بينما يمثل الانفاق المتغير التابع . وتوضيح 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 ويافترض أن هذا المعامل هو

0.844 وهو ما يفيد أن هناك علاقة طردية قوية ما بين الدخل والانفاق . بمعنى 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 بواسطتها 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 مستقلين علي الاقل والتي تؤثر علي المتغير التابع . ويعد الانحدار الخطي البسيط في أغلب الحالات تبسيطا شديدا للواقع بل واغفالا لجانب كبير مما هو كائن في الحياه العملية حيث نجد في الواقع ان المتغير التابع يعتمد في تفسيره وتغييره علي عدد كبير من المتغيرات المستقلة . ولذلك نجد ان الانحدار المتعدد هو الاكثر شيوعا واستخداما في الحياه العلمية فهو يفيد في حل الكثير من مشاكل التحليل الاقتصادي والتجاري والادارة الصناعية وسوف نقتصر في هذا الباب علي دراسة نموذج الانحدار الخطي البسيط بشئ من التفصيل تاركين الانحدار المتعدد لدراسات احصائية فيما بعد أن شاء الله .

الانحدار الخطي البسيط :

بيننا في المقدمة السابقة انه عند دراستنا للانحدار الخطي البسيط لابد من تحديد كل من المتغير المستقل والمتغير التابع . فاذا كان لدينا المتغيران y . x فإن 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 y . X ويمكن توضيح ذلك علي النحو التالي :

الخط المستقيم الاول والذي يكون فيه المتغير (y) هو المتغير التابع ، (x) هو المتغير المستقل ويأخذ الصورة التالية :

$$\hat{Y} = a + bx$$

ويسمى هذا النموذج بخط انحدار y علي X حيث (\hat{y}) هي القيم المستنتجة (المتوقعة) 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 الصورة الخطية التالية :

$$\hat{X} = \hat{a} + \hat{b} y$$

ويسمى هذا النموذج بخط انحدار x علي y . حيث \hat{X} هي القيم المستنتجة (المتوقعة) والمحسوبة من معادلة خط انحدار X علي y وذلك عند قيم y المختلفة .

هذا وكلما زادت تركيز نقاط الشكل الانتشاري 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 العلاقة بين المتغيرين Y, X علاقة تامة (اي ان: $r = \pm 1$) فأن جميع ازواج المشاهدات (Y, x) تحقق ايا من هذين الخطين y علي x او x علي y . وفي تلك الحالة فأن كافة القيم الفعلية تتساوي مع كافة القيم المقدره المقابلة لها اي أن: $X = \hat{X}$ ، $y = \hat{Y}$ وذلك لجميع قيم x, y وهنا يجب التمييز بين كل من \hat{Y}_i, Y_i او بين \hat{X}_i, X_i حيث X_i, Y_i هي مجموعة القيم المشاهدة

(الفعلية) للمتغيرين x, y علي الترتيب أما \hat{X}, \hat{Y} فانهما كما سبق تعريفها يمثلان القيم المتوقعة لكل من Y, X وذلك من معادلتني خطي الانحدار y علي x و x علي y علي الترتيب . هذا وكلما كان الفارق بين القيم المشاهدة والقيم المتوقعة صغيرا كلما زادت قوة العلاقة الارتباطية بين المتغيرين محل الدراسة والعكس صحيح وسوف نتعرض الان لدراسة معادلتني

خطي الانحدار آخذين في الاعتبار البيانات الغير مبوبة والبيانات المبوبة وذلك على النحو المبين التالي :

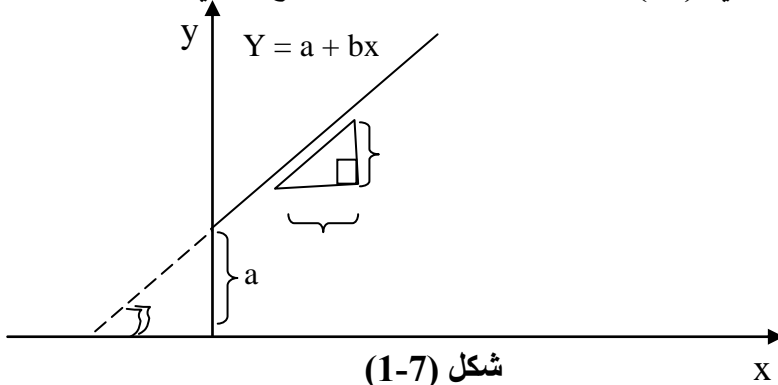
أولاً: معادلة خط انحدار y على x :

معادلة خط انحدار y على x تأخذ الصورة التالية

$$\hat{Y} = a + bX \quad (1-7)$$

حيث :

\hat{Y} تعبر عن القيمة المتوقعة للمتغير (y) عند قيم (x) المختلفة ، a ، b ثابتان . وهذان الثابتان يحددان الخط المستقيم تماما لذا يسمى هذان الثابتان بمعلمتي Parameters خط الانحدار . ومن خلال العلاقة (1-7) فان الثابت a هو عبارة عن قيمة Y عندما $x = 0$ ، اما عن الثابت (b) فهو عبارة معدل التغير في قيمة المتغير (y) بالنسبة الي (x) ويسمي الثابت (a) بثابت معادلة الانحدار اما (b) فتسمي بمعامل خط انحدار y على x . أما X فهي عبارة عن القيم المشاهدة (الفعلية) للمتغير X ويمكن تمثيل معادلة خط انحدار (y) على (x) بيانيا من خلال الشكل الموضح التالي :



شكل (1-7)

خط إنحدار (y) على (X)

وفي حالة تمام العلاقة بين المتغيرين ($y \cdot x$) (اي حينما تكون $r_{xy} = 1 \pm$) فان كافة النقاط الفعلية لازواج المشاهدات تقع علي خط الانحدار المقدر اي ان:

$$\widehat{Y}_i = Y_i = a + b X_i \quad \text{where } i = 1, 2, \dots, n$$

وهو ما معني انعدار ما يسمي بأخطاء التقدير وهي عبارة عن الفرق ما بين القيم الفعلية والمتوقعه اي ان :

$$Y_i - \widehat{Y}_i = 0 \quad \text{for } i = 1, 2, \dots, n$$

الانه من الناحية العملية فانه عادة ما تختلف القيم الفعلية (y) عن المتوقعه (\widehat{Y}) والمحسوبة من معادلة خط الانحدار المقدر . ويقل هذا الاختلاف لاشك كلما زادت قوة العلاقة بين المتغيرين والعكس صحيح تماما . وهذا يعني انه عادة ما تختلف القيمة ($y - \widehat{Y}_i$) عن الصفر وذلك نظرا لأن القيمة المطلقة لمعامل الارتباط الخطي عادة ما تختلف عن الواحد الصحيح . فاذا رمزنا لهذه القيمة بالرز (e_i) اي ان :

$$e_i = Y_i - \widehat{Y}_i \quad \text{for } i = 1, 2, \dots, n$$

فان (e_i) وهي عبارة عن الفرق ما بين القيمة المشاهدة والقيمة المتوقعه أو ما يسمي بالخطأ في تقدير (y_i) . وهذا الفرق يكون سالبا اذا كانت قيمة المشاهدة اسفل خط الانحدار ويكون موجيا اذا كانت قيمة (y) المشاهدة اعلي خط الانحدار .

وبأستخدام المعادلة (1-7) السابقة فانه يمكن التعبير عن الخطأ في التقدير علي الصورة :

$$e_i = Y_i - \widehat{Y}_i = Y_i - a - bx_i$$

1- إيجاد معادلة خط انحدار (y) على (x) للبيانات الغير مبوبة :

إذا كان لدينا عدد (n) من ازواج المشاهدات للمتغيرين x . y ولتكن :

$$(x_1, y_1), (x_2, y_2), (x_3, y_3) \dots \dots \dots, (X_n, Y_n)$$

فإن 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 نعيه بكلمة افضل في شأن المعلمتين a , b كما سبق وان ذكرنا في ان الفروق ما بين القيم المشاهدة والقيم المتوقعة للمتغير (y) يمكن علي اساسها الحكم علي مدي جودة تمثيل خط الانحدار للعلاقة بين المتغيرين Y،X ولذلك فان المعيار الذي يمكن من خلاله الحكم علي افضلية ودقة التقدير لتلك المعلمات سيكون من خلال تقليل الفروق بين القيم الفعلية (المشاهدة) والقيم المتوقعة 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 :

$$\sum |e_i| = \sum |y_i - \hat{y}_i| \quad \text{in its minimum value}$$

* توفيق خط الانحدار بحيث يكون قيم الفروق المطلقة أصغر ما يمكن .

* توفيق خط الانحدار الذي يجعل مجموع مربعات الفروق بين القيم المشاهدة والقيم المتوقعة لـ y اقل ما يمكن اي ان :

$$\sum e_i^2 \quad \text{in its minimum value}$$

هذا وتسمى 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 .

*تقدير معلمتي نموذج خط انحدار Y على X باستخدام طريقة المربعات الصغرى :

وفقا لطريقة المربعات الصغرى فإنه يمكن تقدير معلمتي معادلة خط انحدار Y على X على النحو التالي :

حيث ان معادلة خط انحدار Y على X هي : $Y = a + bx$

فأنه يمكن ايجاد قيمتي a , b والتي تجعل مجموع مربعات الفروق (e_i^2) اقل ما يمكن وذلك من خلال ايجاد النهاية الصغرى للدالة (7 - 3) مع استخدام الشرطين اللازمين لكي تكون هذه الدالة نهاية صغرى عند نقطة وذلك على النحو التالي :

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - a - bx_i)^2 \quad (3 - 7)$$

وهذان الشرطان اللازمان هما :-

✓ التفاضل الاول للدالة عند النقطة يساوي الصفر .

✓ التفاضل الثاني عند هذه النقطة يساوي مقدارا موجبا .

ومن ثم فبتطبيق الشرط الاول على الدالة (7-3) وذلك باعتبار التفاضل الجزئي لهذه الدالة بالنسبة الي كل من a, b مساويا للصفر. فأننا نحصل على 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 فإن :

$$\sum_{i=1}^n (y_i - a - bx_i) = 0$$

$$\sum y_i = na + b \sum x_i \quad \text{أي أن (4-7)}$$

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

$$\sum x_i y_i = a \sum x_i + b \sum x_i^2 \quad \text{(5-7)}$$

هذا وتسمى المعادلتان (4-7) و (5-7) بالمعادلات الطبيعية للنموذج وبحل هاتين المعادلتين معا يمكن الحصول علي تقدير لمعلمتي نموذج خط الانحدار b a . او من هاتين المعادلتين يمكن الوصول الي الشكل التالي :-

$$\hat{b} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$

وهناك صورة أخرى وهي :

$$= \frac{\frac{\sum xy}{n} - \frac{\sum x}{n} \cdot \frac{\sum y}{n}}{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2} \quad (6-7)$$

وقد فضلنا الصيغه (6-7) 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 الاولى عند النقطة (a , b) تحقق الشرط الثاني لاجاد النهاية الصغري للدالة (7 - 3) حيث تتحدد قيمة (a) بالصيغة (7-7) بينما تتحدد قيمة معامل الانحدار b من خلال الصيغة (6-7).

هذا ويمكن ان نيسر علي القارئ كيفية تذكر المعادلتين الطبيعتين (4-7) و (7-5) وذلك علي النحو التالي :

حيث ان معادلة خط انحدار Y علي X تأخذ الشكل التالي :

$$\hat{Y} = a + bx$$

فانه يمكن الحصول علي المعادلة الطبيعية الاولى (4-7) وذلك من خلال اعتبا مجموع طرفي معادلة خط احدار Y علي X مع اعتبار (\hat{Y}) بدلا من (Y) حيث أن مجموع القيم الفعلية يساوي مجموع القيم المتوقعة اي أن $\sum y = \sum \hat{y}$ 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 :-

$$\sum_{i=1}^n y_i = \sum_{i=1}^n \hat{y}_i \quad (8-7)$$

ومن العلاقة (8-7) نخلص إلى :

ان مجموع القيم الفعلية يساوي مجموع القيم المتوقعة . ومن ثم فإن الوسط الحسابي للقيم الفعلية او المشاهدة يساوي الوسط الحسابي للقيم المتوقعة .
ومن خلال استخدامنا لهذه الخاصية لطريقة المربعات الصغرى يمكن استنتاج صورة اخري لمعادلة خط انحدار (y) علي (x) وذلك كما يلي :

فحيث أن :

$$\hat{y} = a + bx$$

لذا فإن

$$\bar{y} = a + b\bar{x}$$

$$\hat{y} - \bar{y} + b(x - \bar{x}) \quad (9-7)$$

ومن المعادلة (7 - 9) يمكن التوصل إلي صورة ثالثة لمعادلة خط انحدار (y) علي X وهي :

$$\hat{y} = \bar{y} + b(x - \bar{x}) \quad (10 - 7)$$

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 بأفترض 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 - بأفترض 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 يجب الحصول علي المجاميع الموضحة 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

$$\begin{aligned} \hat{a} &= \frac{\sum xy}{n} - \frac{\sum x}{n} \cdot \frac{\sum y}{n} \\ &= \frac{1219}{7} - \frac{77}{7} \cdot \frac{105}{7} \\ &= \frac{1219}{7} - \frac{(77)^2}{7} \\ &= \frac{9.143}{6} = 1.524 \end{aligned}$$

أما عن ثابت معادلة انحدار y علي x فهو :

$$\begin{aligned}\hat{b} &= \bar{y} - a\bar{x} \\ &= \frac{105}{7} - (1.524) \left(\frac{77}{7}\right) = -1.769\end{aligned}$$

ومن ثم فإن معادلة خط انحدار Y علي X تأخذ الصورة الخطية التالية :

$$\hat{y} = \hat{a} + \hat{b}x = 1.764 + 1.524x$$

1- لايجاد حجم المدخرات المتوقع اي (\hat{y}) عندما يصل سعر الفائدة اي (x) الي 20 في المائة فإنه يتم التعويض عن قيمة المتغير المستقل (x) في معادلة خط الانحدار المقدرة بالقيمة (20)

اي ان :

$$\hat{y} = 10764 + 1.524(20) = 28.716 \quad (\text{million l.E})$$

2- حينما تقل اسعار الفائدة بمقدار (5) في المائة لكافة النقاط المعطاة فإنه بدلا من أيجاد المعادلة المقدرة من خلال طرح 5 في المائة من كافة قيم (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) 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 علي ثابت معادلة الانحدار أي (a) أما ميل خط الانحدار (b) فهو لم يتأثر بهذا التغير .

أخطاء التقدير :-

بعد تحديد معادلة خط انحدار Y علي X فإنه يمكن ايجاد القيم المتوقعة للمتغير (y) المختلفة وذلك كما هو موضح في الجدول التالي بالتطبيق علي المثال السابق :-

حساب القيم المتوقعة 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 :

Xi	Yi	$\hat{y}_i = -1.764 + 1.524x_i$	$e_i = y_i - \hat{y}_i$
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 العموم الثالث من الجدول السابق وذلك من خلال التعويض المباشر عن قيم (x) الموجودة في العموم الاول ومن خلال الجدول السابق لاحظ أن :

$$\sum y_i = \sum \hat{y}_i \quad ,$$

$$\sum e_i = 0$$

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 بمعنى 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 .

2- ايجاد معادلة خط انحدار Y على X للبيانات المبوبة :

لايجاد معادلة خط انحدار Y على X للبيانات المبوبة من خلال جدول تكراري مزدوج نقوم بأتباع الخطوات التالية :-

- اعتبر (x) المتغير الاول (المستقبل) حيث x هي اقسام هذا المتغير في حالة كونه متغيرا منفصلا او مراكز فئاته اذا كان متغيرا متصلا .
- احسب متوسطات المتغير الثاني (التابع) المقابلة لاقسام هذا المتغير المنفصل او مراكز فئات المتغير (x) المتصل . وهذه المتوسطات تعتبر بمثابة قيم المتغير (y) .

- اوجد معادلة انحدار Y علي X والتي تأخذ الصيغه (7 - 1) وذلك من خلال تقدير معالمها من خلال المعادلتين (6-7) و (7 - 7) السابقتين .
والمثال التالي يوضح كيفية تطبيق الخطواط السابقة :-

مثال 2 :-

فيما يلي الجدول التالي يبين توزيع 50 اسرة حسب فئات الدخل (x) والانفاق (y) الاسبوعي بالجنيه :

$x \backslash y$	60-	70-	80-	90-	100-	110	120-	Σ
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
Σ	3	8	14	12	8	4	1	50

والمطلوب : اوجد معادلة خط انحدار الانفاق (y) علي الدخل (x) . اي
معادلة خط انحدار Y علي X .

الحل :

فيما يلي حساب مراكز فئات الدخل المختلفة (x) وقيم (y) المقابلة لها .
 فمركز الفئة الاولى لـ $x = 65$ فإن قيمة y المقابلة هي :

$$Y = \frac{2(45) + 1(55)}{3} = 48.3$$

ومركز الفئة الثانية لـ $x = 75$ فإن قيمة (y) المقابلة هي :

$$y = \frac{2(45) + 3(55) + 2(65) + 1(75)}{8} = 57.5$$

• ومركز الفئة الثالثة لـ $x = 85$ فإن قيمة (y) المقابلة لها هي :

$$y = \frac{2(55) + 5(65) + 6(75) + 1(85)}{8} = 69.3$$

• ومركز الفئة الرابعة لـ $x = 95$ فإن قيمة (y) المقابلة لها هي :

$$y = \frac{1(55) + 3(65) + 4(75) + 3(85) + 1(95)}{8} = 75$$

• ومركز الفئة الخامسة لـ $x = 105$ فإن قيمة (y) المقابلة لها هي :

$$y = \frac{1(75) + 5(85) + 4(95)}{8} = 86.3$$

• ومركز الفئة السادسة لـ $x = 115$ فإن قيمة (y) المقابلة لها هي :

$$y = \frac{2(95) + 2(105)}{4} = 10$$

وأخيرا فإن مركز الفئة السابع لـ $x = 125$ فإن قيمة (y) المقابلة لها هي :

$$y = \frac{1(105)}{1} = 105$$

وبعد حساب كل من مراكز الفئات المختلفة لـ (X) وما يقابلها من قيم (Y) فإنه يتم اجراء نفس الحسابات الخاصة بمعادلة خط انحدار Y علي X للبيانات الغير مبوبة لذا يلزم الحصول علي المجاميع الموضحة بالجدول التالي :

X	Y	X Y	X ²
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

وبافتراض الصورة الخطية لنموذج خط انحدار y علي (x) هي :

$$\hat{y} = \hat{a} + \hat{b}x$$

حيث أن :

$$\hat{b} = \frac{\frac{\sum xy}{n} - \frac{\sum x}{n} \cdot \frac{\sum y}{n}}{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

اي ان :

$$\hat{b} = \frac{\frac{54154}{7} - \left(\frac{665}{7}\right) \left(\frac{541.4}{7}\right)}{\frac{388.714}{400} - \left(\frac{665}{7}\right)^2} = 0.9718$$

أما عن ثابت معادلة الانحدار فهو :-

$$\hat{a} = \bar{y} - \hat{b}\bar{x} = \frac{\sum y}{n} - \hat{b} \frac{\sum x}{n}$$

$$= \left(\frac{541.4}{7}\right) - 0.9718 \left(\frac{665}{7}\right) = -140978 \quad (l.E)$$

أي أن معادلة خط انحدار (y) على (x) تأخذ الصورة التالية :

$$\hat{y} = -14.978 + 0.9718x$$

• العلاقة بين معامل انحدار (y) على (x) وبين معامل الارتباط الخطي البسيط (r_{XY}) :

من خلال معادلة ميل خط انحدار Y على X الواردة بالمعادلة (7-6) ومعادلة معامل الارتباط الخطي البسيط لبيرسون سنورد العلاقة ما بين معامل خط الانحدار Y على X ومعامل الارتباط الخطي البسيط دون اثبات تسهيلا على القارئ : حيث أن :-

$$\hat{b} = r_{xy} \left(\frac{\sigma_y}{\sigma_x} \right) \quad (11-7)$$

هذا ومن خلال المعادلة او العلاقة (7 - 11) يمكننا استنتاج أن معامل إنحدار y على x يأخذ نفس إشارة معامل الارتباط الخطي البسيط لبيرسون بين المتغيرين حيث أن النسبة ما بين الانحرافات المعيارية للمتغيرين هي دائما ابداء موجب . هذا ويمكن اعادة كتابة المعادلة (7 - 10) لمعادلة انحدار Y على X بدلالة معامل الارتباط الخطي البسيط بين المتغيرين وذلك على النحو التالي :

فحيث ان

$$\hat{y} = \bar{y} + b(x - \bar{x})$$

لذا فإن :

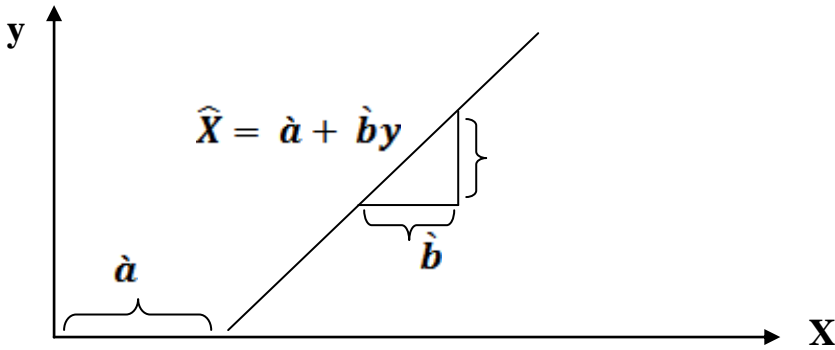
$$= \bar{y} + r_{xy} \frac{\sigma_y}{\sigma_x} (x - \bar{x})$$

ثانيا : معادلة خط انحدار (x) علي (y) :-

تاخذ معادلة خط انحدار x علي y الصورة الخطية التالية :

$$\hat{x} = \hat{a} + \hat{b}y \quad (12 - 7)$$

حيث (\hat{x}) هي عبارة عن القيم المستنتجة (المتوقعة) للمتغير (x) عند القيم المختلفة للمتغير (y) . أما \hat{b} . \hat{a} فهما ثابتان ويسميان بمعلمي Parameters خط انحدار X علي y حيث \hat{a} تسمى بثابت معادلة الانحدار وهي عبارة عن قيمة x عندما تكون قيمة y مساوية للصفر . اما \hat{b} فتسمى 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 y بـ x والعكس صحيح . ووفقا لذلك فإن (x) في معادلة خط انحدار x علي y تمثل المتغير التابع بينما تمثل (y) المتغير المستقل . وهنا سوف نتعرض بدون

أثبتت الي العلاقات المتعلقة بمعادلة خط انحدار x علي y وذلك علي نحو مماثل لما سبق عرضه في حالة خط انحدار Y علي X فخطأ التقدير في قيم (x) يمكن التعبير عنه بالعلاقة

$$e_i = x_i - \hat{x}_i \quad \text{for } i = 1, 2, \dots, n$$

ولإيجاد معادلة خط إنحدار (x) علي (y) فإنه يمكن تحديد معلمتي النموذج أي قيمتي \hat{b}, \hat{a} وذلك من خلال إستنتاج المعادلتين الطبيعييتين المتعلقةتين بخط أنحدار x علي y وذلك علي النحو التالي :

$$\left. \begin{aligned} \sum x &= na + \hat{b} \sum y \\ \sum xy &= \hat{a} \sum y^2 + \hat{b} \sum y \end{aligned} \right\} (13 - 7)$$

وهاتين المعادلتين تحققان ان مجموع مربعات اخطاء الفروق $\sum e_i^2$ يكون اصغر ما يمكن . ولعله من المفيد أن نسهل علي القارئ كيفية تذكر المعادلتين الطبيعييتين الواردتين في العلاقة (7- 13) وذلك علي النحو التالي :

حيث يمكن الحصول علي المعادلة الاولى بأخذ مجموع الطرفين للمعادلة (7-12) وذلك باعتبار ان (x) بدلالة من (\bar{x}) . وأما المعادلة الثانية فيمكن الحصول عليها من خلال ضرب طرفي المعادلة (7-12) في المتغير المستقل (Y) في هذه الحالة ثم إعتبار المجموع بعد الضرب (مع إعتبار x بدلا من \bar{x} أيضا) .

هذا ويمكن من خلال هاتين المعادلتين أستنتاج معلمتي خط الانحدار وذلك علي

$$\hat{b} = \frac{\frac{\sum xy}{n} - \frac{\sum x}{n} \cdot \frac{\sum y}{n}}{\frac{\sum y^2}{n} - \left(\frac{\sum y}{n}\right)^2} \quad \text{النحو التالي :}$$

$$\left. \begin{aligned} \hat{a} &= \bar{x} - \hat{b}\bar{y} \end{aligned} \right\} (14 - 7)$$

وبعد تقدير معلمتي خط انحدار x علي y تصبح صورة نموذج 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 :

$$(x - \bar{x}) = \hat{b}(y - \bar{y}) \quad (15-7)$$

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 تأخذ الصورة التالية :

$$\hat{b} = r_{xy} \left(\frac{\sigma_{\bar{x}}}{\sigma_y} \right) \quad (17-7)$$

ومن ثم يمكن كتابة المعادلة (16-7) علي الصورة :

$$X = \bar{x} + r_{xy} \left(\frac{\sigma_{\bar{x}}}{\sigma_y} \right) (y - \bar{y}) \quad (18-7)$$

كما أن مجموع القيم الفعلية للمتغير التابع x تساوي مجموع القيم المتوقعة \hat{x}

$$\sum x = \sum \hat{x} \quad \text{اي ان :}$$

مثال (3) :

البيانات التالية تمثل تكاليف الاعلان (x) وقيمة المبيعات (y) وذلك بالآلاف الجنيهات خلال عشر سنوات متتالية لإحدى 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 الاعلام (x) علي قيمة المبيعات (y) .
- (2): إذا بلغت قيمة المبيعات في أحد السنوات 160 الف جنيه فما هي القيمة المتوقعة للمنصرف علي الاعلام 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 الاعلام (x) علي قيمة المبيعات (y) تأخذ الصورة الخطية التالية .

$$\hat{x} = \hat{a} + \hat{b} y$$

ولتقدير معلمتي خط الانحدار a , b يتم إيجاد المجاميع الموضحة 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

$$\hat{b} = \frac{\frac{\sum xy}{n} - \frac{\sum x}{n} \cdot \frac{\sum y}{n}}{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

$$= \frac{\frac{25680}{10} - \left(\frac{180}{10}\right) \left(\frac{1370}{10}\right)}{\frac{192100}{10} - \left(\frac{1370}{10}\right)^2}$$

$$\hat{a} = \bar{X} - \hat{b} \bar{y} = \frac{102}{441} = 0.23 \left(\frac{1370}{10}\right) = \left(\frac{180}{10}\right) - 0.23 \left(\frac{1370}{10}\right)$$

ومن ثم فإن خط انحدار x علي y تأخذ الصورة التالية - 13.51

$$\hat{x} = -13.51 + 0.23y$$

(2): بعد إيجاد معادلة خط انحدار x علي y فإن يمكن حساب القيمة المتوقعة

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 فإنه من المتوقع أن يكون المنصرف علي الاعلام في هذا العام هو 23.29 الف

جنيه .

(3): الجدول التالي يوضح القيم المتوقعة لـ x وأخطاء التقدير المناظرة لها :

x_i	y_i	$\hat{x}_i = -13.51 + 0.23 y_i$	$x_i - \hat{x}_i$
10	110	$-13.51 + 0.23 (110) = 11.079$	- 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:

$$\sum x_i = \sum \hat{x}_i = 180 \quad \sum e_i = 0$$

• إيجاد معادلة خط انحدار x على y للبيانات المبوبة :

لايجاد معادلة خط انحدار x على 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 إذا كان متغيرا متصلا .
- (2): يتم حساب متوسطات قيم المتغير الثاني (التابع) المقابلة لمراكز فئات المتغير (y) أو المقابلة لأقسام هذا المتغير . وهذه المتوسطات تمثل قيم المتغير (x) .
- (3): نوجد معادلة انحدار x على y وذلك من خلال أيجاد المجاميع اللازمة لمعلمتي خط الانحدار . والمثال التالي يوضح ذلك .

مثال :-

- باستخدام التوزيع التكراري المزدوج الموضح في مثال (2) من هذا الباب اوجد معادلة خط انحدار الدخل (x) على الانفاق (y) .

الحل :

*فيما يلي حساب مراكز فئات الانفاق (y) المختلفة وقيم (x) المقابلة لها :

1-مركز الفئة الاولى لـ $y = 45$ يقابله قيمة x المقابلة لها هي:

$$x = \frac{2(65) + 2(75)}{4}$$

2-مركز الفئة الثاني لـ $y = 55$ يقابله قيمة x المقابلة لها هي :

$$X = \frac{1(65) + 3(75) + (85) + 1(95)}{7} = 79.3$$

3- مركز الفئة 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 الخامسة 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 فإنه 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 :

$$\hat{b} = \frac{\frac{\sum xy}{n} - \frac{\sum x}{n} \cdot \frac{\sum y}{n}}{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2} = \frac{\frac{50827}{7} - \left(\frac{649.2}{7}\right)\left(\frac{525}{7}\right)}{\frac{42175}{7} - \left(\frac{525}{7}\right)^2}$$

$$= \frac{305.29}{400} = 0.76$$

وكذلك فإن :

$$\hat{a} = \bar{x} - \hat{b}\bar{y} = \frac{649.2}{7} - (0.76)\left(\frac{525}{7}\right) = 92.24 - 54$$

$$= 35.74$$

وعليه تأخذ صورة معادلة خط إنحدار x علي y الصورة التالية :

$$\hat{x} = 35.74 + 0.764 y$$

أما عن العلاقة ما بين ميلي خطي الانحدار اي \hat{b}, \hat{b} ومعامل الارتباط الخطي

البسيط ليبرسون r_{xy} فحيث أن :

$$\hat{b} = r_{xy} \frac{\sigma_{\bar{x}}}{\sigma_y}$$

$$\hat{b} = r_{xy} \frac{\sigma_{\bar{x}}}{\sigma_y}$$

$$\hat{b} \hat{b} r_{xy}^2$$

لذا فإن

$$r_{xy} = \sqrt{\hat{b} \hat{b}}$$

وعليه فإن : (7 - 19)

ويأخذ معامل الارتباط الخطي البسيط لبيرسون إشارة أي من معاملي الانحدار

ومن خلال العلاقة (7 - 19) فإن يمكن القول أن :

$$\left. \begin{aligned} \hat{b} &= \frac{r_{xy}^2}{\hat{b}} \\ \hat{b} &= \frac{r_{xy}^2}{\hat{b}} \end{aligned} \right\} (7 - 20)$$

وأخيرا فإنه 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 فإن الخطان يتقاطعان عند نقطة

الايواساط الحسابية للمتغيرين أي (\bar{x}, \bar{y}) واللذان يحققان معادلتني خطي

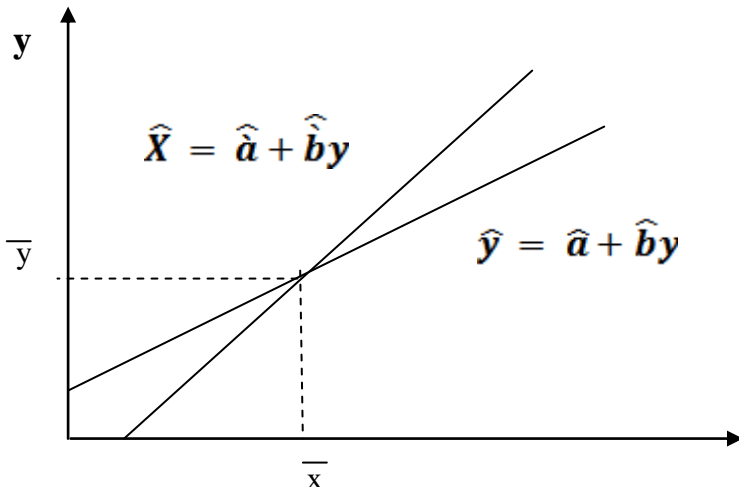
الانحدار بطبيعة الحال . لذا فإن جميع القيم المتوقعة للمتغير (y) أي (\hat{y})

تساوي نظيراتها من القيم المشاهدة (y) لهذا المتغير . وكذلك الامر بالنسبة

للمتغير x وعليه فإن 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 الفعلية 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 إحدار 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 الارتباط الخطى البسيط لبيرسون r_{xy} فكلما اقتربت هذه القيمة من الواحد الصحيح كلما زادت درجة اقتراب الخطية من بعضها والعكس صحيح . وينطبق الخطان علي بعضهما تمام الانطباق ليصبا معادله واحده وذلك في حالة وجود علاقة تامة بين المتغيرين أو الظاهرتين محل الدراسة.

تمارين الفصل السابع

(1)- إذا كانت :

$$X= 52 , y = 7 , r_{xy} = 0.4 , \sigma_x = 2 , \sigma_y = 3$$

اوجد معادلة انحدار y علي x .

(2)- في تحليل 60 مفردة وجد أن معادلتى خطى الإنحدار هي كالآتى :

$$1000y = 768x - 3608$$

$$5x = 64 + 24$$

والمطلوب تحديد قيمة معامل الارتباط من هذه البيانات .

(3)- الجدول التالي يبين الارقام القياسية للانتاج (x) والدخل القومي (y) بالملايين في الفترة من 1941 حتى 1947 في احدي الدول .

الدخل القومي (y)	الرقم القياسي (x)	السنة Year
500	100	1941
520	105	1942
580	107	1943
600	98	1944
610	109	1945
600	110	1946
650	114	1947

والمطلوب :

- 1- ايجاد معامل الارتباط بين الظاهرتين .
- 2- ايجاد معادلتى الانحدار (انحدار y علي x ، انحدار x علي y)
- 3- الحصول علي القيمة المتنبأ بها للدخل القومي عندما يكون الانتاج

4- الحصول علي القيمة المتنبأ بها للإنتاج عندما يكون الدخل القومي
700

(4)- البيانات التالية تمثل الكميات المعروضة (x) عند الاسعار المناظرة (y):

15	20	9	10	28	5	سعر السلعة (y)
153	37	155	120	409	90	الكميات المعروضة (x)
		12	18	26	12	سعر السلعة (y)
		199	250	452	130	الكميات المعروضة (x)

والمطلوب ايجاد معادلتني خطي الإنحدار .

(5)- اذا فرضنا أن لديك البيانات التالية : المتغير (x) يعبر عن أعمار الرجال المتزوجين ، (y) تمثل عدد مالديهم من أطفال . وكانت لديك البيانات التالية:

$$\sigma_y = 1.3535 , \quad \sigma_x = 5.5605 , \quad r_{xy} = 0.484 ,$$

$$\bar{y} = 2.09 , \quad \bar{x} = 33.175$$

والمطلوب

- حساب خط انحدار y علي x .
- حساب خط إنحدار x علي y .
- متوسط عدد الاطفال لرجل عمره 30 سنة .
- متوسط العمر لرجل لديه طفلان .

(6)- الجدول التالي يبين دخل مجموعة من 8 أسر في اليوم (x) وما تنفقه في خلال نفس الفترة (y) :

54	68	56	76	64	84	52	64	X
25	50	42	60	52	60	40	52	Y

والمطلوب :

- 1- معادلة انحدار الدخل علي الانفاق .
- 2- معادلة انحدار الانفاق علي الدخل .
- 3- قدر إنفاق الاسر التي يبلغ دخلها 70 جنيه في الاسبوع .
- 4- قدر دخل الأسرة التي تنفق 48 جنيه في اسبوع .

- 5- معامل الارتباط بالطريقة العادية والتحقق من النتائج عن طريق معاملي خطي الانحدار .
- 6- إيجاد معادلتى خطى الانحدار بمعلومية معامل الارتباط وقران النتائج التي حصلت عليها في (1) ، (2) .

(7) - عرف كل من الارتباط والانحدار موضحا الفرق بينهما .

- (8) - فيما يلى لديك الجدول التالى يوضح بيانات الانتاجية في العمل (Y) والخبرة بالسنوات (X) لعدد من 7 عمال :

الانتاجية y	الخبرة بالسنوات X
70	3.9
49	3.1
97	4.9
71	3.2
45	1.8
61	2.9
39	2.1

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 ؟

- (9) - الجدول التالى يمثل توزيع عينة من 500 شخص تبعا للطول (y) والوزن (x) :

مجموع	190 فاكثر	-180	-170	-160	اقل من 160	
50	-	-	15	20	15	اقل من 6
100	-	-	40	50	10	-60
100	-	10	40	50	-	-65
100	-	30	40	30	-	-70
100	20	40	40	-	-	-75
50	25	15	10	-	-	-80
						فاكثر
500	45	95	185	150	25	

والمطلوب :

- أ- احسب معامل الارتباط الخطي البسيط بين الوزن والطول في هذه البيانات.
 ب- استنتج معادلة الانحدار الخطي للطول علي الوزن بدلالة معامل الارتباط.
 ج- استخدم المعادلة التي حصلت عليها في تقدير طول أحد الاشخاص وزنه 85 كيلو جرام .
 د- استنتج معادلة إنحدار الوزن علي الطول بدلالة معامل الارتباط للبيانات الميوبة .

(10) -البيانات التالية توضح العلاقة بين اطوال وأوزان عشرة أشخاص :

Weight (x) : 66 ، 64 ، 64 ، 76 ، 69 ، 68 ، 59 ، 74 ، 62 ، 72 ،

Length (y) : 134 ، 159 ، 179 ، 168 ، 156 ، 134 ، 181 ، 149 ، 157 ، 143 ،

والمطلوب :

أ – إيجاد معادلة خط انحدار الطول علي الوزن .

ب- القيمة المتوقعة لـ 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 :

X : 78 ، 65 ، 90 ، 83 ، 67 ، 44 ، 53 ، 71 ، 82 ، 98

Y : 52 ، 45 ، 70 ، 49 ، 32 ، 28 ، 25 ، 38 ، 42 ، 53

Z : 8 ، 10 ، 5 ، 13 ، 8 ، 2 ، 12 ، 3 ، 7 ، 10

والمطلوب :

- (أ): حساب الانحراف المعياري لكل من x , y , z .
- (ب): حساب معاملات الارتباط الخطية البسيطة الآتية : r_{yz} , r_{xz} , r_{xy}
- (ج): من نتائجك (ا) ، (ب) اوجد :
- *معادلة خط انحدار حجم المدخرات الاسبوعي للأسرة علي الدخل الاسبوعي لها (أي معادلة خط انحدار z علي x) .
- *معادلة خط انحدار الانفاق الاسبوعي للأسرة على الطعام علي الدخل الاسبوعي (أي معادلة خط انحدار y علي x) .
- *بين ايهما أكثر تأثراً بالدخل الاسبوعي ، الانفاق الاسبوعي علي الطعام أم حجم المدخرات الاسبوعي .
- (د): من نتائجك في (ج) احسب ما يلي :
- *أخطاء التقدير في حساب كل من y , z وذلك عند القيم التالية لـ x : 44 ، 82 ، 20 ،
- *احسب القيم المتوقعه لكل من y , z وذلك عند اليم التالية لـ x : 75 ، 40 ، 120

(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 (بالآلاف الجنيهات) وذلك لإحدى الشركات في ثماني شهور متتالية :

الاعلان (x) : 25 20 35 53 46 20 51 28

المبيعات (y) : 108 127 175 150 113 136 130 104

والمطلوب :

(أ): إيجاد معادلتني خطي إنحدار y علي x وإنحدار 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 لاجاد معادلتني خط

انحدار y 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

$$r_{xy} = 0.72 \quad \frac{\Sigma y}{\Sigma x} = 2.5$$

(15) : اثبت أن معامل الارتباط الخطي بين x ، y (r_{xy}) يمكن كتابته علي الصورة التالية

$$r_{xy} = b \frac{\sigma_y}{\sigma_x}$$

حيث b هي معامل انحدار y علي x ، σ_x ، σ_y 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 :

125	115	105	95	y X
2	5	7	4	52.5
4	7	10	6	57.5
7	10	12	6	62.5
3	6	8	3	67.5

(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 فيما بين المتغيرين x ، y .
 ب- إيجاد معادلة انحدار y علي x ثم حساب خطأ تقدير y عند القيمة $15=x$
 ج- 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 مكونة 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 المتوقعة لـ (y) وذلك باستخدام المعادلة المحسوبة في (أ).

ج- علي ضوء نتائجك في (أ) ، (ب) إستنتج – دون حساب – معامل الارتباط الخطي البسيط x ، y .

د- من نتائجك في أ ، ب ، ج استنتج أي اي من المتغيرين x ، y هو الأكبر 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 مجموعة من الطلاب في مادة الاحصاء (x) ودرجاتهم في مادة المحاسبة (y) ، وكانت معادلتا خط انحدار y علي x و x علي y النحو التالي

$$Y = -5.9 + 1.35x$$

$$X = 4.51 + 0.73y$$

المطلوب :

أ – احسب معامل الارتباط الخطي بين x ، y .

ب – مستعينا بنتائجك في (أ) ، حدد أيهما الأكبر من حيث التشتت النسبي درجات الاحصاء (x) أم درجات المحاسبة (y) ؟

ج – إذا حصل طالب معين من بين مجموعة 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 (ج) 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 الأزواج موزعين حسب عمر الزوج (x) وعمر الزوجه (y) :

X \ Y	25-	35-	45-	55-	65-75	Σ
25-	5	4	3			12
35-	3	6	5			14
45-			6	8	2	16
55-65				2	6	8
Σ	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 40 سنة.
- ج- 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 "

أولاً: مراجع باللغة العربية :

- 1-د. إبراهيم موسى عبدالفتاح "مبادئ الإحصاء الوصفي والتحليلي" المكتبة العلمية بالزقازيق ، 2005 .
- 2-د. أبوبكر عبدالرحمن عبدالمتعال "مبادئ الإحصاء" ،كلية التجارة ،جامعة جنوب الوادي 2008 .
- 3-د. أماني موسى محمد ، د. محمد محمود خطاب "الإحصاء في: إتخاذ القرارات- مدخل حديث" ،كلية التجارة بسوهاج، جامعة سوهاج ، 2008 .
- 4-د. صلاح مهدي محمد " الإحصاء " ،كلية التجارة ، جامعة عين شمس ، 2008 .
- 5-د. محمد جلال الدين أبو الذهب وآخرون " مقدمة في الإحصاء للتجاربيين- 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 في مبادئ الإحصاء"، كلية الإقتصاد والعلوم السياسية، جامعة القاهرة ، 1977 .

ثانياً: 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.