جامعة يحي فارس بالمدية

كلية العلوم الاقتصادية والعلوم التجارية وعلوم التسيير

مطبوعة جامعية بعنوان:

محاضرات في الإحصاء الوصفي

مطبوعة موجهة لطلبة السنة الأولى ل م د لميدان العلوم الاقتصادية

والعلوم التجارية وعلوم التسيير

إعداد

الدكتور: حمداني محي الدين

السنة الجامعية2016-2017

فهرس المحتويات

الفصل الأول:المفاهيم الاساسية لعلم الاحصاء

1.1 :مفهوم الاحصاء.....7

2.1.1.الفرق بين الاحصاء والاحصائيات.....7

2.1. مصطلحات الإحصائية 9

1.2.1.المجتمع الاحصائي 9

2.2.1 العينة الاحصائية 9

1. 2.2.2 أنواع العينات.....10

3.2.1. الوحدة الاحصائسة: 11

4.2.1.المتغيرات الاحصائية 12

3.1.المنهج الاحصائي 12

1.3.1.جمع البيانات 12

2.3.1. مصادر البيانات 13

3.3.1. أساليب جمع البيانات 13

4.3.1.تصنيف البيانات 13

5.3.1.عرض البيانات 14

1.5.3.1.العرض الجدولي 14

2.5.3.1.العرض البياني 21

الفصل الثاني: مقاييس المتغيرات الاحصائية.....34

1.2.مقاييس النزعة المركزية 35

1.2.2.متوسطات القياس.....36

1.1.2.2.المتوسط الحسابي..... 36

2.1.2.2 المتوسط الهندسي 44

3.1.2.2 المتوسط التوافقي.....47

4.1.2.2 المتوسط التربيعي.....50

2.2.2.متوسطات الوضع.....52

1.2.2.2.المنوال.....52

3.2.متوسطات التكرار.....55

1.3.2.الوسيط.....55

.2.3.2.شبيهلت الوسيط.....60

1.2.3.2.الربيعيات.....61

2.3.2..2.العشيريات..... 63

2.3.2..3.المئينيات.....65

الفصل الثالث:مقاييس التشتت والشكل.....84

1.3.**مقاييس التشتت**.....84

1.1.3.مقاييس التشتت المطلق.....85

1.1.1.3.المدى 85

2.1.1.3.المدى الربيعي.....86

3.1.1.3.المدى العشيري.....88

4.1.1.3.المدى المئيني.....88

5.1.1.3.الانحراف المتوسط.....89

6.1.1.3.التباين والانحراف المعياري.....94

2.1.3.مقاييس التشتت النسبي.....111

1.2.1.3.المدى النسبي.....111

2.2.1.3.الانحراف الربيعي النسبي.....112

3.2.1.3.الانحراف المعياري النسبي.....112

2.3.مقاييس الشكل.....116

1.2.3.مقاييس الالتواء.....116

1.1.2.3.معاملالالتواء الاول لبيرسون.....118

2.1.2.3.معاملالالتواء الثاني لبيرسون.....119

3.1.2.3. معامل الالتواء الربيعي 121

4.1.2.3.معامل الالتواءلكندال.....123

2.2.3.مقاييس التفرطح.....123

1.2.2.3.معامل التفرطح لكيلي 124

2.2.2.3. معامل التفرطح لبيرسون.....124

3.2.2.3.معامل التفرطح لفيشر.....125

الفصل الرابع: الارقام القياسية.....129

4-1.مفهوم الرقم القياسي.....129

2.4.أنواع الارقام القياسية.....129

3.4:الأرقام القياسية المرجحة.....133

4.4: خصائص الارقام القياسية.....136

الفصل الخامس:الارتبط والانحدار الخطي 142

5-1: الارتباط الخطي البسيط.....144

2.5":معامل الارتباط.....145

3.5:نموذج الانحدار الخطي.....153

المراجع.....160

تقديم:

هذه المطبوعة هي عبارة عن محاضرات في الاحصاء الوصفي موجهة لطلبة السنة الاولى جذع مشترك مستوى الليسانس في ميدان العلوم الاقتصادية والعلوم التجارية وعلوم التسيير، تتناول وبأسلوب سهل وبسيط النسبة الاعظم من مقياس الاحصاء الوصفي مدعمة بأمثلة بسيطة مرتبطة بكثير من النقاط المقررة في المقياس مضاف إليها مجموعة تمارين للحل في كل فصل .و قد تم تقسيم المطبوعة إلى خمسة فصول، يتضمن الفصل الاول مفاهيم أساسية في علم الاحصاء وكيفية عرض البيانات وتحليلها، و تناول الفصل الثاني مقاييس النزعة المركزية و تعرض الفصل الثالث إلى مقاييس التشتتوالشكل بينما تناول الفصل الرابع إلى الأرقام القياسية، بينما تناول الفصل الخامس والاخير تحليل الارتباط و الانحدار الخطي البسيط.

الفصل الأول:المفاهيم الاساسية لعلم الاحصاء Basic concepts of statistics

1-1:مفهوم الإحصاءThe concept of statistics :هناك تعاريف عديدة للإحصاء،فقديماً عرِّف الإحصاء على انه مجرد جمع البيانات وتبويبها في جداول أو إبرازها في رسومٍ بيانية أو أشكال تصويرية، لكن هذا التعريف بدأ يتلاشى تدريجياً فاسحاً المجال للتعريف الحديث ، الذي يعرف علم الإحصاء على أنه ذلك العلم الذي يبحث في جمع البيانات وتنظيمها وعرضها وتحليلها واستقراء النتائج ، واتخاذ القرارات بناءاً عليها، وغن ذلك يقودنا إلى طرح السؤال التالي:

1-1-1:ما الفرق بين الإحصاء والإحصائيات؟The difference between the census and statistics

عند تصفحنا للقاموس نجد تعريفين بارزين أحدها يتعلق بالإحصاء (في صيغة المفرد)، والآخر بالإحصائيات (في صيغة الجمع).

أ. **الإحصائيات** تعرف بكونها عبارة عن مجموع المعلومات العددية الناتجة عن تعداد السكان، ومعطيات سجلات الأحوال المدنية والدراسات الميدانية الخاصة بها، أي أن مصطلح الإحصائيات يتعلق بالمعطيات الرقمية العددية التي تجمعها الهيئات المختصة.

ب. **الإحصاء statistics** فهو أحد علوم الرياضيات التطبيقية، الذي يختص باستحداث أكثر الطرق فعالية في جمع وتحليل وتفسير البيانات، بغرض استنتاج الحقائق منها على شكل عبارات مبنية على الاحتمال.

ويعرف أيضاً على انه العلم الذي يختص بجمع وعرض وتحليل البيانات الرقمية لعمل استدلالات، واتخاذ قرارات في ظل عدم التأكد، وذلك في مجال الاقتصاد والإعمال والعلوم الاجتماعية والطبيعية الأخرى، فهو مزيج من العمليات الحسابية والمنطق.

فالاحصاء قد بدأ كعلم وصف بحت، ثمَّ تطوَّر إلى اداة قوية لإتخاذ القرارات مع نمو الاستدلال فيه. وبناءًا على ذلك يمكن تقسيم الاحصاء إلى:

1-1-2:أنواع الإحصاء:Types of statistics

أ. **الإحصاء الوصفي:Descriptive statistics :** ويعرف بأنه طريقة وصف كمية، تعتمد على جمع البيانات وتبويبها وعرضها في صورة جداول، ثم تلخيصها بغرض وصف مجتمع ما أي يهدف إلى إعطاء وصف بسيط بواسطة عدد محدد من المقاييس، ومن وسائل الإحصاء الوصفي المتمثلة في:العرض البياني، الدراسة الرياضية.

يعرف الإحصاء على أنه العلم الذي يدرس مختلف طرق ووسائل جمع البيانات الكمية عن مختلف الظواهر الاقتصادية والاجتماعية وغيرها، وترتيب هذه البيانات وتبويبها وتحليلها وتفسيرها وتقديمها بأشكال وصور ملائمة بهدف تسهيل اتخاذ القرار على أساس سليم

عبارة من مجموعة الأساليب الإحصائية التي تعنى بجمع البيانات وتنظيمها وتصنيفها وتلخيصها وعرضها بطريقة واضحة في صورة جداول أو أشكال بيانية وحساب المقاييس الإحصائية المختلفة لوصف متغير ما (أو أكثر من متغير) في مجتمع ما أو عينه منه.

ب. الإحصاء الاستدلالي Inferential statistics

عبارة عن مجموعة من الأساليب الإحصائية التي تستخدم بغرض تحليل بيانات ظاهرة (أو أكثر) في مجتمع ما على أساس بيانات عينة احتمالية تسحب منه وتفسيرها للتوصل إلى التنبؤ واتخاذ القرارات المناسبة.

ويتلخص الأسلوب الإحصائي في الخطوات التالية:

1. جمع البيانات عن طريق التجربة والمشاهدة بوفرة كافية لاستخلاص النتائج منها.
2. عرض هذه البيانات بطريقة تساعد على تفهمها والاستفادة منها حيث أن البيانات الإحصائية في صورتها الأولية لا يمكن الاستفادة أو استخلاص النتائج منها وذلك في حالة وجود عدد كبير من الأرقام أو الصفات.

ومن وسائل الإحصاء التحليلي أو الاستدلالي:

1. التقدير:ويقصد بذلك القيام بتقدير أو حساب قيم تحل محل القيم الحقيقية أو الفعلية(والتي تدعى بالقيم المقدرة أو المحسوبة) التي تمثل الظاهر المختلفة في المجتمع، وذلك لأن تلك القيم الحقيقية قد يكون نمن المتعذر إن لم نقل من المستحيل حسابها.
2. اختبار الفرضيات:وهذا يعني استخدام البيانات الإحصائية التي تجمع في دراسة ما، للوصول إلى قرار بشأن الفرض أو الفرضيات التي افترضت في البداية كتفسير مؤقت لظاهرة ما نحن مهتمون بدراستها، وهذا القرار إما يكون بالقبول أو بالرفض.

ويعتمد الإحصاء الرياضي على نظرية الاحتمالات والتوزيعات الاحتمالية.

1-1-3:مجالات تطبيق الإحصاءFields of application of statistics : يعتبر علم الإحصاء من العلوم الأكثر أهمية و فائدة في كل المجالات، فقد اصبح في وقتنا الحاضر كثير الاستعمال في مختلف المجالات كالصناعة والزراعة والخدمات، مما جعله يُدَرَّسُ في كل الجامعات والمعاهد، وفي القطاعات المختلفة.

1-2:مصطلحات إحصائية:Statistical terms

1-2-1:المجتمع الإحصائيPopulation : ويتشكَّلُ من مجموع الوحدات أو العناصر الإحصائية التي تشترك في صفة أو أكثر، نستطيع من خلالها تحديد عناصر المجتمع الإحصائي.ويمكن أن ندرس المجتمع الإحصائي بصفة واحدة أو أكثر. يمكن أن يكون المجتمع الإحصائي:

1: محدود: عند ما يكون من الممكن حصر مفرداته مثل عدد الطلبة الجامعيين في الجزائر عدد سكان ولايتك، عدد رؤوس الأغنام في السودان،...الخ.

2.غير محدود: عند ما يكون من الصعوبة حصر عدد مفرداتنه مثل عدد النجوم في الفضاء، عدد البكتيريا في كمية من الحليب، عدد النقاط التي تشكل مستقيم،..الخ

فالمجتمع هو عبارة عن مجموعة من العينات، التي تحتوي بدورها مجموعة من الوحدات معروفة ومعرَّفة بصفة واضحة لكي يمكن تمييزها عن الوحدات التي تشكِّل مجتمعاً آخر. فالوحدة في الإحصاء تقابل العنصر في الرياضيات، والعينة تقابل المجموعة الجزئية، والمجتمع يقابل المجموعة الكلية في الرياضيات.

1-2-2:**العينة الاحصائيةStatistical sample** :وهي جزء من المجتمع الإحصائي أُخِذَت منه لتمثيله،لكنها ليست أي جزء، بل الجزء الذي يمثله أحسَن تمثيل. وحتى تكون كذلك يجب أن تكون:

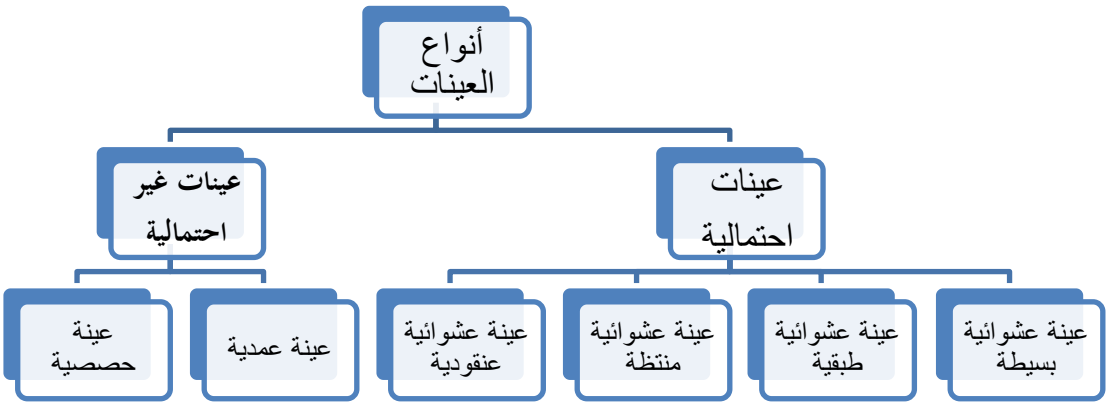
- مختارة بطريقة عشوائية (لكل عنصر من المجتمع نفس الفرصة في أن يكون ضمنها).
- يتناسب حجمها وحجم المجتمع الإحصائي.

غير أن حجم العينة يختلف حسب أهمية الدراسة الإمكانيات المادية والبشرية المتاحة للقيام بالدراسة، فقد تتكون من جميع عناصر المجتمع الإحصائي في حالة المجتمع المحدود.وهي الطريقة العملية لدراسة المجتمعات الإحصائية في حالة المجتمعات الكبيرة. ويتوقف نجاح استخدام أسلوب المعاينة على عدة عوامل هي:

1. كيفية تحديد حجم العينة . 2.طريقة اختيار مفردات العينة . 3. نوع العينة المختارة.

و يمكن تقسيم العينات وفقا لأسلوب اختيارها إلى نوعين كما هو موضح في الشكل اسفله

1-2-2-1:أنواع لعينات Types of samples :



أ.**العينات الاحتمالية أو العشوائية the random sample** هي العينات التي يتم اختيار مفرداتها بطريقة عشوائية، بحيث يكون لكل عنصر فرصة او احتمال ان يتواجد فيها. وأهم انواعها:

1)العينة العشوائية البسيطة:the simple random sample تختار هذه العينة من **المجتمع الاحصائي** المتجانس، أي ان جميع عناصره متماثلة كاختيار عينة من الطلاب كلية الاقتصاد بجامعة المدية

2)العينة العشوائية الطبقية the Stratified random sample: وتستخدم عند ما يكون المجتمع غير متجانس أي مقسما الى مجموعات بحيث تتقاسم كل مجموعة خصائص مشتركة، بحيث نسحب عينة بسيطة من كل مجموعة يتناسب حجمها وحجم عدد افراد المجموعة أو الطبقة، وجموع هذه العينات البسيطة يشكل العينة الكلية

مثال: يراد اختيار عينة مكونة من 40 طالب من طلبة كلية الاقتصاد التي تضم 3400 طالب وهم مقسمين على النحو التالي: 1600 طالب سنة اولى، 1600 طالب سنة ثانية، 800 طالب سنة ثالثة،

$$\text{عدد الطلبة المسحوب من كل سنة} = \frac{\text{حجم العينة} * \text{عدد طلبة السنة}}{\text{العدد الاجمالي للطلبة بالكلية}}$$

ومنه فعدد الطلبة المسحوب من السنة:

$$\text{الاولى:} \frac{1600*40}{4000} = 16 \quad ، \quad \text{الثانية} = \frac{1600*40}{4000} = 16 \quad \text{الثالثة} = \frac{800*40}{4000} = 8$$

3)العينة العشوائية المنتظمةthe Systematic random sample ويشترط فيها ترقيم عناصر المجتمع من **01** الى **N** على أن نعتمد نظاماً محدداً في السحب منها

4)العينة العشوائية العنقودية The cluster random sample : هي عينة يتم تكوينها باتباع عدة مراحل تبدأ في تقسيم المجتمع الى مجموعة جزئية ثم نختار من هذه الاخيرة واحدة بطريقة عشوائية ونقسمها ايضا الى مجموعات جزئية نختار منها ايضا واحدة بطريقة عشوائية وهكذا الى ان نصل الى اصغر مجموعة جزئية والتي نختار منها عناصر العينة المطلوبة **بطريقة عشوائية**.

مثال: سحب عينة وطنية من تلاميذ السنة الثالثة ابتدائي

الحل: العينة يجب ان تكون عنقودية لان هناك :48 ولاية ، في كل ولاية مجموعة من الدوائر، في كل دائرة مجموعة من البلديات، في كل بلدية مجموعة من المدارس، وفي كل مدرسة عدد من أقسام السنة الثالث.

ب.العينات غير الاحتمالية(غير العشوائية)The non-random sample: هي التي يتم اختيار مفرداتها بطريقة غير عشوائية، حيث يقوم الباحث باختيار مفردات العينة بالصورة التي تحقق الهدف وأهم انواع العينات غير الاحتمالية مايلي:

1)**العينة الحصصية**: نختار العينة ليس عشوائيا وإنما بطريقة متعمدة ويشترط فيها الحصص المطلوبة: **مثلا:** ان يختار مدير مدرسة عينة من 9 تلاميذ لإجراء مسابقة ويشترط ان تتكون من 05 اناث و04 ذكور.

2)**العينة العمدية(القصدية)**: كأن نختار عينة من الطلبة لتمثيل الثانوية في مسابقة ما بين الثانويات

3)عينة الصدفة: تتكون من عناصر يتم مقابلتها بالصدفة، **مثلا:** اختيار اشخاص في محل تجاري.

1-2-3:الوحدة الإحصائيةStatistical Unit) :هي العنصر الأول في موضوع الملاحظة.يمكن أن تكون فرداً أو شيئاً ما، حيث يكون هذا الفرد أو الشيء يحمل صفة أو أكثر تحدد ضمن أهداف الباحث، والهدف من جمع البيانات، ففي تعداد السكان يمكن أن نعتبر الأسرة هي الوحدة الإحصائية وفي دراسة عدد الأميين يمكن أن نعتبر الأمي هو الوحدة الإحصائية.والشرط الوحيد في الوحدة الإحصائية أن تكون قابلة للتعريف بشكلٍ دقيق.

1. **البياناتData**: هي مجموعة القيم (المشاهدات أو الملاحظات) التي تُجمع من الوحدة الاحصائية أثناء دراسة معينة، و قد تكون البيانات رقمية (كمية) مثل الأجر الشهري لمجموعة من العمال أو عدد الاطفال لمجموعة من الأسر....، و قد تكون البيانات وصفية (كيفية) مثل المستوى التعليمي لمجموعة من الموظفين في مؤسسة ما ،أو الحالة الاجتماعية لمجموعة من الأشخاص و غيرها

1-2-4:المتغيرات الإحصائية(الصفات الإحصائية المتغيرة)statistical variables :وهي الصفة التي يهتم بها الباحث في الوحدة الإحصائية موضوع الدراسة، والقابلة للتغير من فردٍ لآخر في المجتمع الإحصائي والتي نميز فيها نوعان:

أ. **متغيرات إحصائية نوعيةQualitative variables**: وهي المتغيرات التي لا يمكن قياسها أو غير القابلة للقياس مثل المهنة، الجنسية، الحالة العائلية، مستوى التعليم

ب. **متغيرات إحصائية كمية Quntitative variableSوهي** الخصائص أو المتغيرات التي يمكن قياسها، وهي أكثر المتغيرات انتشاراً واستعمالاً باعتبار أن لغة الإحصاء هي لغة الأرقام، مثل الإنتاج، الاستهلاك، الوزن، الطول .....الخ.وتنقسم بدورها إلى قسمين:

1. متغيرات كمية منقطعة(**discrete variables**):وهي المتغيرات التي تأخذ قيَّماً متباعدة، غالباً ما تكون أعداداً طبيعيةً 1 2 3 4 .....الخ.أي أن المتغير لا يمكن أن يأخذ أي قيمة بين قيمتين مثل عدد أطفال أسرة، عدد عمال مصنع، عدد السيارات المنتجة في مصنع.....الخ.
2. متغيرات كمية متصلة**continuous variables**):وهي الصفات أو المتغيرات التي تأخذ أي قيمة في مجالٍ معيَّنٍ مثل أطوال طلبة هذا التخصص وأعمارهم وأوزانهم.

بصفة عامة، يعتبر المتغير مستمراً إذا كان مرتبطاً بالزمن(السن، السرعة)والكتلة(الوزن، الكثافة)،النقود(السعر، الدخل)، الفضاء(الطول، المساحة).

1-3:**المنهج الاحصائي:Statistical method** ويتمثل في مجموعة الاساليب المستخدمة من أجل جمع وتبويب وتحليل وتفسير البيانات بغرض الوصول إلى نتائج. وتتلخَّصُ فيما يلي:

1-3-1:جمع البيانات Data collection وهي أول خطوة يجب أن يقوم بها الباحث لأنها أساس والدراسة وتتمثل في الحصول على القياسات أو التعدادات أو قيم المشاهدات للتجارب التي يقوم بها الباحث، وكلما كان جمع البيانات دقيقاً كلما زادت ثقة الدارس في الاعتماد عليها في الوصول إلى نتائج موثوقة وصحيحة، وهو ما تفتقد إليه تلك النتائج في حالة ما تكون هناك أخطاء في جمع البيانات. وللحصول على مثل هذه البيانات هناك أساليب وطرق متعددة نتبعها في عملية الجمع تلك والتي تختلف باختلاف البحث المراد انجازه من جهة، وباختيار الباحث في أن يقوم هو بنفسه بعملية الجمع أو أنه يعتمد على البيانات المنشورة سابقاً من جهة أحرى وباختلاف اختيار الباحث في أن يدرس المجتمع الاحصائي كله أو يكتفي بعينة تمثل هذا المجتمع قدر الامكان من جهة أحرى وضرورة انجاز الباحث لبحثه في الآجال المحددة.

ولكن عموماً ومهما اختلفت الأسباب.

1-3-2:مصادر البيانات Data Sources :هناك مصدرين أساسيين للمعلومات هما:

أ. **المصدر المباشر(الطريقة المباشرة)**:وتتمثل في الحصول على البيانات والمعلومات الاحصائية ميدانياً، من خلال الاتصال بعنصر الدراسة مباشرة، حيث يستقي الباحث معلوماته بملاحظته للظاهرة بنفسه،أي القيام باتصالاته مباشرة بهذه الوحدات الاحصائية.إلا أن هذه الطريقة تكون مكلفة وصعبة لأنها تتطلب الكثير من الجهود والأموال، وهو ما قد نكون لا نتوفر عليه، كما أنها تتطلب وقتاً قد يكون كبيراً.

ب. **المصدر غير المباشر(الطريقة غير المباشرة)**:وتتمثل في الحصول على المعلومات الاحصائية من المصادر غير الميدانية، من خلال الاتصال بأشخاص أو هيئات لهم اتصال أقوى من اتصالاته بهذه ألظاهرة مثل : ألمؤسسات ألمراكز الدواوين الوزارات ألمجلات ألدوريات التي تنشرها الأجهزة الرسمية.غير أن هذه الطريقة تعتريها الكثير من السلبيات والنقائص لأن الاحصائيات تكون قد جمعت لغير غرض ألباحث فتكون المعطيات تشوبها النقائص.

1-3-3:أساليب جمع البيانات Data collection methods: بشكل عام هناك أسلوبين لجمع البيانات هما:

أولاً: أسلوب الحصر الشامل Comprehensive inventory method: وفيه تجمع البيانات عن كل مفرده من مفردات المجتمع، وهذا الأسلوب لا يتبع عادة إلا في حالة التعدادات التي تجريها الدول وتدعمها بإمكانيات ضخمه، مثل تعدادات السكان والتعدادات الصناعية والتعدادات الزراعية.

ثانيا: أسلوب المعاينة sampling method: وفيه يتم جمع البيانات عن جزء من مفردات المجتمع يختار بطريقة علمية محددة ويطلق عليه العينة، ثم بعد ذلك يتم تعميم نتائج الدراسة على المجتمع بأكمله.

1-3-4:تصنيف وتبويب البيانات Classification and tabulation of data: وهي العملية التي تأتي بعد جمع البيانات، وتتمثل في تجميع الوحدات الاحصائية في فئات أو مجموعات معينة، طبقاً للبيانات المطلوب تبويبها، وللكيفية التي على أساسها سوف تستخدم بعد تبويبها، لأن المعطيات الخام كثيراً ما تكون فوضوية.لذلك يجب ترتيبها وتصنيفها وتبويبها ليسهل أخذ فكرة عنها، لأن البيانات الخام تكون في صورة غير معبره ويصعب استنتاج معلوماتٍ منها.وتتم عملية التبويب عادة بوضع هذه البيانات في جداول منسقة وعرضها بطرق مناسبة كالأشكال الهندسية والرسوم البيانية والتوزيعات التكرارية. ومن أهم

أساليب التبويب والتصنيف تعتمد غالباً على التصنيف والتبويب:الزمني، الجغرافي ، الكمي ،الكيفي أو النوعي، العشوائي،الخلط أي أكثر من تصنيف واحد من هذه التصنيفات.

ولكي يكون الجدول قابلاً للأستعمال يجب أنم يتضمن العديد من العناصر الأساسية التي سنذكرها لاحقاً.

1-3-5: عرض البيانات

بعد الحصول على المعلومات الاحصائية وتبويبها، لابد من عرضها وفق صفات وخصائص معينة لإبراز الصفات الرئيسية لها، إذ هناك طرق عديدة لعرض الإحصائيات، منها:

العرض الجدولي: ويتخذ عدة أشكال منها:

أ. **طريقة العرض الكتابي** وتتمثل هذه الطريقة في أن يقوم الباحث بعرض المعلومات الاحصائية ضمن نصٍ كتابي.فإذا كنَّا نتكلم عن تغير البطالة في الجزائر خلال الفترة من سنة 1967 إلى سنة 2008 نضع:لقد عرفت نسبة البطالة في الجزائر تغيرات ملحوظة، حيث انتقلت من 33 بالمائة سنة 1967 لتنخفض إلى 17 بالمائة سنة 1987 لترتفع من جديد وتصل إلى 33 بالمائة سنة 1998، لتتراجع 27 يبالمئة سنة 2001 لتنخفض إلى 11.3 بالمائة سنة 2008.

وتعتبر هذه الطريقة مفيدة عند ما يكون لدينا أرقام قليلة، غير انها تصبح غير مفيدة في حالة ماإذا كانت الأرقام كثيرة، حيث تصبح مملة ويصعب على السامع أو القاريء استيعاب المحتوى المعبر عنه كتابياً.

ب. **طريقة العرض نصف الجدولي**:وتتمثل هذه الطريقة في ذكر الأرقام في مكانٍ بارز في الصفحة بشكلٍ متسلسل ومنطقي.فإذا كنَّا نتكلم عن تغير البطالة في الجزائر خلال الفترة 1997-2008 نكتب:

معدل البطالة في الجزائر كان مساوياً لـ33 بالمائة سنة 1967

معدل البطالة في الجزائر كان مساوياً لـ17 بالمائة سنة 1987

معدل البطالة في الجزائر كان مساوياً لـ33 بالمائة سنة 1998

معدل البطالة في الجزائر كان مساوياً لـ27 بالمائة سنة 2001

معدل البطالة في الجزائر كان مساوياً لـ11.3 بالمائة سنة 2008

وتعتبر هذه الطريقة مفيدة عند ما يكون لدينا أرقام قليلة، غير انها تصبح غير مفيدة في حالة ماإذا كانت الأرقام كثيرة، حيث تصبح مملة ويصعب على السامع أو القاريء استيعاب المحتوى المعبر عنه كتابياً.

ت. **طريقة العرض الجدولي:** وتتمثل هذه الطريقة في وضع البيانات الاحصائية في جدول. وغالباً ما تستعمل في عرض تغيُّر الظاهرة مع الزمن أو متغيرات أخرى أو مع كلاهما.فإذا كنَّا نتكلم عن تغيُّر معدل البطالة في الجزائر خلال الفترة 1967-2008، نكتب:

جدول رقم (1-1)تطوُّر معدل البطالة في الجزائرخلال الفترة (1967-2008)

السنوات	1967	1985	1987	1989	1992	1993	1994	1995	1996	1997	1998
معدل البطالة %	33	16,5	17	19	21,5	24,4	27	28	28	28,2	33
السنوات	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	
معدل البطالة %	29,29	20,77	27,3	25,9	23,7	17,7	15,3	12,3	11,8	11,3	

ملاحظة:الارقام الوارد في الجدول أعلاه فعلية وليست افتراضية

المصدر

Nacer-eddine Sadi Nacer-eddine Sadi, La privatisation des entreprises publique En Algérie, Opu, Alger, 2ème édition, 2006, p50. + CNES, 2000 - .ONS . Institut de la méditerranée, France: profil pays Algérie, Janvier 2006, p2

وتعتبر طريقة العرض الجدولي أو استخدام طريقة الجداول الإحصائية هي أسهل وأمثل طريقة لعرض البيانات، نظراً لما تسمح به من تلخيص وتبويب لها. وحتى يكون الجدول قابل للإستعمال يجب أن يتضمن العناصر التالية:

- رقم الجدول.
- عنوان الجدول الذي يجب أن يكون واضحاً وملائماً، يعكس مضمون الجدول باختصار
- وحدة القياس(الوحدة المستعملة).
- عناوين الأسطر والأعمدة التي تدل على محتوى الجدول، والتي يجب أن تكون غاية في الاختصار.
- المصدر، الذي يجب أن يورد بشكلٍ دقيق ، والذي يشار إليه عادة في أسفل الجدول بعد الشروح(إن وجدت) للعودة إليه إن استدعت الضرورة.
- الشروح العامة والخاصة التي تساعد على تفسير المعطيات العددية أو الملاحظات، والتي توضع في اسفل الجدول، كما هو موضح في الجدول أعلاه.

وإذا قمنا بإعداد جدول يتضمن كل قيم المتغير الإحصائي فسيكون هذا الجدول طويلاً جداً خصوصاً إذا كان لدينا سلسلة طويلة من القيم التي تحويها الظاهرة المدروسة.لذلك وردت فكرة التكثيف، حيث يتم تقسيم البيانات إلى فئاتٍ أو قيم، ثم حصر عدد البيانات الواقعة ضمن كل فئة أو المقابلة لأي قيمة.وتتم عملية بناء الجدول الاحصائي بالمراحل التالية:

بناء التوزيع التكراري:

إذا كان عدد البيانات أو مداها كبيراً، فإنه يكون من الضروري تقسيم تلك البيانات إلى فئات، والتي يجبذ أن يتراوح عددها بين 10 و 15 فئة، يجبذ أيضاً أن تكون متساوية الطول، ومنفصلة عن بعضها البعض أي غير متداخلة، وكافية لإحتواء جميع البيانات.

تتلخص خطوات انشاء جدول توزيع تكراري فيما يلي:

- تحديد أو حساب مدى البيانات(Entendue) أو (Rang) أو المجال العام للبيانات، والذي نحصل عليه بايجاد الفرق بين أكبر وأصغر قيمة في البيانات $R=X_{max}-X_{min}=E$
- تحديد عدد الفئات طبقاً للقوانين المعروفة لتحديدها ، والتي منها:

أ. **معادلة Sturges** والمعطاة بالشكل $K=1+3,322\log(N)$

حيث: k: عدد الفئات، N: عدد القيم، $\log(.)$: اللوغاريتم العشري

ب. **معادلة يول (Yule):** التي تحدد الفئات وفق العلاقة التالية: $K=2,5\sqrt[4]{n}$

حيث n هو عدد البيانات

ت. أو طبقاً لما يرى الباحث أنه مفيد له للوصول غلى النتائج التي يرغب فيها.

- تحديد طول الفئة، وذلك بقسمة المدى على عدد الفئات مع التقريب إلى الأعلى.
- تعيين الحد الأدنى للفئة الأولى.
- تعيين الحد الأعلى للفئة الأولى، وذلك باضافة طول الفئة للحد الأعلى.
- تعيين الحدود الدنيا والعليا لبقية الفئات بنفس الطريقة، مع اعتبار الحد الأعلى للفئة هو نفسه الحد الأدنى للفئة التي تليهان وهكذا.
- تعيين مراكز الفئات،والذي نحصل عليه بقسمة مجموع حدي الفئة على 2 أو باضافة نصف المدي إلى الحد الأدنى للفئة.

- تفريغ البيانات على الفئات.
- تسجيل مجموع المشاهدات(تكرار كل فئة) أمامه في عمود التكرارات (ni)
- تجميع التكرارات لجميع البيانات،ومقارنة ذلك بعدد البيانات (المشاهدات) حيث يكون $N=\sum_{i=1}^{n} ni$

ملاحظة:الحد الأدنى للفئة الأولى يجب أن يكون مساويا أو اقل من أصغر قيمة موجودة في البيانات المعطاة، كما أن الحد العلى للفئة الأخيرة يجب أن يكون مساوياً أو أكبر من أكبر قيمة في البيانات.

ملاحظات:

- يمكن أن تكون الفئات غير متساوية.
- ليس هناك طريقة محددة لتحديد العدد المناسب للفئات.
- الطرق الاحصائية المذكورة أعلاه في تحديد عدد الفئات ليست اجبارية، ويمكن للباحث تحديد الطول الذي يراه مناسباً للفئة.
- عدد الفئات يكون مناسباً إذا كان يغطي جميع البيانات أيسمح بتفريغها.
- يجب أن تكون الفئات غير متداخلة.

مثال (1-1):البيانات التالية تمثل عدد الساعات التي عملها 50 عاملاً في مصنعٍ ما خلال اسبوع.

38 33 46 29 36 42 26 24 48 48 43 32 48 27 33 48 41 44 43

41 24 48 39 37 33 35 38 49 32 23 34 23 22 46 28 25 47 46

44 45 43 46 47 44 39 23 36 43 36 24

لعرض هذه البيانات في جدول توزيع تكراري نتبع الخطوات السالفة الذكر وهي:

1. تعيين المدى العام للبيانات، والذي هو الفرق بين أكبر وأصغر قيمة في البيانا $R=X_{max}-X_{min}$

=49-22=27

2. تعيين عدد الفئات، التي يفضل أن تكون متساوية في الطول،حيث نرمز لها بالرمز k ونفترض ان عددها في هذا المثال هو k=6

3. تعيين طول الفئةن والذي نرمز له بالرمز " Δ "حيث هناك علاقة رياضية لحساب طول الفئة مباشرة، والذي هو عبارة عن الفرق بين حديها الأعلى والأذنى، وفق قانون ستورجز "H.A.Sturges" الذي وضعه استناداً إلى قاعدة تجريبية لتحديد طول الفئة، تعتمد هذه القاعدة على مجال الدراسة وحجم المجتمع أو العينة ، حيث $\Delta=\frac{R}{1+3.32Log(N)}$ ou $\Delta = \frac{R}{1+1.32Ln(N)}$

حيث Δ هو طول الفئة، R هو المدى العام للبيانات ، N هو عدد البيانات أو المشاهدات، Log هو اللوغاريتم العشري ، Ln هو اللوغاريتم النيبيري

ومن هذه العلاقة يقول ستورجز أن عدد الفئات يكون مساوياً لـ $K=1+3,322\log(N)$ أو لـ $K=1+1.32Ln(N)$

4. تعيين الحد الأدنى للفئة الأولى، والذي يكون مساوياً أو أقل من أصغر قيمة موجودة في هذه البيانات، والتي هي في هذا المثال هي 22،ولنفترض ان الحد الأدنى هو 20.
5. تعيين الحد الأعلى للفئة الأولى=20+5=25.
6. تعيين الحدود الدنيا والعليا لبقية الفئات على نفس المنوال، مع اعتبار أن الحد الأعلى للفئة السابقة هو عبارة عن الحد الأدنى للفئة اللاحقة، وبذلك يكون الحد الأعلى للفئة الثانية هو 25+5=30.
7. تعيين مراكز الفئات، والذي هو عبارة عن مجموع حدي الفئة مقسوماً على 2 أي ان مركز الفئة الأولى $C_i=\frac{L_i+L_{i+1}}{2}=(20+25)/2=22.5$ حيث: L_i: الحد الأدنى للفئة i و L_{i+1} : الحد الأعلى للفئة i.
8. تفريغ البيانات على الفئات.
9. تسجيل مجموع تكرارات كل فئة أمامه في عمود التكرارات (fi).
10. تجميع التكرارات لجميع الفئات، حيث يكون $N=\sum_{i=1}^{n} ni$ كما هو موضح في الجدول التالي

جدول رقم (1-2): كيفية تشكيل جدول توزيع تكراري

رقم القيم	حدود الفئات	ciمركز الفئة	تفريغ القيم داخل الفئات	niالتكرار
1	]20-25]	22.5	24 24 24 23 23 23 22	07
2	]25-30]	27.5	29 28 27 26 25	05
3	]30-35]	32.5	34 33 33 33 32 32	06
4	]35-40]	37.5	39 38 38 38 37 36 36 36 35	09

5	]45-40]	42.5	44 44 43 43 43 43 42 41 41 41	10
6	]50-45]	47.5	49 48 48 48 48 48 47 47 46 46 46 46 45	13
المجموع				N= $\sum_{i=1}^{n} fi = 50$

إذن فالجدول أعلاه يمثل توزيع تكراري لعدد ساعات العمل التي عملها 50 عاملاً خلال اسبوع، حيث تم ترتيب العاملين حسب عدد ساعات العمل تصاعدياً كما هو موضح في العمود الثاني، بينما يمثل العمود الرابع يمثل افراغ تلك البيانات فيما يمثل العمود الخامس عدد المرات التي ظهرت بها البيانات المقابلة لكل فئة والتي تعرف باسم التكرارات، والتي تقابل القيم المرتبة تصاعدياً أو مراكز الفئات في حالة المتغير المستمر.وعليه فجدول التوزيع التكراري يتألف من:

- فئات قيم المشاهدة أو القياسات xi
- التكرارت المقابلة لهذه القياسات ni

أنواع التكرارات : هناك عدة أنواع للتكرارات، تتمثل في:

1. **التكرار المطلق** (ni): وهو التكرار العادي كما هو مبين في الجدول أعلاه.
2. **التكرار النسبي** (fi):وهو عبارة عن نسبة التكرار المطلق إلى مجموع التكرارات، أي نسبة تكرار تلك الفئة إلى مجموع التكرارات حيث $fi=\frac{ni}{\sum_{i=1}^{n} ni}$، وبذلك يكون مجموع التكرارات النسبية يساوي الواحد الصحيح

$$\sum_{i=1}^{n}\left(\frac{fi}{N}\right)=\frac{1}{N}(\sum fi)=\frac{1}{N}N=\frac{f1}{N}+\frac{f2}{N}+\frac{f3}{N}+\cdots\ldots\ldots\ldots\ldots+\frac{fn}{N}=\frac{f1+f2+f3\ldots\ldots\ldots+fn}{N}=\frac{\sum fi}{N}=\frac{N}{N}=1$$

والجدول الذي يعطينا الفئات مع مراكزها وتكراراتها النسبية يسمى جدول التكرار النسبي.

عرض البيانات باستخدام الجداول التكرارية المزدوجة:

و يستخدم هذا النوع من الجداول عند دراسة متغيرين يعتمد أحدهما على الأخر، كدراسة الطول و الوزن لمجموعة من الأفراد ، أو دراسة الكميات و الأسعار لسلعة ما و غيرها...، و بناء جدول توزيع تكراري مزدوج يتم كما يلي:

- وضع بيانات أحد المتغيرين أفقيا وتخصص الأعمدة لبيانات المتغير الأخر.

- نرمز لقيم المتغير الأول بالرمز X_i حيث i تتراوح من 1 إلى K، و نرمز لقيم المتغير الثاني بالرمز y_j حيث j تتراوح من 1 إلى M.

ويمكن أن نشكل الجدول المزدوج التالي

الجدول رقم (1–3): جدول تكراري مزدوج.

$\sum$	y_m ..... y_3 y_2 y_1 y_4	المتغير y_j / المتغير X_i
	n_{1m} ..... n_{13} n_{12} n_{11} n_{14}	X_1
	n_{1m} ..... n_{23} n_{22} n_{21} n_{24}	X_2
	n_{1m} ..... n_{33} n_{32} n_{31} n_{34}	X_3
	n_{1m} ..... n_{43} n_{42} n_{41} n_{44}	X_4
	. . . .	.
	. . . .	.
	n_{km} ..... n_{k1} n_{k1} n_{k1} n_{k1}	X_n
$\sum\sum$		$\sum$

مثال(1–2): سحبت عينة عشوائية من مجتمع ما تتكون من 100 أسرة قصد دراسة ظاهرتين هما، النفقات الاستهلاكية للأسرة (10^3 دج) والتركيبة الأسرية من حيث عدد الأطفال، فكانت النتائج المتحصل عليها مبينة في الجدول التالي:

الجدول رقم (1–4) توزيع الأسر حسب النفقات الاستهلاكية وعدد الأطفال

$\sum$	3	2	1	0	عدد الأطفال / النفقات
20	3	6	8	3	$]30-20]$
48	13	12	15	8	$]40-30]$
32	10	11	7	4	$]50-40]$
100	26	29	30	15	$\sum$

التكرار المتجمع :**Cumulative Frequency** وينقسم إلى قسمين هما:

أ. **التكرار المتجمع الصاعد frequence cumulative croissante** $\binom{fi}{ni}\uparrow$:التكرار المتجمع الصاعد لأي قيمة هو عبارة عن تكرار هذه القيمة مضافاً إليه تكرار الفئة السابقة، ونستعمل عند حسابه عبارة "أقل من أو يساوي" ونتعامل مع الحدود العليا للفئات.فهو عبارة عن عدد البيانات التي تقل قيمتها عن الحد الأعلى للفئة المقابلة.

ب. **التكرار المتجمع النازل** $\binom{fi}{ni}\downarrow$ **frequence cumulative decroissante**:التكرار المتجمع النازل لأي قيمة هو عبارة عن مجموع التكرارات $N=\sum_{i=1}^{n} ni$ مطروحاً منه تكرارات الفئات السابقة، نستعمل في حسابه عبارة "أكبر من أو يساوي" ونتعامل مع الحدود السفلى ل لفئات.فهو عبارة عن عدد البيانات التي تزيد قيمتها عن الحد الأدنى للفئة المقابلة.

ملاحظات:

- يمكن تكوين التكرار النسبي المتجمع الصاعد $f_i\uparrow$ أو النازل $f_i\downarrow$ بنفس طريقة تكوين التكرارات المطلقة التجميعية.
- أول قيمة في التكرار المتجمع الصاعد هي أول قيمة في التكرار المطلق وآخر قيمة فيه هي مجموع التكرارت.
- أول قيمة في التكرار المتجمع النازل هي مجموع التكرارت وآخر قيمة فيه هي آخر قيمة في التكرار المطلق باستخدام معطيات **المثال (1–1)** نقوم بإيجاد التكرارات النسبية والنسبية المئوية وكذا التكرارات المتجمعة الصاعدة والنازلة:

جدول رقم(1–5):جدول تكراري

حدود الفئات	ciمركز الفئة	niالتكرار	التكرار النسبي f_i	التكرار النسبي المئوي $f_i\%$	التكرار المتجمع الصاعد $n_i\uparrow$	التكرار المتجمع النازل $n_i\downarrow$
]25-20]	22.5	07	0.14	14	07	50
]30-25]	27.5	05	0.1	10	12	43
]35-30]	32.5	06	0.12	12	18	38
]40-35]	37.5	09	0.18	18	27	32

]45-40]	42.5	10	0.2	20	37	23
]50-45]	47.5	13	0.26	26	50	13
المجموع		50	1.00	100%		

من خلال الجدول أعلاه يمكن تقديم الشروح والقراء آت لنتائج الجدولة حيث نقول أن جميع العمال يعملون عدد ساعات يساوي أو يقل عن 50 ساعة أسبوعياً،وأن 13 عامل فقط يعملون عدد ساعات يساوي أو يزيد عن 54 ساعة، وأن 18% منهم يعملون مابين 35 و40 ساعة،.....الخ

1-3-5-2: العرض(التمثيل) البياني

طريقة العرض البياني: تنطلق هذه الطريقة من أهمية خاصية البصر عند الإنسان،فعند ما نحول الارقام المجردة إلى صورٍ وأشكالٍ بيانية تصبح أقرب إلى الحفظ والذاكرة، وتعطينا من خلال النظرة بعض الصفات الرئيسية للظاهرة المدروسة، لأن الطرق البيانية مفيدة للغاية عند تقديم المعلومات الاحصائية لأنها تقدم وصف عام وسريع لتلك المعلومات، لأنه كما يقال:صورة واحدة تعوضك عن ألف كلمة. وهي تقريبية لما تحويه الجداول من أرقام.

تتخذ عملية التمثيل والعروض البيانية أشكالاً وصوراً تختلف باختلاف البيانات، التي سنميز فيها نوعين تبعاً لنوعية المتغير الإحصائي هما المتغير الاحصائي المنقطع والمتغير الاحصائي المتصل أو المستمر..

1. **العرض البياني للبيانات النوعية** :يمكن عرض المتغير النوعي في شكل دوائر أو مربعات أو أعمدة بيانية، تمكن من وصف ومقارنة مجموعات أو مستويات هذا المتغير.

أ. **الدوائر أو المربعات**:تستعمل الدوائر أو المربعات لإظهار تطور قيمة ظاهرة من الظواهر زمنيا أو مكانياً، بحيث نرسم دائرة اختيارية خلال سنة معينة، قطرها اختياري ثم وفق القاعدة الثلاثية نوجد قطر الدائرة في السنة الموالية أوالمكان الآخر.

تستعمل الدوائر أو المربعات إلى جانب الأعمدة البيانية لتمثيل الظاهرة التي تأخذ عددا من القيم من أجل إظهار مدى التباين من مدة لأخرى ومن مكانٍ لآخر.

كما يمكن أن تستعمل الدوائر والمربعات في عرض التوزيع النسبي للأوجه المختلفة لظاهرة معينة في فترة زمنية واحدة (غالباً ما تستعمل هذه الرسوم لتمثيل أجزاء من ظاهرة كلية)، عل أن تكون هذه الأوجه مكونة للمجموع الكلي لهذه الظاهرة.

ولرسم هذا الشكل التوضيحي نرسم دائرة ذات قطرٍ مناسب ونقسمها إلى قطاعات تتناسب مساحتها مع كل مع نسبة كل جزء في المجموع الكلي أي نقسم الزاوية المركزية التي تساوي 360° تبعاً لعدد الأجزاء التي تتكون منها الظاهرة وتبعاً لنسبة كل جزء إلى المجموع الكلي. وعلى أساس هذه الزاوية نقسم مساحة الدائرة إلى قطاعات يمكن تلوينها للتمييز بينها ويكتب داخل كل قطاع اسم الجزء الذي يمثله أو على محيط الدائرة وخارج كل قطاع النسبة التي يمثلها من المجموع الكلي.ويستعمل هذا النوع من العرض البياني غالباً لإظهار التركيب الهيكلي لظاهرة معينة في فترة زمنية واحدة أو مكانٍ واحد

$$\textbf{الزاوية المركزية المقابلة لكل صفة} = \frac{\text{تكرار الصفة}}{\text{مجموع التكرارت}} \times 360$$

مثال(1–2): في مؤسسة للمشروبات كانت نسبة المبيعات الفصلية إلى المبيعات الكلية كما هو موضح في الجدول اسفله.:

الفصول	الصيف	الخريف	الشتاء	الربيع
%المبيعات	59	22	9.5	9.5

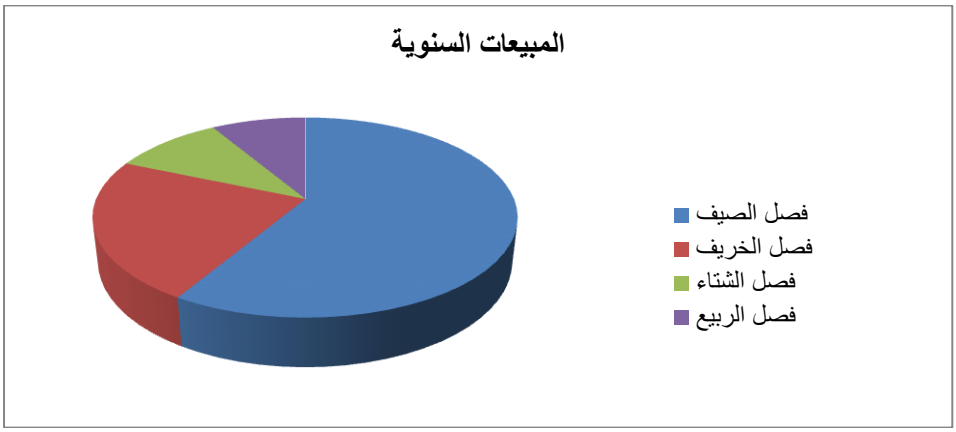
المطلوب: تقديم العرض البياني المناسب

الحل: نقوم بحساب الزاوية المركزية لنسبة المبيعات لكل فصل بتطبيق العلاقة السابقة فنجد:

الفصل	نسبة المبيعات	الزاوية المركزية المقابلة
الصيف	59	212.4
الخريف	22	79.2
الشتاء	9.5	34.2
الربيع	9.5	34.2

وعليه يتخذ العرض الدائري الشكل المبين اسفله

الشكل (1-1): الزوايا الدائرية



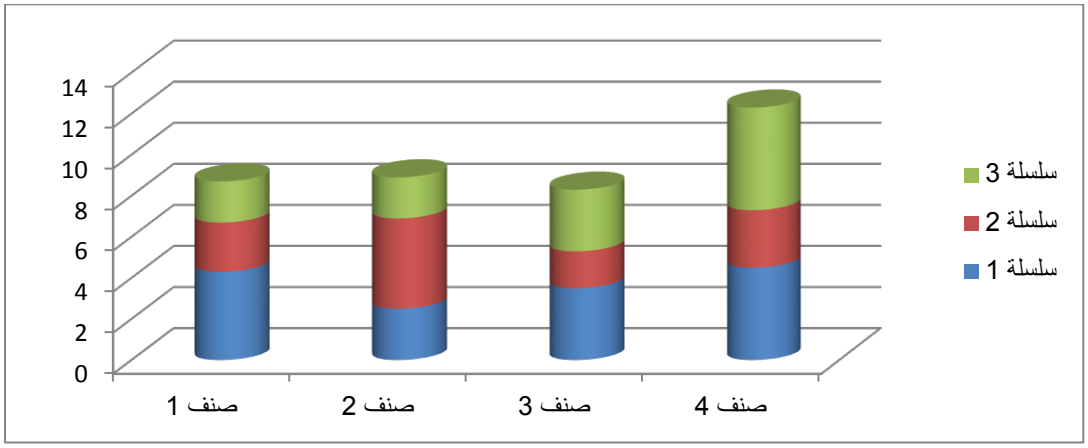
أ. **الأعمدة او المستطيلاتl** تستعمل الأعمدة أو المستطيلات في عرض قيم ظاهرة ما في عدة فترات زمنيه وذلك لإبراز التغير الذي حدث فيها، وكذلك لتوضيح قيم الأوجه المختلفة لظاهرة معينة لإبراز المقارنة بين هذه الأوجه أو في أماكن مختلفة.

يتكون شكل الأعمدة من مستطيلات ذات سمك واحد ومنفصلة عن بعضها البعض، بحيث يتناسب ارتفاعها مع قيم الظاهرة في الفترات الزمنية المختلفة للظاهرة بحيث تكون المستطيلات ذات عرضٍ(سمك) واحد.وإذا كان للظاهرة المدروسة قيماً سالبة واخرى موجبة، ترسم القيم الموجبة في الاتجاه الموجب والقيم السالبة في الاتجاه السالب بحيث يكون محور الصفر فاصلاً بينهما.

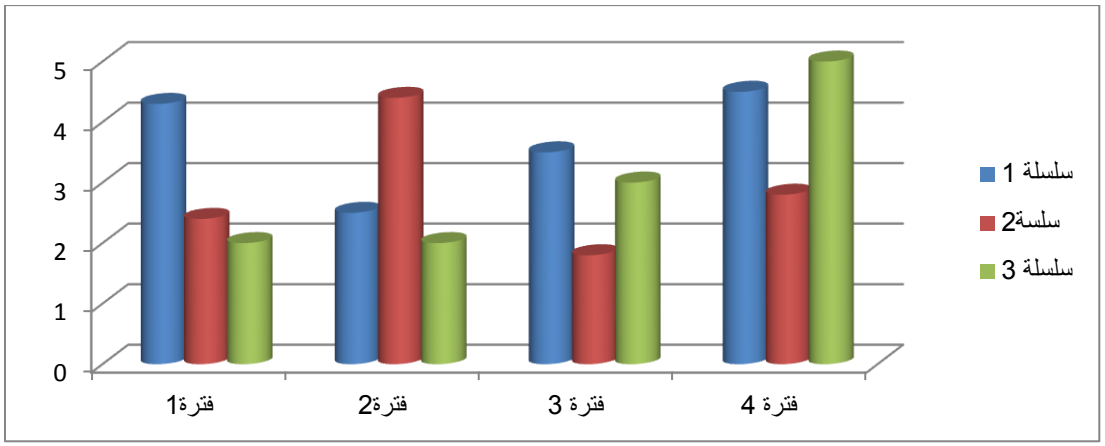
ويمكن ان تكون الظاهرة ذات قيمتين أو أكثر في كل فترة من الفترات الزمنية، ففي هذه الحالة، إذا كان من الممكن جمع هاتين القيمتين بحيث يكون المجموع دالاً على القيمة الكلية للظاهرة، يرسم عمود لكل فترة، ويقسم إلى جزئين بحيث يكون الارتفاع الكلي للعمود معبر عن قيمة الظاهرة، وارتفاع كل جزء منه معبراً ومتناسباً مع قيم كل جزء مهما كان عدد الاجزاء التي تتكون منها الظاهرة.

أما إذا لم يكن من الممكن جمع هذه ألأجزاء فإنه يرسم لكل فترة أعمدة متلاصقة ومتساوية القاعدة،يختلف ارتفاعها تبعاً لاختلاف القيم التي تمثلها هذه ألأعمدة وهي طريقة يمكن استعمالها مهما كان عدد الأجزاء التي تتكون منها الظاهرة.غير أنه لا يجب أن نكثر من الأعمدة في الرسم حتى يبدو مكتظا بها، فيضيع الهدف الأساسي من التوضيح (العرض).

الشكل رقم (1–2): الأعمدة المجزأة



الشكل رقم(1–3) الاعمدة المتلاصقة



العرض البياني للبيانات الكمية المتقطعة (المنفصلة):

يمكن عرض البيانات ذات المتغير الاحصائي المنقطع وفق الطرق الرئيسية التالية:

1. الأعمدة البسيطة: هو عبارة عن أعمدة بسيطة تتناسب أطوالها مع التكرار المقابل لقيمة معينة للمتغير المدروس وتسمى الأعمدة البسيطة.

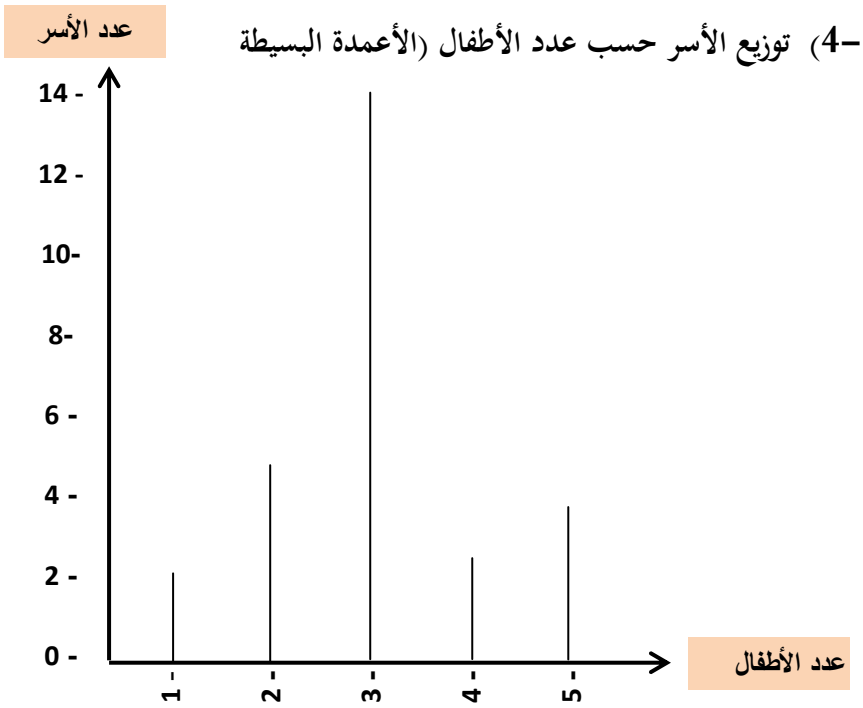
مثال(1–3): يبيّن الجدول التالي يبين توزيع 28 أسرة حسب عدد الأطفال

عدد الأطفال لكل أسرة X_i	1	2	3	4	5	$\sum$
عدد الأسر n_i	2	5	14	3	4	28

المطلوب: عرض هذه البيانات بالطريقة المناسبة.

الحل: أفضل طريقة لعرض هذه البيانات هي الأعمدة البسيطة.

لشكل رقم(1–4) توزيع الأسر حسب عدد الأطفال (الأعمدة البسيطة



2. المنحنى المتجمع السلميLadder curve

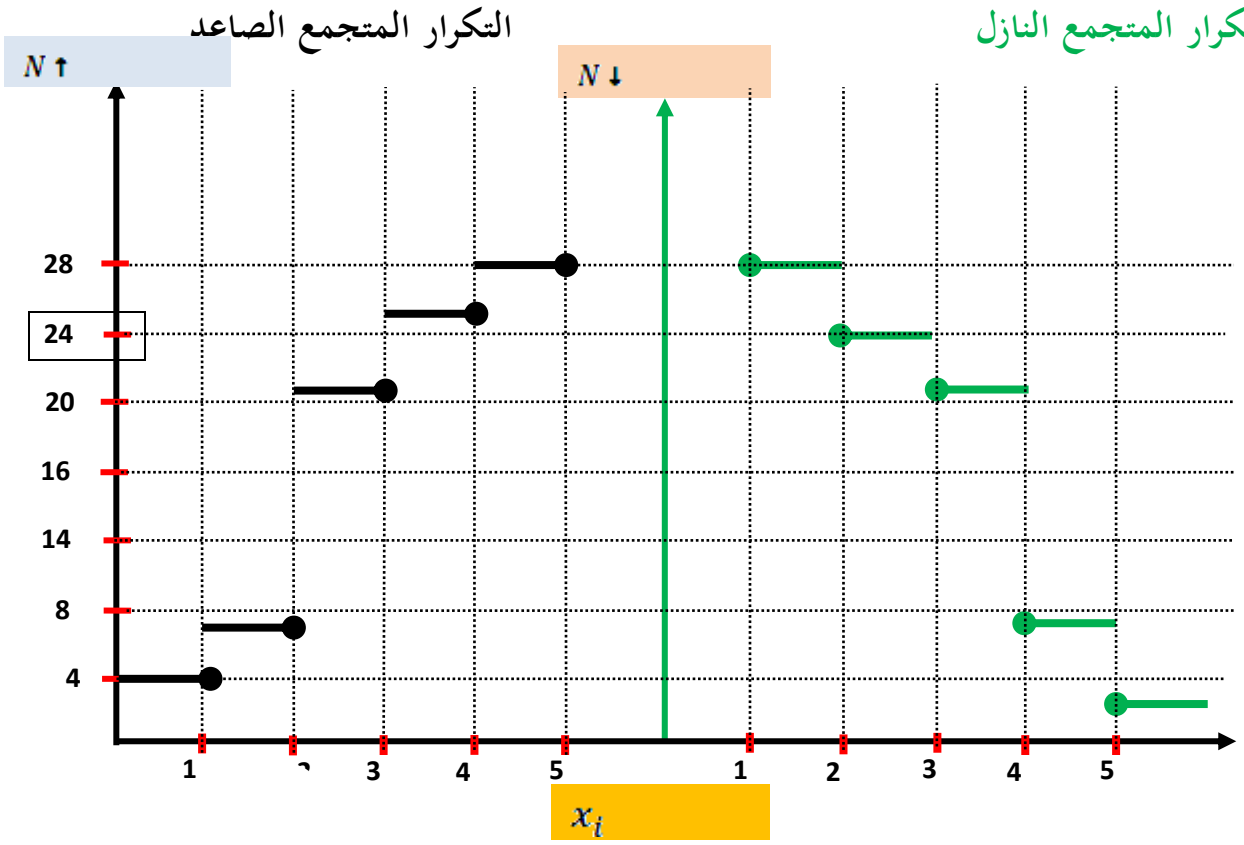
يستعمل لعرض التكرار المتجمع المطلق او النسبي لمتغير منفصل (متقطع) و وهو عبارة عن منحنى في شكل سلم حيث تعبر كل خطوة منه على التكرار التراكمي المقابل لقيمة المتغير.

مثال (1–4) يبيّن الجدول التالي يبين توزيع 28 أسرة حسب عدد الأطفال

عدد الأطفال لكل أسرة X_i	1	2	3	4	5
عدد الأسر ni	2	5	14	3	4
التكرار المتجمع الصاعد ni ↑	2	7	21	24	28
التكرار المتجمع النازل ni ↓	28	26	21	7	4

الحل:

لشكل رقم(1-5) العرض البياني للتكرار المتجمع لمتغير كمي منفص



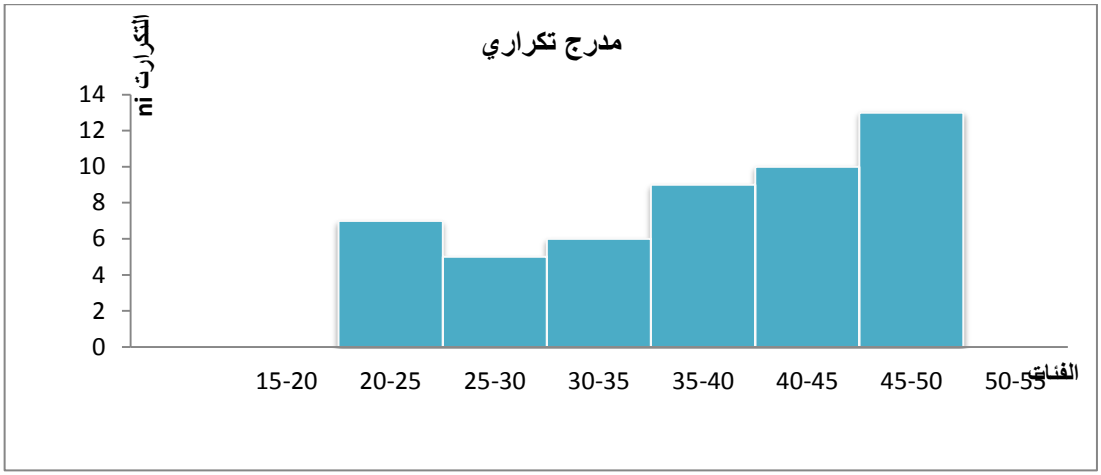
العرض البياني للبيانات الكمية المستمرة (المتصلة):

1. **المدرج التكراري**:histogram :هو عبارة عن مجموعة من المستطيلات المتلاصقة لسلسة من البيانات لظاهرة معينة، تكون مصنفة أو مبوبة إلى فئات تكرارية، بحيث أن كل مستطيل يتناسب طوله مع تكرار تلك الفئة الموضوع على محور الاحداثيات العمودية، ويتناسب عرضه مع طول الفئة الموضوع على الاحداثيات العمودية،والذي نميز فيه نوعين أو حالتين:

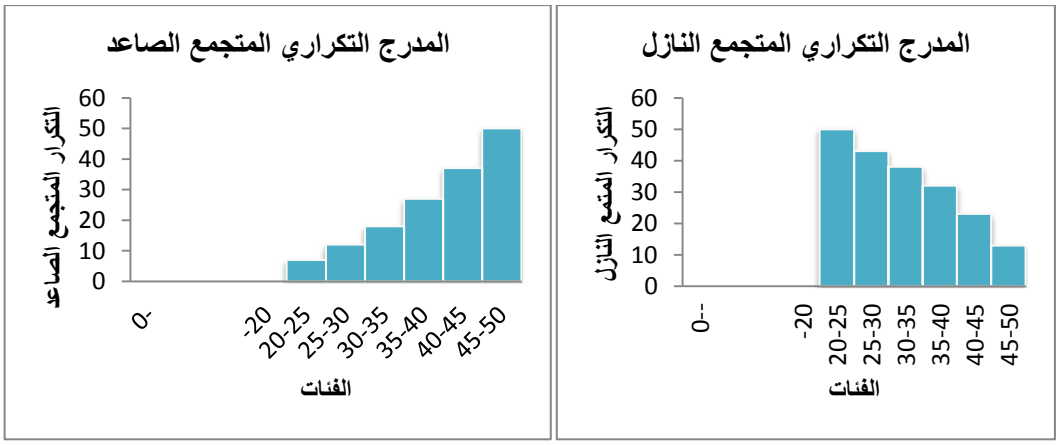
الحالة الأولى: الفئات متساوية الطول

بالعودة إلى الجدول التكراري أعلاه، يمكن انشاء المدرج التكراري له كما هو مبينٌ أدناه

الشكل (1–6): المدرج التكراري



كما يمكن عرض هذه البيانات وفق المدرج التكراري المتجمع الصاعد أو النازل، اللذان هما عبارة عن مجموعة من المستطيلات المتلاصقة ذات عرض يتناسب وطول الفئة وطول يتناسب مع التكرار المتجمع الصاعد أو النازل للفئات كما هو مبين في الشكلين التاليين



الحالة ألثانية: الفئات غير متساوية

غير أن هناك مشكلة تعترضنا عند ما تكون أطوال الفئات غير متساوية، فلو أنشأنا المدرج التكراري بدون أي احتياط أي كما هو موجود لتحصلنا على تمثيل غيرصحيح أي تمثيل مضلل.لذلك يكون من الافضل تصحيح هذا الخطأ عن طريق إعادة الفئات غير المتساوية إلى نفس االقيمة بالاعتماد على

التكرار المعدل الذي يساوي $\frac{\text{التكرار المطلق}}{\text{طول الفئة}}$. طول الفئة المختار كأساس للتعديل $ni^* = \frac{ni}{\Delta} . \Delta^*$

حيث Δi^* هو طول الفئة المختار كأساس لتعديل ألتكرارات والتي هي القاسم المشترك الأكبر لأطوال الفئات

ويكون حساب التكرار المعدل ضرورياً لـ:

- رسم مدرج تكراري صحيح
- تحديد الفئة المنوالية الصحيحة وبالتالي حساب قيمة صحيحة للمنوال.

مثال(1–5) الجدول التالي يبين توزيع عينة من 300 مزرعة حسب المساحة (عدد الهكتارات)

مساحة المزرعة X_i	]35–30]	]45–35]	]50–45]	]65–50]	]85–65]	[90–85]	$\sum$
التكرار n_i	**15**	**45**	**60**	**75**	**90**	**15**	**300**

المطلوب: تمثيل هذه البيانات باستخدام المدرج التكراري؟

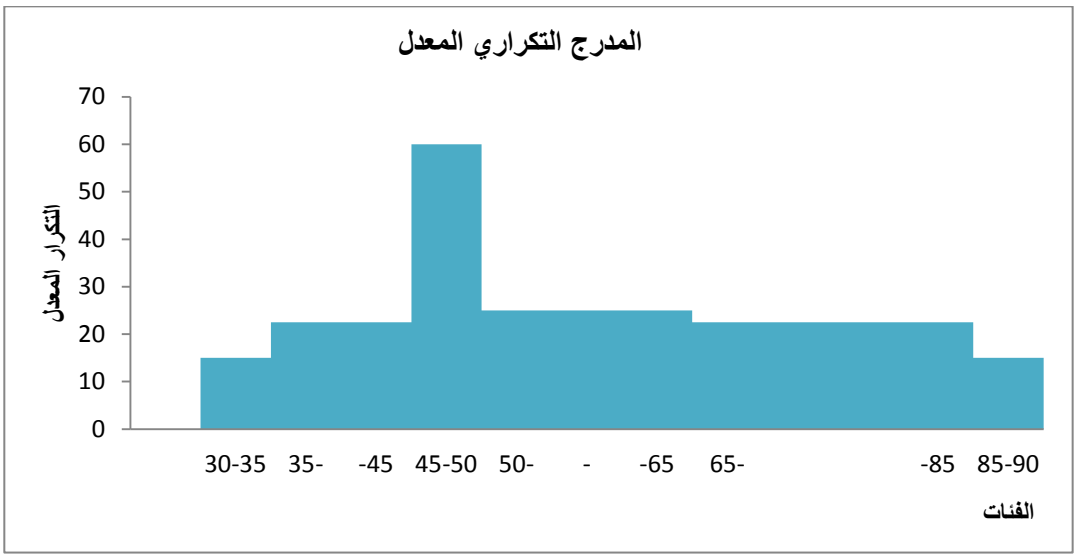
الحل: بما أن الفئات غير متساوي الطول فإنه يكون من الضروري حساب التكرار المعدل للفئات حتى يكون المدرج التكراري يعبر حقيقة عن كيفية توزيع البيانات على مختلف الفئات، ونأخذ طول الفئة المختار والمساوي لـ 5 كأساس لتعديل التكرارات، (طول الفئة المختار هو القاسم المشترك الأكبر لأطوال الفئات أو أصغر طول فئة).

الجدول(1–6): التكرار المعدّل

مساحة المزرعة X_i	التكرار n_i	طول الفئة (Δi)	التكرار المعدل ni'
]35–30]	15	5	$n_1^{'} = \frac{15}{5} \times 15 = 15$
]45–35]	45	10	$n_2^{'} = \frac{45}{10} \times 5 = 22.5$
]50–45]	60	5	$n_3^{'} = \frac{60}{5} \times 5 = 60$
]65–50]	75	15	$n_4^{'} = \frac{75}{15} \times 5 = 25$
]85–65]	90	20	$n_5^{'} = \frac{90}{20} \times 5 = 22.5$
[90–85]	15	5	$n_6^{'} = \frac{15}{5} \times 5 = 15$

فيكون المدرج المعدل بالشكل

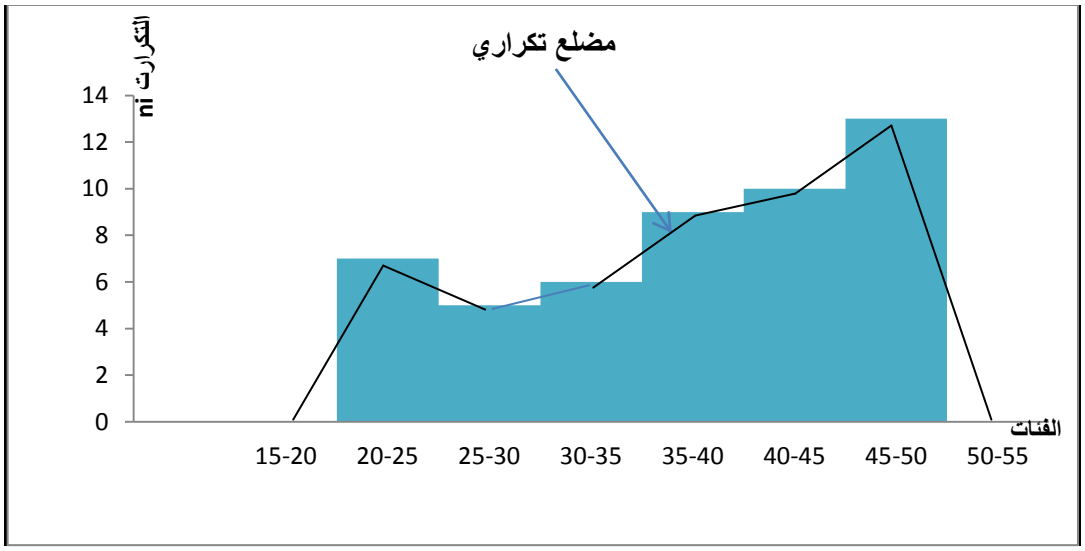
الشكل (1–7): مدرج تكراري معدل



المضلع التكراري *frequency poligon* :

الغرض منه هو إظهار ومقارنة خواص توزيع ألمتغيرات ونحصل عليه بتوصيل مراكز الفئات عند قمة كل مستطيل في المدرج ألتكراري عن طريق قطع مستقيمة، ثم غلق المضلع عند ألطرفين وذلك بافتراض وجود فئة سابقة للفئة الأولى ووجود فئة لاحقة للفئة ألأخيرة تكرار كل منهما معدوم.

الشكل (1–8): المضلع التكراري



ويبدو المضلع التكراري عند رسمه عبارة عن حط منكسر أو مجموعة من الأضلاع والقطع المستقيمة

ملاحظة: مساحة المضلع التكراري تساوي مساحة المدرج التكراري دوماً لأن المضلع التكراري يضيف أجزاء إلى مساحة المدرج التكراري ويستبعد أجزاء أخرى، بحيث أن مجموع الأجزاء المضافة يساوي مجموع الأجزاء المستبعدة، حيث يمكن استخدام النظريات الهندسية للبرهنة على ذلك.

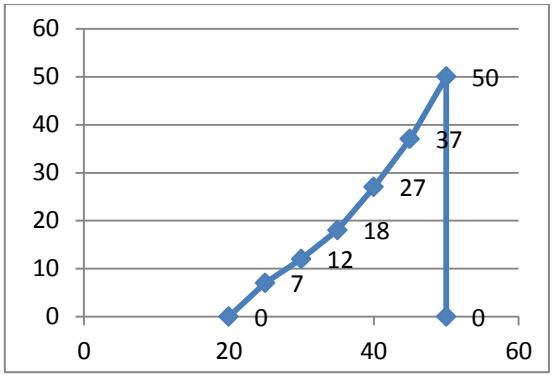
مثال: بالعود إلى المثال(**1-1**) السابق

حدود الفئات	ciمركز الفئة	niالتكرار	التكرار النسبي f_i	التكرار النسبي المئوي $f_i\%$	التكرار المتجمع الصاعد $n_i \uparrow$	التكرار المتجمع النازل $n_i \downarrow$
]25-20]	22.5	07	0.14	14	07	50
]30-25]	27.5	05	0.1	10	12	43
]35-30]	32.5	06	0.12	12	18	38
]40-35]	37.5	09	0.18	18	27	32
]45-40]	42.5	10	0.2	20	37	23
]50-45]	47.5	13	0.26	26	50	13
المجموع		50	1.00	100%		

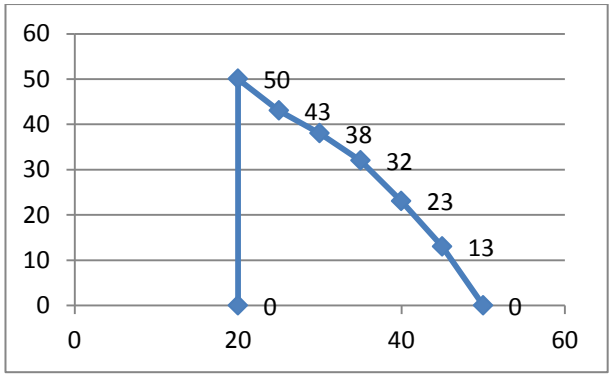
و المطلوب هو تمثيل كل من التكرار المتجمع الصاعد و النازل بيانيا

الحل:

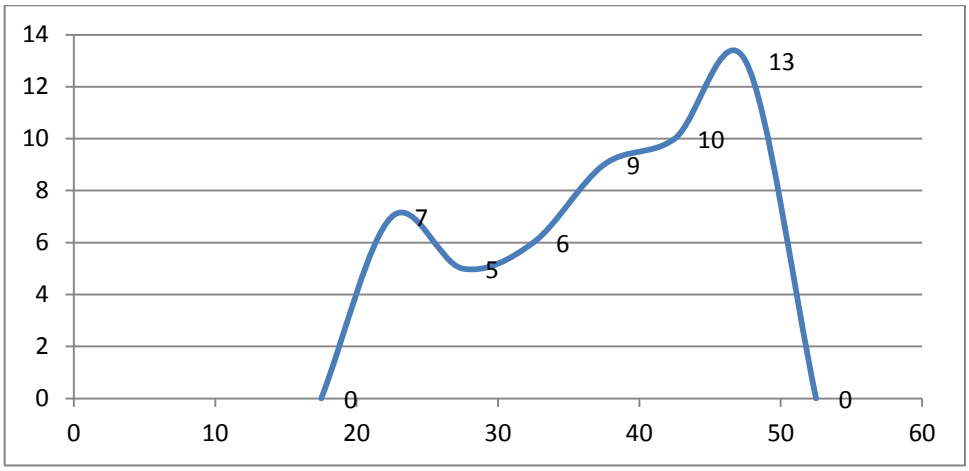
ا لمضلع التكراري المتجمع الصاعد



المضلع التكراري المتجمع النازل



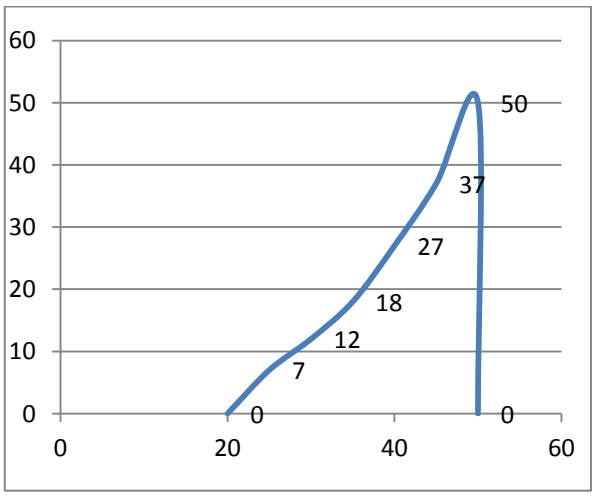
المنحنى التكراري :إذا قمنا بالتوصيل بين النقاط التي تقابل مراكز الفئات بخطوطٍ ممهدة بدلاً من خطوطٍ منكسرة، فنحصل على المنحنى التكراري كما هو مبين في الشكل أدناه.



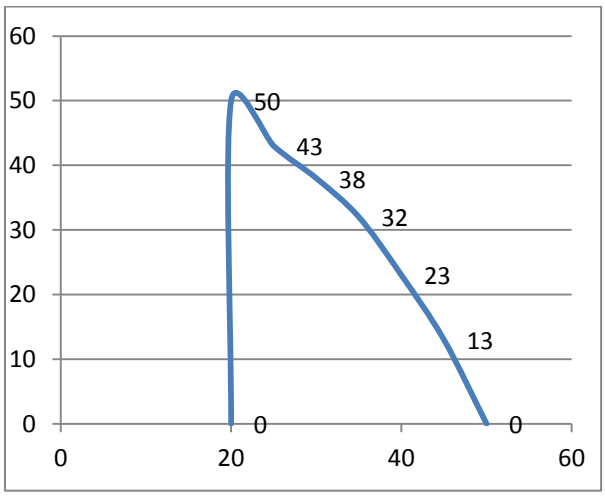
ملاحظة:نلاحظ أن المساحة المحددة بالمضلع التكراري لا تساوي المساحة المحددة بالمنحنى التكراري (المدرج التكراري) لأن المنحنى التكراري يضيف أجزاء ويستبعد اجزاء أخرى لا تساويها، كما أنه ليس من الضروري أن يمر بجميع النقاط.ولا يفضل رسم المنحنى التكراري إلا إذا كانت الفئات كثيرة العدد، ذات أطوال صغيرة، وكان عدد البيانات كبير، وهي من النوع المتصل.

بنفس الطريقة التي مثلنا بها المضلع التكراري المتجمع الصاعد والنازل، يمكن أيضاً تمثيل المنحنى التكراري المتجمع الصاعد والنازل عن طريق التوصيل بين التكرارت المقابلة للحدود العليا للفئات بالنسبة للمنحنى التكراري المتجمع الصاعد، والتوصيل بين التكرارت المقابلة للحدود السفلى للفئات بالنسبة للمنحنى التكراري المتجمع النازل،وكل ذلك بخطوطٍ ممهدة وليس خطوط مستقيمة كما يلي:

لمنحنى التكراري المتجمع الصاعد



المنحنى التكراري المتجمع النازل



التكرارات المتجمعة الصاعدة: هي عبارة عن قطع مستقيمة متصاعدة حسب تصاعد التكرارات التجميعية الصاعدة المقابلة لكل قيمة من قيم المتغير الإحصائي المدروس، تبدأ القطعة من القيمة المقابلة للتكرار وتنتهي عند القيمة السابقة لتلك القيمة.

التكرارات المتجمعة النازلة: هي عبارة عن قطع مستقيمة متنازلة حسب تنازل التكرارات التجميعية النازلة، حيث أن القطعة المستقيمة الأولى تبدأ القطعة من القيمة المقابلة للتكرار وتنتهي عند القيمة الموالية.

ولتوضيح كيفية رسم التكرارات المتجمعة الصاعدة والنازلة نأخذ المثال

الفصل الثاني :مقاييس المتغيرات الاحصائية

مقدمة: إن العرض البياني لا يعتمد بدرجة كبيرة في وصف التوزيعات التكرارية ومقارنتها، وعليه فإنه من المفضل وصف التوزيع التكراري بمجوعة من الأرقام أو المقاييس التي تلخصه أو توضح خصائصه. والغرض من هذه المقاييس هو وصف المجموعة برقم واحد يمثلها معبراً عن مزيد من الوصف والتلخيص، ويفيد هذا الرقم أو المقياس أو المؤشر في المقارنات بين المجموعات أو المجتمعات، كما يفيد في المقارنات التاريخية أو الطولية بما يمكن من وصف التغير أو التطور في الظاهرة عبر الزمن.

فالوسط لأي مجموعة من البيانات الاحصائية هو القيمة التي تعبر عن المجموعة بصفة عامة أو النموذج الذي يمثل مجموعة من القيم أو مفردات الظاهرة، كما أن هذا النموذج أو القيمة تنحرف عنه القيم الآخرى بشيئ من الانتظام.

التحليلات الاحصائية(الطرق العددية)

الطرق البيانية(العرض البياني) مفيد للغاية عند تقديم المعلومات الاحصائية من أجل نقل أو وصف عام وسريع للمعلومات الاحصائية تأكيداً للمثل القائل(صورة واحدة تساوي ألف كلمة).ومع ذلك توجد حدود لاستخدام هذه الطرق في وصف وتحليل المعطيات، فعلى سبيل المثال لا الحصر، نفترض أننا بصدد مناقشة البيانات الاحصائية لمجموعة من الأفراد وأنه ليس لدينا طريقة أخرى غير الطريقة الشفهية، مما يجعل عرض المدرج أو المنحنى أو المضلع التكراري غير متوفر، فنضطر هنا لاستخدام مقاييس وصفية أخرى، تنقل للمستمعين صورة ذهنية عن هذه المجموعة ، كما أنه غير ممكن الاستفادة من استخدام الطرق البيانية في مجال الاستقراء الاحصائي وبالتالي كيفية قياس مدى الاختلاف والتشابه بينهما.

وإذا كان العرض البياني يمدنا بصورة مرئية للبيانات تسعاد على فهما، إلا أن هناك بعض البيانات التي يصعب عرضها بيانياً.لذلك فإن الاحصائيين لا يكتفون باساليب العرض البياني، بل عادة ما يفضلون مقاييس عددية لوصف وتحليل البيانات، تستخدم للتعرف على نمط توزيع قيم الظاهرة موضوع الدراسة، والتي تتكون أساساً من ثلاثة محاور هي:

1. البحث عن القيمة التي تتركز حولها القيم التي تتخذها الظاهرة في تغيرها من وحدة لأخرى، أي ما هي القيمة التي تتركز حولها القيم الأخرى في التوزيع؟.
2. البحث عن مدى تشتت القيم حول القيمة المركزية أو المتوسطة أي معرف إلى أي مدى تتشتت القيم عن القيمة التي اعتبرت متوسطة أو مركزية بالنسبة للبيانات المتعلقة بالظاهرة محل الدراسة؟

3. البحث عن مدى تماثل توزيع القيم على طرفي القيمة المركزية او المتوسطة أي كيف يكون شكل التوزيع النسبي للقيم في فئات القيم، وألى أي مدى تتوزع القيم بانتظام على جانبي القيمة الوسطى؟.

وباختصار فالدراسة الاحجصائية للبيانات يجب أن تركز على:

- البحث عن المقياس المركزي للقيم.
- درجة تشتت وتغير القيم عن هذه القيمة.
- شكل المنحنى البياني الذي يمثل السلسلة (التناظر، التطاول ،التفرطح،...)

وقبل البداية نشير إلى أننا هذه الدروس تعتمد على الوصف العددي لمتغير ذات بعد واحد.

مقاييس النزعة المركزية

1.2.متوسطات القياس

قد نجد عند دراسة ظاهرة ما، بأن قيمها تميل نحو التجمع أو التمركز حول قيمة معينة، فيطلق على هذا النوع من الميل أو الاتجاه اسم النزعة المركزية أو تمركز التوزيع،ى كما يطلق على القيم الوسطى التي تتمركز حولها القيم اسم مقاييس النزعة المركزية أو المتوسطات، التي تستخدم لتحديد مركز البيانات المتاحة.

لقد وضع يول" yule " أسساً يرى أن توفرها أمر مرغوب فيه في المتوسطات لكي تكون جديرة بهذا التمثيل للتوزيعات، وهذه الشروط تتلخص فيما يلي:

1. أن تكون القيمة المتوسطة قيمة موضوعية، محددة وليست مجرد تقدير ذاتي من الباحث، أي يفضل أن تكون طريقة حسابها طريقة رياضية لا يختلف فيها اثنان، كما يفضل أن تكون هذه الطريقة سهلة وغير معقدة.
2. يفضل عند حساب تلك القيمة استخدام جميع قيم المتغير أي يجب أن يبنى تحديدها على جميع القيم، وإلاَّ فإنها لا تمثل مفردات التوزيع كلها.
3. أن تكون القيمة المتوسطة من القيم التي يكون تأثرها بتغير القيمة قليلاً قدر الإمكان، فإذا كان لدينا عدد من العينات المأخوذة من مجتمعٍ واحد فمن النادر أن تتساوى متوسطات العينات مهما كانت صورة هذه المتوسطات، ولكن يحدث أن تكون قيم هذه المتوسطات قريبة من بعضها.
4. أن تكون القيمة المتوسطة صالحة للمعالجة الرياضية أي أنها تشترك في عددٍ كبيرٍ من القوانين الاحصائية، وهذا أهم شرط في الواقع.

وعموماً إذا ما حقق المتوسط كل الصفات الآتية أو معظمها فإنه يعتبر مقبولا، وهذه الصفات هي:

- يجب أن يكون المتوسط معرفاً تعريفاً دقيقاً.
- يجب أن يبنى على جميغ المشاهدات.
- أن يكون من السهل فهمه وتفسيره.
- يمكن حسابه بسهولة وسرعة معقولتين.
- يخضع للعمليات الحسابية الجبرية.
- لا يتأثر كثيراً بالقيم المتطرفة أو الشاذة.
- لا يتأثر كثيراً بتقلبات المعاينة.

وأهم مقاييس النزعة المركزية يمكن ايجازها في ثلاثة أنواع من المتوسطات هي:

1. متوسطات القياس أو الحجم.
2. متوسطات الموضع أو الموقع.
3. متوسطات التكرار.

1.1.2.متوسطات القياس measures means

هناك أربعة أنواع من المتوسطات هي:

أ. المتوسط الحسابي أو الوسط الحسابي
ب. المتوسط الهندسي أو الوسط الهندسي.
ت. المتوسط التوافقي أو الوسط التوافقي.
ث. المتوسط التربيعي أو الوسط التربيعي.

1.1.1.2.المتوسط الحسابي arithemetic meam

هو أشهر هذه المقاييس في العموم وأكثرها استعمالاً لأنه يحقق في الغالب الشروط الواجب توفرها في المتوسط، والتي سبق ذكرها، وهو عبارة عن مجموع المشاهدات أو القياسات على عددها.ويمكن حسابه بطريقتين:

الطريقة المباشرة:

أ. **المتوسط الحسابي البسيط: simple arithemetic meam**

ويستعمل في حالة البيانات غير المبوبة أي عند ما يكون لقيم المشاهدات أو البيانات نفس الأهمية، وهو عبارة عن مجموع هذه القيم على عددها، ويرمز له بالرمز $\bar{x}$ ويقرأ x بار، فإذا كان لدينا مجموعة من القيم هي: x1 x2 x3 x4 .....xn فإن متوسطها الحسابي هو:

$$\bar{x}=\frac{+x1+x2+x3+x4+\cdots\ldots\ldots xn}{N}$$

مثال: أوجد المتوسط الحسابي للسلسلة العددية التالية:2 ، 5 ، 8 ، 13 ، 17 ، 22 ، 33.

$$\bar{X}=\frac{2+5+8+13+17+22+33}{7}=14.28$$

ملاحظة: المتوسط الحسابي لسلسلة متتالية من القيم ، مهما كان أساس التتالي، هو القيمة الوسطى إذا كان عدد مشاهدات السلسلة فردياً، ومتوسط القيمتين الوسطيين إذا كان عدد البيانات زوجياً، أي هو القيمة التي ترتيبها في البيانات المرتبة ترتيباً تصاعدياً هو $\frac{N+1}{2}$ أي $\bar{X}=x_{\frac{n+1}{2}}$ إذا كان عدد البيانات فردياً ، ومتوسط القيمتين اللتين ترتيبهما $\frac{N}{2}$ و$\frac{N+1}{2}$ أي أن $\bar{X}=\frac{X_{\frac{N}{2}}+X_{\frac{N+1}{2}}}{2}$.

مثال: أوجد المتوسط الحسابي للسلسلتين العدديتين التاليتين:

سلسلة1: 5، 10، 15، 20، 25، 30، 35، 40، 45، 50، 55،60 ، 65 ، 70، 75، 80، 85

سلسلة2: 12، 24، 36 ، 48، 60، 72، 84 ، 96 ، 108 ، 120، 132 ، 144، 156 ،168

الحل:

سلسلة1: بما أن عدد بيانات السلسلة فردي (17) فإن متوسط السلسلة هو القيمة الوسطي وهي التي ترتيبها $\frac{N+1}{2}$ أي التي ترتيبها $\frac{17+1}{2}$ أي القيمة التي ترتيبها في البيانات المرتبة ترتيباً تصاعدياً هو 9 وهي $\bar{X}=45$.

سلسلة2:بما أن عدد بيانات السلسلة زوجي (14)، فإن متوسط السلسلة هو القيمة الوسطي للقيمتين التي ترتيبهما في البيانات المرتبة ترتيباً تصاعدياً هو$\frac{N}{2}$ و$\frac{N+1}{2}$ على التوالي أي القيمتين اللتين ترتيبيهما هما 7 و8 ويكون بذلك $\bar{X}=\frac{84+96}{2}=90$

ب. **المتوسط الحسابي المرجح:** ويحسب في حالة بيانات مبوبة، أي إذا لم يكن للبيانات نفس الأهمية، فأذا كان

1. **المتغير منفصل**:وكانت البيانات والتي عددها N هيx_1 x_2 x_3 x_4 .....x_k والتكرارت المقابلة لها هي n_1 n_2 n_3 n_4 ..... n_kفإن متوسطها الحسابي هو

$$\bar{X}=\frac{x1n1+x2n2+x3n3+x4n4+\cdots\ldots\ldots\ldots\ldots+xknk}{n1+n2+n3+n4+\cdots\ldots\ldots\ldots nk}=\frac{\sum_{i=1}^{k} nixi}{\sum_{i=1}^{k} ni}=\frac{\sum_{i=1}^{k} nixi}{N}$$

$$\bar{X}=\frac{\sum_{i=1}^{k} nixi}{N} \Rightarrow \sum_{i=1}^{k} nixi = N\,\bar{X}$$

مثال(2-1)

في امتحان فجائي في مادة ألإحصاء كانت علامات الطلبة كما هو ملخص في الجدول أدناه.

العلاماتxi	13	14	15	16	18	19
عدد الطلبةni	4	3	6	5	4	2

المطلوب:حساب متوسط العلامات التي تحصل عليها طلبة هذا الفوج.

الحل:

العلاماتxi	13	14	15	16	18	19
عدد الطلبةni	4	3	6	5	4	2
xini	52	42	90	80	72	38

$$\bar{X}=\frac{\sum_{i=1}^{k} nixi}{N}=\frac{374}{24}=15.62$$

2. المتغير متصل(مستمر):فإن جدول التوزيع التكراري في هذه الحالة يكون على شكل فئات عددها هوk، فإذا كانت مراكزها هي c_1 c_2 c_3 c_4 .....c_k والتكرارات المقابلة لها هي n_1 n_2 n_3 n_4 ..... n_k فإننا نعتبر أن التكرارات تتجمع حول مراكز الفئات، وعليه نحسب المتوسط الحسابي في هذه الحالة وفق العلاقة

$$\overline{X}=\frac{C1ni+C2n2+C3n3+C4n4+\cdots\ldots\ldots\ldots+Cknk}{n1+n2+n3+n4+\cdots\ldots\ldots\ldots+nk}=\frac{\sum_{i=1}^{k} n_i c_i}{\sum_{i=1}^{k} ni}=\frac{\sum_{i=1}^{k} n_i}{N}\cdot c_i$$

$$\bar{X}=\frac{\sum_{i=1}^{k} n_i}{N}\cdot c_i=\sum f_i c_i$$

مثال (2-2)

ليكن الجدول التالي الذي يمثل أجور عمال مؤسسة

الفئات الأجرية	20-10	30-20	40-30	50-40	60-50	70-60
عدد العمال	15	30	25	20	8	2

المطلوب: حساب متوسط اجور عمال المؤسسة

الحل:

الفئات الأجرية	20-10	30-20	40-30	50-40	60-50	70-60	$\sum$
ciمراكز الفئات	15	25	35	45	55	65	
niعدد العمال	15	30	25	20	8	2	100
ni ci	225	750	875	900	440	130	3320

ويكون بذلك المتوسط الحسابي لأجور عمال المؤسسة هو:

$$\bar{X}=\frac{\sum_{i=1}^{k} nici}{N} = \frac{3320}{100} = 33.2$$

المتوسط الحسابي المرجح لقيم تأخذ متتالية عددية: أي أن الفرق بين القيم يأخذ قيمة ثابتة

فإذا كانت القيم التي يأخذها المتغير الاحصائي هيx_k..... x_1 x_2 x_3 x_4 والتكرارات المقابلة لها هي $n_1\ n_2\ n_3\ n_4 \ldots\ldots\ldots n_k$، وافترضنا أن كل قيمة تزيد عن سابقتها بالمقدار a فإننا نكتب

$$x_1=x_0+a \Rightarrow x_0= x_1-a$$

ومنه نكتب

$$\overline{X}=\frac{\sum_{i=1}^{k} nixi}{\sum ni}=\frac{n1(x0+a)+n2(x0+2a)+n3(x0+3a)+n4(x0+4a)\ldots\ldots\ldots\ldots+nk(x0+ka)}{n1+n2+n3+n4+\cdots\ldots\ldots\ldots\ldots nk}$$

$$=\frac{x0(n1+n2+n3+n4+\cdots\ldots\ldots nk)+a(n1+n2+n3+n4+\cdots\ldots knk)}{n1+n2+n3=n4=\cdots\ldots\ldots nk}=X0+a\frac{\sum_{i=1}^{k} ini}{\sum_{i=1}^{k} ni}=\bar{x}$$

مثال(2−3) أوجد الوسط الحسابي المرجح للتوزيع التكراري التالي

∑	4	3	2	1	الرمز
	40	30	20	10	القيم
17	2	5	7	3	التكرار
40	8	15	14	3	ini

$X_1=X_0+a \Rightarrow X_0 = X_1 - 10 = 10 - 10 = 0$ ومنه $a =X_4-X_3=X_3-X_2=X_2-X_1=X_1-X_0=10$

$$\bar{x} =X_0+a.\frac{\sum ini}{\sum ni}= 0+10.\frac{40}{17} =\frac{400}{17} =25.52$$

الوسط الحسابي المرجح لمتوسطات حسابية: إذا كان لدينا مجموعة من البيانات ذات N1 مشاهدة ومتوسطها الحسابي هو $\overline{x1}$ ومجموعة أخرى من البيانات ذات N2 مشاهدة ومتوسطها الحسابي هو $\overline{x2}$ فإن المتوسط الحسابي للمجموعتين بعد دمجهما يكون هو $\bar{X}=\frac{N1\bar{x}+N2\bar{x}}{N1+N2}$

الطريقة غير المباشرة(الفروق البسيطة، طريقة الانحراف عن متوسط فرضي أو الطريقة المختصرة)

1. **حالة بيانات غير مبوبة:** وكانت البيانات والتي عددها N هي $X_1 \quad X_2 \quad X_3 \quad X_4 \quad \ldots\ldots\ldots X_k$ فإنه يمكن اختيار قيمة من البيانات كتوسط حسابي، وليكن هذا المتوسط هو x_0، فنحسب الفروق بطرح x_0 من كل قيم x_i ويكون الفرق هو z_i حيث:

$$z_i = x_i - x_0 \Rightarrow x_i = x_0 + z_i$$

بجمع طرفي المعادلة بالنسبة لكل قيمة مع مراعاة أن x_0 هو مقادر ثابت فغن:

$$\sum x_i = \sum(x_0 + z_i) = \sum z_i + Nx_0$$

بقسمة الطرفين على N نجد

$$\frac{\sum_{i=1}^{k} x_i}{N} = \frac{\sum_{i=1}^{k} z_i}{N} + \frac{Nx_0}{N} \quad \Rightarrow \bar{X} = \bar{Z} + x_0$$

فإذا كان $\bar{x} < x_0$ فإن $\bar{z}$ يكون سالب وبالتالي فجمعه مع c_0 يعني طرحه منه.

أما أذاكان $\bar{x} > x_0$ فإن $\bar{z}$ يكون موجب وبالتالي نضيف قيمت إلى x_0 ، وعليه نكتب

$$\bar{x} = x_0 + \bar{z}$$

مثال(2–4)أوجد المتوسط الحسابي للسلسلة العددية التالية بطريقة المتوسط الفرضي:2 ، 5 ، 8 ، 13 ، 17 ، 22 ، 33.

الحل: نفترض أن متوسط هذه القيم هو $x_0 = 13$

x_i	2	5	8	13	17	22	33	∑
$x_0 - x_i = z_i$	11–	8–	5–	0	4+	9+	20+	9+

نعلم أن

$$\bar{X} = x_0 + \bar{z} = 13 + \frac{9}{7} = 13 + 1.28 = 14.28$$

2. حالة بيانات مبوبة: فإنه إذا كان

أ. **المتغير منفصل**:وكان عدد البيانات هو N وهي $x_1 \ x_2 \ x_3 \ x_4 \ \ldots\ldots\ldots x_k$ والتكرارات المقابلة لها هي $n_1 \ n_2 \ n_3 \ n_4 \ \ldots\ldots\ldots\ldots n_k$ فإنه يمكن اختيار وسط فرضي من بين تلك البيانات، ثم حساب انحرافات القيم عن هذا المتوسط، وضرب تلك الانحرافات في التكرارات المقابلة، ثم جمع تلك الجداءات وايجاد متوسطها ثم اضافته إلى المتوسط الافتراضي

$$z_i = x_i - x_0 \Rightarrow x_i = x_0 + z_i$$

وبالضرب في التكرارت المقابلة نحصل على:

$$\sum n_i x_i = N x_0 + \sum n_i z_i$$

بالقسمة على N نحصل على

$$\bar{X} = x_0 + \bar{Z}$$

ب.**المتغير متصل**:أما إذا كان التوزيع على شكل فئات عددها N ومراكزها هي

$c_1 \ c_2 \ c_3 \ c_4 \ldots\ldots c_k$ فإنه يمكن اختيار وسط حسابي c_0 هو عبارة عن مركز فئة من الفئآت غالباً تكون الفئة الأكثر تكراراً، ثم نحسب الفروقات بأن نطرح c_0 من مراكز فئات التوزيع

$$z_i = c_i - c_0 \Rightarrow c_i = c_0 + z_i$$

بجمع طرفي المعادلة لكل قيمة مع مراعاة أن c_0 هو قيمة ثابتة نجد

$$\sum n_i c_i = N c_0 + \sum n_i z_i$$

بقسمة الطرفين على N نجد

$$\bar{X} = c_0 + \bar{Z}$$

خواص المتوسط الحسابي :

1. أكثر مقاييس النزعة المركزية استعمالاً.
2. سهل الحساب.
3. يأخذ جميع قيم المشاهدات في الحسبان أثناء حسابه.
4. متوسط انحراف القيم عن وسطها الحسابي يساوي الصفر.

نعلم أن $\bar{x} = \frac{\sum x_i}{N}$ ومنه $\sum_{i=1}^{n} x_i = N\bar{x}$

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

$\sum_{i=1}^{n} ni(xi - \bar{x})$ $\sum_{i=1}^{n} nixi-\sum_{i=1}^{n} nixi-N\bar{x}=\sum_{i=1}^{n} nixi-\sum_{i=1}^{n} nixi=0=$

لأن $\sum_{i=1}^{n} nixi= N\bar{x}$

5. مربع انحرافات القيم عن $\bar{x}$ يكون أصغر ما يمكن.

$$\frac{\sum_{I=1}^{n} ni(xi-\bar{x})^2}{N} = \sum_{i=1}^{n} fi(xi - \bar{x})^2=\text{minimum}$$

نفترض أنه يوجد $\bar{x} \neq a$ تكون عنده المساواة $\sum_{i=1}^{n} fi(xi - \bar{x})^2=\text{minimum}$

$\sum_{i=1}^{n} fi(xi - \bar{x})^2\text{minimum}$ هو ثلاثي حدودي من الدرجة الثانية يكون تمثيلها على شكل قطع مكافيء.والمطلوب هو ايجاد قيمة a التي تجعل ثلاثي الحدود أقل ما يمكن.

انطلاقاً من الشرطين التاليين:1. المشتق الأول بالنسبة لa يساوي الصفر.2- المشتق الثاني موجب.

نضع $T(a)= \sum_{i=1}^{n} fi(xi - \bar{x})^2\text{minimum}$ ثم نحدد القيمة التي تجعل$T(a)$ أقل ما يمكن

$$\sum_{i=1}^{n} fi(xi - \bar{x})^2=\sum_{i=1}^{n} fi(xi^2 - 2axi + a^2) = \sum_{i=1}^{n} fi\, x^2 - 2a\sum_{i=1}^{n} fi\ xi\ + a^2\sum_{i=1}^{n} fi^n = \overline{x^2} - 2a\bar{x}+a^2$$

1. $\frac{\partial T(a)}{\partial a}=0 \Rightarrow \frac{\partial(\overline{x^2}-2a\bar{x}+a^2)}{\partial a}=0 \Rightarrow 2\bar{x}-2a=0 \Rightarrow \bar{x}=a$
2. $\partial\left(\frac{\partial T(a)}{\partial a}\right) = 2 > 0$.ومنه فالقيمة التي تجعل ثلاثي الحدود أصغر ما يمكن هي المتوسط الحسابي.

6. يتأثر الوسط الحسابي بالقيم المتطرفة أو الشاذة.
7. لا يمكن أن يكون لأي توزيع تكراري أكثر من متوسط حسابي.
8. عند اضافة قيمة ثابتة إلى كل قيمة من قيم البيانات ولتكن αفإن المتوسط الحسابي للبيانات بعد الإضافة يساوي المتوسط الحسابي قبل اضافة الثابت مضاف إليه قيمة الثابت(يمكن برهنة ذلك).
9. إذا ضربت كل البيانات في قيمة ثابتة ولتكن α فإن المتوسط الجديد للبيانات الجديدة هو عبارة عن متوسط البيانات السابقة مضروباً في القيمة الثابتة α
10. يكون المتوسط الحسابي ممثلاً جيداً للعينة إذا كانت القيم متقاربة من حيث قيمها.
11. يمكن حساب المتوسط الحسابي في حالة الجداول المفتوحة من الأعلى أو من الأسفل أو من الجهتين.
1. تغيير موقع أي مشاهدة لا يؤثر عن الوسط الحسابي.

متى نستخدم الوسط الحسابي؟

نستخدم الوسط الحسابي حينما يكون هدفنا:

- الحصول على مقياس مستقر لا يختلف كثيراً من عينة لأخرى.
- حساب مقاييس احصائية أخرى.
- حين يكون التوزيع متماثلاً حول مركزه.

2.1.1.2.المتوسط الهندسي:geometric mean

المتوسط الهندسي الذي يرمز له بالرمز"G"هو عبارة عن الجذر النوني لحاصل ضرب القيم الموجبة التي يأخذها المتغير الإحصائي، حيث نميز نوعين من المتوسطات الهندسية

المتوسط الهندسي البسيط simple geometric mean

يحسب في حالة بيانات غير مبوبة أي ليست متكررة، فإذا كانت القيم التي يأخذها المتغير الاحصائي هي $x_1 \ x_2 \ x_3 \ x_4 \dots\dots\dots\dots\dots\dots x_k$ فإن متوسطها الهندسي نحسبه وفق الصيغة التالية

$$G=\sqrt[N]{x_1 x_2 x_3 x_4 \dots\dots\dots\dots x_k}$$

مثال(2–5) أوجد المتوسط الهندسي للقيم: 2 ، 5 ، 8 ، 13 ، 17 ، 22، 33.

الحل

$$G=\sqrt[7]{2.5.8.13.17.22.33}=10.36$$

المتوسط الهندسي المرجح:

ونحسبه عند ما تكون القيم التي يأخذها المتغير الإحصائي لها تكرارات تعبر عن ترجيحاتها، فإذا كان

المتغير منفصل:وكانت القيم هي $x_1 \ x_2 \ x_3 \ x_4 \dots\dots\dots\dots\dots x_k$ والتكرارت الترجيحية المقابلة لها على الترتيب هي $n_1 \ n_2 \ n_3 \ n_4 \dots\dots\dots n_k$ فإن متوسطها الهندسي المرجح يحسب وفق الصيغة التالية

$$G=\sqrt[N]{x_1{}^{n_1}\,x_2{}^{n_2}x_3{}^{n_3}x_4{}^{n_4}\,\dots\dots\dots\dots x_k{}^{n_k}}$$

$$G=(x_1{}^{n_1}\,x_2{}^{n_2}x_3{}^{n_3}x_4{}^{n_4}\,\dots\dots\dots\dots x_k{}^{n_k})^{\frac{1}{N}}$$

باستخدام اللوغاريتم

$$log\,G\;=\frac{1}{N}\;log(x_1{}^{n_1}\,x_2{}^{n_2}x_3{}^{n_3}x_4{}^{n_4}\,\dots\dots\dots\dots x_k{}^{n_k})$$

$$\log G=\frac{1}{N}(n_1\log x_1+n_2\log x_2+n_3\log x_2+n_4\log x_4\,\dots\dots\dots n_k\log x_k)$$

$$\log G=\frac{1}{N}\sum_{i=1}^{k}ni\log xi\;\Rightarrow G=(x_1{}^{n_1}x_2{}^{n_1}x_3{}^{n_3}x_4{}^{n_4}\,\dots\dots\dots\dots x_k{}^{n_k})^{\frac{1}{N}}$$

$$G=x_1{}^{\frac{n_1}{N}}x_2{}^{\frac{n_2}{N}}x_3{}^{\frac{n_3}{N}}x_4{}^{\frac{n_4}{N}}\,\dots\dots\dots\dots x_k{}^{\frac{n_k}{N}}=x_1{}^{f_1}x_2{}^{f_2}x_3{}^{f}x_4{}^{f_4}\,\dots\dots\dots\dots\dots\dots\dots\dots x_k{}^{f_k}$$

مثال(2–6):أحسب المتوسط الهندسي للقيم التالية :7 9 11 13 15 17 19

إذا علمت أن التكرارات المقابلة لها هي على التوالي 3 10 26 33 14 8 6

الحل:

19	17	15	13	11	9	7	x_i
4	8	14	33	26	10	3	n_i
0.06	0.08	0.14	0.33	0.26	0.1	0.03	f_i
7.6725	9.8436	16.4653	36.7601	27.0762	9.5424	2.5353	$n_i\log x_i$
$19^{0.06}$	$17^{0.08}$	$15^{0.14}$	$13^{0.33}$	$11^{0.26}$	$9^{0.1}$	$7^{0.03}$	$x_i{}^{f_i}$

$\log G=\frac{1}{N}\sum_{i=1}^{k}ni\log xi=\;\frac{109.8954}{100}\;=1.098954\Rightarrow G=12.5589$

$G=x_1{}^{f_1}x_2{}^{f_2}x_3{}^{f}x_4{}^{f_4}\,\dots\dots\dots\dots\dots\dots\dots\dots x_k{}^{f_k}=7^{0.03}9^{0.1}11^{0.26}13^{0.33}15^{0.14}17^{0.08}19^{0.06}=12.5585$

المتغير متصل: حيث تكون البيانات عندئذٍ هلى شكل فئات مراكزها هي هي $c_1\ c_2\ c_3\ c_4 \ldots\ldots c_k$ والتكرارات المقابلة لها هي $n_1\ n_2\ n_3\ n_4 \ldots\ldots n_k$ فإننا نعتبر أن التكرارات تتجمع حول مراكز الفئات، وتحسب قيمة المتوسط الهندسي عندئذٍ وفق الصيغة

$$G=\sqrt[N]{{c_1}^{n_1}\,{c_2}^{n_2}{c_3}^{n_3}{c_4}^{n_4}\ldots\ldots\ldots\ldots{c_k}^{n_k}}$$

$$G=({c_1}^{n_1}\,{c_2}^{n_2}{c_3}^{n_3}{c_4}^{n_4}\ldots\ldots\ldots\ldots{c_k}^{n_k})^{\frac{1}{N}}$$

$$G={c_1}^{\frac{n_1}{N}}{c_2}^{\frac{n_2}{N}}{c_3}^{\frac{n_3}{N}}{c_4}^{\frac{n_4}{N}}\ldots\ldots\ldots\ldots{c_k}^{\frac{n_k}{N}}={c_1}^{f_1}{c_2}^{f_2}{c_3}^{f}{c_4}^{f_4}\ldots\ldots\ldots\ldots\ldots\ldots\ldots{c_k}^{f_k}=\left(\prod_{i=1}^{k}{c_i}^{n_i}\right)^{\frac{1}{N}}$$

خواص المتوسط الهندسي: characteristis of geometric mean

1. هو انسب وسط للقيم التي تكون متوالية هندسية
2. أكثر ملاءمة من المتوسط الحسابي للحصول على معدل النسب.
3. قليل التأثر بالقيم المتطرفة أو الشاذة.
4. ليس له وجود إذا كان ضمن القيم قيمة سالبة.
5. يتعذر حسابه في حالة التوزيعات المفتوحة.
6. هو مقياس للقيمة وليس للوضع.
7. يكون أصغر دائماً من الوسط الحسابي لنفس البيانات.

حالات تطبيق الوسط الهندسي:

1. يستخدم في مجال الرياضيات المالية، حيث يعتبر أساس حساب القيمة الحالية.
2. يستعمل في حساب الأرقام القياسية(دراسة التغيرات النسبية).
3. يستعمل في حساب معدلات النمو.
4. معدل نمو الناتج، معدل زيادة الأجور والنمو السكاني

3.1.1.2.الوسط التوافقي:Harmonic mean

يعرف الوسط التوافقي الذي يرمز له بالرمز "H " لمجموعة من القيم $x_1 \ x_2 \ x_3 \ x_4 \dots\dots\dots\dots x_k$ التي يأخذها المتغير الاحصائي بأنه مقلوب الوسط الحسابي لمقلوب قيم المتغير الاحصائي، حيث نميز بين نوعين من المتوسطات التوافقية:

1. **المتوسط التوافقي البسيط: The simp Harmonic mean le**

إذا كانت القيم التي يأخذها المتغير الاحصائي هي $x_1 \ x_2 \ x_3 \ x_4 \dots\dots\dots\dots x_k$ ، ومقلوب تلك القيم هي على التوالي $\frac{1}{x_1} \ \frac{1}{x_2} \ \frac{1}{x_3} \ \frac{1}{x_4} \dots\dots\dots\dots \frac{1}{x_k}$

بجمع هذه القيم نحصل على:

$$\frac{1}{x_1}+\ \frac{1}{x_2}+\ \frac{1}{x_3}+\ \frac{1}{x_4}+\cdots\dots\dots\dots+\frac{1}{x_k}\ =\sum_{i=1}^{k}\frac{1}{x_i}$$

$$\frac{1}{H}=\frac{1}{N}\left(\sum_{i=1}^{k}\frac{1}{x_i}\right)\Rightarrow H=\frac{N}{\sum_{i=1}^{k}\frac{1}{x_i}}\Rightarrow H=\frac{1}{\frac{\sum_{i=1}^{k}\frac{1}{x_i}}{N}}$$

مثال (**2–7**) أوجد الوسط التوافقي للقيم 5 7 9 11

الحل:

$$H=\frac{N}{\sum_{i=1}^{k}\frac{1}{x_i}}=\frac{4}{\frac{1}{5}+\frac{1}{7}+\frac{1}{9}+\frac{1}{11}}=\frac{4}{\frac{693+495+385+315}{3465}}=\frac{13360}{1888}=7.34$$

مثال(**2–8**):

يقطع مسافر مسافة 10 كلم على طريق خارج المدينة بسرعة 60كلم/سا، ومسافة قدرها 10 كلم داخل المدينة بسرعة 15 كلم/سا.فماهو متوسط السرعة في الساعة.

الحل: يستخدم المتوسط التوافقي أساساً لإيجاد متوسط المعدلات

$$H = \frac{N}{\sum_{i=1}^{k}\frac{1}{x_i}} \Rightarrow \mathrm{H} = \frac{2}{\frac{1}{60}+\frac{1}{15}} = \frac{2}{\frac{1+4}{60}} = \frac{120}{5} = 24\ \mathrm{Km/h}1.$$

ومنه فلحساب المتوسط التوافقي يجب:

1. حساب مقلوب كل قيمة.
2. حساب مجموع مقلوبات تلك القيم.
3. تقسيم المجموع على عدد القيم.
4. حساب مقلوب القيمة هو الوسط التوافقي.

2. **المتوسط التوافقي المرجح**: إذا كنت القيم التي يأخذها المتغير الإحصائي هي $x_4..x_k \quad x_3 \quad x_2 \quad 1$ والتكرارات المقابلة هي $n_1\ n_2\ n_3\ n_4 \dots\dots\dots n_k$ فإن وسطها التوافقي هو

$$\frac{1}{H} = \frac{1}{N}\left(\frac{n_1}{x_1} + \frac{n_2}{x_2} + \frac{n_3}{x_3} + \frac{n_4}{x_4} \dots\dots\dots \frac{n_k}{x_k}\right)$$

$$\frac{1}{H} = \frac{1}{N}\sum_{i=1}^{k}\left(\frac{n_i}{x_i}\right) \Rightarrow \mathrm{H} = \mathrm{H} = \frac{N}{\sum_{i=1}^{k}\left(\frac{n_i}{x_i}\right)}$$

مثال (2–9)

قطع سائق المسافة الفاصلة بين مدينتين على أربع مراحل متساوية، المسافة المقطوعة في كل منها 100 كم، فإذا قطع المرحلة الأولى بسرعة 100 كم/ساعة، والمرحلة الثانية بسرعة 120 كم/ساعة والمرحلة الثالثة بسرعة 150 كم/ساعة والمرحلة الرابعة بسرعة 80 كم/ساعة.

المطلوب: أوجد متوسط سرعة هذا السائق على طول المرحلة ؟

الحل:

في هذه الحالة المتوسط الحسابي لا يعطي فكرة صحيحة عن مثل هذه الظواهر، لذا نقوم باستخدام المتوسط التوافقي لأنه أفضل مقياس لقياس متوسط السرعة.

متوسط سرعة السائق على طول المرحلة هو:

$$H = \frac{100+100+100+100}{\frac{100}{100}+\frac{100}{120}+\frac{100}{150}+\frac{100}{80}} = \frac{400}{1+0{,}83+0{,}67+1{,}25} = \frac{400}{3{,}75} = 106{,}67 km/h$$

وعليه فإن متوسط سرعة السائق على طول المرحلة هو: 106,67 كم/ساعة.

أما إذا كانت القيم مجمعة في فئات مراكزها هي $c_1\ c_2\ c_3\ c_4 \dots \dots \dots c_k$ وتكرارتها هي $n_1\ n_2\ n_3\ n_4 \dots \dots \dots n_k$ فإن

$$H = \frac{N}{\frac{n_1}{c_1}+\frac{n_2}{c_2}+\frac{n_3}{c_3}+\frac{n_4}{c_4}+\cdots\cdots\cdots\frac{n_k}{c_k}} = \frac{N}{\sum_{i=1}^{k}\left(\frac{n_i}{c_i}\right)}$$

خواص المتوسط التوافقي characteristis of harmonic mean

1. يستخدم جميع قيم البيانات.
2. يتأثر قليلاً بالقيم المتطرفة.
3. لا يمكن حسابه في التوزيعات المفتوحة.
4. مقياس للقيمة وليس للوضع.
5. يتعذر حسابه إذا كانت إحدى القيم معدومة.
6. يفضل اسيتخدامه عن باقي المتوسطات الأخرى في حالة حساب معدلات السرعة بالنسبة للزمن أو معدلات التغير في الإنتاج ببعض المصانع والآلات أو متوسطات الاسعار بدلالة وحدة النقود.
7. $\bar{X} > G > H$

حالات تطبيق الوسط التوافقي:

- يستخدم في مسائل المعدلات مثل ايجاد متوسط القراءة لمجموعة من الأفراد بدلالة عدد الكلمات في الدقيقة .
- ايجاد متوزسط الأسعار بدلالة سعر السلعة الواحة.

تمرين: بيِّن أن $\frac{2}{\frac{1}{x_1}+\frac{1}{x_2}} < x_1x_2 < \frac{x_1+x_2}{2}$

الحل:

1): إثبات أن $x_1x_2 < \frac{x_1+x_2}{2}$

$x_1، x_2 > 0 \Rightarrow (\sqrt{x_1} - \sqrt{x_2})^2 > 0 \Rightarrow x_1 + x_2 - 2\sqrt{x_1 x_2} > 0$

$\Rightarrow x_1 + x_2 > 2\sqrt{x_1 x_2}$ *وهو المطلوب اثباته*

بالقسمة على 2 نجد

$\Rightarrow \frac{x_1+x_2}{2} > \sqrt{x_1 x_2}$

2):اثبات أن $\frac{2}{\frac{1}{x_1}+\frac{1}{x_2}} < x_1 x_2$

$\frac{2}{\frac{1}{x_1}+\frac{1}{x_2}} = \frac{2}{\frac{x_1+x_2}{x_1 x_2}} = \frac{x_1 x_2}{x_1+x_2} < \sqrt{x_1 x_2}$

لقد سبق إثبات أن $\sqrt{x_1 x_2} < \frac{x_1+x_2}{2}$ وبضرب الطرفين في $2\sqrt{x_1 x_2}$ نجد

$2x_1 x_2 < (x_1 + x_2)\sqrt{x_1 x_2}$

بقسمة الطرفين على $_1 + x_2$ نحصل على

$\frac{2x_1 x_2}{x_1+x_2} < \sqrt{x_1 x_2}$ *وهو المطلوب اثبات. ومنه* $\frac{2}{\frac{1}{x_1}+\frac{1}{x_2}} < x_1 x_2 < \frac{x_1+x_2}{2}$

4.1.1.2.الوسط التربيعي Quadratic mean

الوسط التربيعي لمجموعة من القيم هو عبارة عن الجذرالتربيعي لمجموع مربعات القيم على عددها أي الجذر التربيعي للوسط الحسابي لمربعات القيم، يرمز له بالرمز "Q"، حيث نميز حالتين:

حالة البيانات غير المبوبة(الوسط التربيعي البسيط).

إذا كانت القيم التي يأخذها المتغير الإحصائي هي $x_1 \quad x_2 \quad x_3 \quad x_4 \ldots\ldots\ldots x_k$ *فإن وسطها* التربيعي يعطى بالعلاقة

$$Q^2 = \frac{x_1^2+x_2^2+x_3^2+x_4^2\ldots\ldots\ldots\ldots+x_k^2}{n_1+n_2+n_3+n_4\ldots\ldots\ldots\ldots n_k} = \frac{\sum_{i=1}^{k} x_i^2}{N} \Rightarrow Q = \sqrt{\frac{\sum_{i=1}^{k} x_i^2}{N}}$$

حالة بيانات مبوبة

إذا كانت القيم التي ياخذها المتغير الإحصائي هي $x_1 \ x_2 \ x_3 \ x_4 \dots \dots \dots x_k$ *والتكرارات المقابلة لها هي* $n_1 \ n_2 \ n_3 \ n_4 \dots \dots \dots n_k$ على الترتيب فإن وسطها التربيعي يعرف بالعلاقة

$$Q^2 = \frac{n_1x_1^2+n_2x_2^2+n_3x_3^2+n_4x_4^2 \dots\dots\dots\dots n_kx_k^2}{n_1+n_2+n_3+n_4 \dots\dots\dots\dots n_k} = \frac{\sum_{i=1}^{k} n_i x_i^2}{\sum_{i=1}^{k} n_i} \Rightarrow Q = \sqrt{\frac{\sum_{i=1}^{k} n_i x_i^2}{N}}$$

أما إذا كانت البيانات موزعة ضمن فئات مراكزها هي $c_1 \ c_2 \ c_3 \ c_4 \dots \dots \dots c_k$ وتكرارتها هي $n_1 \ n_2 \ n_3 \ n_4 \dots \dots \dots n_k$ فإن وسطها التربيعي يحسب وفق الصيغة

$$Q^2 = \frac{n_1C_1^2+n_2C_2^2+n_3C_3^2+n_4C_4^2 \dots\dots\dots\dots n_kC_k^2}{n_1+n_2+n_3+n_4 \dots\dots\dots\dots n_k} = \frac{\sum_{i=1}^{k} n_i C_i^2}{\sum_{i=1}^{k} n_i} \Rightarrow Q = \sqrt{\frac{\sum_{i=1}^{k} n_i C_i^2}{N}}$$

خواص الوسط التربيعي:

1. يستخدم جميع القيم.
2. يتأثر بالقيم المتطرفة.
3. لا يمكن حسابه في الجداول المفتوحة.
4. يستخدم في بعض المقاييس الأخرى (التباين، الانحراف المعياري).

نسمي المتوسطات السابقة بمتوسطات القيمة، حيث تنتسب إلى مجموعة المتوسطات المرفوعة غلى قوة ما.حيث تأخذ الصيغة العامة لهذه المتوسطات الصيغة التالية:

$$\bar{Y} = \sqrt[m]{\frac{\sum x_i^m}{N}} = \left(\sum_{i=1}^{n} x_i^m\right)^{\frac{1}{m}}$$

بحيث أن:

- $\bar{Y}$ يمثل الوسط المطلوب حسابه.
- X قيم المتغير
- N عدد وحدات هذا المتغير.

- m تمثل درجة هذا المتوسط والتي تحدد نوعه، فإذا كانت:

$$m=1 \Rightarrow \bar{Y}=\frac{\sum n_i x_i}{N}=\bar{x}$$

$$m=2 \Rightarrow \bar{Y}=\left(\frac{\sum n_i x_i^2}{N}\right)^{\frac{1}{2}}=Q$$

$$m=3 \Rightarrow \bar{Y}=\left(\frac{\sum n_i x_i^3}{N}\right)^{\frac{1}{3}}=$$ الوسط التكعيبي

$$m=-1 \Rightarrow \bar{Y}=\left(\frac{\sum n_i x_i^{-1}}{N}\right)^{\frac{1}{-1}}=\frac{1}{\frac{\sum n_i x_i^{-1}}{N}}=\frac{N}{\sum \frac{n_i}{x_i}}=H$$

2..متوسطات الوضع:measures of location

ويشمل هذا النوع من المتوسطات على:

1.2.2 المنوال The mode

يعرف المنوال الذي يرمز له بالمرمز M_0 بأنه قيمة المتغير الأحصائي التي يقابلها أكبر تكرار مطلق أو نسبي ، ويعرف أيضاً بالقيمة السائدة او الشائعة(dominante) أو الأكثر تكرار، ونميز في حسابه بين:

أ. حالة متغير منفصل:

تكون في هذه الحالة قيمة المنوال دقيقة، وهي القيمة الأكثر تكرارً في مجموع البيانات، وهو ما يعني أن وجود المنوال مرهون بوجود تكرارات مختلفة لكل قيمة من قيم المجتمع الإحصائي، فإذا لم تكن هناك تكرارات أو كانت متساوية فلا يوجد منوال عند ئذٍ

مثال(2–10) لتكن علامات محصل عليها في مقياس الاحصاء لتخصص س.

العلامات	5	6	8	9	10	12	13	14	المجموع
عدد الطلبة	4	8	9	18	10	25	16	10	100

ويكون بذلك منوال هذا التوزيع هو12 = M_0 لكون القيمة 12 يقابلها التكرار الأكبر.

ب. حالة متغير متصل:

في هذه الحالة نتكلم عن الفئة المنوالية La classe modale التي يقابلها أكبر تكرار متوسط للوحدة الطولية، وهي الفئة التي تقع ضمنها قيمة المنوال،وهو النهاية العظمى أو النهايات العظمى للمنحنى أو المدرج التكراري، او المدرج التكراري المعدل، وتكون قيمة المنوال في هذه الحالة تقريبية، ويمكن تحديد قيمة المنوال حسابياً وبيانياً.

1. **أطوال الفئات متساوية:** إذا كانت الفئات متساوية الطوال فإن الفئة المنوالية هي الفئة التي يقابلها أكبر تكرار مطلق أو نسبي في الجدول أو المدرج التكراري.ولحساب قيمة المنوال نستخدم:
   أ. **الطريقة الجبرية:** حيث هناك عدة طرق منها

طريقةالفروق لـ Pearson

حيث أن المنوال يقسم الفئة المنوالية إلى قسمين يتناسبان مع الفرق بين تكرار الفئة المنوالية وتكرار الفئتين السابقة واللاحقة للفئة المنوالية. فإذا كان طول الفئة المنوالية هو Δ فإن المنوال يقسمها إلى المسافة x التي تمثل بعد قيمة المنوال عن الحد الأدنى للفئة المنوالية ، و$\Delta - x$ وهي بعد قيمة عن الحد الأعلى للفئة المنوالية، ومنه فإن

$$\frac{X}{\Delta - x} = \frac{f_m - f_1}{f_m - f_2} \Rightarrow x(f_m - f_2) = (\Delta - x)(f_m - f_1)$$

$$x(f_m - f_2) = \Delta(f_m - f_1) - x(f_m - f_1) \Rightarrow x[(f_m - f_1) + (f_m - f_2)] = \Delta(f_m - f_1) \Rightarrow$$

$$x = \frac{\Delta(f_m - f_1)}{(f_m - f_1) + (f_m - f_2)}$$

وبما أن المنوال هو عبارة عن الحد الأدنى للفئة المنوالية مضافاً إليه البعد x تصبح قيمة المنوال تحسب وفق العلاقة

$$M_0 = L_0 + \frac{\Delta(f_m - f_1)}{(f_m - f_1) + (f_m - f_2)}$$

حيث :

L_0 هو الحد الأدنى للفئة المنوالية.

f_m تكرار الفئة المنوالية

f_1 تكرار الفئة السابقة للفئة المنوالية.

f_2 تكرار الفئة اللاحقة للفئة المنوالية.

هناك طريقة أخرى تعرف بطريقة العزوم أو الرافعة لـ kong حيث تفترض الطذريقة وجود رافعة تتجاذبها قوتنان بحيث يكون: قْ1x ذ1=ق2 x 2ذ ا] القوة في الذراع = المقاومة في الذراع

القوة تمثل الفرق بين تكرار الفئة المنوالية وتكرار الفئة السابقة لها ويعمل عند بداية الفئة

القوة الثانية تمثل الفرق بين تكرار الفئة المنوالية وتكرار الفئة اللاحقة لها ويعمل عند نهايتها أي أن

$$\frac{X}{\Delta - x} = \frac{f_m - f_1}{f_m - f_2} \Rightarrow x(f_m - f_2) = (\Delta - x)(f_m - f_1)$$

$$x(f_m - f_2) = \Delta(f_m - f_1) - x(f_m - f_1) \Rightarrow x[(f_m - f_1) + (f_m - f_2)] = \Delta(f_m - f_1)$$

$$\Rightarrow x = \frac{\Delta(f_m - f_1)}{(f_m - f_1) + (f_m - f_2)}$$

2. **أطوال الفئآت غير متساوية:**

نقوم في هذه الحالة بإجراء التعديل التكراري، حيث تنسب التكرارات إلى نفس وحدات الطول ثم نحسب المنوال بنفس الطريقة السابقة.

مثال(2–11) ليكن المثال التالي الذي يمثل الأجر الساعي لعمال مؤسسة.

الفئآت الأجرية	niعدد العمال	aiطول الفئة	ai/a	التكرار المعدل
20-10	12	10	1	12
30-20	13	10	1	13
50-30	32	20	2	16
80-50	47	30	3	15.66
90-80	15	10	1	15

حيث a هو القاسم المشترك الأكبر لأطوال الفئآت وهو 10.

ب. **الطريقة البيانية**:لتحديد المنوال بيانياً نتبع الخطوات التالية:

- نرسم المدرج التكراري الذي يظهر لنا الفئة المنوالية التي هي ذات التكرار الأكبر.
- نوصل الحد الأعلى للفئة السابقة للفئة المنوالية بالحد الأعلى للفئة المنوالية، ونوصل الحد الأدنى للفئة اللاحقة للفئة اللاحقة بالحد الأدنى للفئة المنوالية.
- نسقط نقطة تقاطع الخطين السابقين على المحور الأفقي فنحصل على المنوال الذي هو بعد نقطة تقاطع الشاقول مع الحط الأفقي

خواص المنوال:Mode characteristics

- المنوال هو أسوأ مقاييس النزعة المركزية لأنه يتأثر بطول الفئة في التوزيع وتتغير قيمته بتغير عدد الفئات.لذلك فهو أقلها استعمالاً.
- يمكن فهمه بسهولة لأنه القيمة الأكثر شيوعا أو الأكثر تكراراً.
- عديم الفائدة في خالة وجود عدد قليل من البيانات.
- لا يعتمد في حسابه على جميع البيانات.
- يمكن وجود أكثر من منوال لنفس المجموعة من البيانات.
- لا يمثل القيمة الوسطى في التوزيع كما هو الأمر بالنسبة للوسيط.
- لا يتأثر بالقيم المتطرفة أو الشاذة.
- تتأثر قيمة المنوال بحدجم العينة وبطول الفئة وبطريقة ترتيب البيانات،فهو مقياس غير ثابت.

2-3:متوسطات التكرار: ويرتبط حسابها بعدد التكرارت، واهمها:

1.3.2. الوسيط The median

الوسيط لمجموعة من البيانات المرتبة ترتيباً تصاعدياً أو تنازلياً هو المشاهدة الوسطى أو العدد الأوسط إذا كان عدد تلك البيانات فردي، أو الوسط الحسابي للقيمتين الوسطيين إذا كان عدد تلك البيانات زوجي، فقيمة الوسيط تقسم البيانات إلى مجموعتين متساويتين من حيث العدد أي هو القيمة التي يقل عنها عدد من البيانات مساوٍ لعدد البيانات التي تكبرها، أي هو المفردة من البيانات التي تقع في منتصف البيانات، حيث يكون عندئذ 50 % من البيانات أقل منها و 50 % من البيانات أكبر منها، ويرمز للوسيط بالرمز M_e.

الوسيط للبيانات الأولية (غير المبوبة):

لحساب الوسيط من البيانات الأولية يجب أولا ترتيب البيانات (القيم) تصاعديا أو تنازليا وهنا نميز حالتين:

الوسيط للبيانات غير المبوبة المفردة:

إذا كان عدد القيم (n) فردي فإن الوسيط هو القيمة التي ترتيبها $\frac{n+1}{2}$، وبالتالي قيمة الوسيط هي المفردة (القيمة) التي ترتيبها يقابل رتبة الوسيط ($M_e = X_{\frac{n+1}{2}}$).أما إذا كان عدد البيانات زوجي فإن قيمة الوسيط هي المتوسط الحسابي للقيمتين اللتين ترتيبهما على التوالي $\frac{n}{2}$ و $\left(\frac{n}{2}+1\right)$ أي أن $M_{e=}\frac{x_{\frac{n}{2}}+x_{\left(\frac{n}{2}+1\right)}}{2}$

مثال(2-12): أوجد الوسيط للبيانات التالية:

1): 8 9 3 6 12 15 10

2): 9 18 17 8 13 14 10 15

الحل: نقوم بترتيب بيانات السلسلتين

x_1	x_2	x_3	x_4	x_5	x_6	x_7
3	6	8	9	10	12	15

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
8	9	10	13	14	15	17	18

1): عدد القيم فردي وعليه فإن رتبة الوسيط هي ، وبالتالي قيمة الوسيط هي المفردة (القيمة) التي ترتيبها يقابل رتبة الوسيط $M_e = X_{\frac{n+1}{2}} = X_{\frac{7+1}{2}} = X_4 = 9$

2) عدد القيم زوجي ، وعليه فإن قيمة الوسيط هي متوسط القيمتين التين ترتيبهما على التوالي $\frac{\frac{n}{2}+\left(\frac{n}{2}+1\right)}{2}$

أي هي متوسط القيمتين x_4 x_5 أي $M_e = \frac{13+14}{2} = 13.5$

الوسيط في حالة بيانات مبوبة: لحساب الوسيط لتوزيع تكراري نتبع الاخطوت التالية:

- نحسب التكرار المتجمع الصاعد أو النازل.
- تحديد ترتيب الوسيط، وذلك بقسمة عدد البيانات على 2 وذلك لتحديد الفئة الوسيطية أي الفئة التي يقع ضمنها الوسيط.ويكون بذلك الوسيط هو:

1. **حالة متغير منفصل:**هو القيمة الأولى التي يعادل أو يزيد تكرارها المتجمع الصاعد عن نصف عدد البيانات
2. **حالة متغير مستمر**: وهنا تكون لدينا مجموعة من المشاهدات المجمعة في فئات مراكزها هي $C_1 \ C_2 \ C_3 \ C_4 \ \dots \dots \dots C_k$ فإن التكرارات المقابلة لها على الترتيب،لإن وسيطها يحدد وفق طريقتين:

أ. **الطريقة الرياضية:** نحسب في هذه الحالة الوسطي عن طريق التكرارات التجميعية الصاعدة أو النازلة كما يلي:يجب تحديد الفئة الوسيطية التي هي الفئة الأولى التي يساوي تكرارها المتجمع الصاعد أو يزيد عن نصف عدد البيانات($\frac{N}{2}$)،حيث يكون الوسيط منتمياً لهذه الفئة الوسيطية حيث يحسب الوسيط في حالة متغير مستمر طبقاً للصيغتين التاليتين:

الوسيط عن طريق التكرارات التجميعية الصاعدة $(\uparrow)$ يعرف بالعلاقة $$M_e = a + \left(\frac{\frac{N}{2} - F_1}{F_m}\right).\Delta$$

الوسيط عن طريق التكرارات التجميعية الهابطة $(\downarrow)$ يعرف بالعلاقة) $$M_e = b - \left(\frac{\frac{N}{2} - F_2}{F_m}\right).\Delta$$

بحيث أن:

Δ طول الفئة الوسيطية

a : الحد الأدنى للفئة الوسيطية.

b :الحد الأعلى للفئة الوسيطية.

$\frac{N}{2}$ ترتيب الوسيط

F_m تكرار الفئة الوسيطية .

F_1 التكرارالمتجمع الصاعد للفئةة السابقة للفئة الوسيطية

F_2 التكرار المتجمع النازل للفئة اللاحقة للفئة الوسيطية

ملاحظة: يحظ في الصيغتين السابقتين أن في حالة التكرارات الصاعدة نضيف إلى الحد الأدنى للفئة الوسيطية المقدار $\Delta.\left(\frac{\frac{N}{2}-F_1}{F_m}\right)$ وفي حالة التكرارات المتجمع النازل ننقص من الحد الأعلى للفئة الوسيطية المقدار $\Delta.\left(\frac{\frac{N}{2}-F_2}{F_m}\right)$

ب. الطريقة البيانية

يمكن تحديد قيمة الوسيط من خلال رسم المنحنى المتجمع الصاعد أو النازل أو كليهما معاً كما يلي:

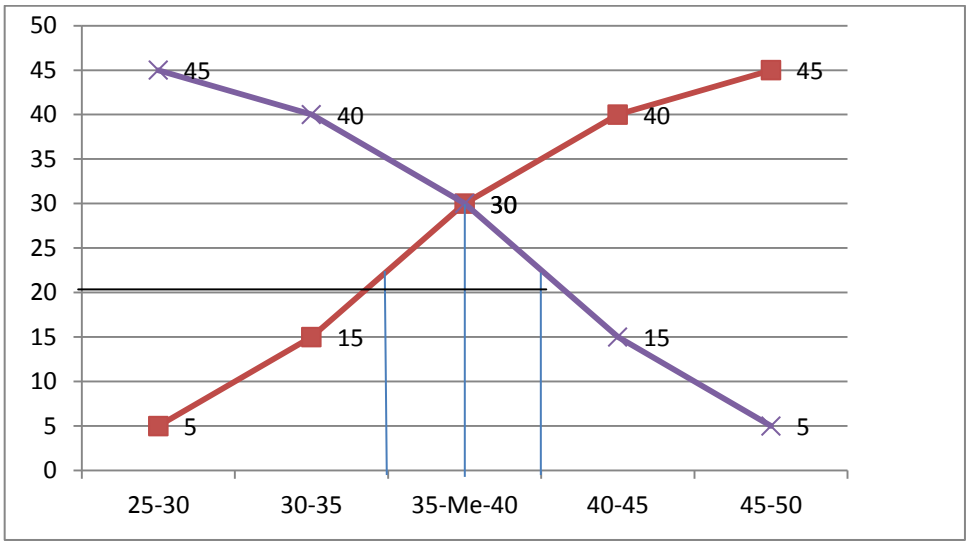
- يمكن تقدير قيمة الوسيط من المنحنى المتجمع الصاعد أو النازل أو كليهما معاً حيث أنه في حالة وجودهما معاً، نقوم باسقاط نقطة تقاطعهما على المحور الأفقي فتكون نقطة تقاطع الشاقول مع المحور الافقي هي الوسيط.كما هو موضع في المثال أدناه

مثال(2-13)لتكن لدينا مجموعة من البيانات المجمعة في الجدول التالي

الفئات	ni	ni↑	ni↓
25-30	05	05	54
30-35	10	15	40
35-40	15	30	30
40-45	10	40	15
45-50	05	45	05

Σ	45	-	-

رسم المنحنى المتجمع الصاعد والنازل



ثم نحدد قيمة نصف عدد البيانات والتي ترتيب الوسيط وهي $\frac{}{2}$ $=\frac{45}{2}=22.5$ ومنه قيمة الوسيك تنتمي إلى الفئة $[35-40]$أي أن $M_e \in [35-40]$ أي أن a=35 و b=40 f_1= 15 f_m= 15 Δ=5

وبذلك تكون قيمة الوسيط هي $M_e=a+\left(\frac{\frac{N}{2}-F_1}{F_m}\right).\Delta = 35 + \frac{22.5-15}{15}. 5 = 37.5$

$$M_e=b-\left(\frac{\frac{N}{2}-F_1}{F_m}\right).\Delta = 40 - \frac{22.5-15}{15}. 5 = 37.5$$

خواص الوسيط: The mean characteristics

1. الوسيط سهل التعريف وسهل الحساب لأنه قيمة موضعية.
2. الوسيط لا يتأثر بالقيم المتطرفة أو الشاذة (عديم التحيز).
3. الوسيط يتأثر بعدد المشاهدات.
4. لا يأخذ بعين الاعتبار جميع قيم المشاهدات أثناء حسابه بل يعتمد على البعض منها فقط.
5. الوسيط لا يخضع للعمليات الجبرية.
6. يأخذ بعين الاعتبار موقع القيم وليس متوسطها.

7. يمكن حساب قيمته في حالة الجداول المفتوحة شريطة عدم وجود قيمته في الفئة المفتوحة.
8. يمكن إيجاد الوسيط من التوزيعات التكرارية غير المنتظمة ذات أطوال الفئات غير المتساوية
9. أقل ثباتاً من الوسط الحسابي.
10. لا يمكن حساب وسيط عام لعدد من العينات المعروفة قيم وسيط كل منها كما هو الحال فى المتوسط الحسابى
11. يمكن إيجاد الوسيط بالرسم .
12. هو قيمة موضعية فقط.
13. الوسيط غير شائع الاستعمال.

متى نستخدم الوسيط؟.

نستخدم الوسيط عند ما:

1. يكون التوزيع التكراري مفتوح من جهة أو من الجهتين شريطة عدم تركز البيانات في الفئة المفتوحة.
2. نحتاج إلى حساب مقياس سريع.
3. يكون التوزيع ملتوٍ بشكلٍ واضحٍ.

2.3.2.شبيهات الوسيط:

يقترن بالوسيط كقيمة موضعية كلٌ من الربيعيات والعشيريات والمئينيات، إذ أنه إذا كنَّا قد قلنا سابقاً بأن الوسيط يقسم البيانات إلى قسمين متساويين من حيث عدد البيانات فإنه يمكننا تعميم فكر الوسيط، بحيث نفكر في ايجاد القيم التي تقسم البيانات إلى أكثر من مجموعتين متساويتين في عدد البيانات،

- فإذا قسمنا هذه البيانات إلى أربعة أقسام متساوية ندعوها بالربيعيات.
- وإذا قسمناها إلى عشرة أقسام متساوية ندعوها بالعشيريات.
- وإذا قسَّمناها إلى مائة قسم متاوية ندعوها بالمئينيات.

فهذه كلها من المقاييس الموضعية، يطلق عليها اسم شبيهات الوسيط *quasi mediane* لتشابهها معه في التعريف وطريقة الحساب، حيث أنه لحساب أيٍ منها نستخدم نفس الصيغة المطبقة في حساب قيمة الوسيط مع تغيير رتبة المقياس فقط، التي ينتج عن تغييرها، تغير الفئة ، وبالتالي التكرارت المقابلة والفئات السابقة واللاحقة.

1.2.3.2.الربيعيات *Quartiles*

هي ثلاثة قيم يأخذها المتغير الإحصائي، تقسم البيانات إلى أربعة أجزاء متساوية من حيث عدد البيانات، بحيث أن الربيعي Q_1 ترتيبه 25 % من البيانات أي $\frac{N}{4}$، حيث يقع في نهاية الربع الأول من البيانات أي أن $\frac{1}{4}$ من البيانات أقل منه و$\frac{3}{4}$ من البيانات أكبر منه،الربيعي الثاني Q_2 ترتيبه 50 % من البيانات أي $\frac{2N}{4}$ ، حيث يقع في منتصف البيانات، ومنه فقيمته مساوية لقيمة الوسيط، الربيعي الثالث Q_3 ترتيبه 75 % اي $\frac{3N}{4}$ أي أن ثلاثة أرباع البيانات أقل من قيمته وربعها أكبر من قيمته. ويمكن حساب الربيعيات وفق الصيغة التالية:

أ. **حالة بيانات غير مبوبة:**

1. **عدد البيانات فردي:** إذا كان عدد البيانات فردي فإن الربيعيات تحدد وفق الصيغة التالية

$$Q_i = X_{i\left(\frac{n+1}{4}\right)} \quad \text{حيث} \quad i \in [1;2,3]$$

$$I=1 \Rightarrow Q_1 = X_{1\left(\frac{n+1}{4}\right)}$$

$$I=2 \Rightarrow Q_2 = X_{2\left(\frac{n+1}{4}\right)} = X_{\left(\frac{n+1}{2}\right)} = M_e$$

$$I=3 \Rightarrow Q_3 = X_{3\left(\frac{n+1}{4}\right)}$$

مثال(2–14)أوجد الربيعيات الثلاث للقيم التالية:7 4 8 14 10 9 13 16 12 11 6

الحل: ترتيب البيانات

X_{11}	X_{10}	X_9	X_{78}	X_7	X_6	X_5	X_4	X_3	X_2	X_1
16	*14*	*13*	*12*	*11*	*10*	*9*	*9*	*7*	*6*	*4*

بما أن عدد البيانات فردي، وهو 11 فإن:

$$I = 1 \Rightarrow Q_1 = X_{1\left(\frac{n+1}{4}\right)} = X_{1\left(\frac{11+1}{4}\right)} = X_3 = 7$$

$I = 2 \Rightarrow Q_2 = X_{2\left(\frac{11+1}{4}\right)} = X_{\left(\frac{11+1}{2}\right)} = X_6 = 10$

$I=3 \Rightarrow Q_3 = X_{3\left(\frac{11+1}{4}\right)} = X_9 = 13$

2. عدد البيانات زوجي: إذا كان عدد البيانات زوجي فإن الربيعيات تحسب وفق الصيغة التالية:

$$\boldsymbol{Q_I = \frac{x_{\left(i\frac{n}{4}\right)} + x_{\left(i\frac{n}{4}+1\right)}}{2}} \qquad \boldsymbol{i \in [1,2,3]}$$

$$I=1 \Rightarrow Q_1 = \frac{x_{\left(\frac{n}{4}\right)} + x_{\left(\frac{n}{4}+1\right)}}{2}$$

$$I=2 \Rightarrow Q_2 = \frac{x_{\left(2\frac{n}{4}\right)} + x_{\left(2\frac{n}{4}+1\right)}}{2} = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2} = M_e$$

$$I=3 \Rightarrow Q_3 = \frac{x_{\left(3\frac{n}{4}\right)} + x_{\left(3\frac{n}{4}+1\right)}}{2}$$

مثال(2–15) أوجد الربيعيات الثلاث للقيم التالية: 4 6 8 12 16 20 24 26 28 32 36 40

الحل: ترتيب البيانات

X_{12}	X_{11}	X_{10}	X_9	X_{78}	X_7	X_6	X_5	X_4	X_3	X_2	X_1
40	36	32	28	26	24	20	16	12	8	6	4

بما أن عدد البيانات زوجي وهو 12 فإنه إذا كان:

$$I=1 \Rightarrow Q_1 = \frac{x_{\left(\left(\frac{n}{4}\right)\right)} + x_{\left(\left(\frac{n}{4}+1\right)\right)}}{2} \Rightarrow Q_1 = \frac{x_{\left(\left(\frac{12}{4}\right)\right)} + x_{\left(\left(\frac{12}{4}+1\right)\right)}}{2}$$

$$\Rightarrow Q_1 = \frac{x_3 + x_4}{2} \Rightarrow Q_1 = \frac{8+12}{2} \Rightarrow Q_1 = \frac{20}{2} = \boldsymbol{10}$$

$$I=2 \Rightarrow Q_2 = \frac{x_{\left(2\frac{n}{4}\right)} + x_{\left(2\frac{n}{4}+1\right)}}{2} \Rightarrow Q_2 = \frac{x_{\left(2\frac{12}{4}\right)} + x_{\left(2\frac{12}{4}+1\right)}}{2}$$

$$\Rightarrow Q_2 = \frac{x_6 + x_7}{2} \Rightarrow Q_2 = \frac{20+24}{2} \Rightarrow Q_2 = \boldsymbol{22}$$

$$I=3 \Rightarrow Q_3 = \frac{x_{\left(3\frac{n}{4}\right)} + x_{\left(3\frac{n}{4}+1\right)}}{2} \Rightarrow Q_3 = \frac{x_{\left(3\frac{12}{4}\right)} + x_{\left(3\frac{12}{4}+1\right)}}{2}$$

$$\Rightarrow Q_3 = \frac{x_9 + x_{10}}{2} \Rightarrow Q_3 = \frac{28+x_{32}}{2} \Rightarrow Q_3 = \boldsymbol{30}$$

$$\mathbf{Q_i}=\boldsymbol{a_i}+\left(\frac{\boldsymbol{iN}}{\boldsymbol{4}}-\boldsymbol{F1}\right).\Delta\boldsymbol{i} \qquad \boldsymbol{i}=\mathbf{1;2,3}$$

ب. حالة بيانات مبوبة: إذا كان :

1. **المتغير منفصل:**فإن قيمة الربيعي هي القيمة المتضمنة لترتيب الربيعي بعد تشكيل التكرار المتجمع الصاعد.
2. **المتغير متصل:**فإن قيمة الربيعيات تحسب بنفس طريقة حساب الوسيط باستثناء تغيير الترتيب الذي يؤدي إلى تغيير الفئة محل الدراسة، ومنه فإن:

حيث أن :

Q_i *رقم الربيعي.*

$\boldsymbol{a_i}$ هو الحد الأدنى للفئة الربيعية$\boldsymbol{i}$.

$\frac{\boldsymbol{iN}}{\boldsymbol{4}}$ هو *ترتيب الربيعي* ($\boldsymbol{i}$).

$\boldsymbol{F1}$ التكرار المتجمع الصاعد للفئة السابقة للفئة الربيعية.

FQi تكرار الفئة الربيعية

$\boldsymbol{\Delta i}$ طول الفئة الربيعية

الطريقة البيانية: لتحديد الربيعيات بيانياً، نتبع الطريقة التالية:

- تحديد التكرارت المتجمعة الصاعدة.
- نرسم المنحنى المتجمع الصاعد.
- نسقط النقاط التي ترتيبها $\frac{\mathrm{N}}{4}$ $\frac{2\mathrm{N}}{4}$ $\frac{3\mathrm{N}}{4}$ على منحنى التكرار المتجمع الصاعد، ثم نسقط نقاط تقاطعها مع المنحنى على المحور الأفقي.
- نقاط تقاطع الاسقاطات مع المحور الأفقي تمثل الربيعيات Q_1 Q_2 Q_3

2. .2.3.2.العشيريات Deciles

وهي تسعة قيم يأخذها المتغير الإحصائي تقسم البيانات إلى عشرة أجزاء متساوية من حيث العدد، ونرمز لها بالرمز D_i، حيث تقع كل قيمة i =1،2،3،4،5،6،7،8،9 في نهاية جزء،بحيث يحدد ترتيب العشيريات وفق الصيغة $\frac{iN}{10}$ ويكون بذلك ترتيب العشيريات على التوالي

هـ $\frac{N}{10}$ $\frac{2N}{10}$ $\frac{3N}{10}$ $\frac{4N}{10}$ $\frac{5N}{10}$ $\frac{6N}{10}$ $\frac{7N}{10}$ $\frac{8N}{10}$ $\frac{9N}{10}$ حيث أن 10 % من البيانات أقل من العشيري الأول و90% أكبر منه، وتكون 20% من البيانات أصغر من العشيري الثاني و80% من البيانات أكبر من العشيري الثاني، وتكون 90 % من البيانات أقل من العشيري التاسع و10% أكبر من العشيري التاسع. ويمكن حساب الربيعيات وفق الصيغة التالية:

أ. **حالة بيانات غير مبوبة:**

1. **عدد البيانات فردي:** إذا كان عدد البيانات فردي فإن الربيعيات تحدد وفق الصيغة التالية

$$D_i = X_{i\left(\frac{n+1}{10}\right)} \qquad i=1\ ;\ 2\ ,3,\ 4\ ;5\ ;6\ ,\ 7\ ;\ 8\ ;\ 9.$$

وتكون بذلك العشيريات التسعة بالصيغ التالية:

$D_1 = X_{1\left(\frac{n+1}{10}\right)}$ $D_2 = X_{2\left(\frac{n+1}{10}\right)}$ $D_3 = X_{3\left(\frac{n+1}{10}\right)}$ $D_4 = X_{4\left(\frac{n+1}{10}\right)}$ وهكذا.

عدد البيانات زودجي: إذا كان عدد البيانات زوجي فإن الربيعيات تحدد وفق الصيغة التالية

$$D_i = \frac{X_{\frac{in}{10}} + X_{\frac{in}{10}+1}}{2} \qquad i=1\ ;\ 2\ ;\ 3\ ;\ 4\ ;\ 5\ ;\ 6\ ;\ 7\ ;\ 8\ ;9$$

وتكون بذلك العشيريات التسعة بالصيغ التالية

$D_1 = \frac{X_{\frac{1n}{10}} + X_{\frac{1n}{10}+1}}{2}$ $D_2 = \frac{X_{\frac{2n}{10}} + X_{\frac{2n}{10}+1}}{2}$ $D_3 = \frac{X_{\frac{3n}{10}} + X_{\frac{3n}{10}+1}}{2}$ وهكذا.

ب. **حالة بيانات مبوبة:**إذا المتغير منفصل فإن قيمة العشيري هي القيمة المتضمنة لترتيب العشيري بعد تشكيل التكرار المتجمع الصاعد. أما إذا كان المتغير متصل فإن العشيريات تحسب بنفس طريقة حساب الوسيط باستثناء تغيير الترتيب الذي يؤدي غلى تغيير الفئة العشيرية محل الدراسة، وتحسب قيمة العشيري في هذه الحالة وفق الصيغة العامة التالية:

$$\mathbf{D_i = a_i + \left(\frac{\frac{iN}{10} - F1}{FDi}\right).\Delta i} \qquad \boldsymbol{i = 1; 2, 3}$$

حيث أن:

D_i *رقم العشيري*

a_i هو الحد الأدنى للفئة العشيريةi.

$\frac{iN}{10}$ هو *ترتيب العشيري* (i).

$F1$ التكرار المتجمع الصاعد للفئة السابقة للفئة العشيرية.

FDi تكرار الفئة العشيرية.

Δi طول الفئة العشيرية.

3. .2.3.2.المئينيات:Centiles

هي 99 قيمة تقسم البيانات إلى 100 جزء متساوي من حيث عدد البيانات، يقع كل مئيني بعد 1% من البيانات، وبالتالي تكون باقي البيانات أكبر منه، ونرمز لها بالرمز P_i وتحسب بنفس كيفية حساب الوسيط وبقية اشباهه.

أ. حالة بيانات غير مبوبة:

1. **عدد البيانات فردي**: إذا كان عدد البيانات فردي فإن الربيعيات تحدد وفق الصيغة التالية

$p_i = X_{i\left(\frac{n+1}{100}\right)}$ i=1 ; 2, 3, 4 ; 5 ; 6, ... ; 97 ; 98 ; 99

2. **عدد البيانات زوجي**: إذا كان عدد البيانات زوجي فإن الربيعيات تحدد وفق الصيغة التالية

$$p_i = \frac{X_{\frac{in}{100}} + X_{\frac{in}{100}+1}}{2}$$ i=1 ; 2, 3, 4 ; 5 ; 6, ... ; 97 ; 98 ; 99

ب. حالة بيانات مبوبة:

إذا المتغير منفصل فإن قيمة المئيني هي القيمة المتضمنة لترتيب المئيني بعد تشكيل التكرار المتجمع الصاعد. أما إذا كان المتغير متصل فإن المئينيات تحسب بنفس طريقة حساب الوسيط باستثناء تغيير الترتيب الذي يؤدي غلى تغيير الفئة المئينية محل الدراسة، وتحسب قيمة المئينيات في هذه الحالة وفق الصيغة العامة التالية:

$$\mathbf{p}_i = a_i + \left(\frac{\frac{iN}{100} - F1}{FPi}\right).\Delta i \qquad i = 1; 2, 3 \ldots\ldots\ldots 97, 98, 99$$

حيث أن:

$\mathrm{p_i}$ رقم المئيني

$\mathbf{a_i}$ هو الحد الأدنى للفئة المئينية$\mathbf{i}$.

$\frac{\mathbf{iN}}{\mathbf{100}}$ هو ترتيب المئيني (i).

$\mathbf{F1}$ التكرار المتجمع الصاعد للفئة السابقة للفئة المئينية$\boldsymbol{i}$.

FDi تكرار الفئة المئينية.

$\mathbf{\Delta i}$ طول الفئة المئينية.

مقارنة بين المتوسطات:

مقارنة بين الوسط الحسابي والوسيط والمنوال

أ. في البيانات المتماثلة تتساوى هذه المقاييس الثلاثة أي يكون $\bar{X} = \mathrm{M_e} = \mathrm{M_o}$.

ب. في التوزيعات القريبة من التماثل تكون هذه المتوسطات الثلاثة مرتبطة مع بعèضها البعض بالعلاقة التي اكتشفها كارل بيرسون K.Pearson من كثر تجاربه، وهذه العلاقة هي

$(\bar{X} - \mathrm{M_o}) = 3(\bar{\mathrm{X}} - \mathrm{M_e})$ حيث تفيد هذه العلاقة في ايجاد المتوسط الحسابي إذا كان التزويع مفتوح.

ت. في التزويعات غير المتماثلة تكون هذه المتوسطات محتلفة عن بعضها البعض $\mathrm{M_e} \neq \mathrm{M_o} \neq \bar{X}$

ث. في التوزيعات الملتوية نحو اليمين(ذات الالتواء الموجب) تكون المتوسطات الثلاثة تحقق المتراجحة التالية

$\bar{X} > \mathrm{M_e} > \mathrm{M_o}$

ج. في التوزيعات الملتوية نحو اليسار(ذات الإلتواء السالب) تكون المتوسطات الثلاثة محققة للمتراجحة التالية

$\bar{X} < \mathrm{M_e} < \mathrm{M_o}$

1. مقارنة بين الوسط الحسابي والهندسي والتوافقي والتربيعي:

هناك علاقتان بارزتان بين هذه المتوسطات وهما:

أ. الوسط التوافقي لمجموعة من القيم الموجبة هو أقل أو يساوي وسطها الهندسي، الذي بدوره أقل أو يساوي وسطها الحسابي، وهذا الأخير أقل أو يساوي وسطها التربيعي أي أن $\dot{Q} \geq \overline{\mathrm{X}} \geq \dot{\mathrm{G}} \geq \dot{\mathrm{H}}$

ب. مربع الوسط الهندي لمجموعة من القيم الموجبة يساوي حاصل جداء وسطها التوافقي في وسطها الحسابي

$$G^2 = \dot{\bar{X}}.\dot{H}$$

2. مقارنة بين المتوسطات بصفة عامة:

- الوسط الحسابي هو أكثر المتوسطات فائدة وأعمقها وأكثرها استخداماً، وهو الذي يستخدم عادة كمقياس للنزعة المركزية إلا إذا كان هناك ما يمنع استخدامه كما في حالة التوزيعات المفتوحة أو التوزيعات التي تتغير بنسبة تكاد تكون ثابتة أو في التوزيعات التي تحتوي على قيم متطرفةٍ.
- الوسط الحسابي هو الوحيد الذي يحقق شروط يول، لذلك فهو صالح للمعالجة الرياضية أي أن يستخدم في القوانين الإحصائية الأخرى وهو مقياس غير متحيز.
- الوسيط يستخدم في التوزيعات النوعية، حيث لا نستطيع القياس في درجات عددية وأنما نستطيع الترتيب حسب النوع والحجم.

تمرين أول: إذا كان $0< x1$ و $0< x2$. بين أن $G2= H.X$.

الحل

إذا كان x_1 x_2 < 0 فإن:

$$\bar{X}= \frac{x_1+x_2}{2} \qquad \dot{H}=\frac{2}{\frac{1}{x_1}+\frac{1}{x_2}} =\frac{2}{\frac{x_1+x_2}{x_1.x_2}} =\frac{2x_1x_2}{x_1+x_2}$$

$$\bar{X}.\dot{H} = \frac{x_1+x_2}{2} . \frac{2x_1x_2}{x_1+x_2} =x_1x_2$$

$$\bar{X}.\dot{H}=x_1x_2=\sqrt{x_1x_2}\sqrt{x_1x_2}=\dot{G}\dot{G}=G^2$$

$$\bar{X}.\dot{H}=G^2$$

تمرين ثانٍ: إذا كان متوسط أجور عمال مؤسسة ما مساويا لـ: 5000 و.ن ، وكان متوسط أجور العمال (الذكور) هو 5200 و. ن ، ومتوسط أجور العاملات هو 4200 و.ن .

1. فماهي نسبة العمال الذكور والإناث من مجموع العمال؟.

2. إذا أردنا سحب عينة من هذا المجتمع ، فماهي الطريقة المفضلة للسحب حتى تكون العينة ممثلة لهذا المجتمع ؟ ماهي خصائص أو مواصفات هذا النوع من العينات .

الحل

. متوسط أجور مجموع العمال هو $\overline{X}$ =5000 DA

. متوسط اجور العمال الذكور هو $\overline{X_A}$ =5200 DA

. متوسط أجور العمال الإناث هو $\overline{X_B}$ =4200 DA

1. المطلوب هو حساب نسبة العمال والذكور والإناث من مجموع العمال.

إن مجموع العمال هو 100% .

إذا فرضنا أن نسبة أو عدد العمال الذكور هو x فإن N_A =x وعليه فإن نسبة أو عدد الإناث هو

N_B=100- x ونحن نعلم أن :

$$\bar{X}=\frac{N_A.\overline{X_A}+N_B.\overline{X_B}}{N_A+N_B}=\frac{N_A.\overline{X_A}+N_B.\overline{X_B}}{N} \qquad N=N_A+N_B$$

$$\bar{X}=\frac{X.5200+(100-X).4200}{100}=\frac{5200X+420000-4200X}{100}$$

$$\bar{X}=\frac{420000+1000X}{100}=5000\Rightarrow 500000=420000+1000x$$

$$\Rightarrow 1000x=80000\Rightarrow x=80=N_A\Rightarrow N_B=20$$

وعليه فإن نسبة الذكور هي 80% من مجموع العمال ونسبة الإناث هي 20% من مجموع العمال

2. الطريقة المفضلة للسحب هي السحب العشوائي الطبقي أي نسحب من كل طبقة عدد معين بصفة عشوائية يتناسب وحجم الطبقة في المجتمع الإحصائي. تتصف العينات التي تسحب وفق هذه الطريقة بمجموعة من المواصفات نذكر منها:

- ارتفاع مستوى تمثيل العينة للمجتمع الذي سحبت منه.
- تتطلب معرفة جيدة بالمجتمع المدروس ليتم تحديد طبقاته بشكل جيد.

- أكثر كفاءة من غيرها في الحالات التي تكون فيها مجموعة معينة من المجتمع ذات مواصفات خاصة هي السائدة.
- يمكن الحصول على نتائج جيدة الدقة من عينه حجمها صغير جداً.
- قد تتصف بشيء من التعقيد إذا كان عدد الطبقات كبيراً.

تمرين ثالث: ليكن جدول التوزيع التكراري التالي الذي يمثل الأجر الساعي لـ: 75 عامل

مراكز الفئات	53	59	65	71	77	83	؟
عدد العمال	01	05	؟	15	22	11	09

المطلوب:

1. تحديد المتغير الإحصائي المدروس ، ونوعيته والعرض البياني المناسب له؟.
2. إكمال الجدول السابق إذا علمت أن $x=75.04$ ثم إيجاد الحدود الدنيا والعليا للفئات.
3. تحديد مختلف التكرارات.

تحديد بقية متوسطات القيمة(الحجم) ، ثم المقارنة بين متوسطات القيمة كلها.

4. تحديد عدد العمال الذين تقل أجورهم الساعية أو تساوي 71 و ن ، وعدد العمال الذين تساوي أجورهم الساعية أو تزيد عن 68 و ن.
5. تحديد نسبة العمال الذين تنتمي أجورهم إلى المجال[59-77].
6. تقديم العرض البياني المناسب للتكرار المطلق المتجمع الصاعد والنازل.

الحل

1): اتمام الجدول

الفئات	$[56-50[$	$[62-56[$	$[68-62[$	$[74-68[$	$[80-74[$	$[86-80[$	$[98-86[$	Σ
التكرار المطلق	1	5	12	15	22	11	9	75
التكرار النسبي	0.0175	0.066	0.16	0.2	0.293	0.146	0.12	01
التكرار المئوي	1.75%	6.6%	16%	20%	29.3%	14.6%	12%	100
↑التكرار المتجمع	01	06	18	33	55	66	75	-
↓التكرار المتجمع	75	74	69	57	42	20	09	-

2): المتغير الاحصائي المدروس هو الأجر الساعي، وهو متغير كمي مستمر.

- العرض البياني المناسب له هو المدرج التكراري أو المنحنى التكراري.

3): بما أن مجموع التكرارات هو 75 فإن تكرار الفئة $[62-68[$ هو 12.

حيث أن C_7 هو مركز الفئة7 $\bar{X}=\frac{\sum_{i=1}^{k} nici}{N} \Rightarrow \bar{X}=\frac{1.53+5.59+12.65+15.71+22.77+11.83+9C_7}{75}$

$\Rightarrow 5628=4800+9C_7 \Rightarrow 9C_7=828 \Rightarrow C_7=92$

4):حساب بقية المتوسطات

- $G=\sqrt[N]{c_1{}^{n_1}\, c_2{}^{n_2} c_3{}^{n_3} c_4{}^{n_4} \dots\dots\dots\dots c_k{}^{n_k}}$

$G=53^{\frac{1}{75}}59^{\frac{5}{75}}65^{\frac{12}{75}}71^{\frac{15}{75}}77^{\frac{22}{75}}83^{\frac{11}{75}}92^{\frac{9}{75}}$

$G=(1.0543)(1.3123)(1.9501)(2.3455)(3.5757)(1.9118)(1.7204)=74.4256$

$G=74.4256$

- $\dot{H}=\frac{N}{\sum\frac{n_i}{c_i}}=\frac{75}{\sum\frac{n_i}{c_i}}=\frac{75}{1.0175} \cong 73.85$

- $Q=\sqrt{\frac{\sum_{i=1}^{k} n_i C_i^2}{N}} = \sqrt{\frac{428922}{75}}=75.62$

نلاحظ أن المتوسطات الاربعة ترتبط مع بعضها وفق العلاقة التي تم التطرق إليها وهي $\dot{Q} \geq \overline{X} \geq \dot{G} \geq \dot{H}$

$$75.62 > \overline{75.04} > 74.\dot{4}2 > 73.\dot{8}5$$

5):تحديد عدد العمال الذين تقل أجورهم عن 71= عدد العمال الذين يقل أجرهم عن 68+ عدد العمال الذين يتراوح أجرهم بين 68 و71=18+10=28 العمال من $15 \leftarrow \underbrace{]74-68]}_{6}$

العمال من $x \leftarrow \underbrace{[71-68]}_{4}$

$$x=\frac{15.4}{6}=10$$

عدد العمال الذين تساوي أجورهم أو تزيد عن 68 هو عدد العمال الإجمالي ناقص عدد العمال الذين تقل أجورهم عن 68 أي 75-18=57 عامل.

6): عدد العمال الذين تتراوح أجورهم بين 62 و68 هو 12.

عدد العمال الذين تتراوح أجورهم بين 68 و74 هو 15(سبق حسابه).

عدد العمال الذين تتراوح اجورهم بين 59 و62 هو 2.5.

عدد العمال الذين تتراوح اجورهم بين 74 و77 هو 15.

وعليه فعدد العمال الذين تتراوح اجورهم بين 59 و77 هو 44 ومنه فنسبة العمال الذين تنتمي أجورهم إلى المجال $[59-77]$ هي 52.66% .

يمكن ايضاً استخدام النسب مباشرة إذا استخدمنا التكرار النسبي في حساب ذلك.

تمرين رابع:بيّن أن

1) $\sum_{i=1}^{N} (X_i - \overline{X}) = 0$

2) $\frac{1}{N}\sum_{I=1}^{N}(X_i - \overline{X})^2 = \frac{1}{N}\sum_{I=1}^{N} X_I^2 - \overline{X}^2$

3) $\sum_{i=1}^{N}(Xi - \overline{X})^2 < \sum_{i=1}^{N}(X_i - c)^2 \quad \forall \overline{X} \neq c$ مربع انحرافات القيم عن متوسطها الحسابي اقل مايمكن

الحل

$$\sum_{i=1}^{N}(xi - \bar{x}) = 0$$

$$\sum_{i=1}^{N}(xi - \bar{x}) = \sum_{i=1}^{N} xi - N\bar{X} = 0$$

نعلم ان $\bar{x} = \frac{\sum xi}{N} \Rightarrow \sum xi = N\bar{X}$ ومنه فالمساواة $\sum_{i=1}^{N}(xi - \bar{x}) = 0$ محققة.

2) إثبات أن $\frac{1}{N}\sum_{I=1}^{N}(X_i - \overline{X})^2 = \frac{1}{N}\sum_{I=1}^{N} X_I^2 - \overline{X}^2$

$$\frac{1}{N}\sum_{I=1}^{N}(X_i - \overline{X})^2 = \frac{1}{N}\sum_{I=1}^{N}(x^2 - 2\bar{x}xi + \bar{X}^2) = \frac{1}{N}\sum_{i=1}^{N} x^2 - 2\bar{X}\frac{1}{N}\sum_{i=1}^{N} xi - \frac{1}{N}N\bar{X}^2 =$$

$$= \frac{1}{N}\sum_{i=1}^{N} x^2 - 2\bar{x}\bar{x} + \bar{X}^2 = \frac{1}{N}\sum_{i=1}^{N} x^2 - 2\bar{X}^2 + \bar{X}^2 = \frac{1}{N}\sum_{i=1}^{N} x^2 - \bar{X}^2$$

3): إثبات أن : $\sum_{i=1}^{N}(Xi - \overline{X})^2 < \sum_{i=1}^{N}(X_i - c)^2 \quad \forall \overline{X} \neq c$

سبق اثبات ذلك سابقاً.

<u>تمرين خامس:</u>

إذا كان متوسط أجور عمال مؤسسة ما مساويا لـ: 5000 و.ن، وكان متوسط أجور العمال (الذكور) هو 5200 و. ن ، ومتوسط أجور العاملات هو 4200 و.ن.

فماهي نسبة العمال الذكور والإناث من مجموع العمال؟.

إذا أردنا سحب عينة من هذا المجتمع، فماهي الطريقة المفضلة للسحب حتى تكون العينة ممثلة لهذا المجتمع ؟ ماهي خصائص أو مواصفات هذا النوع من العينات.

الحل

. متوسط أجور مجموع العمال هو $\overline{X} = 5000\ DA$

. متوسط اجور العمال الذكور هو $\overline{X_A} = 5200\ DA$

. متوسط أجور العمال الإناث هو $\overline{X_B} = 4200\ DA$

3. المطلوب هو حساب نسبة العمال والذكور والإناث من مجموع العمال.

إن مجموع العمال هو 100% .

إذا فرضنا أن نسبة أو عدد العمال الذكور هو x فإن $N_A = x$ وعليه فإن نسبة أو عدد الإناث هو $N_B = 100 - x$ ونحن نعلم أن :

$$\bar{X} = \frac{N_A.\overline{X_A} + N_B.\overline{X_B}}{N_A + N_B} = \frac{N_A.\overline{X_A} + N_B.\overline{X_B}}{N} \qquad N = N_A + N_B$$

$$\bar{X} = \frac{X.5200 + (100 - X).4200}{100} = \frac{5200X + 420000 - 4200X}{100}$$

$$\bar{X} = \frac{420000 + 1000X}{100} = 5000 \Rightarrow 500000 = 420000 + 1000x$$

$$\Rightarrow 1000x = 80000 \Rightarrow x = 80 = N_A \Rightarrow N_B = 20$$

وعليه فإن نسبة الذكور هي 80% من مجموع العمال ونسبة الإناث هي 20% من مجموع العمال

4. الطريقة المفضلة للسحب هي السحب العشوائي الطبقي أي نسحب من كل طبقة عدد معين بصفة عشوائية يتناسب وحجم الطبقة في المجتمع الإحصائي. تتصف العينات التي تسحب وفق هذه الطريقة بمجموعة من المواصفات نذكر منها:

- ارتفاع مستوى تمثيل العينة للمجتمع الذي سحبت منه.
- تتطلب معرفة جيدة بالمجتمع المدروس ليتم تحديد طبقاته بشكل جيد.
- أكثر كفاءة من غيرها في الحالات التي تكون فيها مجموعة معينة من المجتمع ذات مواصفات خاصة هي السائدة.
- يمكن الحصول على نتائج جيدة الدقة من عينه حجمها صغير جداً.

- قد تتصف بشيء من التعقيد إذا كان عدد الطبقات كبيراً.

تمرين سادس: ليكن جدول التوزيع التكراري التالي الذي يمثل الأجر الساعي لـ: 102عامل

مراكز الفئات الأجرية	10	20	*C3*	*C4*	*70*	*80*	*Σ*
عدد العمال	*15*	*N2*	*N3*	*N4*	*8*	*N6*	102

المطلوب

احسب $C3$ و $C4$ إذا علمت أن متوسطهما الحسابي هو 40 ومتوسطهما التوافقي هو 37.5 وأن $C3 < C4$

احسب $N2$ و $N4$ إذا علمت أن متوسطهما الحسابي هو 25 ومتوسطهما التوافقي هو 24 و أن $N4 < N2$.

احسب $N3$ و $N6$ إذا علمت أن متوسطهما الحسابي هو 14.5 ومتوسطهما الهندسي هو 10و أن $N6 < N3$.

حدد الحدود السفلى والعليا لجميع الفئات أذا علمت أن الحد الأسفل للفئة الأولى هو 5.

احسب مختلف التكرارات.

احسب كل من: المتوسط الحسابي، المتوسط الهندسي، المتوسط التوافقي. ماذا تلاحظ.

ماهو عدد العمال الذين يتقاضون أجرا ساعيا يزيد عن 25 و ن، ونسبة العمال الذين يتقاضون أجرا ساعيا يقل عن 35 و ن

الحل

1): حساب c_3 و c_4

$$\frac{c_3+c_4}{2}=40 \Rightarrow C_3+C_4=80 \Rightarrow C_3=80-C_4 \qquad (1)$$

$$\frac{2}{\frac{1}{c_3}+\frac{1}{c_4}}=37.5 \Rightarrow 2c_3c_4=37.5(c_3+c_4) \qquad (2)$$

بتعويض (1) في (2) نجد

$$2(80-C_4)C_4=37.5(80-C_4+C_4) \Rightarrow 160C_4-2C^2{}_4=3000 \Rightarrow -2C^2{}_4+160C_4-3000=0$$

$\Delta = B^2 - 4AC = (160)^2 - 4(-2)(-3000) = 1600 \Rightarrow \sqrt{\Delta} = 40$

$c_{4_1} = \frac{-B-\sqrt{\Delta}}{2A} = \frac{-160-40}{-4} = 50 \Rightarrow c_{3_1} = 30$

$c_{4_2} = \frac{-B+\sqrt{\Delta}}{2A} = \frac{-160+40}{-4} = 30 \Rightarrow c_{3_2} = 50$

وبما أن $c_3 < c_4$ فإن $c_3 = 30$ و $c_4 = 50$

2): حساب N_2 و N_4

لدينا $\frac{N_2+N_4}{2} = 25 \Rightarrow N_2 + N_4 = 50 \Rightarrow N_2 = 50 - N_4$ (1)

$\frac{2}{\frac{1}{N_2}+\frac{1}{N_4}} = 24 \Rightarrow 2N_2N_4 = 24(N_2 + N_4)$ (2)

بتعويض (1) في (2) نجد

$2(50 - N_4)N_4 = 24(50 - N_4 + N_4)$

$100N_4 - 2N_4^2 = 1200 \Rightarrow -2N_4^2 + 100N_4 - 1200 = 0$

$\Delta = B^2 - 4AC = 100^2 - 4(-2)(-1200) = 10000 - 9600 = 400 \Rightarrow \sqrt{\Delta} = 20$

$N_{3_1} = \frac{-B-\sqrt{\Delta}}{2A} = \frac{-100-20}{-4} = 30 \Rightarrow N_{4_1} = 20$

$N_{3_2} = \frac{-B+\sqrt{\Delta}}{2A} = \frac{-100+20}{-4} = 20 \Rightarrow N_{4_2} = 20$

وبما أن $N_4 < N_2$ فإن $N_2 = 30$ و $N_4 = 20$

3): حساب N_3 و N_6

لدينا $\frac{N_3+N_6}{2} = 14.5 \Rightarrow N_3 + N_6 = 29 \Rightarrow N_3 = 29 - N_6$ (1)

$\sqrt{N_3N_6} = 10 \Rightarrow N_3N_6 = 100$ (2)

بتعويض (1) في (2) نجد

$(29 - N_6)\, N_6$=100 $\Rightarrow -{N_6}^2+29N_6$ -100=0

$\Delta= B^2$-4AC=$(29)^2$-4$(-1)(-100)$ = 441 $\Rightarrow \sqrt{\Delta}$ =21

$N_{6_1}=\frac{-B-\sqrt{\Delta}}{2A}=\frac{-29-21}{-2}$ =25 $\Rightarrow$ N_{3_1} =4

$N_{6_2}=\frac{-B+\sqrt{\Delta}}{2A}=\frac{-29+21}{-2}$ = 4 $\Rightarrow$ N_{3_2} =25

وبما أن $N_6 < N_3$ فإن N_3=25 و N_6=4

	85-75	75-65	65-35	35-25	25-15	15-5	الفئات
	80	70	50	30	20	10	مراكز الفئات C_I
	4	8	20	25	30	15	التكرار n_i
	102	98	90	70	45	15	التكرار المتجمع↑
	02	12	32	57	87	102	التكرار المتجمع↓
	0.032	0.078	0.196	0.245	00294	0.147	f_i
3380	320	560	1000	750	600	150	$c_i n_i$
	00.05	0.1142	0.4	0.8333	1.5	1.5	$\frac{n_i}{c_i}$
	1.1505	1.3928	2.1527	2.3008	2.4126	1.4028	$c_i^{f_i}$

5):المتوسط الحسابي $\bar{X}=\frac{\sum_{i=1}^{N} c_i n_i}{N}=\frac{3380}{102}$ =33.13

المتوسط الهندسي:بما أن المتغير متصل، والبيانات مبوبة في فئات فإن قيمة المتوسط الهندسي تحسب وفق الصيغة

G=$c_1^{\frac{n_1}{N}} c_2^{\frac{n_2}{N}} c_3^{\frac{n_3}{N}} c_4^{\frac{n_4}{N}} \dots\dots\dots\dots c_k^{\frac{n_k}{N}} = {c_1}^{f_1} {c_2}^{f_2} {c_3}^{f} {c_4}^{f_4} \dots\dots\dots\dots\dots\dots\dots {c_k}^{f_k}$

G=(1.4028)(2.4126)(2.3008)(2.1527)(1.3928)(1.1505) =26.86

المتوسط التوافقي: بما أن المتغير متصل، والبيانات مبوبة في فئات فإن قيمة المتوسط التوافقي تحسب وفق الصيغة

$$H = \frac{N}{\sum_{i=1}^{k}\left(\frac{n_i}{c_i}\right)} = \frac{102}{3.3976} = 23.19$$

نلاحظ أن $H < G < \bar{}$

6): عدد العمال الذين يتقاضون أجراً ساعياً يزيد عن 25 و ن هو 57 عامل.

عدد العمال الذين يتقاضون أجراً ساعياً يقل عن 35 و ن هو 70 عامل.

تمرين سابع :قام أحد المعاهد المتخصصة بسبرٍ للاراء حول المصاريف الأسرية السنوية لمجتمع يتكون من 100 أسرة

فكانت النتائج كما يلي: المصاريف: بآلاف الوحدات النقدية

فئات المصاريف	4-0	8-4	12-8	-12L_4	$L_4$22-	30-22	42-30	$\sum$
التكراراتni	6	N_2	N_3	17	14	11	3	100

المطلوب:

1. حدد الوحدة الإحصائية، المتغير الإحصائي، نوعية المتغير.
2. ماهو العرض البياني المناسب لهذا التوزيع.
3. أحسب N_2 و N_3 علماً ان العشير الرابع $D_4 = 9.5$.
4. احسب الحد الأعلى للفئة الربيعية 4 إذا علمت أن المتوسط الحسابي $\bar{X} = 13$.
5. احسب المتوسط الهندسي والمتوسط التوافقي.
6. ماهو عدد الاسر التي تقل مصاريفها النقدية عن 22 الف وحدة نقدية.
7. حدد المصاريف السنوية للأسرة ذات المرتبة 25 وللأسرة .ات المرتبة 75.
8. حدد حسابياً وبيانياً كل من M_e M_o

الحل

1. الوحدة الاحصائية هي الأسرة ، المتغير الاحصائي هو المصاريف السنوية، نوعيته:كمي مستمر.
2. العرض البياني المناسب هو المدرج التكراري، المضلع التكراري و المنحنى التكراري.
3. حساب N_2 و N_3 علماً ان العشير الرابع $D_4 = 9.5$.

$N_2 + N_3 = 49 \Rightarrow N_3 = 49 - N_2$

ترتيب D_4 هو $40 = \frac{4N}{10} = \frac{4.100}{10}$ وبما أن $D_4 = 9.5$ فأن الفئة العشرية هي 8-12

الجدول التكراري

فئات الانفاق	C_i	n_i	$n_i^{\uparrow}$	f_i	$n_i C_i$
0-4	2	6	6	0.06	12
4-8	6	$N_2 = 25$	$6 + N_2 = 31$	0.25	150
8-12	10	$N_3 = 24$	$6 + N_2 + N_3 = 55$	0.24	240
L_4-12	14	17	$23 + N_2 + N_3 = 72$	0.17	$17.\frac{(12+L_4)}{2}$
22-L_4	19	14	$37 + N_2 + N_3 = 86$	0.14	$14.\frac{(L_4+22)}{2}$
22-30	26	11	$48 + N_2 + N_3 = 97$	0.11	286
30-42	36	3	$51 + N_2 + N_3 = 100$	0.03	108

$$\mathbf{D_i = a_i + \left(\frac{\frac{iN}{10} - F1}{FDi}\right).\Delta i} \Rightarrow D_4 = \mathbf{a_4 + \left(\frac{\frac{4N}{10} - F1}{FD4}\right).\Delta 4} \Rightarrow \mathbf{9.5 = 8 + \frac{40 - (6 + N_2)}{N_3}.4}$$

$$\Rightarrow \mathbf{1.5 = \frac{(40 - N_2 - 6)}{N_3} 4 \Rightarrow 1.5 N_3 = 160 - 4N_2 - 24}$$

بتعويض **3** بما تساويه نجد

$$1.5(49 - \mathbf{N_2}) = \mathbf{136 - 4N_2} \Rightarrow \mathbf{73.5 - 1.5 N_2 = 136 - 4N_2} \Rightarrow \mathbf{62.5 = 2.5 N_2} \Rightarrow \mathbf{N_2 = 25} \Rightarrow \mathbf{N_3 = 24}$$

4. حساب الحد الأعلى للفئة الربيعية L_4 إذا علمت أن المتوسط الحسابي $\bar{X} = 13$.

لدينا

$$\sum n_i C_i = N \bar{X} \Rightarrow \mathbf{1300 = 1052 + 15.5\, L_4} \Rightarrow \mathbf{248 = 15.5\, L_4} \Rightarrow \mathbf{L_4 = 16}$$

5. حساب المتوسط الهندسي والمتوسط التوافقي

المتوسط الهندسي: $G=c_1^{\frac{n_1}{N}} c_2^{\frac{n_2}{N}} c_3^{\frac{n_3}{N}} c_4^{\frac{n_4}{N}} \dots\dots\dots\dots c_k^{\frac{n_k}{N}} = c_1^{f_1} c_2^{f_2} c_3^{f} c_4^{f_4} \dots\dots\dots\dots\dots\dots\dots c_k^{f_k}$

$$G=2^{0.06}6^{0.25}10^{0.24}14^{0.17}19^{0.14}26^{0.11}36^{0.03}=10.68$$

المتوسط التوافقي: $H=\frac{N}{\frac{n_1}{c_1}+\frac{n_2}{c_2}+\frac{n_3}{c_3}+\frac{n_4}{c_4}+\cdots\dots\dots\dots\frac{n_k}{c_k}}=\frac{N}{\sum_{i=1}^{k}\left(\frac{n_i}{c_i}\right)}$

$$H=\frac{100}{12.0242047}=8.31$$

6. عدد الأسر التي تقل مصاريفها عن 22 ألف و ن هو 86 أسرة.

7. دخل الأسرة ذات المرتبة 25 هو قيمة الربيعي الأول التي هي نفسها قيمة المئيني 25

ترتيب $P_{25}=Q_1$ هو $\frac{25N}{100}=\frac{N}{4}=25$ ومنه فالفئة الربيعية هي الفئة 4-8 ومنه:

$$Q_1=4+\frac{(25-6)}{25}.4=7.04$$

ترتيب $P_{75}=Q_3$ هو $\frac{75N}{100}=\frac{3N}{4}=75$ ومنه فالفئة الربيعية (المئينية) هي 16-22 وعليه فإن

$$Q_3=16+\frac{(75-72)}{14}.6=17.28$$

8. حساب الوسيط والمنوال حسابياً

ترتيب الوسيط هو $\frac{N}{2}=\frac{100}{2}=50$ ومنه فالفئة الوسيطية هي 8-12 وعليه فقيمة الوسيط هي

$$M_e=a+\left(\frac{\frac{N}{2}-F_1}{F_m}\right).\Delta \Rightarrow M_e=8+\frac{(50-31)}{14}.4=11.16$$

الفئة الأكثر تكراراً هي الفئة 4-8 وعليه فقيمة المنوال هي

$$M_o=L_\circ+\frac{\Delta(f_m-f_1)}{(f_m-f_1)+(f_m-f_2)} \Rightarrow M_o=4+\frac{(25-6)}{(25-6)+(25-24)}.4=4+\frac{19*4}{20}=7.8$$

تمرين ثامن: الجدول التالي يمثل توزيع طلبة جامعة على التخصصات الموجودة بها

التخصصات	طب	صيدلة	آداب	قانون	لغات أجنبية	اقتصاد
عدد الطلبة	700	500	1300	1700	800	5000

نريد سحب عينه حجمها 500 طالب من بين هؤلاء الطلبة.

- ماهي الطريقة المفضلة في عملية السحب حتى تكون العينة ممثلة أحسن تمثيل للطلبة.
- ماهي خصائص أو مواصفات هذه الطريقة.
- احسب العدد المسحوب من كل تخصص حتى تكون العينة كذلك.

الحل

- الطريقة المفضلة للسحب هي السحب العشوائي الطبقي.
- يتصف هذا النوع من العينات بـ:
  - ارتفاع مستوى تمثيل العينة للمجتمع.
  - يمكِّن من الحصول على نتائج عالية الدقة.
  - يتطلب معرفة جيدة بالمجتمع حتى يتم تحديد الطبقات.
  - تتصف بشيء من التعقيد إذا كان عدد الطبقات كبير.
- حتى تكون العينة المكونة من 500 طالب ممثلة أحسن تمثيل للطلبة، يجب أن نسحب وبطذريقة عشوائية من طلبة:

الطب: $\frac{500*700}{10000}$ =35 طالب.

ا لقانون $\frac{500*1700}{10000}$ = 85 طالب.

الصيدلة $\frac{500*500}{10000}$ =25 طالب.

الأدب $\frac{500*1300}{10000}$ =65 طالب

لغات أجنبية $\frac{500*800}{10000}$ = 40 طالب.

الاقتصاد $\frac{500*5000}{10000}$ =250 طالب

تمرين تاسع: ليكن جدول التوزيع التكراري التالي الذي يمثل الأجر الساعي لـ: 200عامل

مراكز الفئات	15	25	C3	60	75	C6	Σ
عدد العمال	30	N2	50	N4	16	04	200

المطلوب:

1- احسب C3 وC6 اذا علمت أن متوسطهما الحسابي هو 65 ومتوسطهما الهندسي هو 60 وأن C3 < C6

2- احسب N2 وN4 أذا علمت أن متوسطهما الحسابي هو 50 ومتوسطهما التوافقي هو 48 و أن N2 < N4

3- حدد الحدود السفلى والعليا لجميع الفئات أذا علمت أن الحد الأسفل للفئة الأولى هو 10.

4- احسب مختلف التكرارات.

5- احسب كل من: المتوسط الحسابي ، المتوسط الهندسي ، المتوسط التوافقي . ماذا تلاحظ.

6- احسب كل من: القيمة التي تقسم البيانات إلى قسمين متساويين، القيمة الشائعة.

7- حدد شكل التوزيع باستعمال المقاييس التي تم حسابها.

8- ماهو عدد العمال الذين يتقاضون أجرا ساعيا اقل من 30 و ن ، وعدد العمال الذين يتقاضون أجرا ساعيا يزيد عن 50 و ن

الحل

1. حساب c_3 و C_6

$$\frac{c_3+c_6}{2}=65 \Rightarrow c_3+c_6=130 \Rightarrow c_3=130-c_6 \qquad (1)$$

$$\sqrt{c_3c_6}=60 \Rightarrow c_3c_6=3600 \qquad (2)$$

بتعويض (1) في (2) نجد

$$(130-c_6).c_6=3600 \Rightarrow 130c_6-{c_6}^2-3600=0$$

معادلة من الدرجة الثانية

$$\Delta=B^2-4AC=(130)^2-4(-1)(-3600)=16900-14400=2500 \Rightarrow \sqrt{\Delta}=50$$

$$C_{6_1}=\frac{-B-\sqrt{\Delta}}{2A}=\frac{-130-50}{-2}=90 \Rightarrow C_{3_1}=40$$

$$C_{6_2}=\frac{-B+\sqrt{\Delta}}{2A}=\frac{-130+50}{-2}=40 \Rightarrow C_{3_2}=90$$

وبما أن $c_3 < c_6$ فإن $c_6 = 90$ و $c_3 = 40$

2. حساب N_2 و N_4

$$\frac{N_2+N_4}{2}=50 \Rightarrow N_2+N_4=100 \Rightarrow N_2=100-N_4 \qquad (1)$$

$$\frac{2}{\frac{1}{N_2}+\frac{1}{N_4}}=48 \Rightarrow \frac{2N_2N_4}{N_2+N_4}=48 \Rightarrow 2N_2N_4=48(N_2+N_4) \qquad (2)$$

بتعويض (1) في (2) نجد

$2(100 - N_4)N_4$=48$(100 - N_4 + N_4) \Rightarrow 200N_4 - 2{N_4}^2 = 4800 \Rightarrow -2{N_4}^2 + 200N_4 - 4800 = 0$

$\Delta = B^2 - 4AC = (200)^2 - 4(-2)(-4800) = 40000 - 38400 = 1600 \Rightarrow \sqrt{\Delta} = 40$

$N_{4_1} = \frac{-B-\sqrt{\Delta}}{2A} = \frac{-200-40}{-4} = 60 \Rightarrow N_{2_1} = 40$

$N_{4_2} = \frac{-B+\sqrt{\Delta}}{2A} = \frac{-200+40}{-4} = 40 \Rightarrow N_{2_2} = 60$

وبما أن $N_4 < N_2$ فإن $N_2 = 60$ و $N_2 = 40$

3. 4) تحديد حدود الفئات ومختلف التكرارات

$\sum$	100-80	80-70	70-50	50-30	30-20	20-10	حدود الفئات
	90	75	60	40	25	15	مراكز الفئات c_i
200	04	16	40	50	60	30	التكرار المطلق n_i
	200	196	180	140	90	30	التكرار المتجمع ↑
	04	20	60	110	170	200	التكرار المتجمع ↓
	0.02	0.08	0.20	0.25	0.3	0.15	التكرار النسبي f_i
	2 %	8 %	20 %	25 %	30 %	15 %	التكرار المئوي %
	0.2	1.6	2	2.5	6	3	التكرار المعدل
39.55	1.5	6	12	10	7.5	2.25	$f_i C_i$
34.75	1.09417	1.41255	2.26793	2.51186	2.62652	1.50111	$c_i^{f_i}$
6.5744	0.044	0.213	0.666	1.25	2.4	2	n_i/c_i

4. حساب المتوسطات

$\bar{X} = \frac{\sum_{i=1}^{k} n_i c_i}{N} = \frac{\sum_{i=1}^{k} n_i}{N} . c_i = \sum f_i c_i \Rightarrow \bar{X} = 39.55$

$G=\sqrt[N]{c_1{}^{n_1}\, c_2{}^{n_2} c_3{}^{n_3} c_4{}^{n_4} \dots\dots\dots\dots . c_k{}^{n_k}}$

$G=(c_1{}^{n_1}\, c_2{}^{n_2} c_3{}^{n_3} c_4{}^{n_4} \dots\dots\dots\dots . c_k{}^{n_k})^{\frac{1}{N}}$

$G=c_1{}^{\frac{n_1}{N}} c_2{}^{\frac{n_2}{N}} c_3{}^{\frac{n_3}{N}} c_4{}^{\frac{n_4}{N}} \dots\dots\dots\dots . c_k{}^{\frac{n_k}{N}} = c_1{}^{f_1} c_2{}^{f_2} c_3{}^{f} c_4{}^{f_4} \dots\dots\dots\dots . . c_k{}^{f_k} = 34.75$

$$H = \frac{N}{\frac{n_1}{c_1}+\frac{n_2}{c_2}+\frac{n_3}{c_3}+\frac{n_4}{c_4}+\cdots\cdots\cdots\cdots\frac{n_k}{c_k}} = \frac{N}{\sum_{i=1}^{k}\left(\frac{n_i}{c_i}\right)}$$

$$H = \frac{200}{6.5744} = 30.42$$

نلاحظ أن $H < G < \bar{}$

5. حساب القيمة الوسيطية أي القيمة التي تقسم البيانات إلى قسمين متساويين

رتبة الوسيط هي $\frac{N}{2} = \frac{200}{2} = 100$، ومنه فقيمة الوسيط تقع ضمن الفئة 30-50

$$M_e = a+\left(\frac{\frac{N}{2}-F_1}{F_m}\right).\Delta \Rightarrow M_e = 30 + \frac{(100-90)}{50}.20 = 30 + \frac{200}{50} = 34$$

حساب القيمة الشائعة أي الأكثر تكراراً وهي تقع ضمن الفئة الأكثر تكرار معدل لأن الفئات غير متساوية، وهي الفئة 20-30.وتحسب وفق الصيغة:

$$M_o = L_\circ + \frac{\Delta(f_m - f_1)}{(f_m - f_1)+(f_m - f_2)} \Rightarrow M_o = 20 + \frac{(6-3)}{(6-3)+(6-2.5)}.10$$

$$\Rightarrow M_o = 20 + \frac{30}{6.5} \Rightarrow M_o = 20 + 4.61 = 24.61$$

6. بما أن $M_o < \bar{X}$ فإن التوزيع ملتوي نحو اليمين.

7. عدد العمال الذين يتقاضون أجراً يقل عن 30 ألف و ن هو 90 عامل.

عدد العمال الذين يتقاضون أجراً يزيد عن 50 الف و ن هو 60 عامل

الفصل الثالث:مقاييس التشتت و الالتواء Measures of dispersion and skewness

1.3 مقاييس التشتتMeasures of dispersion

إن مقاييس النزعة المركزية لا تعطينا صورة عن مدى تقارب أو تباعد البيانات، وعليه فإن مقاييس الوضع المركزية لا تكتفي لإعطاء صورة واضحة عن الظاهرة ما لم تكن لدينا فكرة عن مدى تشتت هذه القيم أو تفرقها حول وسطها. ولهذا نحتاج إلى مؤشرات حاصة لقياس التشتت تسمى مقاييس التشتت.

مثال(3–1):لتكن المجموعتان التاليتان من البيانات لتساقط الأمطار خلال الأسبوع الأول من شهر جانفي لسنتي 1997 و 1998بالمليمتر

المجموعة الأولى:10 20 30 40 50 60 70.

المجموعة الثانية: 10 30 30 40 50 50 70

فلو أردنا المقارنة بين المجموعتين من حيث القيم المركزية لوجدنا أن

المتوسط الحسابي للمجموعة الأولى يساوي المتوسط الحسابي للمجموعة الثانية =40= $\overline{X}_A = \overline{X}_B$

الوسيط للمجموعة الأولى يساوي الوسيط للمجموعة الثاني =40= $M_{e_A} = M_{e_B}$

مع أن قيم المجموعة الأولى تختلف عن قيم المجموعة الثانية.

مثال(3–2):في أحد الاقسام كانت علامات الطلبة إما 9 أو 3 أو حولهما فكان متوسط العلامات هو 6.5 من عشرة، بينما في قسم آخر كانت كلامات الطلبة بين 6 و7 من عشرة فكان متوسط الاعلامات هو 6.5 من عشرة.تلاحظ رغم الفروق الواضحة في اعلامات إلا أن المجموعتين لهما نفس الوسط الحسابي

مثال(3–6):العينات الاتية

(1): 8 8 8 8 8

(2): 7 5 8 12

(3): 4 3 6 16 11

هي عينات مختلفة لكن جميعها لها نفس المتوسط الحسابي وهو 8= $\bar{X}$.

والعينتان (1): 14 15 18 25 30 35 38

(2): 18 20 22 25 29 30 31

لهما نفس المتوسط الحسابي ونفس الوسيط لكنهما مختلفتان

والعينتان (1): 78 79 79 80 80 80 81 81 82

(2): 40 60 60 80 80 80 100 100 120

لهما نفس المتوسط الحسابي $\overline{X}_A = \overline{X}_B = 80$

ولهما نفس المنوال $M_{o_A} = M_{o_B} = 80$

ولهما نفس الوسيط $M_{e_A} = M_{e_B} = 80$

ورغم ذلك فالعينتان غير متماثلتان، ففي العينة الاولى نجد البيانات مرصوصة حول القيمةالمعيارية (80) بينما نجد القيم في العينة الثانية مشتتة على مدى واسع.

لذلك كان لابد من استعمال مقاييس أخرى تبين لنا مدى اختلاف البيانات بينها، ومدى التفاوت والتغير بين مفرداتها فيما إذا كانت متقاربة من بعضها البعض أم متباعدة عن بعضها البعض؟ بمعنى آخر لا نكتفي بالبحث عن متوسط هذه القيم لتمثيلها، ولكن لا بد أن نبحث أيضاً عن مدى بعدها عن هذه القيمة المركزية، ومدى بعدها عن القيم المتطرفة وبعدها عن بعضها البعض، وهو ما نسميه بالتبعثر أو التشتت

تعريف التشتت:يشير مفهوم التشتت أو التبعثر إلى احتلاف وانتشار البيانات، بحث إذا كبر مقياس التشتت كثيراً دلَّ على تباعد البيانات عن بعضها، ويفقد عندئذٍ المتوسط اهميته كقيمة ممثلة لمستوى القيم موضوع الدراسة.أما إذا صغر المقياس فإن اهمية المتوسط تزداد في التحليل الاحصائي الذي يرتبط به.هناك مقاييس عديدة للتشتت يمكن أن نصنفها في صنفين:هما:

3-1-1:مقاييس التشتت المطلقAbsolute measures of dispersion

وهي عبارة عن مقاييس لدرجة التشتت (مقدار التشتت) مقدراً بالوحدات، فإذا كانت الظاهرة المدروسة مقاسة بالدينار، فمقياس التشتت المطلق يقاس عندئذٍ بالدينار، وإذا كانت مقاسة بالمتر، فمقياس التشتت المطلق يقاس أيضاً بالمتر.و أهم مقاييس التشتت المطلق:

3-1-1-1:المدى Range

يسمى أيضاً بمجال التغيير، وهو أبسط مقاييس التشتت المطلق، وهو عبارة عن الفرق بين أكبر قيمة وأصغر قيمة في البيانات، نرمز له بالرمز R حيث:

حالة بيانات غير مبوبة:

حيث: R هو المدى (مدى التغير) $R=X_{max} - X_{min}$

X_{max} أكبر قيمة في البيانات

X_{min} أصغر قيمة في البيانات

مثال(3–4): إذا كانت لدينا سلسلة البيانات التالية:5 8 2 18 11 14 3 118 4 9 21

احسب المدى العام للبيانات.

الحل

$R=X_{max} - X_{min} = 118 - 2=116$ وهو المجال الذي تنتشر فيه البيانات

حالة بيانات مبوبة:أي يكون هناك توزيع تكراري ذي فئات، فيكون المدى أو مجال التغير هو الفرق بين الحد الأعلى للفئة العليا (الأخيرة) والحد الأدنى للفئة الدنيا (الأولى).

خواص المدى:للمدى مزايا وعيوب نذكرها كالتالي:

مزايا المدى

- سهل الحساب والفهم.
- يستعمل في حالات قليلة لأن اعتماده على القيم الطرفية يجعله حساساً لحجم العينة.

عيوب المدى

- لا يعتمد في حسابه على جميع القيم، فهو يأخذ بعين الاعتبار القيمة الكبرى والقيمة الصغرى فقط.
- هو مقياس غير مستقر تماماً لأنه يتأثر بالقيم المتطرفة، مما يجعله يعطي فكرة خاطئة عن تشتت البيانات.
- لا يمكن حسابه في حالتي التوزيعات المفتوحة.

3–1–2:المدى الربيعي:interquartile range

قد تكون القيم متطرفة ، ولتفادي وجود القيم الشاذة عند طرفي المجموعة، نقوم بترتيب البيانات تصاعدياً أو تنازلياً، ثم نقوم بحذف ربع القيم عند كلا الطرفين، ونكتفي بحساب المدى للنصف الأوسط لمجموعة القيم

فنضطر إلى أخذ مجال جزئي من المجال الكلي، حيث نعمد لأخذ الفرق بين قيمتين أخريين من المتغير هما: الربيعي الأول والربيعي الثاني.

فالمدى الربيعي عندئذٍ هو عبارة عن الفرق بين الربيعي الثالث(الاعلى) والربيعي الأول (الأدنى)، ويرمز اللمدى الربيعي بالرمز R_Q

المدى الربيعي أو الانحراف الربيعي	$R_Q = Q_3 - Q_1$

مثال (3–5):لتكن لدينا البيانات التالية

5 4 7 8 1 2 12 4 0 7

احسب كل من المدى العام للبيانات وكذا المدى الربيعي لها.

الحل: نقوم أولاً بترتيب البيانات تصاعدياً

0 12 8 7 7 5 4 4 2 1

- مدى البيانات هو $R = X_{max} - X_{min} = 12 - 0 = 12$
- المدى الربيعي: $R_Q = Q_3 - Q_1$

نحسب كل من Q_3 Q_1

بما أننا أمام بيانات غير مبوبة، وبما أن عدد البيانات فردي ($N=11$) فإن رتبة الربيعي الأول هي $\frac{N+1}{4}$

$$Q_1 = X_{\frac{(N+1)}{4}} = X_{\frac{(11+1)}{4}} = X_3 = 2 = Q_1$$

رتبة الربيعي الثالث تساوي $\frac{3(N+1)}{4}$

$$Q_3 = X_{\frac{3(n+1)}{4}} = X_{\frac{3(11+1)}{4}} = X_9 = 7 = Q_3$$

$$R_Q = Q_3 - Q_1 \Rightarrow R_Q = 7 - 2 = 5$$

نستعمل أحياناً نصف المدى الربيعي أو الانحراف الربيعي، وهو عبارة عن نصف المدى الربيعي أي

$$R_{Q}/_{2} = \frac{(Q_3 - Q_1)}{2}$$

ملاحظات:

نلاحظ أن المدى الربيعي يعتمد على قيمتين فقط هما الربيعي الثالث (الأعلى) والربيعي الأول (الأدنى).
وبنفس الطريقة يمكن حساب المدى العشيري والمدى المئيني ويفضل استعمال المدى الربيعي على المدى لأنه يمكن حسابه من التوزيعات التكرارية المفتوحة

3-1-1-3:المدى العشيري:interdeciles range هو عبارة عن الفرق بين العشيري التاسع (D_2) والعشيري الأول(D_1) ، ويدعي بمجال Kelly ويرمز له عادة بالرمز R_D أو R_K ويكتب على الشكل التالي

$$R_k = R_D = D_9 - D_1$$

4-1-1-3:المدى المئيني :intercentiles range :هو عبارة عن الفرق بين المئين التسعين (P_{90}) والمئين العاشر (P_{10}) وهو نفسه المدي العشيري.يرمز له بالرمز (R_P) ويكتب على الشكل:

$$R_P = P_{90} - P_{10}$$

غير أن المقاييس السابقة (المديات) قليلة الاستعمال لوجود مقياس أفضل منها جميعاً ، يسهل معالجته حسابياً وتتوفر فيه خواص التقدير الاحصائي الجيد وهو الانحراف المعياري الذي سنتطرق إليه فيما بعد.

متى نستخدم المدى (R)؟

نستخدم المدى حين:

- نرغب في قياس سريع لمدى تشتت المفردات دون إعطاء أهمية كبيرة لدقة القياس.
- يكون للمفردات المتطرفة أهمية خاصة.

متى نستخدم المدى الربيعي (R_Q)؟

نستخدم المدى الربيعي حين:

- يكون الوسيط هو مقياس القيمة الوسطى المحسوبة من البيانات.
- يكون التوزيع مفتوحاً أو مبتوراً.
- يكون هناك مفردات قليلة متطرفة أو يكون التوزيع شديد الإلتواء.
- حالة البيانات الوصفية التي تتطلب الترتيب.

3-1-1-5:الانحراف المتوسط average deviation أو متوسط الانحرافات المطلقة

يقاس الانحراف المتوسط لأية مجموعة من القيم سواء كانت مبوبة أو غير مبوبة في توزيع تكراري (بسيط أو ذي فئات)بحساب متوسط انحرافات هذه القيم عن وسطها الحسابي أي بحساب متوسط المجموع الكلي للانحرافات عن هذا المتوسط، يرمز له بالرمز $(\mathrm{E}\bar{X})$ ويقيس مدى تباعد أو قرب المفردات من قيمة وسطى"الوسط الحسابي"

1. حالة البيانات غير المبوبة:

لتكن لدينا مجموعة القيم التالية: $x_1 \quad x_2 \quad x_3 \quad x_4 \dots\dots\dots\dots\dots \quad x_k$ فإن انحرافها المتوسط يكتب بالشكل:

$$\mathrm{E}\bar{X} = \frac{\sum_{i=1}^{k}(X_I - \bar{X})}{N}$$

ولكن وبما أن $\sum_{i=1}^{k}(X_I - \bar{X}) = 0$ (هذا المجموع يساوي الصفر دائماً) لأن مجموع الانحرافات الموجبة عن الوسط الحسابي يساوي مجموع الانحرافات السالبة، لذلك لابد من حذف الإشارة السالبة لنحصل على مقياس ذي معنى.

إن إحدى الطرق للتخلص من الإشارة السالبة هي بأخذ القيمة المطلقة، فنحصل على الانحراف المتوسط المطلق، وبالتالي يصبح الانحراف المتوسط هو عبارة عن المتوسط الحسابي لمجموع الفوارق المطلقة لمختلف القيم أو البيانات عن وسطها الحسابي في هذه الحالة

$$\mathrm{E}\bar{X} = \frac{\sum_{i=1}^{k}(X_I - \bar{X})}{N} = \frac{\sum X_I - \sum \bar{X}}{N} = \frac{\sum X_I - \bar{X}\sum(1)}{N} = \frac{\sum X_I - N\bar{X}}{N} = \frac{\sum X_I}{N} - \frac{N\bar{X}}{N} = \bar{X} - \bar{X} = 0$$

$$\Rightarrow \sum_{i=1}^{k}(X_I - \bar{X}) = 0$$

لذلك نأخذ الفروقات بالقيمة المطلقة كما هو موشح في الصيغة $\mathrm{E}\bar{X} = \frac{\sum_{i=1}^{k}|X_I - \bar{X}|}{N}$

مثال(3–6):أوجد الانحراف المتوسط للبيانات التالية: 7 9 4 5 8 10 7 6

الحل: نحسب أولاً الوسط الحسابي للبيانات

$$\bar{X} = \frac{\sum X_I}{N} = \frac{7+9+4+5+8+10+7+6}{8} = \frac{56}{8} = 7 = \bar{X}$$

ومنه نحسب الانحراف المتوسط

$$E\bar{X} = \frac{\sum_{i=1}^{k} |X_I - \bar{X}|}{N}$$

$$\Rightarrow E\bar{X} = \frac{|7-7|+|9-7|+|4-7|+|5-7|+|8-7|+|10-7|+|7-7|+|6-7|}{8}$$

$$= \frac{0+2+3+2+1+3+0+1}{8} = \frac{12}{8} = 1.5 \quad \Rightarrow E\bar{X} = 1.5$$

وهو ما يعني أن القيم تنحرف عن متوسطها في المتوسط بـ 1.5 زيادة أو نقصاناً.

2. **حالة البيانات المبوبة في توزيع تكراري بسيط:**

وكانت البيانات والتي عددها N هيx_k..... x_1 x_2 x_3 x_4 والتكرارت المقابلة لها هي

$n_1\ n_2\ n_3\ n_4 \dots\dots\dots n_k$ فإن الانحراف المتوسط يكتب على الشكل التالي:

$$E\bar{X} = \frac{\sum_{i=1}^{k} n_i(X_i - \bar{X})}{N} \qquad N = \sum_{i=1}^{k} n_i$$

وبما أن $\sum_{i=1}^{k} {}_i(X_i - \bar{X}) = 0$ فإننا نأخذ القيمة المطلقة للانحرافات

$$E\bar{X} = \frac{\sum_{i=1}^{k} n_i(X_i - \bar{X})}{N} = \frac{\sum_{i=1}^{k} n_i X_i - \bar{X}\sum_{i=1}^{k} n_i}{N} = \frac{\sum_{i=1}^{k} n_i X_i - N\bar{X}}{N} = \frac{\sum_{i=1}^{k} n_i X_i}{N} - \frac{N\bar{X}}{N} =$$

$$\bar{X} - \bar{X} = 0 \Rightarrow \sum_{i=1}^{k} n_i(X_i - \bar{X}) = 0 \quad \Rightarrow \quad E\bar{X} = \frac{\sum_{i=1}^{k} n_i|X_i - \bar{X}|}{N} \qquad \text{حيث} \quad N = \sum_{i=1}^{k} n_i$$

مثال(3–7): احسب الانحراف المتوسط للتوزيع التكراري التالي:

x_i	03	07	11	15	19	$\sum$
n_i	2	5	9	7	2	N=25

الحل: بما اننا امام توزيع تكراري بسيط فإن الانحراف المتوسط يكتب كما يلي

$E\bar{X} = \frac{\sum_{i=1}^{k} n_i|X_i-\bar{X}|}{N}$ مع $\bar{X} = \frac{\sum_{i=1}^{k} n_i X_i}{N}$ حيث $N=\sum_{i=1}^{k} n_i$

$n_i\|X_i - \bar{X}\|$	$\|X_i - \bar{X}\|$	$(X_i - \bar{X})$	$x_i n_i$	n_i	x_i
16.64	8.32	3-11.32=-8.32	06	2	03
21.60	4.32	7-11.32=-4.32	35	5	07
2.88	0.32	11-11.32=-0.32	99	9	11
25.76	3.68	15-11.32=3.68	105	7	15
15.36	7.68	19-11.32=7.68	38	2	19
82.24	----	---------	283	N=25	$\sum$

الوسط الحسابي

$\bar{X} = \frac{\sum_{i=1}^{k} n_i X_i}{N}$=11.32

$E\bar{X} = \frac{\sum_{i=1}^{k} n_i|X_i-\bar{X}|}{N} = \frac{82.24}{25}$ =3.29

وهو ما يعني أن القيم تنحرف عن متوسطها في المتوسط بـ 3.29 زيادة أو نقصاناً.

3. حالة البيانات المبوبة في توزيع تكراري ذي فئات

إذا كانت لدينا مجموعة من البيانات مجمعة في فئات مراكزها هي $c\ c_2\ c_3\ c_4 \dots \dots \dots c_k$ وتكرارتها هي $n_1\ n_2\ n_3\ n_4 \dots \dots \dots n_k$ فإن انحرافها المتوسط يكتب على الشكل التالي:

$$E\bar{X} = \frac{\sum_{i=1}^{k} n_i(C_i-\bar{X})}{N} \qquad N=\sum_{i=1}^{k} n_i$$

وبما أن $\sum_{i=1}^{k} {}_i(C_i - \bar{X}) = 0$ فإن الانحراف المتوسط يكتب بالشكل

$$E\bar{X} = \frac{\sum_{i=1}^{k} n_i |C_i - \bar{X}|}{N}$$

مثال(3–8):ايجاد الوسط الحسابي والانحراف المعياري للبيانات التالية

x_i فئات الأجر	60-50	70-60	80-70	90-80	100-90	110-100	120-110	Σ
n_i عدد العمال	8	10	16	14	10	05	02	65

الحل:بما أن البيانات مبوبة في توزيع تكراري ذي فئات فإننا نكتب:

$$\bar{X} = \frac{\sum_{i=1}^{k} n_i c_i}{N} \quad \text{et} \quad E\bar{X} = \frac{\sum_{i=1}^{k} n_i |C_i - \bar{X}|}{N} \quad N = \sum_{i=1}^{k} n_i$$

$$i = 1 ; 2 ; 3 \ldots 7$$

$n_i \lvert C_i - \bar{X} \rvert$	$\lvert C_i - \bar{X} \rvert$	$(C_i - \bar{X})$	$n_i c_i$	c_i	n_i	x_i
198.16	*24.77*	*55-79.77=-24.77*	*440*	*55*	*08*	*60-50*
147.70	*14.77*	*65-79.77=-14.77*	*650*	*65*	*10*	*70-60*
76.32	*4.77*	*75-79.77=-04.77*	*1200*	*75*	*16*	*80-70*
73.22	*5.23*	*85-79.77=+5.23*	*1190*	*85*	*14*	*90-80*
152.30	*15.23*	*95-79.77=15.23*	*950*	*95*	*10*	*100-90*
126.15	*25.23*	*105-79.77=25.23*	*525*	*105*	*05*	*110-100*
70.46	*35.23*	*115-79.77=35.23*	*230*	*115*	*02*	*120-110*
835.31	-----	----------	*5185*		*65*	

$$\bar{X} = \frac{\sum_{i=1}^{k} n_i c_i}{N} = \frac{5185}{65} \cong 79.77$$

$$E\bar{X} = \frac{\sum_{i=1}^{k} n_i |C_i - \bar{X}|}{N} = \frac{835.31}{65} = 12.85$$

وهو ما يعني أن القيم تنحرف عن متوسطها في المتوسط بـ 12.85 زيادة أو نقصاناً.

مثال(3–9):لنفترض أنه لدينا عينة تتكون من 100 طالب،مبوبة في توزيع تكراري ذي فئات حسب الوزن

$\sum$	74-72	71-69	68-65	65-63	62-60	الفئات(الوزن بـ كلغ) x_i
100	08	27	42	18	05	عدد الطلبة n_i

المطلوب: حساب الانحراف المعياري لأوزان هؤلاء الطلبة

الحل:

بما أن البيانات مبوبة في توزيع تكراري ذي فئات فإن:

$\bar{X} = \frac{\sum_{i=1}^{k} n_i c_i}{N}$ et $E\bar{X} = \frac{\sum_{i=1}^{k} n_i |c_i - \bar{X}|}{N}$ $N = \sum_{i=1}^{k} n_i$

$n_i \lvert C_i - \bar{X} \rvert$	$\lvert C_i - \bar{X} \rvert$	$(C_i - \bar{X})$=(C_i-67.45)	$n_i c_i$	c_i	n_i	x_i
32.25	6.45	61-67.45=- 6.45	305	61	05	62-60
62.10	3.45	64-67.45=-3.45	1152	64	18	65-63
18.90	0.45	67-67.45=-0.45	2814	67	42	68-66
68.85	2.55	70-67.45=2.55	1890	70	27	71-69
44.50	5.55	73-67.45=5.55	554	73	08	74-72
226.50	--------	----------	6745	----	100	------

$\bar{X} = \frac{\sum_{i=1}^{k} n_i c_i}{N} = \frac{6745}{100} = 67.45$

$E\bar{X} = \frac{\sum_{i=1}^{k} n_i |c_i - \bar{X}|}{N} = \frac{226.5}{100} = 2.265$ kg

وهو ما يعني أن أوزان الطلبة تنحرف عن متوسطها في المتوسط بـ 2.265 كلغ زيادة أو نقصاناً.

خصائص الانحراف المتوسط: يلاحظ على هذا المقياس أنه:

1. يأخذ بعين الاعتبار جميع القيم.
2. يقيس تشتت القيم باستخدام الانحرافات عن مقياس النزعة المركزية.
3. قيمته تصغر كلما كبر حجم العينة وهو ما يعني زيادة تجانس التوزيع بزيادة حجم العينة، وهي خاصية للمقياس الجيد للتشتت.
4. يحسب على أساس الانحرافات المطلقة للمفردات عن وسطها الحسابي.
5. لا يمكن حسابه لعينات مجتمعة انطلاقاً من الانحراف المتوسط لكل مجموعة.
6. يعد إهمال إشارة الفروق أكبر عيوبه

متى يستخدم الانحراف المتوسط؟

يفضل استخدام الانحراف المتوسط حين:

1. تكون هناك مفردات متطرفة تؤثر بشكلٍ واضِحٍ على التباين بسبب تربيع الفروق بين القيم المختلفة ووسطها الحسابي.
2. يراد حساب مقياس مناسب للتشتت بديلاً للإنحراف المعياري خاصة وأنه متى كان التوزيع معتدلاً أو قريباً من الاعتدال، فإنه يمكن تقدير الانحراف المعياري بمعلومية متوسط الانحرافات المطلقة(الانحراف المتوسط).

ملاحظة: بنفس الطريقة يمكن حساب الانحراف الوسيطي كما يلي:

$$\mathrm{EM_e} = \frac{\sum_{i=1}^{k}|X_I - \mathrm{Me}|}{N} \qquad \mathrm{EM_e} = \frac{\sum_{i=1}^{k} n_i|X_i - \mathrm{Me}|}{N} \qquad \mathrm{EM_e} = \frac{\sum_{i=1}^{k} n_i|C_i - \mathrm{Me}|}{N}$$

كما يمكن وبنفس الطريقة حساب الانحراف المنوالي كما يلي:

$$EM_{\mathrm{o}} = \frac{\sum_{i=1}^{k}|X_I - \mathrm{M_o}|}{N} \qquad EM_{\mathrm{o}} = \frac{\sum_{i=1}^{k} n_i|X_i - \mathrm{M_o}|}{N}$$

$$EM_{\mathrm{o}} = \frac{\sum_{i=1}^{k} n_i|C_i - \mathrm{M_o}|}{N}$$

3-1-1-6:التباين والانحراف المعياري variance and standard deviation

رأينا في حساب الانحراف المتوسط، انه للتخلص من الاشارة السالبة للانحرافات عن الوسط الحسابي، نأخذ القيم المطلقة. غير أن هناك طريقة أخرى للتخلص من الاشارة السالبة وهي تربيع تلك الانحرافات عن الوسط الحسابي.

انطلاقاً من ذلك، يمكن تعريف التباين ومن ثم الانحراف المعياري كما يلي:

التباين هو عبارة عن الوسط الحسابي "المتوسط الحسابي" لمربعات الانحرافات عن الوسط الحسابي، ويرمز له بالرمز $\nu_{(x)}$ أو $\sigma^2(x)$ أو S^2

في حين أن الانحراف المعياري هو عبارة عن الجذر التربيعي لمتوسط مربعات انحراف المعطيات عن الوسط الحسابي للسلسلة. بعبارة أخرى الانحراف المعياري هو عبارة عن الجذر التربيعي للتباين ويرمز له عادة بالرمز $\sigma(x)$ أي ان الانحراف المعياري يمثل مؤشراً لدرجة تفاوت أو اختلاف أو تشتت القيم في البيانات، حيث يقوم حسابه أساساً على الانحراف بين الدرجات الأصلية ووسطها الحسابي أي $\sigma(x) = \sqrt{\sigma^2(x)} = \sqrt{\nu_{(x)}}$ ويمكن حساب التباين والانحراف المعياري بعدة طرق كما يلي:

أ. **الطريقة المباشرة:**

1. **حالة البيانات غير المبوبة:**

لتكن لدينا مجموعة القيم التالية: $x_1 \; x_2 \; x_3 \; x_4 \ldots\ldots\ldots\ldots \; x_k$ حيث وسطها الحسابي $\bar{x}$ ، فنحصل على التباين و من ثم الانحراف المعياري كما يلي

$$V(x) = \sigma^2(x) = = \frac{\sum_{i=1}^{k}(X_I - \bar{X})^2}{N} = \left[\frac{1}{N}\sum_{i=1}^{k}(X_I - \bar{X})^2\right]$$

ومنه فالانحراف المعياري يساوي

$$\sigma(x) = \sqrt{\nu_{(x)}} = \sqrt{\frac{\sum_{i=1}^{k}(X_I - \bar{X})^2}{N}} = \left[\frac{1}{N}\sum_{i=1}^{k}(X_I - \bar{X})^2\right]^{1/2}$$

مثال(3–10):لتكن لدينا السلسلة العددية التالية:20 18 15 12 10

المطلوب: حساب كل من التباين والانحراف المعياري.

الحل: نحسب الوسط الحسابي $\bar{x}$

$$\bar{x} = \frac{20+18+15+12+10}{5} = \frac{75}{5} = 15$$

نحسب انحرافات قيم السلسلة عن وسطها الحسابي فتكون كالتالي: +5 ، +3 ، 0 ، -3 ، -5

نحسب مربعات هذه الانحرافات فتكون كالتالي: 25 9 0 9 25 فيكون التباين هو

$$(x) = \sigma^2(x) = \frac{\sum_{i=1}^{k}(X_I - \bar{X})^2}{N} = \frac{25+9+0+5+25}{5} = \frac{68}{5} = 13.6 = (x)$$

ومنه فالانحراف المعياري يساوي

$$\sigma(x) = \sqrt{(x)\ \nu} = \sqrt{13.6} = 3.68 = \sigma(x)$$

2. حالة بيانات مبوبة في توزيع تكراري بسيط:

إذا كانت لدينا مجموعة من القيم (kقيمة) هيx_k..... x_1 x_2 x_3 x_4 والتكرارت المقابلة لها هي $n_1\ n_2\ n_3\ n_4 \dots \dots \dots n_k$ فإن تباينها يكتب على الشكل التالي:

$$v(x) = \sigma^2(x) = \frac{\sum_{i=1}^{k} n_i(X_i - \bar{X})^2}{N} \qquad N = \sum_{i=1}^{k} n_i \qquad \bar{X} = \frac{\sum_{i=1}^{k} n_i X_i}{N}$$

وبالتالي فانحرافها المعياري يكتب على الشكل التالي:

$$\sigma(x) = \sqrt{(x)\ \nu} = \sqrt{\frac{\sum_{i=1}^{k} n_i(X_i - \bar{X})^2}{N}}$$

مثال(3-11): احسب التباين والانحراف المعياري للتوزيع التكراري البسيط التالي:

$\sum$	19	15	11	07	03	x_i
25	02	07	09	05	02	n_i

الحل: بما أن البيانات موزعة في شكل توزيع تكراري بسيط فإنه لحساب التباين نتبع الخطوات التالية:

- نحسب الفروقات بين $(X_I - \bar{X})$
- نربع كل واحد من هذه الفروقات $(X_i - \bar{X})^2$
- نضرب مربعات هذه الفروقات في التكرارات المقابلة لها $n_i(X_i - \bar{X})^2$
- نجمع هذه الفروقات المضروبة في التكرارات المقابلة لها $\sum_{i=1}^{k} n_i(X_i - \bar{X})^2$
- نقسم المجموع على عدد المشاهدات N فنحصل على قيمة التباين، وبجذره نحصل على قيمة الانحراف المعياري

الجدول ادناه يلخص كل هذه الخطوات.

$n_i(X_i - \bar{X})^2$	$(X_i - \bar{X})^2$	$(X_I - \bar{X})$	$n_i x_i$	n_i	x_i
138.4448	69.2224	3-11.32 = - 8.32	06	02	03
93.3120	18.6624	7-11.32 = - 4.32	35	05	07
0.9216	00.1024	11-11.32 = - 0.32	99	09	11
94.7968	13.5424	15-11.32 = +3.68	105	07	15
117.9648	58.9824	9-11.32 = + 7.32	38	02	19
445.44	--------	----------	283	25	$\sum$

$$\bar{X} = \frac{\sum_{i=1}^{k} n_i X_i}{N} = \frac{283}{25} = 11.32$$

$$v(x) = \sigma^2(x) = \frac{\sum_{i=1}^{k} n_i(X_i - \bar{X})^2}{N} = \frac{445.44}{25} = 17.82 = v(x) \Rightarrow \sigma(x) = \sqrt{(x)\ v} = \sqrt{17.82} = 4.22 = \sigma(x)$$

3. حالة بيانات مبوبة في توزيع تكراري ذي فئات

إذا كانت لدينا مجموعة من البيانات مجمعة في فئات مراكزها هي $c\ c_2\ c_3\ c_4 \dots\dots\dots c_k$ وتكرارتها على الترتيب هي $n_1\ n_2\ n_3\ n_4 \dots\dots\dots n_k$ فإن تبتاينها يعطى بالعلاقة التالية:

$$v(x)= \sigma^2(x) = \frac{\sum_{i=1}^{k} n_i(C_i-\bar{X})^2}{N} \qquad N=\sum_{i=1}^{k} n_i \qquad \bar{X} = \frac{\sum_{i=1}^{k} n_i C_i}{N}$$

وبالتالي يعطى انحرافها المعياري بالعلاقة التالية:

$$\sigma(x)=\sqrt{(x)\ v} = \sqrt{\frac{\sum_{i=1}^{k} n_i(C_i-\bar{X})^2}{N}}$$

مثال(3-12): إذا كانت لدينا مجموعة من البيانات موزعة في توزيع تكراري ذي فئات كما يلي:

الفئات x_i	10-0	20-10	30-20	40-30	$\sum$
التكرار n_i	1	2	3	4	10

المطلوب: احسب كلا من التباين والانحراف المعياري لهذه البيانات.

الحل:

الفئات x_i	التكرار n_i	مراكز الفئات c_i	$n_i c_i$	$(C_I - \bar{X})$	$(C_i - \bar{X})^2$	$n_i(C_i - \bar{X})^2$
10-0	1	5	05	5-25=-20	400	400
20-10	2	15	30	15-25=-10	100	200
30-20	3	25	75	25-25=0	0	0
40-30	4	35	140	35-25=+10	100	400
-----	10	----	250		600	1000

$$\bar{X} = \frac{\sum_{i=1}^{k} n_i c_i}{N} = \frac{250}{10} = 25 = \bar{X}$$

$$v(x)= \sigma^2(x) = \frac{\sum_{i=1}^{k} n_i(c_i-\bar{X})^2}{N} = \frac{1000}{10} = 100 = v(x) \Rightarrow \sigma(x)=\sqrt{(x)\ v} = \sqrt{100} = 10 = \sigma(x)$$

ملاحظة:تعترض تطبيق هذه القوانين أحياناً حسابات طويلة ومتعددة، خاصة عند ما يتعلق الأمر بحساب مجموع مربعات الانحرافات عن الوسط الحسابي حالة ما إذا تضمن الوسط كسوراً عشرية.فهناك بعض الطرق التي تسمح باختصار هذه الحسابات منها:

أ. **الطريقة غير المباشرة:**

1. الطريقة الأولى: الطريقة المختصرة

أ. **حالة بيانات غير مبوبة:**

ب. $v(x)= \sigma^2(x) = \frac{\sum_{i=1}^{k}(X_I - \bar{X})^2}{N} = \frac{\sum(x_i^2 - 2x_i\bar{x} + \bar{x}^2)}{N} = \frac{\sum x_i^2 - \sum 2x_i\bar{x} + \sum\bar{x}^2}{N}$

$= \frac{\sum x_i^2 - 2\bar{x}\sum x_i + \bar{x}^2\sum(i)}{N} = \frac{\sum x_i^2}{N} - 2\bar{x}\frac{\sum x_i}{N} + \bar{x}^2\frac{N}{N} = \frac{\sum x_i^2}{N} - 2\bar{x}^2 + \bar{x}^2$

$v(x)= \frac{\sum x_i^2}{N} - \bar{x}^2 = Q^2 - \bar{x}^2 \Rightarrow \sigma(x) = \sqrt{v(x)} = \sqrt{\frac{\sum x_i^2}{N} - \bar{x}^2} = \sqrt{Q^2 - \bar{x}^2} = \sigma(x)$

مثال: بالعودة للمثال (3-10) الخاص بالبيانات غير المبوبة التالية:

x_i :	20	18	15	12	10
$x_i^2 =$	400	324	225	144	100

$\bar{x} = 15 \Rightarrow \bar{x}^2 = 225$

$v(x)= \frac{\sum x_i^2}{N} - \bar{x}^2 = \frac{1193}{5} - (25)^2 = 238.6 - 225 = 13.6 = v(x) \Rightarrow \sigma(x) = \sqrt{v(x)}$

$= \sqrt{13.6} = 3.68 = \sigma(x)$ وهي نفس النتيجة المحسوبة سابقاً

ب. **حالة بيانات مبوبة في توزيع تكراري بسيط:**

$$v(x)= \sigma^2(x) = \frac{\sum_{i=1}^{k} n_i(X_i - \bar{X})^2}{N} = \frac{\sum_{i=1}^{k} n_i(x_i^2 - 2x_i\bar{x} + \bar{x}^2)}{N} = \frac{\sum_{i=1}^{k}(n_i x_i^2 - 2n_i x_i\bar{x} + \bar{x}^2 n_i)}{N}$$

$$\Rightarrow v(x)=\frac{\sum n_i x_i^2-2\bar{x}\sum n_i x_i+\bar{x}^2\sum n_i}{N}=\frac{\sum n_i x_i^2}{N}-\ 2\bar{x}\frac{\sum n_i x_i}{N}+\bar{x}^2\frac{\sum n_i}{N}$$

$$\Rightarrow\ v(x)=\frac{\sum n_i x_i^2}{N}\text{ - }2\bar{x}\bar{x}\ +\bar{x}^2\frac{N}{N}=\frac{\sum n_i x_i^2}{N}\text{ - }2\bar{x}^2+\bar{x}^2$$

$$\Rightarrow v(x)=\frac{\sum n_i x_i^2}{N}\text{ - }\bar{x}^2 \qquad \text{مع} \qquad \text{N=}\sum_{i=1}^{k} n_i \qquad \bar{X}=\frac{\sum_{i=1}^{k} n_i x_i}{N}$$

$$v(x)=\frac{\sum n_i x_i^2}{N}\text{ - }\bar{x}^2\ \Rightarrow\ \sigma(x)=\sqrt{(\text{x})\ \text{v}}\ =\sqrt{\frac{\sum n_i x_i^2}{N}-\ \bar{x}^2}\ =\sqrt{Q^2\ -\ \bar{x}^2}$$

مثال: بالعودة للمثال(3–11)(حالة البيانات المبوبة في توزيع تكراري بسيط)

$n_i x_i^2$	x_i^2	$n_i x_i$	n_i	x_i
18	09	06	02	03
245	49	35	05	07
1089	121	99	09	11
1575	225	105	07	15
722	361	38	02	19
3649	------	283	25	$\sum$

$$\bar{x}=\frac{\sum_{i=1}^{k} n_i x_i}{N}=\frac{283}{25}=11.32=\ \bar{x}$$

$$v(x)=\frac{\sum n_i x_i^2}{N}\text{ - }\bar{x}^2\ =\frac{3649}{25}-(11.32)^2=145.96-128.14=17.82=v(x)$$

$$\sigma(x)=\sqrt{(\text{x})\ \text{v}}\ =\sqrt{17.82}=4.\ 22=\sigma(x)$$

ت. حالة بيانات مبوبة في توزيع تكراري ذي فئات:

$$v(x)=\ \sigma^2(\text{x})\ =\frac{\sum_{i=1}^{k} n_i(C_i-\bar{X})^2}{N}=\frac{\sum_{i=1}^{k} n_i\left(C_i^2-2C_i\bar{x}+\bar{x}^2\right)}{N}=\frac{\sum_{i=1}^{k}\left(n_i C_i^2-2n_i C_i\bar{x}+\bar{x}^2 n_i\right)}{N}$$

$$\Rightarrow v(x)=\frac{\sum n_i C_i^2-2\bar{x}\sum n_i C_i+\bar{x}^2\sum n_i}{N}=\frac{\sum n_i C_i^2}{N}-\ 2\bar{x}\frac{\sum n_i C_i}{N}+\bar{x}^2\frac{\sum n_i}{N}$$

$$v(x) = \frac{\sum n_i C_i^2}{N} - 2\bar{x}\bar{x} + \bar{x}^2 \frac{N}{N} = \frac{\sum n_i C_i^2}{N} - 2\bar{x}^2 + \bar{x}^2$$

$$\Rightarrow v(x) = \frac{\sum n_i C_i^2}{N} - \bar{x}^2 \qquad \text{مع} \qquad N = \sum_{i=1}^{k} n_i \qquad \bar{X} = \frac{\sum_{i=1}^{k} n_i C_i}{N}$$

$$v(x) = \frac{\sum n_i C_i^2}{N} - \bar{x}^2 \Rightarrow \sigma(x) = \sqrt{v(x)} = \sqrt{\frac{\sum n_i C_i^2}{N} - \bar{x}^2}$$

مثال: بالعودة إلى المثال (3-12): الخاص بالبيانات المبوبة في جدول توزيع تكراري ذي فئات

الفئات x_i	التكرار n_i	مراكز الفئات c_i	$n_i c_i$	c_I^2	$n_i c_I^2$
10-0	1	5	05	25	25
20-10	2	15	30	225	450
30-20	3	25	75	625	1875
40-30	4	35	140	1225	4900
-----	10	----	250	------	7250

$$\bar{X} = \frac{\sum_{i=1}^{k} n_i c_i}{N} = \frac{250}{10} = 25 = \bar{X}$$

$$v(x) = \frac{\sum n_i C_i^2}{N} - \bar{x}^2 = \frac{7250}{10} - (25)^2 = 725 - 625 = 100 = v(x)$$

$v(x) = 100$ et $\sigma(x) = \sqrt{v(x)} = \sqrt{100} = 10$ وهي نفس النتيجة المتوصل إليها بالطريقة السابقة

2. الطريقة الثانية: طريقة الانحرافات عن وسط فرضي x_o

أ. **حالة البيانات غير المبوبة:** $v(x) = \sigma^2(x) = \frac{\sum_{i=1}^{k} (X_I - \bar{X})^2}{N}$

نضع

$$(x_I - \bar{x}) = (x_i - x_o) - (\bar{x} - x_o) = \Rightarrow$$

$$v(x)=\sigma^2(x)=\frac{\sum_{i=1}^{k}(X_I-\bar{X})^2}{N}=\frac{\sum[(x_i-x_o)-(\bar{x}-x_o)]^2}{N}=\frac{\sum[(x_i-x_o)^2-2(x_i-x_o)(\bar{x}-x_o)]+(\bar{x}-x_o)^2}{N}$$

$$=\frac{\sum(x_i-x_o)^2-2\sum(x_i-x_o)(\bar{x}-x_o)+\sum(\bar{x}-x_o)^2}{N}=\frac{\sum(x_i-x_o)^2-2(\bar{x}-x_o)\sum(x_i-x_o)+(\bar{x}-x_o)^2.N}{N}$$

$$=\frac{\sum(x_i-x_o)^2}{N}-2(\bar{x}-x_o)\frac{\sum(x_i-x_o)}{N}+(\bar{x}-x_o)^2\frac{N}{N}$$

$$=\frac{\sum(x_i-x_o)^2}{N}-2(\bar{x}-x_o)(\bar{x}-x_o)+(\bar{x}-x_o)^2 \qquad \text{لأن} \quad \frac{\sum(x_i-x_o)}{N}=(\bar{x}-x_o)$$

$$v(x)=\frac{\sum(x_i-x_o)^2}{N}-(\bar{x}-x_o)^2 \;\; et \;\; \sigma(x)=\sqrt{v(x)}=\sqrt{\frac{\sum(x_i-x_o)^2}{N}-(\bar{x}-x_o)^2}$$

مثال: بالعودة إلى **المثال(3–10)** الخاص بالبيانات غير المبوبة السابقة نجد:

x_i : 20 18 15 12 10 حيث $\bar{x}$ =15

نضع مثلاً x_o =12 $\Leftarrow$

(x_i-x_o) : 8+ 6+ 3+ 0 2–

$(x_i-x_o)^2$: 64 36 9 0 4

$$(\bar{x}-x_o)=(15-12)=3 \Rightarrow (\bar{x}-x_o)^2=9$$

$$v(x)=\frac{\sum(x_i-x_o)^2}{N}-(\bar{x}-x_o)^2=\frac{64+36+9+0+4}{5}-(3)^2=\frac{113}{5}-9=22.6-9=13.6=v(x)$$

$$\sigma(x)=\sqrt{(x)\;v}=\sqrt{13.6}=3.68=\sigma(x)$$

ب. حالة بيانات مبوبة في توزيع تكراري بسيط

كانت لدينا صيغة التباين في هذه الحالة هي

$$v(x)=\sigma^2(x)=\frac{\sum_{i=1}^{k}n_i(X_i-\bar{X})^2}{N}$$

$$(x_I-\bar{x})=(x_i-x_o)-(\bar{x}-x_o)= \Rightarrow$$

$$v(x)=\sigma^2(x)=\frac{\sum_{i=1}^{k}n_i(X_i-\bar{X})^2}{N}=\frac{\sum_{i=1}^{k}n_i\left((x_i-x_o)-(\bar{x}-x_o)\right)^2}{N}$$

$$v(x)=\sigma^2(x)=\frac{\sum_{i=1}^{k}n_i\left((x_i-x_o)-(\bar{x}-x_o)\right)^2}{N}=\frac{\sum_{i=1}^{k}n_i[(x_i-x_o)^2-2(x_i-x_o)(\bar{x}-x_o)+(\bar{x}-x_o)^2]}{N}=$$

$$v(x) = \frac{\sum_{i=1}^{k} n_i (x_i - x_o)^2 - 2(\bar{x} - x_o) \sum_{i=1}^{k} n_i (x_i - x_o) + (\bar{x} - x_o)^2 \sum_{i=1}^{k} n_i}{N}$$

$$v(x) = \frac{\sum_{i=1}^{k} n_i (x_i - x_o)^2 - 2(\bar{x} - x_o) \sum_{i=1}^{k} n_i (x_i - x_o) + (\bar{x} - x_o)^2 N}{N}$$

$$v(x) = \frac{\sum_{i=1}^{k} n_i (x_i - x_o)^2}{N} - 2(\bar{x} - x_o) \frac{\sum_{i=1}^{k} n_i (x_i - x_o)}{N} + (\bar{x} - x_o)^2 \frac{N}{N}$$

$$v(x) = \frac{\sum_{i=1}^{k} n_i (x_i - x_o)^2}{N} - 2(\bar{x} - x_o)(\bar{x} - x_o) + (\bar{x} - x_o)^2 \quad \text{مع} \quad \frac{\sum_{i=1}^{k} n_i (x_i - x_o)}{N} = (\bar{x} - x_o)$$

$$v(x) = \frac{\sum_{i=1}^{k} n_i (x_i - x_o)^2}{N} - (\bar{x} - x_o)^2 \quad et\ \sigma(x) = \sqrt{v(x)} \sqrt{\frac{\sum_{i=1}^{k} n_i (x_i - x_o)^2}{N} - (\bar{x} - x_o)^2}$$

مثال: بالعودة إلى **المثال(3–11)** الخاص بالبيانات المبوبة في توزيع تكراري بسيط نجد:

نفترض أن $x_o = 11$ فيكون:

x_i	n_i	$n_i x_i$	$(x_i - x_o)$	$(x_i - x_o)^2$	$n_i (x_i - x_o)^2$
03	02	06	3 – 11 = -8	64	128
07	05	35	7 – 11 = - 4	16	80
11	09	99	11 - 11 =0	0	0
15	07	105	15 – 11 = 4	16	112
19	02	38	19 – 11 =8	64	128
∑	25	283	--------	----	448

$$\bar{x} = \frac{\sum_{i=1}^{k} n_i x_i}{N} = \frac{283}{25} = 11.32 = \bar{x}$$

$$(\bar{x} - x_o) = (11.32 - 11) = 0.32 \Rightarrow (\bar{x} - x_o)^2 = 0.1024$$

$$v(x) = \frac{\sum_{i=1}^{k} n_i (x_i - x_o)^2}{N} - (\bar{x} - x_o)^2 = \frac{445}{25} - (0.32)^2 = 17.92 - 0.1024 = 17.82$$

$$Et\ \sigma(x) = \sqrt{v(x)} = \sqrt{17.82} = 4.22 = \sigma(x)$$

ت. حالة بيانات مبوبة في توزيع تكراري ذي فئات: $v(x)$

بنفس الطريقة لدينا: $v(x)= \sigma^2(x) = \frac{\sum_{i=1}^{k} n_i(C_i-\bar{X})^2}{N}$

$$(x_I - \bar{x}) = (x_i - x_o) - (\bar{x} - x_o) = \Rightarrow$$

$$v(x)=\sigma^2(x) = \frac{\sum_{i=1}^{k} n_i(C_i-\bar{X})^2}{N} = \frac{\sum_{i=1}^{k} n_i\left((C_i-x_o)-(\bar{x}-x_o)\right)^2}{N}$$

$$v(x)=\sigma^2(x) = \frac{\sum_{i=1}^{k} n_i\left((C_i-x_o)-(\bar{x}-x_o)\right)^2}{N} = \frac{\sum_{i=1}^{k} n_i[(C_i-x_o)^2- 2(C_i-x_o)(\bar{x}-x_o)+(\bar{x}-x_o)^2]}{N} =$$

$$v(x) = \frac{\sum_{i=1}^{k} n_i (C_i-x_o)^2-2(\bar{x}-x_o)\sum_{i=1}^{k} n_i(C_i-x_o)+(\bar{x}-x_o)^2\sum_{i=1}^{k} n_i}{N}$$

$$v(x) = \frac{\sum_{i=1}^{k} n_i (C_i-x_o)^2-2(\bar{x}-x_o)\sum_{i=1}^{k} n_i(C_i-x_o)+(\bar{x}-x_o)^2 N}{N}$$

$$v(x) = \frac{\sum_{i=1}^{k} n_i (C_i-x_o)^2}{N} - 2(\bar{x} - x_o)\frac{\sum_{i=1}^{k} n_i(C_i-x_o)}{N} +(\bar{x} - x_o)^2\frac{N}{N}$$

$$v(x) = \frac{\sum_{i=1}^{k} n_i (C_i-x_o)^2}{N} - 2(\bar{x} - x_o)(\bar{x} - x_o) +(\bar{x} - x_o)^2 \quad \text{مع} \quad \frac{\sum_{i=1}^{k} n_i(C_i-x_o)}{N} = (\bar{x} - x_o)$$

$$v(x) = \frac{\sum_{i=1}^{k} n_i (C_i-x_o)^2}{N} - (\bar{x} - x_o)^2 \quad et\ \sigma(x)=\sqrt{v(x)}\ \sqrt{\frac{\sum_{i=1}^{k} n_i (C_i-x_o)^2}{N} - (\bar{x} - x_o)^2}$$

مثال: بالعودة إلى **المثال (3–12)** الخاص بالبيانات المبوبة في توزيع تكراري ذي فئات نجد:

الحل: نفترض أن x_o =15 فيكون:

$n_i(C_i-\bar{X})^2$	$(C_i-\bar{X})^2$	$(C_I - \bar{X})$	$n_i c_i$	مراكز الفئات c_i	التكرار n_i	الفئات x_i
100	100	5–15= – 10	05	5	1	10–0
00	00	15–15 = 0	30	15	2	20–10
300	100	25 – 15=10	75	25	3	30–20
1600	400	35 –15 =20	140	35	4	40–30
2000	600		250	----	10	-----

$$\bar{X} = \frac{\sum_{i=1}^{k} n_i c_i}{N} = \frac{250}{10} = 25 = \bar{X}$$

$$(\bar{x} - c_o) = (25 - 15) = 10 \Rightarrow (\bar{x} - x_o)^2 = 100$$

$$v(x) = \frac{\sum_{i=1}^{k} n_i\,(c_i - x_o)^2}{N} - (\bar{x} - c_o)^2 = \frac{2000}{10} - (10)^2 = 200 - 100 = 100 = v(x)$$

$$Et\ \sigma(x) = \sqrt{v(x)} = \sqrt{100} = 10 = \sigma(x)$$

3. الطريقة الثالثة: طريقة تغيير المتغير

نستطيع القيام بتبديل المتغير مثلما تم التطرق إليه في كيفية حساب المتوسط الحسابي، حيث تقوم هذه الطريقة بتبديل المتغير x_i بالمتغير z_i كالتالي:

أ. حالة بيانات غير مبوبة

نضع $a\ et\ x_o$ ثابتان حيث أن $z_i = \frac{x_i - x_o}{a} \Rightarrow x_i = x_o + az_i \Rightarrow \bar{x} = x_o + a\bar{z}$

و أن $\bar{z}$ يسمى الوسط الحسابي المؤقت.

لدينا التباين في حالة بيانات غير مبوبة مساوٍ ل:

$$v(x) = \sigma^2(x) = \frac{\sum_{i=1}^{k}(X_I - \bar{X})^2}{N}$$

نعوض كل من x_i و $\bar{x}$ بما يساويهما فنحصل على:

$$v(x) = \frac{\sum_{i=1}^{k}(X_I - \bar{X})^2}{N} = \frac{\sum_{i=1}^{k}(x_o + az_i - x_o - a\bar{z})^2}{N} = \frac{\sum_{i=1}^{k}(az_i - a\bar{z})^2}{N} = \frac{\sum_{i=1}^{k} a^2(z_i - \bar{z})^2}{N} = a^2\frac{\sum_{i=1}^{k}(z_i - \bar{z})^2}{N}$$

$$v(x) = a^2\ v(z)$$

$$v(x) = \sigma^2(x) = a^2 v(z) = a^2\ \sigma^2(z) \Rightarrow \sigma(x) = a\,\sigma(z)$$

ومنه يمكن كتابة التباين والانحراف المعياري كما يلي:

$$v(x) = \sigma^2(x) = \frac{\sum_{i=1}^{k}(X_I - \bar{X})^2}{N}$$

$$v(x) = a^2 v(z) = a^2\frac{\sum_{i=1}^{k}(z_i - \bar{z})^2}{N} \qquad et\ \sigma(x) = a\,\sigma(z) \quad et\ a \text{ عدد ثابت}$$

مثال: بالعودة إلى **المثال(3-10)** الخاص بالبيانات غير المبوبة السابقة نجد:

x_i : 20 18 15 12 10

الحل: نضع a=2 و $15 = x_o$

$(z_i - \bar{z})^2$	$(z_i - \bar{z})$	$z_i = \frac{x_i - x_o}{a}$	x_i
6.25	2.5	$\frac{20-15}{2}$ =2.5	20
2.25	1.5	$\frac{18-15}{2}$ =1.5	18
00	00	$\frac{15-15}{2}$ =00	15
2.25	-1.5	$\frac{12-15}{2}$ =- 1.5	12
6.25	- 2.5	$\frac{10-15}{2}$ =- 2.5	10

$\bar{z} = \frac{\sum_{i=1}^{k} z_i}{N} = \frac{00}{5} = 0 = \bar{z}$

$\therefore v(z) = \frac{\sum_{i=1}^{k}(z_i - \bar{z})^2}{N} = \frac{17}{5} = 3.4 = v(z) \Rightarrow \sigma(z) = \sqrt{v(z)} = \sqrt{3.4} = 1.84 = \sigma(z)$

$v(x) = a^2 v(z) = 2^2(3.4) = 13.6 = v(x)$

$\sigma(x) = a\,\sigma(z) \Rightarrow \sigma(x) = 2.1(1.84) = 3.68 = \sigma(x)$

ب. حالة بيانات مبوبة في توزيع تكراري بسيط:

بنفس الطريقة نضع: ثابتان $a \ et x_o$ حيث أن $z_i = \frac{x_i - x_o}{a} \Rightarrow x_i = x_o + az_i \Rightarrow \bar{x} = x_o + a\bar{z}$

$v(x) = \sigma^2(x) = \frac{\sum_{i=1}^{k} n_i (X_i - \bar{X})^2}{N} = \frac{\sum_{i=1}^{k} n_i (x_o + az_i - x_o - a\bar{z})^2}{N}$

$= \frac{\sum_{i=1}^{k} n_i (az_i - a\bar{z})^2}{N} = a^2 \frac{\sum_{i=1}^{k} n_i (z_i - \bar{z})^2}{N} = v(x) = a^2\ v(z)$

$v(x) = \sigma^2(x) = a^2 v(z) = a^2\ \sigma^2(z) \Rightarrow \sigma(x) = a\,\sigma(z)$

$$v(x) = \frac{\sum_{i=1}^{k} n_i (X_i - \bar{X})^2}{N} = a^2 \frac{\sum_{i=1}^{k} n_i (z_i - \bar{z})^2}{N}$$

$$v(x) = \sigma^2(x) = a^2 v(z) = a^2\ \sigma^2(z) \Rightarrow \sigma(x) = a\,\sigma(z)$$

مثال: بالعودة إلى **المثال (3–11)** الخاص بالبيانات المبوبة في توزيع تكراري بسيط، نضع a=2 = $11 x_o$

$n_i(z_i-\bar{z})^2$	$(z_i-\bar{z})^2$	$(z_i-\bar{z})$	$n_i z_i$	$z_i=\frac{x_i-x_o}{a}$	n_i	x_i
34.6112	17.3056	- 4.16	-8	$(3-11)/2=-4$	02	03
23.3330	4.6666	-2.16	-10	$(7-11)/2=-2$	05	07
0.2304	0.0256	-0.16	00	$(11-11)/2=00$	09	11
23.6992	3.3856	+1.84	+14	$(15-11)/2=+2$	07	15
29.4912	14.7456	+3.84	+18	$(19-11)/2=-4$	02	19
111.365		-----	+4	---------	25	$\sum$

$\bar{z}=\frac{\sum_{i=1}^{k} n_i z_i}{N}=\frac{4}{25}=0.16=\bar{z}$

$v(z)=\frac{\sum_{i=1}^{k} n_i(z_i-\bar{z})^2}{N}=\frac{111.365}{25}=4.4546=v(z)\Rightarrow\sigma(z)=\sqrt{v(z)}=\sqrt{4.4546}=2.11=\sigma(z)$

$v(x)=a^2 v(z)=2^2(4.4546)=17.82=v(x)\ \ et$

$\sigma(x)=a\,\sigma(z)\Rightarrow\sigma(x)=2(2.11)=4.22=\sigma(x)$

ت. حالة بيانات مبوبة في توزيع تكراري ذي فئات

$z_i=\frac{C_i-x_o}{a}\Rightarrow C_i=c_o+az_i\Rightarrow\bar{x}=c_o+a\bar{z}$ حيث أن $a\ et\ x_o$ ثابتان بنفس الطريقة

نضع:

$v(x)=\sigma^2(x)=\frac{\sum_{i=1}^{k} n_i(C_i-\bar{X})^2}{N}=\frac{\sum_{i=1}^{k} n_i(c_o+az_i-c_o-a\bar{z})^2}{N}$

$=\frac{\sum_{i=1}^{k} n_i(az_i-a\bar{z})^2}{N}=a^2\frac{\sum_{i=1}^{k} n_i(z_i-\bar{z})^2}{N}=v(x)=a^2\ v(z)$

$v(x)=\sigma^2(x)=a^2 v(z)=a^2\ \sigma^2(z)\ \Rightarrow\sigma(x)=a\,\sigma(z)$

$v(x)=\frac{\sum_{i=1}^{k} n_i(C_i-\bar{X})^2}{N}=a^2\frac{\sum_{i=1}^{k} n_i(z_i-\bar{z})^2}{N}$

$v(x)=\sigma^2(x)=a^2 v(z)=a^2\ \sigma^2(z)\ \Rightarrow\sigma(x)=a\,\sigma(z)$

مثال: بالعودة إلى **المثال(3–12)** الخاص بالبيانات المبوبة في توزيع تكراري ذي فئات نجد:

أولاً: نضع a =10 و x_o = 15

$n_i(z_i-\bar{z})^2$	$(z_i-\bar{z})^2$	$(z_i-\bar{z})$	$n_i z_i$	$z_i=\frac{Ci-x_o}{a}$	C_i	n_i	x_i
4	**4**	**–1–1=–2**	**–1**	$\frac{(5-15)}{10}=-1$	**5**	**1**	**10–0**
2	**1**	**0–1=–1**	**00**	$\frac{(15-15)}{10}=0$	**15**	**2**	**20–10**
0	**0**	**1–1=0**	**+3**	$\frac{(25-15)}{10}=1$	**25**	**3**	**30–20**
4	**1**	**+2–1=1**	**+8**	$\frac{(35-15)}{10}=2$	**35**	**4**	**40–30**
10	**------**	**------**	**+10**	**-------**	**----**	**10**	**-----**

$$\bar{z}=\frac{\sum_{i=1}^{k} n_i z_i}{N}=\frac{10}{10}=1=\bar{z}$$

$$v(z)=\frac{\sum_{i=1}^{k} n_i (z_i-\bar{z})^2}{N}=\frac{10}{10}=1=v(z)\Rightarrow \sigma(z)=\sqrt{v(z)}=\sqrt{1}=1=\sigma(z)$$

$$v(x)=a^2 v(z)=10^2(1)=100=v(x)\ \ e$$

$$\sigma(x)=a\,\sigma(z)\Rightarrow \sigma(x)=10(10)=10=\sigma(x)$$ وهو المطلوب

ملاحظة: من أهم خواص التباين

1. إذا كان تباين مجموعة من البيانات σ_1^2 وعدد مفرداتها N_1 و تباين مجموعة أخرى مستقلة عن المجموعة الأولى σ_2^2 وعدد مفرداتها N_2 فإن تباين المجموعة الناتجة عن دمج المجموعتين يكونا مساوياً تقريباً لـ: $\sigma^2(x)=\frac{N_1\sigma_1^2+N_2\sigma_2^2}{N_1+N_2}$. وهذا فقط في حالة تساوي وسطهما الحسابي.

مثال: أجرى أحد الطلبة 10 تجارب لتعيين معامل الاحتكاك، فكان التباين في قياساته 0.4، وأجرى طالب آخر 7 تجارب فكان تباين قياساته 0.3 . فماهو التباين للقياسات من التجارب جميعها.

$$\begin{bmatrix} N_1=10 & \sigma_1^2=0.4 \\ N_2=7 & \sigma_2^2=0.3 \end{bmatrix}\Rightarrow \sigma^2(x)=\frac{N_1\sigma_1^2+N_2\sigma_2^2}{N_1+N_2}=\frac{10(0.4)+7(0.3)}{10+7}=\frac{4+2.1}{17}=0.36=\sigma^2$$

2. إذاكان الانحراف المعياري لمجموعتين من القيم تحوي N_1 و N_2 من المفردات هو σ_1 σ_2 على الترتيب مأخوذة حول الوسطين $\bar{x}_1$ و $\bar{x}_2$ لكل منهما.فإنه بضم المجموعتين لبعضهما في مجموعة واحدة تحوي (N_1+N_2) مفردة فإن الانحراف المعياري لهما σ مقاساً من المتوسط العام لهما يحسب وفق المعادلة التالية:

$$\sigma^2 = \frac{N_1\sigma_1^2+N_2\sigma_2^2}{N_1+N_2} + \frac{N_1N_2(\bar{x}_1+\bar{x}_2)}{N_1+N_2}$$

خصائص التباين ومن ثم الانحراف المعياري:

1. إن التباين وبالتالي الانحراف المعياري معم جداً في التحليلات الاحصائية، فالتباين مفيد عند مقارنة تشتت مجموعتين من البيانات حيث ان صغر قيمة التباين يدل على قلة تشتت البيانات.
2. التباين ومن ثم الانحراف المعياري يخضع للعمليات الجبرية.
3. يعتبر التباين افضل وأدق مقاييس التشتت.
4. إذا اضفنا أو حذفنا ثابت من كل قيمة من قيم مجموعة البيانات فإن المتوسط الحسابي لتلك البيانات يتأثر بتلك القيمة زيادة أو نقصاناً بينما التباين لا يتأثر.
5. إذا ضربنا أو قسمنا كافة البيانات بقيمة معينة أو على قيمة معينة تساوي b فإن قيمة التباين تزيد أو تنقص بمقدار b^2 .
6. تباين التوزيع حول متوسطه الحسابي أصغر من تباين هذه القيم حول أية قيمة أخرى بمقدار بعادل مربع الفرق بين المتوسط الحسابي الحقيقي وعذه القيمة
7. غذا كانت قيم التباين ومن ثم الانحراف المعياري صغيرة دلت على تركز المعطيات حول الوسط الحسابي والعكس صحيح أي انه إذا كانت قيمته كبيرة دلت على عدم التجانس وشدة تباعد قيم البيانات عن بعضها.
8. الانحراف المعياري أكثر مقاييس التشتت استخداماً، حيث يدخل في عدة حسابات منها الارتباط، كما يعد مؤشراً لعدد من التوزيعات النظرية مثل التوزيع الطبيعي الذي يكون فيه 50% من مفردات العينة تنحصر قيمتها بين قيمة المتوسط الحسابي لهذه المجموعة ناقص أو زائد 0.67 الانحراف المعياري للمجموعة أي

$$50\%\ \text{data} \in [\bar{x} \pm 0.67\,\sigma]$$

ملاحظات:

1. التباين ومن ثم الانحراف المعياري يأخذ بعين الاعتبار جنميع المفردات إلا أنه يعطي وزناً أو اهمية كبيرة للمفردات المتطرفة، وذلك بالمقارنة بمتوسط الانحرافات المطلقة،وهذا يعني أنه يتأثر كثيراُ بالقيم المتطرفة، وهو لهذا السبب قد بفضل على بقية المقاييس عند قياس تشتت توزيع شديد الإلتواء.
2. الانحراف المعياري للعينة يعد أحسن تقدير للانحراف المعياري الحقيقي للمجتمع، وذلك لأنه أكثر استقراراً من بقية المقاييس، وذلك لأن قيمته لا تختلف كثيراً من عينة لأخرى من نفس المجتمع.
3. الانحراف المعياري هو أكثر مقاييس التشتت استخداماً في الاحصاء، وأكثرها اهمية لإعتبارات متعددة تتعلق بعملية الاستنتاج الاحصائي سواء في التقديرات أو في اختبار الفرضيات.
4. يدخل الانحراف المعياري في تركيب العديد من المقاييس الاحصائية الهامة مثل مقاييس الالتواء والتباين النسبي والدرجات المعيارية ومعامل الارتباط.
5. هناك علاقة تجريبية تربط بين الانحراف المعياري وبين متوسط الانحرافات المطلقة ونصف المدى الربيعي ويشترط لتحقيق هذه العلاقة أن يكون التوزيع متماثلاً أو محدود الالتواء.وهذه العلاقة التجريبية هي

$$\begin{cases} E_{\bar{x}} = \frac{4}{5}\sigma(x) \\ R_Q/_2 = \frac{2}{3}\sigma(x) \quad ou \quad R_Q = \frac{4}{3}\sigma(x) \end{cases}$$

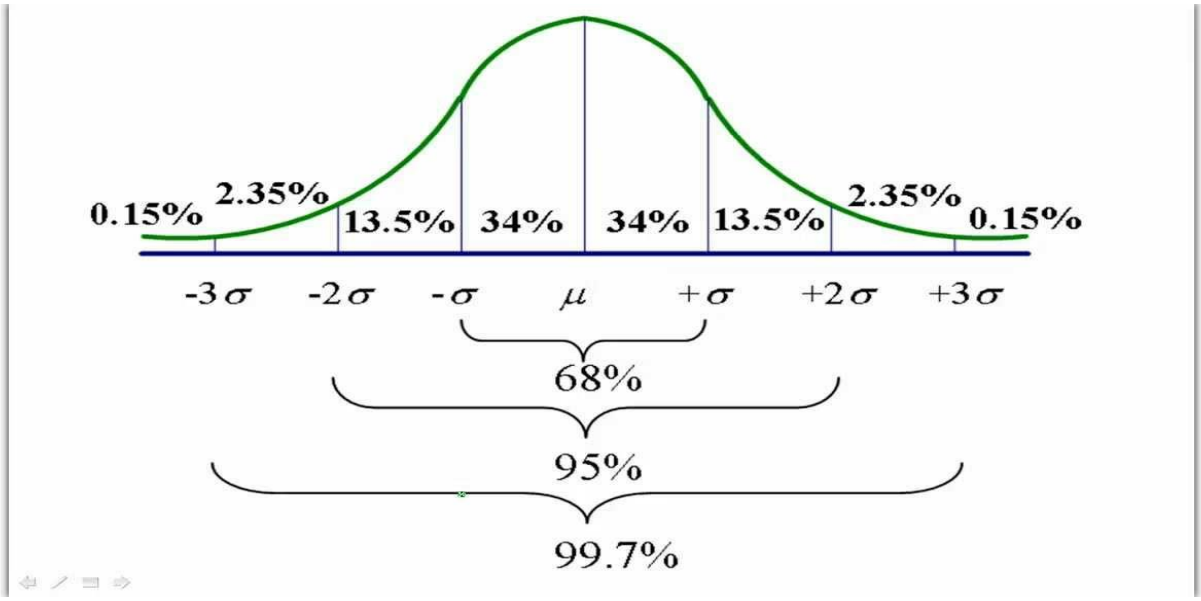
متى يستخدم الانحراف المعياري؟.

يفضل استخدام الانحراف المعياري في الحالات التالية:

1. حين نود حساب أفضل المقاييس للتشتت.
2. حين لا يكون قياس تشتت الظاهرة هو نهاية التحليل الاحصائي بل أنه بداية لعمليات احصائية أخرى أكثر اهمية، ونعني بها الاستنتاج الاحصائي بشقيه التقديرات والاختبارات الاحصائية.
3. هو مقياس لتحديد المساحة التي تقع تحت المنحنى، فهو منتشر الاستخدام لقياس درجة الثقة والاعتماد، حيث أنه في التوزيعات الطبيعية أو قليلة الالتواء يكون:

$$\begin{cases} 50\%\ \text{data} \in [\bar{x} \pm 0.67\ \sigma] \\ 68.28\%\ \text{data} \in [\bar{x} \pm 1\ \sigma] \\ 95.45\%\ \text{data} \in [\bar{x} \pm 2\ \sigma] \\ 99.73\%\ \text{data} \in [\bar{x} \pm 3\ \sigma] \\ 99.98\%\ \text{data} \in [\bar{x} \pm 4\ \sigma] \end{cases}$$

كما هو مبين في الشكل أدناه، حيث أن σ هو الانحراف المعياري



أي أن المجالات المشار إليها أعلاه تكون تحصر تلك النسب من المجموع الكلي من أو من المساحة الكلية تحت المنحنى.فكلما صغرت قيمة الانحراف المعياري كلما زادت امكانية الحصول على مفردات قريبة من الوسط الحسابي، والعكس صحيح.

3-1-2::مقاييس التشتت النسبي relative measures of dispersion

وهي عبارة عن أعداد مجردة من وحدة القياس لأنها عبارة عن نسبة التشتت المطلق إلى ‘حدى القيم المركزية أي أن

$$\text{مقياس التشتت} = \frac{\text{مقياس التشتت المطلق}}{\text{مقياس التشتت الذي حُسِب حوله التشتت المطلق}} \times 100.$$

هناك عدة مقاييس للتشتت النسبي من أهمها:

3-1-2-1المدى النسبي (معامل المدى): ويعطى بالعلاقة $C_{V_R} = V_R = \frac{R}{\bar{X}} .100$

بحيث:

V_R يمثل المدى النسبي
R المدى $Range$
$\overline{X}$ الوسط الحسابي

وقد نستعمل معامل المدى الربيعي النسبي الذي يكتب بالشكل التالي:

$$V_{R_Q} = \frac{Q_3 - Q_1}{Q_2} \cdot 100 = \frac{Q_3 - Q_1}{M_e} \cdot 100$$

بحيث :

V_{R_Q} المدى الربيعي النسبي (معامل المدى الربيعي).

Q_1 الربيعي الأول Q_2 الربيعي الثاني Q_3 الربيعي الثالث

3-1-2-2:الانحراف الربيعي النسبي (معامل التغير الربيعي):Quartile variation coefficient

ويعرف أيضاً باسم معامل الاختلاف الربيعي ويعتمد في حسابه على المدى الربيعي، ويرمز له بالرمز V_Q

ويكتب على الشكل $V_Q = \frac{\frac{Q_3 - Q_1}{2}}{\frac{Q_3 + Q_1}{2}} .100 = \frac{Q_3 - Q_1}{Q_3 + Q_1} .100 = \frac{\frac{Q_3 - Q_1}{2}}{M_e} .100 = \frac{\frac{Q_3 - Q_1}{2}}{Q_2} .100$

لأن $M_e = Q_2 = \frac{Q_3 + Q_1}{2}$

مثال(3-13): لدينا توزيع تكراري فيه $Q_{1=}$ 23.17 $Q_{3=}47.43$

وتوزيع تكراري آخر فيه $Q_{1=}$ 11.3 $Q_{3=}29.8$

فأي التوزيعين أكثر تغيراً؟.

الحل: نحسب معامل التغير الربيعي لكل من التوزيعين

$$V_{Q_1} = \frac{Q_3 - Q_1}{Q_3 + Q_1} \sigma(x) .100 = \frac{47.43 - 23.17}{47.43 + 23.17} .100 = 34.36\ \% = 34.36$$

$$V_{Q_2} = \frac{Q_3 - Q_1}{Q_3 + Q_1} .100 = \frac{29.8 - 11.3}{29.8 + 11.3} .100 = 45.01\ \% = 45.01$$

إذن بما أن $V_{Q_2} > V_{Q_1}$ ⇐ ان تغير التوزيع الثاني أكبر من تغير التوزيع الأول بمعنى آخر أن تشتت التوزيع الثاني أكبر من تشتت التوزيع الاول.

3-1-2-3: الانحراف المعياري النسبي (معامل الاختلاف) Variation coefficient

يعرف على أنه الانحراف المعياري معبراً عنه كنسبة مئوية من الوسط الحسابي، حيث نحصل عليه بقسمة الانحراف المعياري على الوسط الحسابي مضروباً في 100.ويكتب على الشكل التالي:

$$V=C.V=\frac{\sigma(x)}{\bar{x}} . 100.$$

غير أن ما يعاب على هذا المقياس أنه لا يصلح عي حالة التوزيعات التكرارية المفتوحة لتعذر حساب الوسط الحسابي وبالتالي الانحراف المعياري، كما يعاب عليه أيضاً تأثره بالقيم المتطرفة.

مثال(3-14): لتكن لدينا مجموعتان من العمال حيث أن الوسط الحسابي للأجر اليومي للمجموعتين والانحراف المعياري لهما معطى كما يلي:

المجموعة الأولى: $\bar{x}_1 = 50$ u.m $\quad\quad \sigma_1 = 10$ u.m

المجموعة الثانية: $\bar{x}_2 = 36$ u.m $\quad\quad \sigma_2 = 9$ u.m

فأي المجموعتين تظهر أكثر تشتتاً؟.

الحل: نحسب معامل الاختلاف لكلا المجموعتين:

$$V_1=C.V_1 = \frac{\sigma(x)}{\bar{x}} . 100 = \frac{10}{50} .100 = 20\ \% = 0.20$$

$$V_2=C.V_2 = \frac{\sigma(x)}{\bar{x}} . 100 = \frac{9}{36} .100 = 25\ \% = 0.25$$

بما أن $_1 \quad < \quad _2$ ⇐ أن عمال المجموعة الثانية أكثر تشتتاً في أجورهم.

مثال(3-15): لتكن لدينا مجموعتين من المعلومات، أولاهما تتعلق بطول فئة من العمال، حيث تظهر البيانات أن الوسط الحسابي لاطوال العمال هو $\bar{x}$ =170 cm بانحرافٍ معياري يقدر بـ $\sigma(x) = 17$ cm ،

بينما تتعلق المجموعة الثانية من المعلومات بأوزان العمال، حيث تظهر البيانات أن الوسط الحسابي للأوزان هو $\bar{x}$ =65kg

بانحرافٍ معياري يقدر بـ $\sigma(x)$ = 5 kg.

هل العمال يختلفون أكثر بالنسبة للطول أم بالنسبة للوزن.

الجواب: للإجابة على ذلك نحسب معامل الاختلاف بالنسبة للكول والوزن.

بالنسبة للطول V_1=C.V_1 = $\frac{\sigma(x)}{\bar{x}}$. 100 =$\frac{17}{170}$ .100 = 10 % = 0.1

بالنسبة للوزن V_2=C.V_2 = $\frac{\sigma(x)}{\bar{x}}$. 100 =$\frac{5}{65}$ .100 = 7.69 % = 0.0769

نلاحظ أن $V_1 > V_2$ ⇐ نستخلص أن مجموعة العمال أكثر اختلافاً من حيث الطول منه من حيث الوزن.

ملاحظات:

1. معامل الاختلاف (الانحراف المعياري النسبي) لا يستعمل إلا في الحالات التي تكون فيها المعطيات منسوبة إلى أصلٍ ثابتٍ، وكلها موجبة مثل الأطوال و الأوزان والإثمان وهو معامل صالح للمقارنة بين المتغيرات النسبية لمجموعتين من البيانات فأكثر، ولا فرق في ذلك إذا كانت البيانات المستعملة مختلفة أو نفسها.
2. يعاب على هذا المقياس(معامل الاختلاف) أنه لا يصلح للاستخدام في حالة التوزيعات التكرارية ألمفتوحة حيث يتعذر حساب الوسط الحسابي $\bar{X}$، والانحراف المعياري $\sigma(x)$ بدقة. وفي هذه الحالة يمكن استخدام المقياس الثاني أي الانحراف الربيعي النسبي أو معامل التغير الربيعي، حيث لا نجد صعوبة في أغلب ألحيان في حساب كل من Q_1 ،Q_3
3. يصبح هذا المقايس عديم الفائدة عند ما تكون $\bar{X}$ قريبة من الصفر.

تمارين للحل

التمرين الأول:

من أجل تقدير وسطي لسسن الزواج الأول بين الفتيات ، أخذت عينة من 230 فتاة في إحدى المدن تزوجن حديثا لأول مرة ، فوجد أن أعمارهن عند الزواج كانت كالتالي:

العمر	15	16	17	18	19	20	21	22	23	24	25	30	35
العدد	2	4	8	13	20	35	40	45	30	20	10	2	1

المطلوب: إيجاد:1- المدى العام 2- المدى الربيعي -3 الانحراف المتوسط 4- الانحراف الوسيطي

5- الانحراف ألمنوالي 6- التباين والانحراف المعياري.

التمرين الثاني:_ ليكن لديك جدول التوزيع التكراري التالي:

الفئات	[30-50 [	[50-70 [	[70-90 [	[90-110 [	[110-130 [
التكرارات	18	33	45	38	26

المطلوب :احسب ما يلي:

1- الوسيط 2- الربيعي الأول والثالث 3- المدى العام 4- المدى الربيعي

5- الانحراف المتوسط والانحراف ألمنوالي 6- احسب الانحراف الهندسي والانحراف التوافقي

– الانحراف المعياري والتباين.

التمرين ألثالث_:ليكن لديك جدول التوزيع التكراري التالي:

الفئات	[10-20 [	[20-30 [	[30-40 [	[40-50 [	[50-60 [	Σ
التكرارات	10	20	35	23	12	100

المطلوب:

1- إيجاد المدى العام.

2- إيجاد المدى الربيعي والمدى العشري .

3- إيجاد الانحراف المتوسط والانحراف التوافقي.

4- إيجاد التباين والانحراف المعياري.

التمرين الرابع:بالعودة للتمارين السابقة ، المطلوب:

1- إيجاد المدى النسبي.

2- إيجاد الانحراف الربيعي النسبي .

3- إيجاد الانحراف المتوسط النسبي.

4- إيجاد الانحراف المعياري النسبي.

التمرين الخامس: يبين الجدول التكراري التالي توزيع 100 طالب حسب علاماتهم في مقياسي الإحصاء والاقتصاد الجزئي على الترتيب:

x_i	2	4	5	7	9	10	11	13	14	15	16
n_{i_1}	1	3	6	8	14	20	25	18	2	2	1
n_{i_2}	2	3	5	7	10	15	20	22	10	5	1

المطلوب:

1- حساب المدى النسبي في الحالتين . ماذا تلاحظ ؟.

2- حساب المدى الربيعي النسبي في الحالتين. ماذا تلاحظ ؟.

3- حساب الانحراف المتوسط النسبي في الحالتين. ماذا تلاحظ ؟.

4- حساب الانحراف المعياري النسبي في الحالتين . ماذا تلاحظ ؟.

5- قارن تشتت العلامات في الحالتين بناء على الأسئلة 1 ،2 ،3،4.

6- حساب معامل الاختلاف في الحالتين. ماذا تلاحظ ؟

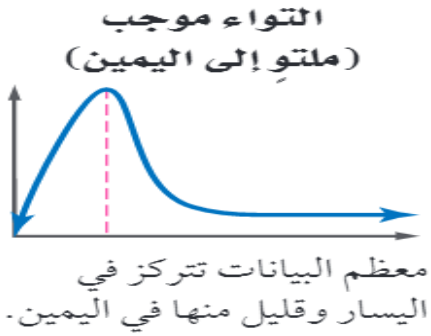
3-2: مقاييس الشكل Form measures

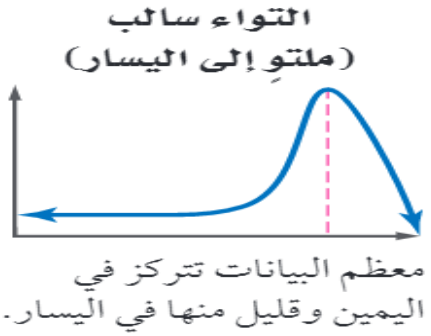
لقد رأينا فيما سبق أن التوزيع التكراري قد يمثل بقيمة مركزية، ثم رأينا أن هذه القيمة المركزية لا تدل دلالة كافية على التوزيع وصفاته، فأرفقنا القيمة المركزية (المتوسطة) بمقياس للتشتت،ولإتمام وصف التوزيع التكراري، يرفق ما سبق بمقاييس أخرى تدعى مقاييس الشكل، وأهمها:

3-2-1: مقاييس الالتواء measures of skewness

يقال عن توزيع تكراري أنه متناظر "متماثل" إذا كانت القيم موزعة بشكلٍ متساوٍ على جانبي القيمة المركزية أي إذا كان عدد القيم التي يقابلها انحراف عن القمة المركزية قدره $(-e)$ يساوي عدد القيم التي يقابله انحراف عن القيمة المركزية قدره $(+e)$. وفي الحالة المعاكسة يقال أن التوزيع غير متناظر "غير متماثل" أي ملتوٍ. وبناءً على هذا نميز نوعين من الالتواء:

أ. **توزيع ذو إلتواء معتدل:** وفيه تتوزع البيانات حول القيمة المركزية بشيئ من الالتواء، ويكون المنحنى التكراري الخاص بها له فرعان، فيكون متصاعداً حتى يصل إلى القمة(المنوال)ثم يبدأ بالتناقص فقد يكون منحنى التوزيع ممتد ناحية اليسار ونقول أنه سالب الإلتواء، وتكون عند ئذٍ القيم الصغيرة تقابلها تكرارات صغيرة والقيم الكبيرة تقابلها تكرارات كبيرة، أو ممتد ناحية اليمين ونقول أنه موجب الألتواء، حيث تكون القيم الصغيرة تقابلها تكرارات كبيرة والقيم الكبيرة تقابها تكرارات صغيرة، كما يوضح الشكلان أدناه.





ب. **توزيع ذو إلتواء شديد:** وفيه تكون البيانات متمركزة في الفئة المتطرفة الدنيا أو العليا ثم يتناقص عددها بالتتابع، ويكون منحنى التوزيع في هذه الحالة له فرع واحد، إما صاعد نحو اليمين عند ما تتمركز البيانات في الفئة العليا، أو نازل نحو اليمين عند ما تتمركز البيانات في الفئة الدنيا أي إما متزايداً من اليمين إلى اليسار أو متناقص من اليمين إلى اليسار كما يوضح الشكل أدماه.

الهدف من دراسة الالتواء:

تهدف دراسة الالتواء إلى تحديد شكل التوزيع من حيث درجة ابتعاده عن التوزيع الطبيعي أو المتماثل أو المتناظر، وهل هو موجب أم سالب. ومنه فقياس الالتواء ينبني على صفات التوزيع المعتدل "المتماثل " بحيث أنه إذا كان:

1. التوزيع التكراري معتدل أي طبيعي أو متناظر أو متماثل فإن: $\bar{X} = M_e = M_o$

مفهوم أساسي

خصائص التوزيع الطبيعي



- التمثيل البياني له منحنى يشبه الجرس، ومتماثل حول المستقيم الرأسي المار بالمتوسط.
- يتساوى فيه المتوسط والوسيط والمنوال.
- المنحنى متصل.
- يقترب المنحنى من المحور x في جزأيه الموجب والسالب، ولكنه لا يمسه.

ويكون كذلك بعد الوسيط عن الربيعي الأول متساوٍ مع بعد الوسيط عن الربيعي الثالث أي

$$(M_e - Q_1) = (Q_3 - M_e)$$

وبما أن $M_e = Q_2$ فإن:

$$(Q_2 - Q_1) = (Q_3 - Q_2)$$

وبشكلٍ مشابهٍ يكون بعد الوسيط عن العشير الأول يساوي بعد الوسيط عن العشير التاسع.

$$(M_e - D_1) = (D_9 - M_e)$$

وبما أن $M_e = D_5$ فإن:

$$(D_5 - Q_1) = (D_9 - D_5)$$

وبشكلٍ مشابهٍ أيضاً يكون بعد الوسيط عن المئين العاشر يساوي بعد الوسيط عن المئين التسعين

$$(M_e - P_{10}) = (P_{90} - M_e)$$

وبما أن $M_e = P_{50}$ فإن:

$$(P_{50} - P_{10}) = (P_{90} - P_{50})$$

2. **التوزيع التكراري غير متماثل "غير معتدل"** فإن: $\bar{X} \neq M_e \neq M_o$ بمعنى أنه كلما إلتوى التوزيع التكراري كلما اختلفت هذه المتوسطات عن بعضها البعض. وفي هذه الحالة(حالة الالتواء) نميز نوعان من الالتواء هما:

أ. **التوزيع ملتوي نحو اليمين:** يكون الوسط الحسابي أكبر من الوسيط الذي يكون بدوره أكبر من المنوال، وهذا يعني أن الطرف الأيمن ممتد أكثر، وبالتالي يكون الالتواء نحو اليمين(موجب الالتواء)، وتميل فيه التكرارات الكبيرة إلى التركز عند فئات التوزيع الدنيا $\bar{X} > M_e > M_o$

ب. **التوزيع ملتوي نحو اليسار** : يكون الوسط الحسابي أصغر من الوسيط الذي يكون بدوره أصغر من المنوال،وهذا يعني أن الطرف الايسر ممتد أكثر، وبالتالي يكون الالتواء نحو اليسار(سالب الالتواء) ، وتميل فيه التكرارات الكبيرة إلى التركز عند فئات التوزيع العليا، $\bar{X} < M_e < M_o$

ومن أهم مقاييس الالتواء نجد:

3-2-1-1:معامل الالتواء لبيرسون:Pearson's coefficient of skewness

إن ظهور الفوارق بين المتوسطات الثلاث السابقة يدل على عدم التماثل (الالتواء)، وتزداد هذه الفوارق تبعاً لزيادة شدة الالتواء، ومن ثم يقترح **Pearson** قياس الالتواء بالفرق بين الوسط الحسابي والمنوال للتوزيع التكراري موضوع الدراسة $(\bar{x} - M_o)$فهذا الفرق هو مؤشر للالتواء على اعتبار تساوي قيمتيهما في التوزيع المعتدل أو المتماثل.لذلك يقترح **Pearson** حساب معامل الالتواء بقسمة الالتواء على الانحراف المعياري وهو ما يعرف بمعامل **Pearson**الأول الذي يعطى بالعلاقة: $P_1 = \frac{\bar{X} - M_o}{\sigma(x)}$ حيث:

P_1 معامل الالتواء الأول لـ **Pearson** . $\bar{X}$ الوسط الحسابي للتوزيع التكراري.

M_o قيمة المنوال للتوزيع التكراري. $\sigma(x)$ الانحراف المعياري للتوزيع التكراري.

حيث أن قيمة P_1 تكون محصورة بين القيمتين (-1) و$(+1)$ أي $-1 \leq P_1 \leq +1$

وهناك أيضاً علاقة تجريبية تجمع بين المتوسطات الثلاث السابقة الذكر بالشكل التالي:

$$(\bar{X} - M_o) = 3(\bar{X} - M_e) \Rightarrow \frac{\bar{X} - M_o}{\sigma(x)} = \frac{3(\bar{X} - M_e)}{\sigma(x)}$$

ونميز بين ثلاثة خالات:

$P_1 = 0 \Leftarrow$ التوزيع متمثل أو متناظرا.

$P_1 > 0 \Leftarrow$ هذا يعني أن التوزيع سالب الألتواء أي ملتوي أو ممتد إلى اليسار.

$P_1 < 0 \Leftarrow$ هذا يعني أن التوزيع موجب الألتواء أي ملتوي أو ممتد إلى اليمين.

3-2-1-2:معامل الالتواء الثاني لـ Pearson (معامل عدم التماثل).

هناك صيغة أخرى لبيرسون يستخرج بها معامل الالتواء والذي يعرف أيضاً بمعامل عدم التماثل لبيرسون، ويعتمد هذا المقياس في تعريفه (تحديده) على العزم الثالث حول الوسط الحسابي(العزم المركزي الثالث)مقسوماً على الانحراف المعياري مكعب.لذلك لا بد من التطرق إلى العزوم.

العزوم:عزم اي قوة هو عبارة عن مقدار العمل الذب تحدثه.يتوقف عذا العزم، أي هذا العمل على القوة نفسها وعلى المسافة بين هذه القوة والنقطة التي تحدث عندا الأثر. فعزم قوة مقدارها 10كلغ على بعد 2 م من نقطة الأصل يعادل عزم قوة مقدارها 5 كلغ على بعد 4 م من هذه النقطة اي ان التوازن يتحقق عند تساوي الناتج الموجب مع الناتج السالب.

بالنسبة للتوزيع التكراري، تكون تكرارات التوزيع هي القوة المؤثرة عليع، وعزم أي تكرار يقاس بحاصل ضرب تكرار كل فئة في انحراف مركزها عن نقطة الأصل في التوزيع التي يمثلها الوسط الحسابي. ويمكن تحديد حساب العزوم كما يلي:

أ. **حالة بيانات غير مبوبة**:لدينا N من البيانات $x_1 \ x_2 \ x_3 \ x_4 \ \dots\dots\dots x_k$ ،وليكن a عدد ثابت، فالوسط الحسابي لإنحرافات هذه القيم عن العدد الثابت من المرتبة k يدعى بالعزم، ويرمز له عادة بالرمز M_k أي العزم من المرتبة k.

$$M_k = \frac{(x_1-a)^k+(x_2-a)^k+(x_3-a)^k+\cdots\dots\dots\dots\dots+(x_k-a)^k}{N} = \frac{\sum_{i=1}^{k}(x_i-a)^k}{N}$$

ب. **حالة بيانات مبوبة في توزيع تكراري بسيط**: فإذا كانت القيم التي يأخذها المتغير الاحصائي هي $x_1 \ x_2 \ x_3 \ x_4 \ \dots\dots\dots x_k$ والتكرارات المقابلة لها هي $n_1 \ n_2 \ n_3 \ n_4 \dots\dots\dots n_k$، وليكن a عدد ثابت، فعزمها من المرتبة k يكتب على الشكل التالي:

$$M_k = \frac{n_1(x_1-a)^k+n_2(x_2-a)^k+n_3(x_3-a)^k\dots\dots\dots\dots+n_k(x_k-a)^k}{N} = \frac{\sum_{i=1}^{k} n_i(x_i-a)^k}{N}$$

ت. **حالة بيانات مبوبة في توزيع تكراري ذي فئات**: إذا كانت لدينا مجموعة من البيانات عددها N مبوبة في فئات عددها k ومراكزها هي $c_1 \ c_2 \ c_3 \ c_4 \dots\dots\dots\dots\dots c_k$والتكرارات المقابلة لها هي $n_1 \ n_2 \ n_3 \ n_4 \dots\dots\dots n_k$ ، وليكن a عدد ثابت، فعزمها من المرتبة k يكتب على الشكل التالي:

$$M_k = \frac{n_1(c_1-a)^k+n_2(c_2-a)^k+n_3(c_3-a)^k\dots\dots\dots\dots+n_k(c_k-a)^k}{N} = \frac{\sum_{i=1}^{k} n_i(c_i-a)^k}{N}$$

وفقاً لقيمة المقدار a نميز ثلاثة أنواع من العزوم هي:

1. إذا كان a = 0 نحصل على مجموعة من العزوم تدعى بالعزوم الابتدائية (العزم k حول نقطة الأصل)، وتعطى بالعلاقة التالية:

** حالة البيانات غير المبوبة

$$M_k = \frac{\sum_{i=1}^{k}(x_i-a)^k}{N} = \frac{\sum_{i=1}^{k} x_i^k}{N}$$

** حالة البيانات في توزيع تكراري بسيط $M_{\mathrm{k}} = \frac{\sum_{i=1}^{k} n_i(x_i - a)^k}{N} = \frac{\sum_{i=1}^{k} n_i x_i^k}{N}$

** حالة البيانات في توزيع تكراري ذي فئات $M_{\mathrm{k}} = \frac{\sum_{i=1}^{k} n_i(c_i - a)^k}{N} = \frac{\sum_{i=1}^{k} n_i c_i^k}{N}$

2. إذا كان $a = \bar{x}$ فنحصل على مجموعة العزوم تدعى بالعزوم المركزية (العزم k حول الوسط الحسابي ونرمز له بالرمز m_k ويعطى بالعلاقة التالية

** حالة البيانات غير المبوبة $\mathrm{m_k} = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^k}{N}$

** حالة البيانات في توزيع تكراري بسيط $m_{\mathrm{k}} = \frac{\sum_{i=1}^{k} n_i(x_i - \bar{x})^k}{N}$

** حالة البيانات في توزيع تكراري ذي فئات $m_{\mathrm{k}} = \frac{\sum_{i=1}^{k} n_i(c_i - \bar{x})^k}{N}$

ومنه فإن $P_2 = \frac{m_3^2}{\sigma^3(\ )}$

حيث P_2هو معامل الالتواء لبيرسون، m_3 هو العزم الثالث حول الوسط الحسابي،$\sigma^3(X)$ مكعب الانحراف المعياري، وباعتبار أن $m_2 = \sigma^2(X) \Rightarrow \sigma(x) = \sqrt{m_2} \Rightarrow P_2 = \frac{m_3^2}{(\sqrt{m_2})^3}$

3-1-2-3:معامل الالتواء الربيعي (معامل الالتواء ليول Yule)

يقترح يول مقياساً أخر للالتواء يقوم على أساس العلاقة بين الربيعي الأول والثالث والوسيط بحيث إذا كان التوزيع معتدلاً فإن المسافة بين الربيعي الاعلى تتساوى مع المسافة بين الوسيط والربيعي الأدنى أي

$$\begin{cases} (\mathrm{Q_3} - \mathrm{M_e}) = (\mathrm{M_e} - \mathrm{Q_1}) \\ \text{أو} \\ (\mathrm{Q_3} - \mathrm{Q_2}) = (\mathrm{Q_2} - \mathrm{Q_1}) \end{cases}$$

وبالتالي إذا طرح الطرفين من بعضهما كان الرق صفراً، ونقول عندئذٍ أن التوزيع معتدل، أي:

$$\begin{cases} (\mathrm{Q_3} - \mathrm{M_e}) - (\mathrm{M_e} - \mathrm{Q_1}) = \mathrm{Q_3} + \mathrm{Q_1} - 2\mathrm{M_e} = 0 \\ \text{أو} \\ (\mathrm{Q_3} - \mathrm{Q_2}) - (\mathrm{Q_2} - \mathrm{Q_1}) = \mathrm{Q_3} + \mathrm{Q_1} - 2\mathrm{Q_2} = 0 \end{cases}$$

اما إذا كان التوزيع غير معتدل "أي ملتوي" فلا تتساوى المسافة بين الربيع الثالث والوسيط مع المسافة بين الوسيط والربيع الأول أي أن

$$\begin{cases}(Q_3 - M_e) \neq (M_e - Q_1) \\ \text{أو} \\ (Q_3 - Q_2) \neq (Q_2 - Q_1)\end{cases}$$

وبالتالي إذا طرح الطرفين من بعضهما كان الفرق يختلف عن الصفر، ونقول عندئذٍ أن التوزيع ملتوي، أي:

$$\begin{cases}(Q_3 - M_e) - (M_e - Q_1) \neq 0 \\ \text{أو} \\ (Q_3 - Q_2) - (Q_2 - Q_1) \neq 0\end{cases}$$

ومن هذه العلاقة الأخيرة يكون:

أ. إذا كان لدينا $(Q_3 - M_e) - (M_e - Q_1) > 0$

يكون التوزيع ملتوي نحو اليمين، حيث تتجه البيانات نحو القيم العليا، لذلك تكبر قيمة الربيع الأعلى فيكون الفرق بينه وبين الوسيط أكبر من الفرق بين الوسيط والربيع الأدمى، فيكون الفرق موجباً.

ب. إذا كان لدينا $(Q_3 - M_e) - (M_e - Q_1) < 0$

يكون التوزيع ملتوي نحو اليسار، حيث تتجه البيانات نحو فئات القيم الدنيا ، لذلك تبعد قيمة الربيع الأدنى عن الوسيط فيكون الفرق بين الوسيط والربيع الأدنى أكبر من الفرق بين الربيع الأعلى والوسيط، فيكون الفرق سالباً

ت. إلتواء التوزيع يكبر كلما كبرت قيمة المعامل، ويصغر كلما صغرت قيمته.

تقوم هذه الطريقة على مقارنة انتشار منحنى التوزيع على يسار الوسيط مقاساً بـ $(M_e - Q_1)$ بالانتشار على يمين الوسيط مقاساً بـ $(Q_3 - M_e)$، وتقسم الفرق بينهما على مجموعهما فنحصل من ذلك هلى معامل الالتواء لـ يول Yule كما يلي

$$\alpha = \frac{(Q_3 - M_e) - (M_e - Q_1)}{(Q_3 - M_e) + (M_e - Q_1)} = \frac{Q_3 + Q_1 - 2M_e}{Q_3 - Q_1} = \alpha$$

$\alpha = 0$ يعبر عن تناظر تام أو تماثل في توزيع البيانات.

$\alpha > 0$ التواء موجب (التواء نحو اليمين).

$\alpha > 0$ إلتواء سالب (إلتواء نحو اليسار)

3-2-1-4:معامل الالتواء لـ يول وكندال Yule و Kendall .

ويأخذ الشكل التالي $C_{yk} = \frac{Q_3+Q_1-2M_e}{Q_3-Q_1}$

بشكلٍ مشابهٍ لاستعمال معامل الاختلاف الربيعي، يمكن استعمال كذلك معامل الاختلاف العشيري وكذل المئيني بالشكل التالي:

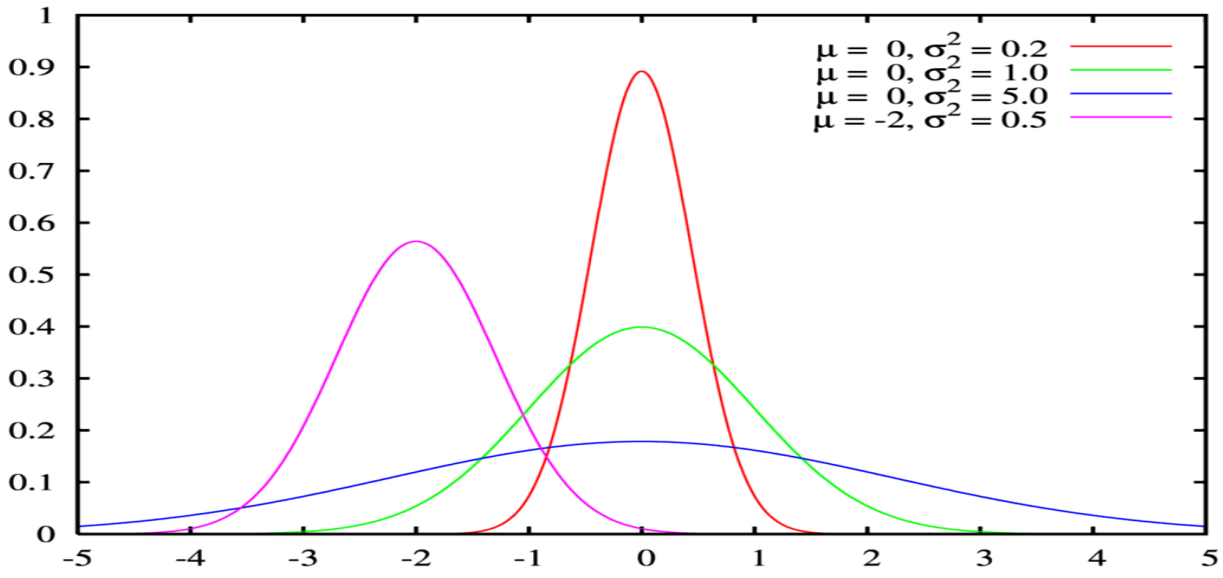
معامل الالتواء العشيري $\alpha=\frac{(D_9-M_e)-(M_e-D_1)}{(D_9-M_e)+(M_e-D_1)} = \frac{D_9+D_1-2M_e}{D_9-D_1} =\alpha$

معامل الالتواء المئيني $\alpha=\frac{(P_{90}-M_e)-(M_e-P_{10})}{(P_{90}-M_e)+(M_e-P_{10})} = \frac{P_{90}+P_{10}-2M_e}{P_{90}-P_{10}} =\alpha$

3-2-2:مقاييس التفرطح Kurtosis measures

المنحنى التكراري المعتدل هو المنمنى المثالي للتوزيعات المنسجمة، المتدرجة في التزايد تدرجاً منتظما أو التي كما ذكرنا سابقاً فيها يتساوى المتوسط الحسابي والوسيط والمنوال، بحيث يكون المنحنى على شكل جرس(ناقوس)، فهو ذا قيمة معتدلة في الارتفاع.لذلك يتخذ كأساس للمقارنة لإظهار علو أو انخفاض قيمة التوزيع المدروسة. ومنا نميز بين ثلاث حالات:

1. أن تكون قمة المنحنى للمثل للتوزيع المدروس أعلى من قمة المنحنى الممثل للتوزيع الطبيعي أو المعتدل، وندعوه عندئذ بالمنحنى المدبب **leptocurtique**، حيث يضيق في الوسط وترتفع قمته عن قمة التوزيع الطبيعي.
2. تكون قمة المنحنى للمثل للتوزيع المدروس منطبقة على قمة المنحنى الممثل للتوزيع الطبيعي أو المعتدل **.mesocurtique**
3. تكون قمة المنحنى الممثل للتوزيع المدروس أدنى من قمة المنحنى الممثل للتوزيع المعتدل أو الطبيعي، ويدعى عندئذٍ بالمنحنى المفرطح، وهو الذي يتسع في الوسط وتنخفض قمته عن قمة المنحنى الطبيعي كما هو موضح في الشكل أدناه



التفرطح يقيس درجة تدبب قمة المنحنى الذي يمثل التوزيع التكراري وذلك بالمقارنة مع التوزيع المعتدل المي لا تكون قمته لا مدببه ولا مفرطحة.

ملاحظة: المنحنى التكراري يمكن تحويله إلى منحنى أكثر تفرطحاً بتكبير وحدات القياس المأخوذة على المحور الأفقي، كما يمكن تحوبل نفس المنحنى إلى منحنى أكثر تدبباً أي ذي قمة حادة بتصغير وحدات القياس على المحور الأفقي.

إذن فالتفرطح هو عبارة عن قياس درجة علو قمة التوزيع بالنسبة للتوزيع الطبيعي عادة. يقاس بعدة مقاييس أهمها:

3-2-2-1:معامل التفرطح لكيلي Kelly (معامل التفرطح المئيني) :

يستعمل هذا المقياس كلاً من الربيعيات والمئينيات، ويسمى لذلك بمعامل التفرطح المئيني، ويكتب على الشكل التالي:

$$K_1=\frac{\frac{1}{2}(Q_3-Q_1)}{(P_{90}-P_{10})}=\frac{1}{2}.\frac{(Q_3-Q_1)}{(P_{90}-P_{10})}\Rightarrow\frac{1}{2}.\frac{R_Q}{R_P}=K_1$$

3-2-2-2:معامل التفرطح لبيرسون "معامل التفرطح العزمي"

يعتمد هذا المقياس في تعريفه على كل من العزم الرابع حول المتوسط الحسابي "العزم المركزي" والانحراف المعياري، أي أن معامل التفرطح هذا هو عبارة عن حاصل قسمة العزم المركزي من المرتبة الرابعة على الانحراف المعياري مرفوعاً للقوة الرابعة أي أن:

$$K_2 = \frac{m_4}{\sigma^4(x)} = \frac{m_2}{(m_2)^2}$$

حيث نميز ثلاث حالات:

$K_2 = 3 \Leftarrow$ التوزيع معتدل.

$K_2 < 3 \Leftarrow$ التوزيع مدبب (قمته أعلى من قمة التوزيع الطبيعي).

$K_2 > 3 \Leftarrow$ التوزيع منبسط أو مفرطح (قمته أدنى من قمة التزيع الطبيعي).

3-2-2-3:معامل التفرطح لفيشر (Fisher).

هذا المقياس يعطى بالعلاقة التالية ويعتمد في حسابه على المقياس السابق لبيرسون

$$K_3 = \frac{m_4}{\sigma^4(x)} - 3 = \frac{m_2}{(m_2)^2} - 3$$

حيث نميز ثلاث حالات:

$K_3 = 0$ أي $K_3 = 3 \Leftarrow$ التوزيع معتدل.

$K_3 < 0$ أي $K_3 < 3 \Leftarrow$ التوزيع مدبب (قمته أعلى من قمة التوزيع الطبيعي).

$K_3 > 3$ أي $K_3 > 3 \Leftarrow$ التوزيع منبسط أو مفرطح (قمته أدنى من قمة التزيع الطبيعي).

ملاحظات عامة حول معاملات الإلتواء: يلاحظ على معاملات الإلتواء السابقة مايلي:

1. تتوقف إشارة معامل الالتواء على إشارة البسط، وذلك لأن المقام دائماً موجب.
2. قيمة معامل الالتواء هي قيمة مجردة من وحدات القياس، لذلك تصلح عذه المعاملات لمقارنة توزيع ظواهر تختلف وحدات قياسها فمثلاً يمكن مقارنة توزيع الأطوال بتوزيع الأوزان بتوزيع الأعمار.....
3. قيم المعاملات الثلاث يمكن أن تكون مختلفة.

مثال(3-16):أحسب مقياس الالتواء الثاني لبيرسون(P_2) ومقياس التفرطح للتوزيع التكراري التالي:

$n_i(x_i - \bar{x})^4$	$(x_i - \bar{x})^4$	$n_i(x_i - \bar{x})^3$	$(x_i - \bar{x})^3$	$n_i(x_i - \bar{x})^2$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})$	$n_i x_i$	n_i	x_i

6480	1296	-1080	-216	180	36	-6	25	5	05
486	81	-162	-27	54	9	-3	48	6	08
0	0	0	0	0	0	0	66	6	11
324	81	108	27	36	9	+3	56	4	14
3888	1296	648	216	108	36	+6	51	3	17
13122	6561	1458	729	162	81	+9	40	2	20
24300	-----	972	----	540	-----	---	286	26	

$$\bar{x} = \frac{\sum_{i=1}^{k} n_i x_i}{N} = \frac{286}{26} = 11 = \bar{x}$$

$$v(x) = \frac{\sum_{i=1}^{k}(X_I - \bar{X})^2}{N} = \frac{540}{26} = 20.77 = \sigma^2(x) \Rightarrow \sigma(x) = \sqrt{(x)\ v} = \sqrt{20.77} = 4.56 = \sigma(x)$$

$$m_3 = \frac{\sum_{i=1}^{k} n_i (x_i - \bar{x})^3}{N} = \frac{792}{26} = 37.38$$

$$m_4 = \frac{\sum_{i=1}^{k} n_i (x_i - \bar{x})^4}{N} = \frac{24300}{26} = 934.62 = m_4$$

4. ومنه فمعامل الالتواء الثاني لـ Pearson يكون:

$$P_2 = \frac{m_3^2}{\sigma^3(x)} = \frac{37.38}{(4.56)^3} = \frac{37.38}{94.652168} = 0.395 = P_2$$

يتضح من ذلك أن التوزيع قليل الالتواء (إلتواء موجب معتدل نحو اليمين لأن $0 < P_2$.

حساب مقياس التفرطح (معامل التفرطح لـ Pearson) :

$$K_2 = \frac{m_4}{\sigma^4(x)} = \frac{m_2}{(m_2)^2} \Rightarrow K_2 = \frac{934.62}{(20.77)^2} = \frac{934.62}{431.36091} = 2.166 \cong 2$$

يتضح من ذلك أن التوزيع مفرطح نوعاً ما وتكون قمة منحناه البياني أدنى من قمة التوزيع الطبيعي

تمارين للحل

التمرين الاول :

يبين التوزيع التكراري التالي توزيع المداخيل النسبية لمجتمع x(المداخيل بـ 1000د ج)

الفئات	...-10[	10-20[	20-30[	30-40[	40-50[	50-60[	60-70[	70-80[	80-90[	90-100[	Σ
%	17.2	11.7	12.1	14.8	15.9	11.9	05	04	3.7	3.6	100

1- هل يمكن حساب الانحراف المعياري لهذا التوزيع ؟ ولماذا ؟.

2- ماهو مقياس التشتت المناسب لهذا التوزيع ؟ احسبه .

3- حدد المجال $[P_{20} - p_{80}]$ ثم حدد نسبة البيانات التي يحتويها هذا المجال.

4-حدد حدود ونسبة المجال $[x \pm 0.67\delta]$ وقارنها بالنسبة النظرية التي يحتويها المجال المشار اليه.

التمرين الثاني

يبين الجدول التكراري التالي توزيع 100 طالب حسب علاماتهم في مقياسين مختلفين.

العلامات	1	4	5	7	9	10	11	12	14	16	Σ
المقياس1	3	3	6	8	14	20	25	18	2	1	100
المقياس2	4	3	5	7	10	16	20	22	10	3	100

1- احسب معامل الاختلاف في الحالتين . ماذا تلاحظ ؟.

2- احسب معامل الاختلاف الربيعي في الحالتين.

3- قارن تشتت العلامات في الحالتين بناءا على السؤالين 1 و2.

4- قار شكل التوزيعين باستعمال معامل Pearson ثم باستعمال معامل Fisher.

التمرين الثالث

لتكن لديك البيانات التالية:

28	27	26	25	4	28	15	19	13	33	29	16	28	19	34
11	22	22	17	15	10	23	21	14	10	11	24	18	11	20
						17	12	30	16	18	21	23	18	12

المطلوب:

1- ضع البيانات السابقة في جدول تكراري ذي 6 فئات متساوية مع تحديد مختلف التكرارات.

2- احسب المتوسط الحسابي والانحراف المعياري من البيانات الأولية ثم من جدول التوزيع التكراري.ماذا تلاحظ

3- احسب الانحراف المتوسط للبيانات ثم للتوزيع التكراري.

4- ما نسبة البيانات الموجودة ضمن انحراف معياري واحد من الوسط الحسابي أي المجال $[\bar{x} \pm \delta]$

والموجودة ضمن انحرافين من الوسط الحسابي أي التي يتضمنها المجال $[X \pm 2\delta]$.

5- إيجاد العزم الأول والثاني والثالث والرابع حول المتوسط الحسابي للتوزيع التكراري.

6- إيجاد العزم الأول والثاني والثالث والرابع حول الصفر للتوزيع التكراري.

7- حساب معامل الالتواء ثم معامل لتفرطح.

الفصل الرابع: الأرقام القياسية

تمهيد:

يرجع تاريخ استخدام الأرقام القياسية إلى أكثر من قرنين من الزمن، حيث استخدمها الإحصائي الإيطالي كارلي (1764) لمقارنة الأسعار في إيطاليا لسنة 1750 بالأسعار في سنة 1500 . ثم شاع استخدامها بصورة أوسع منذ ذلك الحين، حيث اهتمت الحكومات بتركيب وحساب بعض الأرقام القياسية. ومن الأمور الهامة عند تركيب الرقم القياسي اختيار فترة الأساس أو مكان الأساس التي تعتمد لتركيب الرقم. وعادة يفضل أن تكون فترة الأساس سابقة لفترة المقارنة وتتميز متميزة باستقرار أكثر للظاهرة محل الدراسة

4-1:مفهوم الرقم القياسي:

الرقم القياسي هو أداة لقياس التغير النسبي الحاصل في قيم الظاهرة محل الدراسة من وضعية مكانية أو زمانية تسمى وضعية الأساس إلى وضعية مكانية أو زمانية أخرى تسمي وضعية المقارنة، حيث أن قيمة الظاهرة في وضعية الاساس تعتبر مساوية لـ 100 ، لذلك نميز بين:

الرقم القياسي الزمني: يعطى الرقم القياسي الزمني لأي ظاهرة بالعلاقة التالية:

$$\text{الرقم القياسي للظاهرة} = \frac{\text{قيمة الظاهرة في سنة المقارنة}}{\text{قيمة الظاهرة في سنة الأساس}} \times 100$$

الرقم القياسي المكاني: وهو معطى بالعلاقة التالية:

$$\text{الرقم القياسي للظاهرة} = \frac{\text{قيمة الظاهرة في مكان المقارنة}}{\text{قيمة الظاهرة في مكان الأساس}} \times 100$$

وتستخدم الأرقام القياسية في دراسة تطور الظواهر الاقتصادية والاجتماعية كالإنتاج، الاستهلاك، الصادرات...إلخ، غير أنها أكثر استخداما في دراسة تطور الأسعار والنفقات الاستهلاكية، ولا يقتصر استخدامها كأساس لمقارنة التغير في ظاهرة ما زمنيا أو مكانيا بل يمكن استخدامها للمقارنة بين ظاهرتين مختلفتين أو أكثر.

4-2:أنواع الأرقام القياسية:

يمكن التمييز بين عدة أنواع من الأرقام القياسية، ومن أهمها:

4-2-1:الأرقام القياسية البسيطة:

يقيس الرقم القياسي البسيط تطور ظاهرة معينة بين فترتين زمنيتين مختلفتين أو مكانين مختلفين، وهو عبارة عن النسبة بين قيمة المتغير في فترة المقارنة t_1 إلى قيمة نفس المتغير في فترة الأساس t_0.

4-2-1-1:الرقم القياسي البسيط للأسعار:

هو عبارة عن نسبة السعر في زمن أو مكان المقارنة إلى السعر في زمن أو مكان الأساس والناتج مضروب في 100%، ويعبر عن هذا الرقم بالصيغة التالية:

الرقم القياسي البسيط للأسعار= $\frac{\text{السعر في سنة المقارنة}}{\text{السعر في سنة الأساس}} . 100$ $I_{P_{(t_2/t_1)}} = \frac{P_2}{P_1} .100$

حيث 2 هو السعر في زمن أو مكان المقارنة و P_1 هو السعر في زمن أو مكان المقارنة الأساس.

4-2-1-2:الرقم القياسي البسيط للكميات:

هو عبارة عن نسبة الكمية في زمن أو مكان المقارنة إلى الكمية في زمن أو مكان الأساس والناتج مضروب في 100%، ويعبر عن هذا الرقم بالصيغة التالية:

الرقم القياسي البسيط للكميات = $\frac{\text{الكمية في سنة المقارنة}}{\text{الكمية في سنة الأساس}} . 100$ $I_{Q_{(t_2/t_1)}} = \frac{Q_2}{Q_1} .100$

حيث 2 الكمية في زمن أو مكان المقارنة و Q_1 هي الكمية في زمن أو مكان في سنة الأساس.

4-2-1-3:الرقم القياسي البسيط للقيمة:

إن القيمة تمثل حاصل ضرب سعر السلعة **P** في الكمية المنتجة أو المباعة **Q**، فالصيغة المستخدمة لحساب الرقم القياسي للقيمة **V** هي:

$$I_{V_{(t_2/t_1)}} = \frac{V_2}{V_1} .100 = \frac{Q_2P_2}{Q_1P_1} .100$$

مثال(4-1)

الجدول التالي يمثل تطور أسعار السلع A B C خلال فترتين مختلفتين، المطلوب هو حساب الارقام القياسية لتغير الأسعار والقيمة خلال الفترتين 1 و2.

السلع	فترة أو مكان الاساس			فترة أو مكان المقارنة		
	الكمية	السعر	القيمة	الكمية	السعر	القيمة
A	15	24	360	20	32	640
B	12	20	240	25	24	600
C	18	40	720	45	50	2250
$\sum$	45	84	1320	90	106	3490

من الجدول نستنتج نلاحظ أن:

1. كمية السلعة A قد تغيرت ارتفاعاً بنسبة 33% أي أن كمية مكان أو زمان المقارنة تمثل 133% من كمية زمن أو مكان الأساس حيث أن مقدار التغير يكون 133-100 =33 %
2. كمية السلعة B قد تغيرت ارتفاعاً بنسبة 108 % أي أن كمية مكان أو زمان المقارنة تمثل 208% من كمية زمن أو مكان الأساس حيث أن مقدار التغير يكون 208-100 =108 %
3. كمية السلعة C قد تغيرت ارتفاعاً بنسبة 150% أي أن كمية مكان أو زمان المقارنة تمثل 250% من كمية زمن أو مكان الأساس حيث أن مقدار التغير يكون 250-100 =150 %
4. سعر السلعة A قد تغير ارتفاعاً بنسبة 33% أي أن سعر مكان أو زمان المقارنة يمثل 133% من سعر زمن أو مكان الأساس حيث أن مقدار التغير يكون 133-100 =33 %
5. سعر السلعة B قد تغير ارتفاعاً بنسبة 20% أي أن سعر مكان أو زمان المقارنة يمثل 120% من سعر زمن أو مكان الأساس حيث أن مقدار التغير يكون 120-100 =20 %

6. سعر السلعة C قد تغير ارتفاعاً بنسبة 25% أي أن سعر مكان أو زمان المقارنة يمثل 125% من سعر زمن أو مكان الأساس حيث أن مقدار التغير يكون 125-100 =25 %
7. قيمة الانفاق على السلعة A قد تغيرت ارتفاعاً بنسبة 77.7% أي أن قيمة الانفاق على السعلة A في مكان أو زمان المقارنة تمثل 177.7% من قيمة الانفاق على نفس السلعة في زمن أو مكان الأساس حيث أن مقدار التغير يكون 177.7-100 =25
8. قيمة الانفاق على السلعة B قد تغيرت ارتفاعاً بنسبة 150% أي أن قيمة الانفاق على السعلة B في مكان أو زمان المقارنة تمثل 250% من قيمة الانفاق على نفس السلعة في زمن أو مكان الأساس حيث أن مقدار التغير يكون 250-100 =150
9. قيمة الانفاق على السلعة C قد تغيرت ارتفاعاً بنسبة 212% أي أن قيمة الانفاق على السعلة C في مكان أو زمان المقارنة تمثل 312% من قيمة الانفاق على نفس السلعة في زمن أو مكان الأساس حيث أن مقدار التغير يكون 312-100 =212

4-2-2:الأرقام القياسية التجميعية:

يقيس هذا الرقم النسبة بين مجموع قيم الظاهرة محل الدراسة في سنة أو مكان المقارنة إلى قيم نفس الظاهرة في سنة أو مكان الأساس مضروبة في 100، ويحسب الرقم القياسي التجميعي حسب الصيغ أدناه:

4-2-2-1:لرقم القياسي التجميعي للأسعار:

يقيس تغير أسعار مجموعة سلع مقاسة بأسعارها الحقيقية لفترتين وقد تكون السلع متجانسة أو غير متجانسة، ويحسب بموجب الصيغة التالية: $I_{P_{(t_2/t_1)}} = \frac{\sum P_{2\,i}}{\sum P_{1\,i}} .100$

4-2-2-2:الرقم القياسي التجميعي للكميات: يحسب وفق الصيغة التالية $I_{Q_{(t_2/t_1)}} = \frac{\sum Q_{2\,i}}{\sum Q_{1\,i}} .100$

4-2-2-3: الرقم القياسي التجميعي للقيم:يحسب وفق الصيغة:$I_{V_{(t_2/t_1)}} = \frac{\sum V_{2\,i}}{\sum V_{1\,i}} .100 = \frac{\sum Q_{2\,i}P_{2\,i}}{\sum Q_{1\,i}P_{1\,i}} .100$

مثال بالعودة للجدول **المثال(4–1)** اعلاه فإن:

الرقم القياسي التجميعي للأسعار:

$$I_{P_{(t_2/t_1)}} = \frac{\sum P_{2\,i}}{\sum P_{1\,i}} .100 \Rightarrow I_{P_{(t_2/t_1)}} = \frac{32+24+50}{24+20+40} * 100 = \frac{106}{84} * 100 = 126.2\%$$

أي متوسط اسعار سنة أو مكان المقارنة أصبح يساوي 126.2% من متوسط أسعار سنة أو مكان الأساس،

حيث أن مقدار متوسط التغيرات في أسعار المواد هو $126.2-100 = 26.2\%$

الرقم القياسي التجميعي للكميات:

$$I_{Q_{(t_2/t_1)}} = \frac{\sum Q_{2\,i}}{\sum Q_{1\,i}} .100 \Rightarrow I_{P_{(t_2/t_1)}} = \frac{20+25+40}{15+12+18} * 100 = \frac{90}{45} * 100 = 200\%$$

أي متوسط كميات سنة أو مكان المقارنة أصبح يساوي 200 % من متوسط كميات سنة أو مكان الأساس،

حيث أن مقدار متوسط التغيرات في كميات المواد هو $200-100 = 100\%$

الرقم القياسي التجميعي للقيم:

$$I_{V_{(t_2/t_1)}} = \frac{\sum V_{2\,i}}{\sum V_{1\,i}} 100 = \frac{\sum Q_{2\,i} P_{2\,i}}{\sum Q_{1\,i} P_{1\,i}} .100 \Rightarrow I_{V_{(t_2/t_1)}} = \frac{600+640+2250}{360+240+720} * 100$$

$$= \frac{3490}{1320} * 100 = 264.4\%$$

4–3:الأرقام القياسية المرجحة:

إن الرقم القياسي التجميعي للأسعار والكميات لا يأخذ بعين الاعتبار أوزان كل سلعة وأهميتها النسبية، لذلك جاءت الصيغ المرجحة للأرقام القياسية للتغلب على هذه العيوب حيث تعتمد الأرقام القياسية المرجحة على ترجيح أسعار أو كميات كل سلعة باستخدام معامل معيّن، وتستخدم عادة كمية السلعة المباعة أو سعرها خلال سنة الأساس أو خلال سنة المقارنة، وهذه الأوزان تشير إلى الأهمية النسبية للسلعة.

توجد عدة أنواع من الأرقام القياسية المرجحة، ولكن سوف نهتم بأكثرها استخداما:

4-3-1:الرقم القياسي المرجح للاسبير:

الرقم القياسي المرجح لـ **Laspeyres** يتم فيه الترجيح بأوزان سنة الأساس، حيث يتم ترجيح الأسعار بكميات سنة الأساس بالنسبة للرقم القياسي للأسعار أو يتم ترجيح الكميات بأسعار سنة الأساس بالنسبة للرقم القياسي للكميات، ويأخذ هذا الرقم الصيغتين التاليتين:

صيغة الرقم القياسي المرجح للأسعار لـ Laspeyres هي: $\mathbf{IL}P_{t_2/t_1} = \frac{\sum Q_{1i} P_{2i}}{\sum Q_{1i} P_{1i}} .100$

صيغة الرقم القياسي المرجح للكميات لـ Laspeyres هي $\mathbf{IL}Q_{t_2/t_1} = \frac{\sum Q_{2i} P_{1i}}{\sum Q_{1i} P_{1i}} .100$

مثال (4-2): الجدول التالي يمثل نمط الاستهلاك لعائلة خلال فترتين مختلفتين 1 و 2

الحل:

السلع	الفترة الأولى		الفترة الثانية					
	الكمية Q_{1i}	السعر P_{1i}	الكمية Q_{2i}	السعر P_{2i}	$Q_{1i} P_{1i}$	$Q_{2i} P_{2i}$	$Q_{1i} P_{2i}$	$Q_{2i} P_{1i}$
A	4	4	5	5	16	25	20	20
B	2	7	3	9	14	27	18	21
C	1	8	1.5	10	8	15	10	12
∑					38	67	48	53

الرقم القياسي المرجح للكميات لـ Laspeyres:

$$\mathrm{IL}Q_{t_2/t_1} = \frac{\sum Q_{1i} P_{2i}}{\sum Q_{1i} P_{1i}} .100 \Rightarrow \mathrm{IL}P_{2/1} = \frac{4*5+7*3+8*1.5}{4*4+7*2+8*1} * 100 = \frac{53}{38} * 100 = 139.47\%$$

الرقم القياسي المرجح للأسعار لـ Laspeyres:

$$\mathrm{IL}P_{t_2/t_1} = \frac{\sum Q_{2i} P_{1i}}{\sum Q_{1i} P_{1i}} .100 \Rightarrow \mathrm{IL}Q_{2/1} = \frac{5*4+9*2+10*1}{4*4+7*2+8*1} * 100 = \frac{48}{38} * 100 = 126.31\%$$

4-3-2: الرقم القياسي المرجح لـ Paasche:

يعتمد **Paasche** في حساب رقمه القياسي على الترجيح بأوزان سنة المقارنة، حيث يتم ترجيح الأسعار بكميات سنة المقارنة بالنسبة للرقم القياسي للأسعار أو يتم ترجيح الكميات بأسعار سنة المقارنة بالنسبة للرقم القياسي للكميات، ويأخذ هذا الرقم الصيغتين التاليتين:

صيغة الرقم القياسي المرجح للأسعار لـ Paasche هي: $\mathbf{IP}P_{t_2/t_1} = \frac{\sum Q_{2i} P_{2i}}{\sum Q_{2i} P_{1i}} .100$

صيغة الرقم القياسي المرجح للكميات لـ Paasche هي: $\mathbf{IP}Q_{t_2/t_1} = \frac{\sum Q_{2i} P_{2i}}{\sum Q_{1i} P_{2i}} .100$

مثال: بالعودة إلى االمثال(4-2) أعلاه فإن

الرقم القياسي المرجح للأسعار لـ Paasche هي:

$$IPP_{t_2/t_1} = \frac{\sum Q_{2i} P_{2i}}{\sum Q_{2i} P_{1i}} .100 \Rightarrow IPP_{t_2/t_1} = \frac{5*5+9*3+10*1.5}{5*4+7*3+8*1.5} * 100 = \frac{67}{53} * 100 = 126.41\%$$

الرقم القياسي المرجح للكميات لـ Paasche هي:

$$IPQ_{t_2/t_1} = \frac{\sum Q_{2i} P_{2i}}{\sum Q_{1i} P_{2i}} .100 \Rightarrow IPQ_{t_2/t_1} = \frac{5*5+9*3+10*1.5}{5*4+2*9+1*10} * 100 = \frac{67}{48} * 100 = 139.58\%$$

4-3-3:الرقم القياسي المرجح لFischer

هو من أفضل الأرقام القياسية المرجحة، يستخدم للتخلص من المشكلة التي يعاني منها كل من رقم **Laspeyres** ورقم **Paasche** (التخفيض والتضخيم)، ويقوم على حساب الوسط الهندسي للرقمين القياسيين التجميعيين المرجحين ل **Laspeyres** و **Paasche**، وتكتب علاقته على الشكل التالي:

$$IF_{2/1} = \sqrt{I_L \ I_P}.100$$

يمكن حساب رقم **Fischer** الأمثل كما يلي:

صيغة الرقم القياسي المرجح للأسعار لFischer هي:

$$IF_{P_{2/1}} = \sqrt{\frac{\sum Q_{1i} P_{2i}}{\sum Q_{1i} P_{1i}} * \frac{\sum Q_{2i} P_{2i}}{\sum Q_{2i} P_{1i}}}.100$$

صيغة الرقم القياسي المرجح للكميات لFischer هي:

$$IF_{Q_{2/1}} = \sqrt{\frac{\sum Q_{2i} P_{1i}}{\sum Q_{1i} P_{1i}} * \frac{\sum Q_{2i} P_{2i}}{\sum Q_{1i} P_{2i}}}.100$$

مثال نأخذ معطيات المثال (4-2)والمطلوب حساب الرقم القياسي المرجح لFischer ؟

الحل:

الرقم القياسي المرجح للكميات لـFischer : توصلنا سابقا من خلال حساب رقمي **Laspeyres** و**Paasche** للكميات إلى أنّ:

$IL_{Q_{t_2/t_1}}$ =139.47 $IP_{Q_{t_2/t_1}}$=139.58 وعليه فإن:

$$IF_{Q_{2/1}} = \sqrt{(139.47)(139.58)}.100 = 139.52\%$$

الرقم القياسي المرجح للاسعار لـFischer : كانت نتائج حساب رقمي **Laspeyres** و**Paasche** للأسعار كما يلي:

$IL_{P_{t_2/t_1}}$=126.31% **و** $IP_{P_{t_2/t_1}}$ 126.41% وعليه نجد:

$$IF_{P_{2/1}} = \sqrt{(126.31)(126.41)}.100 = 126.35\%$$

4-4: خصائص الأرقام القياسية:

حتى يكون الرقم القياسي أكثر موضوعية يجب أن يكون قابلاً لتحقيق بعض الخصائص الرياضية المهمة والتي من بينها:

4-4-1:خاصية المطابقة:

الرقم القياسي لفترة بالنسبة للفترة نفسها يجب أن يكون مساوياً لـ1 أو 100%.

$$I_{(t/t)} = \frac{G_t}{G_t} = 1 \text{ ou } 100\%$$

4-4-2: **خاصية الانعكاس:**

وتعني أن مقلوب رقم قياسي هو رقم قياسي أيضاً، ويكون بذلك أن حاصل جداء رقم قياسي في معكوسة يكون مساويا لرقم سنة الأساس فإذا كان الرقم القياسي لظاهرة معينة بين الفترتين t_1 وt_2 هو

$$I_{\left(t_2/t_1\right)}=\frac{G_{t_2}}{G_{t_1}}100\Rightarrow\frac{1}{I_{\left(t_2/t_1\right)}}100=\frac{1}{\frac{G_{t_2}}{G_{t_1}}}100=\frac{G_{t_1}}{G_{t_2}}100=I_{\left(t_1/t_2\right)}\Rightarrow$$

$$\left(I_{\left(t_2/t_1\right)}\right)\left(I_{\left(t_1/t_2\right)}\right)=\left(\frac{G_{t_2}}{G_{t_1}}\right)\left(\frac{G_{t_1}}{G_{t_2}}\right)=(100)^2$$

والرقم القياسي الذي يحقق هذه الخاصية يعتبر من أفضل الأرقام القياسية.

4-4-3:خاصية الدوران:

تعرف هذه الخاصية أيضاً باسم خاصية التحويل أو التعدي، وهي مهمة لكونها تسمح بالتحكم في بعض الاساسية،فإذا كانت لدينا ظاهرة تأخذ القيم $G_1\ G_2\ G_3\ G_4 \dots\dots\dots G_n$ خلال الفترات 1،2،3،4...n فإن المؤشر البسيط خلال الفترة كلها يجب أن يحقق العلاقة التالية

$$I_{\left(t_n/t_1\right)}=\frac{\left(I_{\left(t_2/t_1\right)}\right).\left(I_{\left(t_3/t_2\right)}\right).\left(I_{\left(t_4/t_3\right)}\right).\left(I_{\left(t_5/t_4\right)}\right).....\left(I_{\left(t_{n-1}/t_{n-2}\right)}\right).\left(I_{\left(t_n/t_{n-1}\right)}\right)}{100^{n-1}}$$

حيث: n عدد الأرقام القياسية في البسط.

مثال (3-3):

لتكن لدينا الأرقام القياسية التالية والتي تبيّن التغير في أسعار مادة لخمس فترات:

$$I_{\left(t_2/t_1\right)}=107\ \%\ ;\ I_{\left(t_3/t_2\right)}=111\ \%\ ;\ I_{\left(t_4/t_3\right)}=108\ \%\ ;\ I_{\left(t_5/t_4\right)}=115\ \%\ ;\ I_{\left(t_6/t_5\right)}=117\ \%$$

والمطلوب تحديد الرقم القياسي لأسعار هذه المادة بين فترة الأساس والفترة الخامسة ؟

الحل:

من أجل الحصول على الرقم القياسي لهذه المادة نقوم في هذه الحالة باستخدام خاصية التحويل:

$$I_{\left(t_6/t_1\right)}=\frac{\left(I_{\left(t_2/t_1\right)}\right).\left(I_{\left(t_3/t_2\right)}\right).\left(I_{\left(t_4/t_3\right)}\right).\left(I_{\left(t_5/t_4\right)}\right)\left(I_{\left(t_6/t_5\right)}\right).}{100^{5-1}}=\frac{(107)(111)(108)(115)(117)}{100^4}=172.59\ \%$$

نلاحظ أن أسعار هذه المادة في الفترة الخامسة تمثل 172.59 % من أسعارها في الفترة الأساسية أي أن مقدار التغير خلال الفترة المعتبرة هو 72.59= 100- 172.59 أي أن أسعار فترة المقارنة أعلى من أسعار فترة الأساس بـ 72.59 %.

تمرين

من أجل حساب الرقم القياسي المرجح لفيشر لابد من حساب الرقم القياسي المرجح لكل من لاسبير وباش.

السلعة	الأسعار		الكميات				
	2000	2014	2000	2014	P_0Q_0	P_0Q_1	P_1Q_0	P_1Q_1
الأرز	40	90	16	20	640	800	1440	1800
العدس	120	220	10	12	1200	1440	2200	2640
الحمص	180	300	8	9	1440	1620	2400	2700
الفاصولياء	120	250	12	14	1440	1680	3000	3500
المجموع	460	860	46	55	4720	5540	9040	10640

الجدول يوضح كيفية حساب الأرقام القياسية

حساب الرقم القياسي المرجح للأسعار لفيشر

$$IF_{P_{2/1}} = \sqrt{\frac{\sum Q_{1i} P_{2i}}{\sum Q_{1i} P_{1i}} * \frac{\sum Q_{2i} P_{2i}}{\sum Q_{2i} P_{1i}}}.100$$

الرقم القياسي المرجح للأسعار للاسبير:

$$\mathrm{IL}P_{t_2/t_1} = \frac{\sum Q_{2i} P_{1i}}{\sum Q_{1i} P_{1i}}.100 \Rightarrow = \frac{9040}{4720} \times 100 = 191{,}52\%$$

الرقم القياسي المرجح للأسعار لباش:

$$\text{IP}_{t_2/t_1} = \frac{\sum Q_{2i} P_{2i}}{\sum Q_{2i} P_{1i}} .100 = \frac{10640}{5540} \times 100 = 192,06\%$$

ومنه نجد أنّ:

$$IF_{P_{2/1}} = \sqrt{\frac{\sum Q_{1i} P_{2i}}{\sum Q_{1i} P_{1i}} * \frac{\sum Q_{2i} P_{2i}}{\sum Q_{2i} P_{1i}}}.100 = \sqrt{1,9152 \times 1,9206} \times 100 = 191,79\%$$

وعليه نقول أن أسعار السلع قد ارتفعت بنسبة $91,79\%$ في سنة 2014 مقارنة بسنة الأساس 2000.

حساب الرقم القياسي المرجح للكميات لفيشر: $IF_{Q_{2/1}} = \sqrt{\frac{\sum Q_{2i} P_{1i}}{\sum Q_{1i} P_{1i}} * \frac{\sum Q_{2i} P_{2i}}{\sum Q_{1i} P_{2i}}}.100$

الرقم القياسي المرجح للكميات للاسبير:

$$\text{IL}_{Q_{t_2/t_1}} = \frac{\sum Q_{1i} P_{2i}}{\sum Q_{1i} P_{1i}} .100 = \frac{5540}{4720} \times 100 = 117,37\%$$

الرقم القياسي المرجح للكميات لباش:

$$\text{IP}_{Q_{t_2/t_1}} = \frac{\sum Q_{2i} P_{2i}}{\sum Q_{1i} P_{2i}} .100 = \frac{10640}{9040} \times 100 = 117,7\%$$

ومنه فإن:

$$IF_{Q_{2/1}} = \sqrt{\frac{\sum Q_{2i} P_{1i}}{\sum Q_{1i} P_{1i}} * \frac{\sum Q_{2i} P_{2i}}{\sum Q_{1i} P_{2i}}}.100 = \sqrt{1,1737 \times 1,177} \times 100 = 117,53\%$$

تمارين مقترحة:

التمرين الاول: يبين الجدول التالي تطور أسعار وكميات ثلاثة سلع خلال فترتين مختلفتين 1 و2.

الفترات	السلع	السعر (د ج /كلغ)	الكمية (كلغ)
الفترة1	A	24	15
	B	20	12
	C	40	18
الفترة 2	A	32	20

	B	24	25
	C	50	45

احسب الرقم القياسي للأسعار والكميات لكل من :Laspeyres ، paasche و Fischer .

التمرين الثاني:

الجدول أدناه يمثل أسعار مجموعة من السلع التي يستخدمها الطلبة في دراستهم في العامين 2010 و 2014، والمطلوب حساب الرقم القياسي التجميعي لهذه السلع باعتبار سنة الأساس سنة 2010.

السلعة	أسعار سنة 2010	أسعار سنة 2014
أقلام	15	25
دفاتر	100	160
آلات حاسبة	80	140
أوراق	250	300

التمرين الثالث: يبين الجدول التالي تطور أسعار وكميات ثلاثة سلع خلال سنتي 1992 و 1996

الفترات	السلع	السعر (د ج /كلغ)	الكمية (كلغ)
1992	A	10	15
	B	12	20
	C	15	08
1996	A	15	16
	B	14	21
	C	20	10

المطلوب: باعتبار سنة 1992هي سنة الأساس، باعتبار سنة 1996 سنة الاساس، احسب:

1- الرقم القياسي التجميعي للأسعار.

2- الرقم القياسي التجميعي للكميات.

3- الرقم القياسي التجميعي للقيم.

4- احسب الرقم القياسي للأسعار والكميات لكل من :Laspeyres ، paasche و Fischer

التمرين الثالث: الجدول التالي يبيّن أسعار وكميات السيارات المباعة لدى وكالة بيع السيارات في الفترة 2010-2014.

نوع السيارة	سعر السيارة 10^3 دج		عدد السيارات المباعة	
	2010	2014	2010	2014
النوع A	650	880	30	30
النوع B	880	950	40	52
النوع C	1050	1200	50	60

المطلوب: باعتبار سنة 2010 هي سنة الأساس، أحسب:

1- الرقم القياسي التجميعي للأسعار والكميات.

2- الرقم القياسي المرجح للأسعار للاسبير.

3- الرقم القياسي المرجح للأسعار لباش.

4- الرقم القياسي المرجح للأسعار لفيشر.

الفصل الخامس:الارتباط والانحدار الخطي

مقـدمة

في الفصول الثلاث السابقة تم عرض بعض المقاييس الوصفية، مثل مقاييس النزعة المركزية، والتشتت، ومقاييس الالتواء والتفرطح، وغيرها من المقاييس الأخرى والتي يمكن من خلالها وصف شكل توزيع البيانات التي تم جمعها عن متغير واحد، وننتقل من التعامل مع متغير واحد إلى التعامل مع متغيرين أو أكثر، ويتناول هذا الفصل دراسة وتحليل العلاقة بين متغيرين، وذلك باستخدام بعض طرق التحليل الإحصائي مثل تحليل الارتباط، والانحدار الخطي البسيط، فإذا كان اهتمام الباحث هو دراسة العلاقة بين متغيرين استخدم لذلك أسلوب تحليل الارتباط، وإذا كان اهتمامه بدراسة أثر أحد المتغيرين على الآخر استخدم لذلك أسلوب تحليل الانحدار، ومن الأمثلة على ذلك:

1. الإنفاق، والدخل العائلي.
2. سعر السلعة، والكمية المطلوبة منها.
3. الفترة الزمنية لتخزين الخبز، وعمق طراوة الخبز.
4. تقديرات الطلاب في مقرر الإحصاء، وتقديراتهم في مقرر الرياضيات.
5. كميات السماد المستخدمة، وكمية الإنتاج من محصول معين تم تسميده بهذا النوع من السماد.
6. عدد مرات ممارسة نوع معين من الرياضة البدنية، ومستوى الكلسترول في الدم.
7. وزن الجسم، وضغط الدم.

الارتباط هو علاقة بين متغيرين يمثل كل متغير ظاهرة معينة فإن تغيرت إحدى الظاهرتين في اتجاه معين فالثانية تتغير في اتجاه الأولى أو في اتجاه معاكس للأولى.

والتغير للظاهرتين في نفس الاتجاه بمعنى الزيادة في الأولى يقابله زيادة في الثانية أو العكس نقص في الأولى يقابله نقص في الثانية فالعلاقة تكون طردية أو متزايدة (موجبة) وإن كان الزيادة في الأولى يقابله نقص في الثانية أو العكس النقص في الظاهرة الأولى يقابله زيادة في الثانية فنقول أن الارتباط عكسي أو متناقص (سالباً).

الارتباط يكون تاماً بين المتغيرين إذا عرفنا قيمة أحد المتغيرين نعرف قيمة المتغير الآخر تماماً.

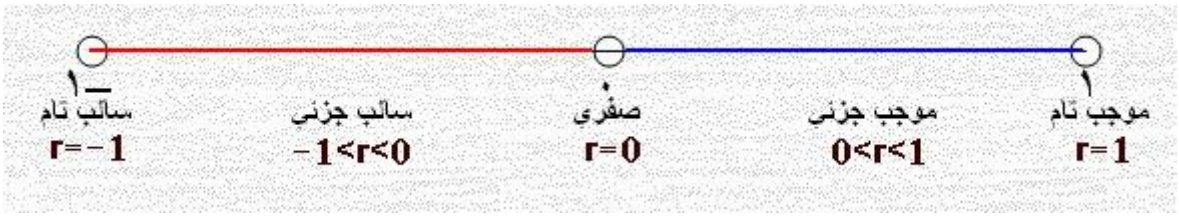
تقاس قوة الارتباط بين متغيرين بمعامل الارتباط ولاستقصاء تأثير أحد المتغيرين على الآخر فنوجد علاقة جبرية تعرف بمعادلة الانحدار (سندرسها لاحقاً) وهي معادلة خط مستقيم (علاقة خطية) بين المتغيرين x ، y مثلاً بالشكل $y = a x + b$ حيث a , b ثابتان يعبر a عن الاتجاه (الميل) و b الجزء المقطوع من محور الصادات (-1) وإن كانت a موجبة كانت العلاقة طردية وإن كانت سالبة فإنَّ العلاقة عكسية والارتباط هنا تاماً لأن النقاط (x ، y) يمر بها خط مستقيم واحد فمثلاً العلاقة $1 - y = 2 x$ تحدد قيم y بمعرفة قيم x وبتمثيلها بيانياً .

يقاس الارتباط بين متغيرين بمقياس كمي يعرف بمعامل الارتباط Coefficient Correlation

معامل الارتباط قد يساوي الصفر (منعدم) بين المتغيرين لعدم وجود علاقة بينهم فمعرفة قيمة أحدهم لا يعني معرفة قيمة الآخر فالمتغيرين الطول ودرجة الامتحان مثلاً وإن تمثيلها بيانياً يظهر مجموعة من النقاط المبعثرة.

إن معرفة العلاقة بين المتغيرات من حيث النوع والقوة والاتجاه يعتبر هدفاً مهماً من أهداف البحث العلمي.

تكون درجة العلاقة قوية عندما تكون قيمة معامل الارتباط ± 1 وتكون العلاقة هنا تامة في حين تضعف درجة العلاقة كلما اقتربت القيمة من الصفر، ونوضح قيمة معامل الارتباط بالشكل الآتي:

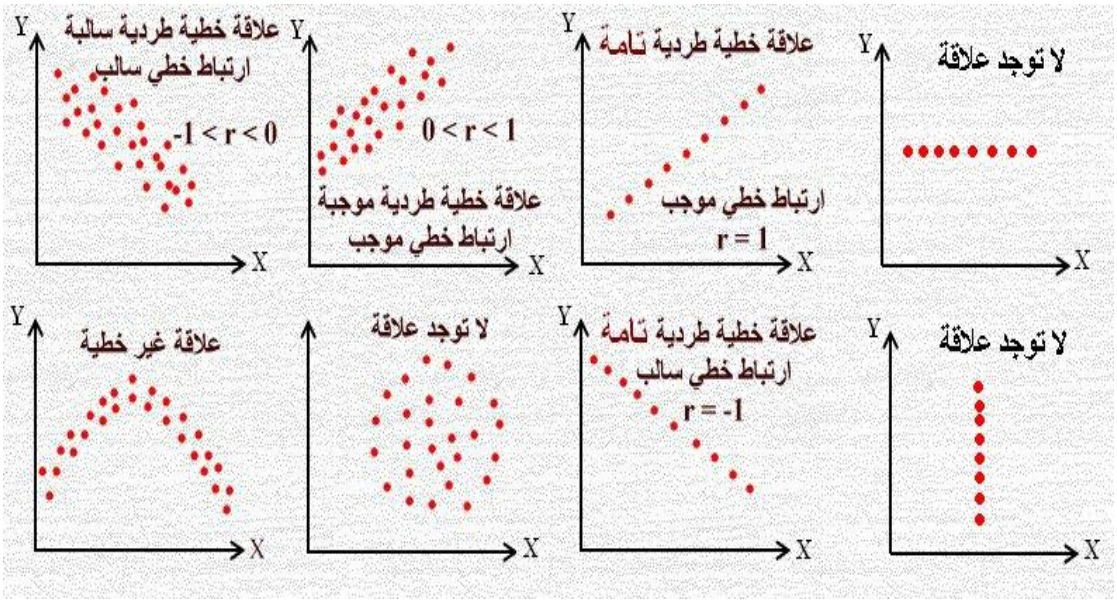


الشكل ألانتشاري

إذا أخذنا (x ، y) كقيم متناظرة لمتغيرين وقمنا بتمثيلها في مستوى الإحداثيات وحصلنا على الأشكال الواردة أسفله، حيث لكل قيمة للمتغير x توجد قيمة تقابلها للمتغير y وإن الأزواج المرتبة (x , y) تكون مجتمع ذو بعدين ويعرف الزوج المرتب بمتغير عشوائي ذو بعدين وللمجتمع ذو البعدين نطرح سؤالاً عل توجد علاقة بين المتغيرين؟ وإن وجدت فكيف نعبر عنها بمعادلة؟ فوجود المعادلة يعني معرفة أحد المتغيرين من معرفة الآخر فالمتغير الأول يعرف بالمتغير المستقل في حين الآخر يعرف بالمتغير التابع، الشكل المرفق هنا يعرف بلوحة الانتشار وكل نقطة هنا تمثل زوج مرتب بالصورة (x ، y).

والأمثلة على ذلك في المجال التطبيقي كثيرة، فإذا كان لدينا المتغيرين $(x\ ;\ )$ ، وتم جمع بيانات عن أزواج قيم هذين المتغيرين، وتم تمثيلها بيانيا فيما يسمى بشكل الانتشار، فإن العلاقة بينها تأخذ أشكالا مختلفة على النحو التالي :

شكل الانتشار لبيان نوع العلاقة بين x , y



5-1:الارتباط الخطى البسيط Simple Correlation

إذا كان الغرض من التحليل هو تحديد نوع وقوة العلاقة بين متغيرين ، يستخدم تحليل الارتباط ، وأما إذا كان الغرض هو دراسة وتحليل أثر أحد المتغيرين على الآخر ، يستخدم تحليل الانحدار، وفي هذا الفصل يتم عرض أسلوب تحليل الارتباط الخطي البسيط، أي في حالة افتراض أن العلاقة بين المتغيرين تأخذ الشكل الخطي ، وسوف يجرى حسابه في حالة البيانات الكمية ، والبيانات الوصفية المقاسة بمعيار ترتيبي.

5-1-1:الغرض من تحليل الارتباط الخطى البسيط

الغرض من تحليل الارتباط الخطي البسيط هو تحديد نوع وقوة العلاقة بين متغيرين، ويرمز له بالرمز r ، ويركز على نقطتين هما:

5-1-2:نوع العلاقة:. وتأخذ ثلاث أنواع حسب إشارة معامل الارتباط كما يلي:

1. إذا كانت إشارة معامل الارتباط سالبة ($r < 0$) توجد علاقة عكسية بين المتغيرين، بمعنى أن زيادة أحد

المتغيرين يصاحبه انخفاض في المتغير الثاني، والعكس.

2. إذا كانت إشارة معامل الارتباط موجبة ($r > 0$)توجد علاقة طردية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحبه زيادة في المتغير الثاني، والعكس .
3. إذا كان معامل الارتباط قيمته صفرا ($r = 0$) دل ذلك على انعدام العلاقة بين المتغيرين.

3-1-5:**قوة العلاقة**:قوة العلاقة تعتمد على قيمة r ق=فإذا كانت $r = 0$ يكون الارتباط منعدم

إذا كانت $0 < r < 0.3$ يكون الارتباط ضعيف جداً.

إذا إذا كانت $0.3 < r < 0.5$ يكون الارتباط ضعيف.

كان إذا كانت $0.5 < r < 0.7$ يكون الارتباط متوسط.

إذا كانت $0.7 < r < 0.9$ يكون الارتباط قوي.

إذا إذا كانت $0.9 < r < 1$ يكون الارتباط قوي جداً.

إذا كانت $r= 1$ يكون الارتباط تام

2-5:**معامل الارتباط الخطى البسيط " لبيرسون"** Pearson

يمكن قياس الارتباط بين متغيرين كميين ، (x ، y)، باستخدام طريقة "بيرسون" Pearson. ومن الأمثلة على ذلك: قياس العلاقة بين الوزن والطول، والعلاقة بن الدرجة التي حصل عليها الطالب وعدد ساعات المراجعة، وهكذا الأمثلة على ذلك كثيرة.

ولحساب معامل الارتباط في العينة ، تستخدم صيغة " بيرسون" التالية :

$$r_p = \frac{\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{N}}{\sqrt{\frac{\sum(x_i-\bar{x})^2}{N}}\sqrt{\frac{\sum(y_i-\bar{y})^2}{N}}} = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2\sum(y_i-\bar{y})^2}}$$

ويمكن اختصار الصيغة السابقة على النحو التالي:

$$r_p = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2}\sqrt{\sum(y_i-\bar{y})^2}} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{(\sum x_i^2 - n\bar{x}^2)(\sum y_i^2 - n\bar{y}^2)}}$$

مثال (5-1)

فيما يلي المساحة المنزرعة بالأعلاف الخضراء بالألف هكتار، وإجمالي إنتاج اللحوم بالألف طن، خلال الفترة من 1995حتى عام 2002.

السنة	1995	1996	1997	1998	1999	2000	2001	2002
المساحة	305	313	297	289	233	214	240	217
الكمية	592	603	662	607	635	699	719	747

والمطلوب: حساب معامل الارتباط بين المساحة والكمية، وما هو مدلوله ؟

الحل

بفرض أن (X) هي المساحة المنزرعة، (Y) هي الكمية، ولحساب معامل الارتباط بين(X ,Y) يتم تطبيق المعادلة السابقة وذلك على النحو التالي:

حساب الوسط الحسابي لكل من المساحة، والكمية($\bar{x}$; $\bar{y}$).

$$\bar{x} = \frac{\sum x_i}{N} = \frac{2108}{8} = 263.5 \qquad ; \bar{y} = \frac{\sum y_i}{N} = \frac{5264}{8} = 658$$

حساب المجاميع

X	Y	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
305	592	41.5	1722.25	-66	4356	-2739
313	603	49.5	2450.25	-55	3025	-2722.5
297	662	33.5	1122.25	4	16	134
289	607	25.5	650.25	-51	2601	-1300.5
233	635	-30.5	930.25	-23	529	701.5
214	699	-49.5	2450.25	41	1681	-2029.5
240	719	-23.5	552.25	61	3721	-1433.5
217	747	-46.5	2162.25	89	7921	-4138.5
2108	5264	0	12040	0	23850	-13528

$\sum(x_i - \bar{x})^2 = 12040$, $\sum(y_i - \bar{y})^2 = 23850$; $\sum (x_i - \bar{x})(y_i - \bar{y}) = -13528$

ومنه فقيمة معامل الارتباط هي:

$$r_p = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2}\sqrt{\sum(y_i-\bar{y})^2}} = \frac{-13528}{\sqrt{12040}\sqrt{23850}} = \frac{-13528}{(109.727)(154.434)} = \mathbf{-0.798}$$

يوجد ارتباط عكسي قوي بين المساحة المنزرعة، وكمية إنتاج اللحوم أي كلما زادت المساحة المزروعة بالأعلاف الخضراء كلما قل انتاج اللحوم

ويمكن تبسيط الصيغة السابقة لتصبح على الشكل

$$r_p = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x_i)^2}{n}\right)\left(\sum y^2 - \frac{(\sum y_i)^2}{n}\right)}} = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{(n\sum x^2 - (\sum x)^2)(n\sum y^2 - (\sum y)^2)}}$$

حيث:

$\sum x_i y_i$ مجموع حاصل ضرب x_i في y_i

$\sum x_i$ مجموع قيم المتغير x_i

$\sum y_i$ مجموع قيم المتغير y_i

$\sum x_i^2$ مجموع مربعات قيم المتغير x_i

$\sum y_i^2$ مجموع مربعات قيم المتغير y_i

وبالتطبيق على بيانات المثال السابق ، يتبع الآتي :

حساب المجاميع:

x	y	xy	x^2	y
305	592	180560	93025	350464
313	603	188739	97969	363609
297	662	196614	88209	438244
289	607	175423	83521	368449
233	635	147955	54289	403225
214	699	149586	45796	488601
240	719	172560	57600	516961
217	747	162099	47089	558009
2108	5264	1373536	567498	3487562

المجاميع المطلوبة
$\sum x_i$ =2108 ;$\sum y_i$ =5264 $\sum x_i y_i$=1373536 $\sum x^2$ =567498 $\sum y^2$= 3487562

حساب معامل الارتباط:

باستخدام المجاميع السابقة، وبالتطبيق على المعادلة أعلاه، نجد أن معامل الارتباط قيمته هي:

$$r_p = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x_i)^2}{n}\right)\left(\sum y^2 - \frac{(\sum y_i)^2}{n}\right)}} = \frac{1373536 - \frac{(2108)(5264)}{8}}{\sqrt{\left[567498 - \frac{(2108)^2}{8}\right]\left[3487562 - \frac{(5264)^2}{8}\right]}}$$

$$= \frac{-13528}{\sqrt{12040}\sqrt{23850}} = \frac{-13528}{(109.727)(154.434)} = -0.798$$

وهي نفس النتيجة السابقة

مثال(5-2):

البيانات التالية تمثل أطوال وأوزان عدد من الطلبة ، حيث الأطوال بعشرات السنتمترات والأوزان بعشرات الكيلوغرامات

الطول	12	13	14	15	16	17
الوزن	6	7	8	9	10	11

هل هناك علاقة بين الطول والوزن؟

الحل:

لدينا بيانات لظاهرتين كميتين، لذلك نستخدم معامل الارتباط لـ Pearson فنسمي الطول x والوزن y .

X	Y	X.Y	X^2	Y^2
12	6	72	144	36
13	7	91	169	49
14	8	112	196	64
15	9	135	225	81
16	10	160	256	100
17	11	187	289	121
87	51	757	1279	451

نحسب معامل Pearson

$$r_p = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{(n\sum x^2 - (\sum x)^2)(n\sum y^2 - (\sum y)^2)}} = \frac{6(757) - (87)(51)}{\sqrt{(6(1279) - (87)^2)(6(451) - (51)^2)}} = \frac{105}{105} = 1$$

هناك ارتباط تام بين وزن طوللا الطالب ووزنه

5-2-1: معامل ارتباط الرتب لسبيرمان Spearman rank correlation coefficient

إذا كانت الظاهرة محل الدراسة تحتوي على متغيرين وصفيين ترتيبين، ومثال على ذلك قياس

العلاقة بين تقديرات الطلبة في مادتين ، أو العلاقة بين درجة تفضيل المستهلك لسلعة معينة ، ومستوى الدخل، فإنه يمكن استخدام طريقة "بيرسون" السابقة في حساب معامل ارتباط يعتمد على رتب مستويات المتغيرين كبديل للقيم الأصلية ، ويطلق على هذا المعامل " معامل ارتباط سبيرمان " **Spearman** ، الذي تم الوصول إلى صيغته اعتماداً على قوانين رياضية تطبق على الأعداد الطبيعية التي تعتبر في هذه الحالة عبارة عن ترتيب للبيانات، إذا أننا نعلم رياضياً ان:

1) $: \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i = \frac{n(n+1)}{2}$

2) $: \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i^2 = \frac{n(n+1)(2n+1)}{6}$

3) $\bar{x} = \bar{y} = \frac{n+1}{2}$

4) $: \sum(x_i - \bar{x})^2 = \sum(y_i - \bar{y})^2 = \sum x_i^2 - \left(\frac{\sum x_i}{N}\right)^2 = \sum y_i^2 - \left(\frac{\sum y_i}{n}\right)^2$

$$= \frac{n(n+1)(2n+1)}{6} - \frac{\left(\frac{n(n+1)}{2}\right)^2}{n} = \frac{n(n+1)(2n+1)}{6} - \frac{n^2(n+)^2}{4n} = \frac{n(n+1)(2n+1)}{6} - \frac{n(n+)^2}{4}$$

$$= n(n+1)\left[\frac{2n+1}{6} - \frac{n+1}{4}\right] = n(n+1)\left(\frac{(2n+)n - 3(n+1)}{12}\right) = n(n+1)\left[\frac{2n+2-3n-3}{12}\right]$$

$$= n(n+1)\left(\frac{n-1}{12}\right) = \frac{n(n^2-1)}{12} = \sum x_i^2 - \left(\frac{\sum x_i}{N}\right)^2 = \sum y_i^2 - \left(\frac{\sum y_i}{n}\right)^2$$

$$d_i = x_i - y_i \Rightarrow d_i^2 = (x_i - y_i)^2 = x_i^2 + y_i^2 - 2x_i y_i \Rightarrow \sum d_i^2 = \sum(x_i - y_i)^2$$

$$= \sum\left(x_i^2 + y_i^2 - 2x_i y_i\right) = \sum x_i^2 + \sum y_i^2 - 2\sum x_i y_i \Rightarrow d_i^2 = 2\sum x_i^2 - 2\sum x_i y_i \Rightarrow$$

$$\sum x_i y_i = \sum x_i^2 - \frac{\sum d_i^2}{2}$$

ونعلم سابقاً أن:

$$r_p = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{\left(\sum x_i^2 - n\bar{x}^2\right)\left(\sum y_i^2 - n\bar{y}^2\right)}} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\left(\sum x_i^2 - n\bar{x}^2\right)}$$

بتعويض رتب x_i و y_i نجد:

$$r_p = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{(\sum x_i^2 - n\bar{x}^2)(\sum y_i^2 - n\bar{y}^2)}} = \frac{\frac{n(n+1)(2n+1)}{6} - \frac{\sum d_i^2}{2} - n\left(\frac{n+1}{2}\right)^2}{\frac{n(n^2-1)}{12}}$$

بالقسمة على n نجد:

$$r_p = \frac{\frac{(n+1)(2n+1)}{6} - \frac{\sum d_i^2}{2n} - \left(\frac{n+1}{2}\right)^2}{\frac{(n^2-1)}{12}} = \frac{(n+1)\left[\frac{2n+1}{6} - \frac{n+1}{4}\right] - \frac{\sum d_i^2}{2n}}{\frac{(n^2-1)}{12}}$$

$$= \frac{(n+1)\left[\frac{4n+2-3n-3}{12}\right] - \frac{\sum d_i^2}{2n}}{\frac{(n^2-1)}{12}} = \frac{(n+)\left(\frac{n-1}{12}\right)}{\frac{(n^2-1)}{12}} - \frac{\sum d_i^2}{2n} \Rightarrow r_p = \frac{\frac{(n^2-1)}{12} - \frac{\sum d_i^2}{2n}}{\frac{(n^2-1)}{12}}$$

$$= 1 - \frac{\sum d_i^2}{2n} * \frac{12}{n^2-1} = 1 - \frac{6\sum d_i^2}{n(n^2-1)} = r_p$$

ومنه فصيغة معامل **Spearman** تكون كالتالي:

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

حيث d هي الفرق بين رتب مستويات المتغير الأول x ،ورتب مستويات المتغير الثاني y أي أن $d = R_x - R_y$

وتتراوح قيمته بين -1 و $+1$.

مثال (3-5) فيما يلي تقديرات 10 طلاب في مادتي الإحصاء، والاقتصاد:

تقديرات إحصاء	أ	جـ+	د	د+	ب+	جـ+	أ+	ب	ب+	ب+
تقديرات اقتصاد	أ+	د	جـ	جـ	أ	ب	ب+	ب	جـ	ب

والمطلوب:

1. احسب معامل الارتباط بين تقديرات الطلبة في المقررين.
2. وما هو مدلوله ؟

الحـل

بفرض أن X هي تقديرات الرياضيات،Y هي تقديرات الإحصاء يمكن حساب معامل الارتباط بينهما باستخدام المعادلة وذلك بإتباع الآتي:

الرتب	1	2	3	4	5	6	7	8	9	10
تقدير احصاء x	أ+	أ	ب+	ب+	ب+	ب	جـ+	جـ+	د+	د

رتب x	1	2	4=(3 + 4 + 5)/3			6	7.5=(7 + 8)/2	9	10
تقدير احصاء y	+أ	أ	+ب	ب	ب	ب	جـ	جـ	جـ	د
رتب y	1	2	3	5=(4 + 5 + 6)/3			8 =(7 + 8 + 9)/3		10

ومنه يمكن حساب المجموع $\sum d_i^2$ كما يلي

x	y	رتبة x	رتبة y	d	d^2
أ	أ+	2	1	1	1
جـ+	د	7.5	10	-2.5	6.25
د	جـ	10	8	2	4
د+	جـ	9	8	1	1
ب+	أ	4	2	2	1
جـ+	ب	7.5	5	2.5	6.25
أ+	ب+	1	3	-2	4
ب	ب	6	5	1	1
ب+	جـ	4	8	-4	16
ب+	ب	4	5	-1	1
					44.5

$$\sum d_i^2 = 44.5$$

- **معامل الارتباط هو:**

$$r = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

$$r = 1 - \frac{6(44.5)}{10(10^2-1)} = 1 - \frac{267}{990}$$

$$r = 1 - 0.2697 = 0.7303$$

5–2–2:مدلول معامل الارتباط :

بما أن $r = 0.703$ ، ويدل ذلك على وجود ارتباط طردي قوي بين تقديرات الطالب في مادة الرياضيات ، ومادة الاحصاء .

ملاحظة: يمكن استخدام صيغة معامل ارتباط "سبيرمان" في حساب الارتباط بين متغيرين كميين، حيث يتم استخدام رتب القيم التي يأخذها المتغير، ونترك للطالب القيام بحساب معامل ارتباط الرتب بين المساحة والكمية في مثال السابق، وعليه أن يقوم بتفسير النتيجة: (معاونة : $\sum d^2 = 148$)

مثال (5–4):عند تقييم مجموعة من الناقدين الرياضيين لعشرة لا عبين تبعاً لدرجة التحمل التدريبي قبل المسابقة وترتيب هؤلاء اللاعبين بعد المسابقة تحصلنا على الجدول التالي

اللاعب	A	B	C	D	E	F	G	H	I	J
رتبة التحمل التدريبي	5	9	10	2	8	7	4	1	6	3
رتبة اللاعب النهائية	4	8	10	2	9	6	3	1	7	5

المطلوب حساب معامل الارتباط لدراسة العلاقة بين التحمل التدريبي والترتيب النهائي:

الحل

اللاعب	A	B	C	D	E	F	G	H	I	J
رتبة التحمل التدريبي R_x	5	9	10	2	8	7	4	1	6	3
رتبة اللاعب النهائية R_y	4	8	10	2	9	6	3	1	7	5
$d = R_x - R_y$	+1	+1	0	0	−1	+1	+1	0	−1	−2
d^2	1	1	0	0	1	1	1	0	1	4

$\sum d^2 = 10$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} \Rightarrow r_s = 1 - \frac{6(10)}{10(99)} = 1 - 0.06 = 0.94$$

معامل الارتباط $r_s = 0.94$ أي ان هناك ارتباط طردي قوي جداً بين التحمل التدريبي والنتيجة النهائية، فكلما كان المتدرب أكثر تحملاً كلما كان تدريبه متقدم ونتائجه أعلى وأحسن.

مثال(5–5):لتكن لدينا التقدير التالية التي حصل عليها 10 طلبة في مادتي التسويق وإدارة المشتريات

التسويق	**ض جداً**	**جيد**	**مقبول**	**ممتاز**	**جيد**	**جيد**	**جيد**	**ج جداً**	**جيد**	**ض جداً**
إدارة المشتريات	**ض جدا**	**مقبول**	**جيد**	**ممتاز**	**مقبول**	**ج جدا**	**مقبول**	**ج جدا**	**ج جدا**	**ضعيف**

المطلوب: حساب معامل الارتباط بين تقديرات المادتين.

الحل:

الدرجة التقديرية في التسويق	الدرجة التقديرية في إدارة المشتريات	رتب التسويق	رتب إدارة المشتريات	رتب التسويق بعد التعديل x_i	رتب إدارة المشتريات بعد التعديل **y**	**d =x-y**	d^2
ض جداً	**ض جدا**	**9**	**10**	**9.5**	**10**	**0.5–**	**0.25**
جيد	**مقبول**	**3**	**6**	**5**	**7**	**2–**	**4**
مقبول	**جيد**	**8**	**5**	**8**	**5**	**3**	**9**
ممتاز	**ممتاز**	**1**	**1**	**1**	**1**	**0**	**0**
جيد	**مقبول**	**4**	**7**	**5**	**7**	**2–**	**4**
جيد	**جيد جداً**	**5**	**2**	**5**	**3**	**2**	**4**
جيد	**مقبول**	**6**	**8**	**5**	**7**	**2–**	**4**
جيد جداً	**جيد جداً**	**2**	**3**	**2**	**3**	**1–**	**1**
جيد	**جيد جداً**	**7**	**4**	**5**	**3**	**2**	**4**
ض جداً	**ضعيف**	**10**	**9**	**9.5**	**9**	**0.5**	**0.25**

من أجل تعديل الترتيب يتوجب حساب متوسط رتب مختلف الملاحظات حسب ورودها في الجدول

متوسط ضعيف جداً $= \frac{10+9}{2} = 9.5$

متوسط جيد $= \frac{7+6+5+4+3}{5} = 5$

متوسط مقبول $= \frac{8+7+6}{3} = 7$

متوسط دجيد جداً $= \frac{4+3+2}{3} = 3$

حسب علاقة **Spearman**

$$r = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6*30.5}{10(10^2-1)} = 1 - \frac{183}{990} = 0.81$$

نلاحظ أن العلاقة بين الدرجات التقديرية للمادتين قوية

5-3:الانحدار الخطى البسيط Simple Regression

إن الغرض من استخدام أسلوب تحليل الانحدار الخطي البسيط، هو دراسة وتحليل أثر متغير كمي على متغير كمي آخر، ومن الأمثلة على ذلك دراسة أثر الإنتاج على التكلفة وأثر الدخل على الإنفاق الاستهلاكي. وهناك أمثلة في كثير من المجالات.

5-3-1:نموذج الانحدار الخطي

في تحليل الانحدار البسيط، نهتم بدراسة أثر أحد المتغيرين ويسمى بالمتغير المستقل أو المتنبأ منه، على المتغير الثاني ويسمى بالمتغير التابع أو المتنبأ به، ومن ثم يمكن عرض نموذج الانحدار الخطي في شكل معادلة خطية من الدرجة الأولى، تعكس المتغير التابع كدالة في المتغير المستقل كما يلي:

$$y_i = \beta_o + \beta_1 X_I + e$$

حيث:

y هو المتغير التابع أي الذي يتأثر.

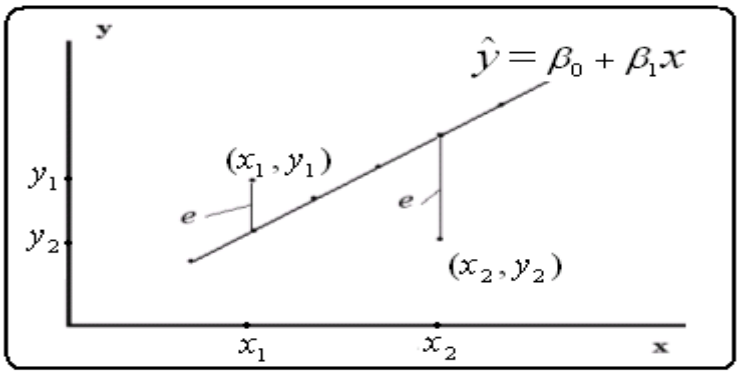
X هو المتغير المستقل الذي يؤثر.

β_o هو النقطة التي يتقاطع فيها الخط المستقيم مع المحور الرأسي، ويعكس قيمة المتغير التابع في حالة انعدام قيمة المتغير النمستقل x أي في حالة $x=0$.

β_1 ميل الخط المستقيم $(\beta_o+\beta_1 X)$ ويعكس مقدار التغير في y إذا تغيرت x بوحدة واحدة.

e هو الخطأ العشوائي، الذي يعبر عن الفرق بين القيمة الفعلية y ، والقيمة المقدرة $\hat{y} = \widehat{\beta_o} + \widehat{\beta_1} X_I$ اي أن $e = y_i - \hat{y}$ يجب أن يكون اقرب ما يمكن للصفر حتى يكون التقدير سليم أي أن مجموع البواقي يؤول

إلى الصفر ويمكن توضيح هذا الخطأ على الشكل التالي لنقط الانتشار



5-3-2:تقدير نموذج الانحدار الخطي البسيط

يمكن تقدير معاملات الانحدار β_1 β_o في النموذج أعلاه باستخدام طريقة المربعات الصغرى، وهذا التقدير هو الذي يجعل مجموع مربعات الأخطاء العشوائية $\sum e^2 = \sum(y_i - (\widehat{\beta_0} + \widehat{\beta_1} X_I))^2$ أقل ما يمكن، أي أن مجموع البواقي يؤول إلى الصفر

$\sum_{i=1}^{n} e = \sum_{i=1}^{n}(y_i - \hat{y})^2 \to 0$(أدنى قيمة)

وبتعويض $\hat{y}$ بقيمتها السابقة نجد

$\sum_{i=1}^{n} e = \sum_{i=1}^{n}(y_i - \widehat{\beta_0} - \widehat{\beta_1} X_I)^2 \to$ عند أدنى قيمة

وحتى يتحقق هذا الشرط يجب أن يتحقق الشرطان التاليان:

1. المشتق الأول بالنسبة للمعالم يكون معدوماً.
2. والمشتق الثاني بالنسبة للمعالم يكون أكبر من الصفر.

إذا وضعنا $L = \sum_{i=1}^{n}(y_i - \widehat{\beta_0} - \widehat{\beta_1} X_I)^2$

بالاشتقاق لـ L بالنسبة لـ β_oنجد

$$\frac{\partial k}{\partial \beta_0} = -2\sum_{i=1}^{n}(y_i - \widehat{\beta_0} - \widehat{\beta_1} X_I) = 0 \Rightarrow \sum_{i=1}^{n}(y_i - \widehat{\beta_0} - \widehat{\beta_1} X_I) = 0 \Rightarrow$$
$$\sum_{i=1}^{n} y_i - N\widehat{\beta_0} - \widehat{\beta_1}\sum_{i=1}^{n} X_i = 0$$

بالقسمة على N نجد:

$$\bar{y} - \widehat{\beta_0} - \beta_1 \bar{X} = 0 \Rightarrow \widehat{\beta_0} = \bar{y} - \beta_1 \bar{X}$$

باشتقاق L بالنسبة لـ β_1 نجد:

$$\frac{\partial k}{\partial \beta_1} = -2\sum_{i=1}^{n}(y_i - \widehat{\beta_0} - \widehat{\beta_1} X_I) X_I = 0 \Rightarrow \sum_{i=1}^{n}(y_i - \widehat{\beta_0} - \widehat{\beta_1} X_I) X_I = 0$$
$$\Rightarrow \sum_{i=1}^{n} y_i X_i - \widehat{\beta_0}\sum_{i=1}^{n} X_i - \widehat{\beta_1}\sum_{i=1}^{n} X_i^2 = 0 \Rightarrow \sum_{i=1}^{n} y_i X_i = \widehat{\beta_0}\sum_{i=1}^{n} X_i + \widehat{\beta_1}\sum_{i=1}^{n} X_i^2$$

بتعويض $\widehat{\beta_o}$ بما تسويه نجد:

$$\sum_{i=1}^{n} y_i X_i = \left(\bar{y} - \widehat{\beta_1}\bar{X}\right)\sum_{i=1}^{n} X_i + \widehat{\beta_1}\sum_{i=1}^{n} X_i^2$$

$$\sum_{i=1}^{n} y_i X_i = \bar{y}\sum_{i=1}^{n} X_i - \widehat{\beta_1}\bar{X}\sum_{i=1}^{n} X_i + \widehat{\beta_1}\sum_{i=1}^{n} X_i^2$$

باخراج β_1 عامل مشترك مع نقل $\bar{y}$ $\sum_{i=1}^{n} X_i$ للطرف الأخر نحصل على:

$$\sum_{i=1}^{n} y_i X_i - \bar{y}\sum_{i=1}^{n} X_i = \widehat{\beta_1}\left(\sum_{i=1}^{n} X_i^2 - \bar{X}\sum_{i=1}^{n} X_i\right) \Rightarrow \widehat{\beta_1} = \frac{\sum_{i=1}^{n} y_i X_i - \bar{y}\sum_{i=1}^{n} X_i}{\left(\sum_{i=1}^{n} X_i^2 - \bar{X}\sum_{i=1}^{n} X_i\right)}$$

$$\Rightarrow \widehat{\beta_1} = \frac{\sum_{i=1}^{n} y_i X_i - \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} X_i}{n}}{\sum_{i=1}^{n} X_i^2 - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} X_i}{n}} = \frac{n\sum_{i=1}^{n} y_i X_i - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} X_i}{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}$$

ومنه فإن:

$$\widehat{\beta_1} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

$$\widehat{\beta_o} = \bar{y} - \widehat{\beta_1}\bar{x}$$

حيث أن $\bar{x}$ هو الوسط الحسابي لقيم x، $\bar{y}$ هو الوسط الحسابي لقيمy ، وتكون القيمة المقدرة للمتغير التابع هو: $\hat{y} = \widehat{\beta_o} + \widehat{\beta_1}x$ ، ويطلق على هذا التقدير " تقدير معادلة انحدارy علىx.

مثال (5–6): فيما يلي بيانات عن كمية البروتين اليومي بالغرام التي يحتاجها العجل الرضيع، ومقدار الزيادة في وزن العجل بالكلغ وذلك لعينة من العجول الرضيعة حجمها 10.

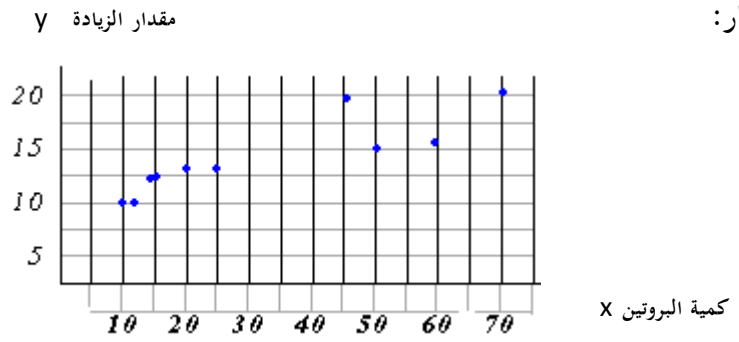
كمية البروتين	10	11	14	15	20	25	46	50	59	70
الزيادة في الوزن	10	10	12	12	13	13	19	15	16	20

والمطلوب :

1. ارسم نقط الانتشار، وما هو توقعاتك لشكل العلاقة ؟
2. قدر معادلة انحدار الوزن على كمية البروتين.
3. فسر معادلة الانحدار.
4. ما هو مقدار الزيادة في الوزن عند إعطاء العجل 50 غرام من البروتين ؟ وما هو مقدار الخطأ العشوائي؟
5. ارسم معادلة الانحدار على نقط الانتشار في المطلوب .

الحل

1- رسم نقط الانتشار:



من المتوقع أن يكون لكمية البروتين أثر طردي (إيجابي) على مقدار الزيادة في الوزن.

2- **تقدير معادلة الانحدار.**

بفرض أن x هي كمية البروتين، y هي مقدار الزيادة في الوزن، يمكن تطبيق المعادلات السابقة ومن ثم حساب المجاميع التالية:

كمية البروتين x	الزيادة في الوزن y	X y	x^2
10	10	100	100
11	10	110	121
14	12	168	196
15	12	180	225
20	13	260	400
25	13	325	625
46	19	874	2116
50	15	750	2500
59	16	944	3481
70	20	1400	4900
320	140	5111	14664

المجاميع المطلوبة
$\sum x_i = 320$ $\sum y_i = 140$ $\sum x_i y_i = 5111$ $\sum x_i^2 = 14664$
حساب الوسطين الحسابيين
$\bar{x} = \frac{\sum x_i}{N} = \frac{320}{10} = 32$
$\bar{y} = \frac{\sum y_i}{N} = \frac{140}{10} = 14$

- بتطبيق المعادلة الأولى في يمكن حساب β_1 كما يلي:

$$\widehat{\beta_1} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2} = \frac{10(5111) - (320)(140)}{10(14664) - (320)^2} = \frac{6310}{44240} = 0.1426$$

- بتطبيق المعادلة $\hat{\beta}_0 = \bar{y} - \widehat{\beta_1}\bar{x}$ التالية يمكن حساب $\hat{\beta}_0$ كما يلي:

$$\widehat{\beta_o} = \bar{y} - \widehat{\beta_1}\bar{x} \Rightarrow \widehat{\beta_o} = 14 - (0.1426)\,32 = 9.4368$$

- إذا معادلة الانحدار المقدرة تكون على الشكل:

$$\hat{y} = 9.4368 + (0.1426)\bar{x}$$

3- **تفسير المعادلة:**

الثابت $\widehat{\beta_o} = 9.4368$: يدل على أنه في حالة عدم استخدام البروتين قي التغذية، فإن الوزن يزيد 9.44 كلغ

معامل الانحدار $\widehat{\beta_1} = 0.1426$ يدل على أنه كلما زادت كمية البروتين غرام واحد، حدث زيادة في وزن العجل بمقدار 0.143 كلغ أى زيادة مقدارها 143 غرام.

4- مقدار الزيادة في الوزن عند $x = 50$ هو:

$$\hat{y} = 9.4368 + (0.1426)(50) = 16.59$$

$$\hat{y} = 9.44 + 0.143(50) = 16.59$$

- وأما ومقدار الخطأ العشوائي هو:

$$\hat{e}_{x=50} = y_{x=50} - \hat{y}_{x=50} = 15 - 16.59 = -1.59$$

5- **رسم معادلة الانحدار على نقط الانتشار.**

- يمكن رسم معادلة خط مستقيم إذا علم نقطتين على الخط المستقيم.

x	50	10
$\hat{y}$	16.59	10.87

تمارين مقترحة

التمرين الأول: لتكن البيانات الآتية عن المتغيرين x و y .

$\sum_{i=1}^{10} x_i = 92$ $\sum_{i=1}^{10} y_i = 86$ $\sum_{i=1}^{10} x_i^2 = 982$ $\sum_{i=1}^{10} y_i^2 = 783$ $\sum_{i=1}^{10} x_i y_i = 865.5$

المطلوب:

- ايجاد معادلة خط انحدار y على x ، ومعادلة خط انحدار x على y.

- ايجاد معامل الارتباط بين قيم x و y.
- إذا كانت قيمة x=10 فما هي القيمة التقديرية لy عندئذٍ.

التمرين الثاني:

البيانات التالية تبين الدخل والاستهلاك لمجموعة من الموظفين حيث الدخل والاستهلاك بعشرات آلاف الدينارات

الدخل	**11**	**8**	**7**	**5**	**3**	**2**
الاستهلاك	**5**	**4**	**3**	**3**	**2**	**1**

.المطلوب:

- قدِّر نموذج الانحدار الخطي البسيط.
- كم يكون الدخل إذا كان الاستهلاك 8 د ج.
- كم يكون الاستهلاك إذا كان الدخل 15 د ج.

التمرين الثالث:

البيانات التالية تمثل مصاريف الاشهار والأرباح المحققة في مؤسسة تجارية، حيث تعطى الارقام بعشرات آلاف الدينارات

X	2	6	7	8	10	11	12
y	3	4	5	6	10	8	13

المطلوب:

- قدِّرْ نموذج الانْحدار الخطي البسيط.
- كم تكون قيمة x إذا كانت قيمة yهي 18.
- كم تكون قيمة y إذا كانت قيمة xهي 15.

التمرين الرابع:البيانات التالية تمثل تقديرات 8 طلبة في مادتي التاريخ والجغرافيا

التاريخ	جيد	راسب	مقبول	مقبول +	جيد جداً	ممتاز	ممتاز +	جيد +
الجغرافيا	جيد جداً	جيد	راسب	مقبول	ممتاز	مقبول +	جيد +	ممتاز +

المطلوب:

هل هناك علاقة بين تقدير الطلبة في المادتين.

التمرين الخامس: إذا كانت Xعدد سنوات الدراسة، وy هي الدخل بآلاف الوحدات النقدية،وكانت:

$\sum_{i=1}^{12} x_i y_i$ =**1334** $\sum_{i=1}^{12} y_i^2$=**990** $\sum_{i=1}^{12} x_i^2$=**1902** $\sum_{i=1}^{12} y_i$ =**102** $\sum_{i=1}^{12} x_i$ =**146**

المطلوب:

- اوجد معامل الارتباط الخطي البسيط بين عدد سنوات الدراسةى والدخل السنوي.
- اوجد معادلة خط انحدار y علىx.
- اوجد معادلة خط انحدار x علىy.
- اوجد الدخل السنوي لفرد يقضي 20 سنة في مؤسسات الدراسة والتكوين.

التمرين السادس:

لدراسة العلاقة بين العمر x وضغط الدم y لعينة مكونة من 12 امراة تحصلنا على البيانات التالية

X	52	42	72	36	63	47	55	49	38	42	68	60
Y	147	125	160	118	149	128	150	145	115	140	152	155

المطلوب:

1. ارسم الشكل الانتشاري للمتغيرين x وy.
2. إذا كان الشكب يشير إلى وجود علاقة خطية بين x وy
   أ. اوجد معامل الارتباط الخطي البسيط.
   ب. قارن قيمته بقيمة معامل ارتباط الرتب لنفس البيانات.
3. باستخدام معامل الارتباط:
   أ. أوجد معامل انحدار y على x.
   ب. أوجد معادلة خط الانحدار.
   ت. ماهو العمر التقدير لامراة ضغط دمها 172، وماهو ضغط دم امرأة عمرها 32.

المراجع:

- أحمد شيبات، " الإحصاء الوصفي"،دار الهدى للطباعة و النشر،عين مليلة،الجزائر،2005.
- سالم عيسى بدر،" مبادئ الاحصاء الوصفي و الاستدلالي"، دار الميسرة للنشر و التوزيع ،عمان، الاردن،ط2010.
- عبد القادر حلمي:مدخل إلى الاقتصاد،ديوان المطبوعات الجامعية،الجزائر،2004.
- عبد الكريم بوحفص:الاحصاء المطبق في العلوم الاجتماعية والانسانية،ديوان المطبوعات الجامعية، الجزائر،2006.
- جيلالي جلاطو: " الإحصاء مع تمارين ومسائل محلولة "، ديوان المطبوعات الجامعية، الجزائر، 2002.
- جاك لوكاين+كريستيان لا بروس:الاحصاء الوصفي،ديوان المطبوعات الجامعية، الجزائر 1985.
- عبد الرزاق عزوز: " الكامل في الإحصاء" الجزء الأول، ديوان المطبوعات الجامعية، الجزائر، 2010.
- عبد الرزاق عزوز: " الكامل في الإحصاء" الجزء الثاني، ديوان المطبوعات الجامعية، الجزائر، 2011.
- علي لزعر:تمارين مصححة ومسائل في الاحصاء، منشورات جامعة باجي مختار، عنابة الجزائر،2006
- فتحي حمدان وكامل فليفل: " مبادئ الإحصاء"،دار المناهج، الأردن، الطبعة الأولى 2006.
- محمد بوهزة: " محاضرات في الإحصاء الوصفي"، دار المحمدية العامة، الجزائر، 2011.
- محمد راتول: " الإحصاء الوصفي"، ديوان المطبوعات الجامعية، الجزائر، الطبعة الثانية 2006.

1- Amany Mousa: Cairo (2005), Statistical Data Analysis, Center for Advancement of Postgraduate Studies and Research, Faculty of Engineering,Cairo University.
2- Bowers .D : Statistics for economics and business , MacMillan education L.T.D ,London ,1991 .
3- Chauvat Gérard : Statistique descriptive , Paris ,1992 .
4- Dubos .J :Statistiques descriptives en sciences économique avec exercices corrigés , Dunod , Paris , 1980
5- Hocine Hamdani ;statistiques descriotive,OPU ?2001
6- Malika Boukella-Bouzouane ;statistiques descriptive ;Casbah edition ;Alger ;2001