الجمهورية الجزائرية الديمقراطية الشعبية
People's Democratic Republic of Algeria
وزارة التعليم العالي والبحث العلمي
Ministry of Higher Education and Scientific Research
جامعة الجزائر 3 University of Algiers 3
معهد التربية البدنية والرياضية Sport and Physical Education Institute

مطبوعة محاضرات مقياس الاحصاء الوصفي

- المستوى : السنة الثانية ليسانس
- إعداد الأستاذة: زاوي عقيلة
- البريد الالكتروني:.lamari zaoui @gmail.com

السنة الجامعية: 2021 / 2022

تمهيد :

يشير مفهوم الاحصاء عند عامة الناس الى مجموعة من الاقام والبيانات فقط كأعداد السكان واعداد الطلبة واعداد المواليد وغيرها من الاعداد ،مفهوم الاحصاء بهذا المعنى يقتصر على اعتباره عد او حساب الاشياء ،ومن هنا فهو من هذا المنظور مفهوم محدود .اما عند المتخصصين فيدل مفهوم الاحصاء كعلم ،على الميدان الذي يهتم بوصف وتلخيص وتفسير الظواهر الطبيعية والاجتماعية والنفسية والتربوية التي تعتبر التغير ميزتها الاساسية ،انه ذلك الفرع من الرياضيات الذي يقدم بأقصى ما يمكن من الدقة والصرامة ،عناصر التقدير اللازمة لتفسير هذه الظواهر والتنبؤ بما قد يطرا عليها من تغيرات .

اضافة الى ذلك يقدم الاحصاء اساليب محددة ومنظمة لجمع البيانات وتبويبها وتلخيصها بأشكال مختلفة كالجداول التكرارية والرسوم البيانية التي تساعد على وصف البيانات ومعالجتها وتحليلها للوصول الى استخلاص نتائج تساعد على اتخاذ قرارات سليمة بخصوص الظواهر التي يدرسها الباحثون او بخصوص فرضيات البحوث العلمية التي يطرحونها .

من هنا يعد الاحصاء الركيزة الاساسية في مجال البحث العلمي باعتبار الادوار الكثيرة التي يقدمها في هذا المجال ،فهو اضافة الى ما سبق ذكره :

- ✓ يساعد على تطوير وتقدم البحث العلمي .
- ✓ يسمح بإصدار احكام على الظواهر بشكل موضوعي .
- ✓ يكشف عن مستويات العلاقات بين الظواهر ودلالتها.
- ✓ يكشف عن الفروق بين المجموعات ودلالتها الاحصائية .

سوف نتطرق في هذا المقياس إلى الإحصاء الوصفي الذي يعالج المرحلة الأولى من الطريقة الإحصائية التي تتضمن ثلاث مراحل:

- ✓ المرحلة الوصفية.
- ✓ المرحلة التنبؤية.
- ✓ مرحلة اتخاذ القرار.

اقتصرنا على المحاور أساسية منها: العرض الجدولي والبياني، خصائص النزعة المركزية، خصائص التشتت والشكل، واخيرا الانحدار الخطي البسيط والارتباط . إن الهدف من اختيار هذه المحاور، هو تمكين القارئ او الطالب من التكيف والتعامل مع الدراسة الإحصائية بسهولة وبساطة . هذه الدراسة التي يحتاج اليها كل طالب مهما كان تخصصه.

لقد عمدت في نهاية كل فصل الى عرض مجموعة من التمارين التطبيقية مرفقة بحلول مختصرة، والتي قد قدمت معظمهما في امتحانات طور (الليسانس) لمعهد التربية البدنية والرياضية (جامعة الجزائر3)، والمعهد الوطني للتخطيط و الإحصاء بالجزائر .

إن موضوع الإحصاء هو جمع وترتيب وتنسيق وتنظيم مجموعة من حوادث متجانسة أو من نفس الصنف، يمكن من خلالها الحصول على علاقات عددية مستقلة عن انحرافات الصدفة والتي تبين وجود أسباب منتظمة بتأثيرات منظمة.

أما الإحصاء فهو علم جمع وترتيب معلومات خاصة بظاهرة معينة وقياس الوقائع كأساس للإستقراء، فأما الوظيفة الاساسية للإحصاء هي جمع وعرض وتحليل المعطيات والتنبؤ بقيم الظاهرة المدروسة مستقبلا وفق شروط معينة ثم اتخاذ القرار المناسب.

تستعمل الطريقة الإحصائية في كل المجالات و تخصصات في العلوم الاجتماعية والعلوم الدقيقة والعلوم الاقتصادية والتجارية وعلوم التسيير...الخ خاصة إذا تعلق الأمر بالظواهر القابلة للقياس. يقول Jean-Jacques Droesberke " إن الأرقام لا تحكم العالم فقط بل تبين كيف نتحكم فيه".

تمكن الدراسة الإحصائية من تبسيط ظاهرة معينة حلال فترة زمنية محدودة على شكل جدول (العرض الجدولي) أو على شكل بياني (العرض البياني) أو على شكل عدد (العرض العددي).

الطريقة الإحصائية تهدف إلى دراسة معطيات ظواهر معينة في فترات سابقة لإيجاد حلول مناسبة لمشاكل الحاضر والتنبؤ بقيمها في المستقبل، إذا بقت نفس الظروف على حالها أو وفق فرضيات معينة.

المحاضرة الاولى :مدخل الى علم الاحصاء الوصفي

- ✓ تحديد المفاهيم والمصطلحات الاساسية
  - الوحدة الاحصائية
  - المجتمع الاحصائي
  - الظاهرة الاحصائية
  - العينة الاحصائية
- ✓ انواع المتغيرات الاحصائية
  - متغيرات كيفية
  - متغيرات كمية
- ✓ مصادر المعلومات الاحصائية

الهدف منها ترسيخ المعارف المسبقة والمطلوبة للمقياس

الفصل الأول : عرض البيانات الاحصائية

1- **بعض المفاهيم الأساسية** (جلاطو جيلالي،2005،ص05)

- **الوحدة الإحصائية :(Unité statistique)** هي الوحدة الأساسية لتكوين المجتمع الإحصائي.
- **المجتمع الاحصائي: (Population)** هو مجموعة المشاهدات والقياسات الخاصة بمجموعة من الوحدات الإحصائية والتي تخص ظاهرة من الظواهر القابلة للقياس: مجتمع من الطلبة، مجتمع من الأسر، مجتمع من المؤسسات ...إلخ.
- **الظاهرة الإحصائية: (Phénomène statistique)** هي الخاصية المدروسة في المجتمع الإحصائي مثلا: طول القامة ، السن، الوزن، العلامة المتحصل عليها في إمتحان معين، الإنتاج، الإدخار، الإستثمار، الاستهلاك،...إلخ.
- **العينة: (Echantillon)** هي جزء من المجتمع الإحصائي ، ولكن ليس أي جزء، إنه الجزء الذي يمثل المجتمع أحسن تمثيل. يختلف حجم العينة حسب أهمية الدراسة وحسب الإمكانيات المادية و البشرية المتاحة للقيام بهذه الدراسة. إن الاعتماد على أسلوب العينة متبع في أغلب الدراسات الميدانية و هذا لإستحالة جمع المعلومات الإحصائية من كل الوحدات والتي تشكل المجتمع المدروس أو بما يسمى بالحصر الشامل.
- **أنواع العينات:** تنقسم العينة العشوائية حسب طريقة سحبها الى نوعين :

✓ **عينة عشوائية :** يقصد بالعشوائية المبدأ الذي يعني اتاحة فرص متساوية لكل مفردات المجتمع للظهور في العينة ؛فاختيار العشوائي اذن هو الاختيار الذي يتحقق وفق مبدا العشوائية ،بينما الاختبار المتحيز هو الاختبار الذي لا يتحقق فيه مبدا العشوائية ؛أي لا تعطي فيه فرص متساوية امام جميع مفردات المجتمع لكي تكون ممثلة في العينة .وعادة ما تتحقق العشوائية باستعمال ما يسمى "الكيس المثالي " ويستخدم في حالة صغر حجم العينة حيث يكتب رقم لكل مفردة من مفردات في اوراق صغيرة متماثلة تم توضع في صندوق و تخلط جيدا وبعدها تسحب عدد من الاوراق يكون مطابقا لعدد مفردات العينة المطلوبة سحبه من المجتمع أي اتاحة الفرصة نفسها لجميع مفردات المجتمع لكي تكون ممثلة في العينة اما اذا كان المجتمع كبيرا فعادة ما تستخدم جداول الارقام العشوائية ،وهي جداول احصائية تم تكوينها ليتسنى للباحثين استخدامها عند اختيار العينات العشوائية .

1- **العينة العشوائية البسيطة : (Echantillon Aléatoire)**تعتبر العينة العشوائية من أكثر العينات إستعمالا، وهي جزء لا على التعيين من المجتمع المدروس، وتخص العينة العشوائية المجتمعات المتجانسة أي تلك التي لها خصائص متشابهة أو متقاربة مثلا: مجتمع من الإطارات المتوسطة يشكل مجتمع متجانس من حيث المداخيل المتحصل عليها، مجتمع من الطلبة يشكل مجتمعا متجانسا من حيث السن وهكذا... إذن تسحب العينة العشوائية من مجتمع متجانس.

يتم الحصول على العينة البسيطة بإجراء القرعة ،حيث يعطي الباحث ارقاما للقائمة التي حصل عليها ويضع الارقام في صندوق او كيس ويسحب الارقام تلو الاخر ،حتى يستكمل الحجم المناسب للعينة ؛فاذا كان عدد افراد

مجتمع البحث محدودا فانه يمكن كتابة اسمائهم على بطاقة ،ثم وضعها في صندوق وخلطها وبعد يأخذ الباحث العدد المراد من البطاقات ،اما اذا كانت الاعداد ضخمة فيتم اللجوء او الاستعانة بجداول "ارقام الاختبار العشوائي " التي وضعها علماء الاحصاء لهذا الغرض .

2- العينة الطبقية:(Echantillon Stratifié) تخص المجتمعات غير المتجانسة أ ي المجتمعات المتكونة من عدة فئات إجتماعية اي عندما يكون مجتمع البحث موزعا توزيعا جغرافيا متنوعا او ينتمي افراده الى طبقات اجتماعية مختلفة او الى مستويات تعليمية متفاوتة لا بد من ان نأخذ بعين الاعتبار هذه الخصائص لنضمن حضورها في العينة بنسب حقيقة حتى تكون ممثلة .يمكن ان نتوخى الدقة في تحديد هذه الخصائص والصفات بحساب ما تمثله احصائيا في المجتمع ،وبعد ذلك تنتقي عينة مطابقة احصائيا لنسبة تلك الخصائص فاختيار عينة طبقية ينبغي تقسيم المجتمع الاحصائي الى طبقات حسب متغير من المتغيرات المدروسة ثم يستخرج من كل طبقة عينة عشوائية وبعدها تدمج العينات الجزئية في عينة واحدة .

3- العينة ذات المراحل المتعددة (العنقودية): يتم اختيار ها عندما تكون مفردات المجتمع على شكل تجمعات او مجموعات او عناقيد ،بحيث يحتوي كل عنقود على الكثير من مفردات المجتمع .والتسمية هنا جاءت من "عناقيد العنب او التمر "حيث توجد المفردات على شكل عناقيد وكل عنقود يحتوي على العديد من حبات او مفردات ،ففي العينة العنقودية يتم اختيار عدد من العناقيد بطريقة عشوائية بسيطة تم نقوم بدراسة كل المفردات داخل هذه العناقيد التي تم اختيارها "او معظمها".

مثال نستعمل هذه العينات عند دراسة المستوى التعليمي في بلد ما: مثلا لدراسة المستوى التعليمي للمرحلة النهائية للتعليم الثانوي نتبع الخطوات التالية:

أ- نسحب ولاية بطريقة عشوائية من أصل مجموع الولايات.

ب- نسحب دائرة من أصل الدوائر التي تتكون منها الولاية المسحوبة.

ج- نسحب ثانوي من أصل الثانويات المتواجدة في الدائرة المختارة.)

د- يسحب قسم من الأقسام النهائية على مستوى الثانوية المسحوبة لتجرى عليه الدراسة.

4. عينة عشوائية منتظمة :_ لا يتم اختيار افراد هذا النوع الا في حالة تجانس المجتمع الاحصائي ،ويمكن ان نسوق مثالا يوضح كيف يتم اختيار عينة منتظمة :لنفترض ان المجتمع الاحصائي متكون 400 طالبا ،ويريد الباحث اختيار عينة عشوائية منه يكون حجمها 40 طالب ،يجب عليه في هذه الحالة تحقيق المراحل التالية :

1. الحصول على قائمة مرقمة للطلبة .
2. قيام بعملية تقسيم حجم المجتمع على حجم العينة أي 40/400.
3. جعل المسافة بين الرقم الاول الذي يختاره والرقم الذي يليه حاصل العملية السابقة أي 10.
4. يختار الرقم الاول عشوائيا وليكن رقم 05. بهذه العملية تكون العينة مكونة من الطلبة الذين يأخذون الارقام التالية (05،15،25،35،.....).

✓ **عينة غير عشوائية:** هي التي اختيار مفرداتها بطريقة غير عشوائية حيث يقوم الباحث باختبار مفردات العينة بالصورة التي تحقق الهدف من المعاينة ،مثل اختبار عينة من الطلبة المتفوقين من عدة كليات واهم انواع العينات غير الاحتمالية .

1. **عينة قصدية :** وتعني ان اساس الاختبار هو خبرة الباحث ومعرفته بان هذه المفردة او تلك تمثل مجتمع البحث ،فالباحث مثلا عندما يختار المدارس التي يعرفها لتمثيل جميع المدارس يعد اختبار عمديا .وينصح الباحث عندما يضطر الى تطبيق هذا الاسلوب ان يبرره تبريرا علميا حتى لا يتهم بالحيز
2. **عينة حصصية :** وسميت بالحصصية لان مجتمع الدراسة يقسم الى فئات (حصص)طبقا لصفاته الرئيسية ،وتمثل كل فئة في العينة بنسبة وجودها في المجتمع ؛فمثلا اذا كان مجتمع البحث طلبة كلية فينبغي اولا تصنيفهم طبقا لتخصصاتهم ،ثم يقرر الباحث النسبة المئوية المطلوب سحبها من كل تخصص ويبدا بحسب النسب وبهذا يتدرج حجم العينة طبقا لعدد الطلاب في كل تخصص ،فالتخصصات ذات الاعداد الكبيرة يكون تمثيلها في العينة اكبر من تمثيل التخصصات ذات الاعداد الصغيرة.

توجد انواع اخرى من العينات لها علاقة بفرضيات الباحثين وهي العينات المستقلة والعينات غير المستقلة ؛فقبل اية معالجة احصائية يجب على الباحث ان يجيب على السؤال .

2- **<u>أنواع الخصائص أو المتغيرات:</u>** يتضمن مصطلح متغير شيئا يتغير ،ويأخذ قيما مختلفة او صفات متعددة فهو مفهوم يعبر عن الاختلافات بين عناصر فئة معينة مثل الجنس ، التحصيل ، الحالة العائلية .

تنقسم المتغيرات احصائية إلى قسمين (محمد خير سليم ابوزيد ، 2018، ص 23-27) :

1-متغيرات كيفية : (variables qualitatives) هي تلك المتغيرات التي لا يمكن قياسها أو غير قابلة للقياس مثل الجنسية، الحالة العائلية، الحالة المدنية...إلخ.

<u>2</u>-متغيرات كمية:(Variables Quantitative) هي تلك الخصائص التي يمكن قياسها وهي أكثر المتغيرات إنتشارا وإستعمالا لأن لغة الإحصاء هي لغة الأرقام، مثال لذلك: الإنتاج، الأستهلاك، الإستثمار، الوزن...إلخ. وتنقسم المتغيرات الكمية إلى قسمين: متغيرات منقطعة ومتغيرات مستمرة.

أ- **متغيرات منقطعة:(variables discrètes)** وهي تلك المتغيرات التي تأخذ قيما صحيحة لا يمكن تجزئتها مثلا عدد الأطفال في العائلة، عدد قطع الغيار المنتجة...إلخ.

ب- **متغيرات مستمرة: variables continues)** وهي تلك المتغيرات التي تأخذ كل القيم الممكنة في مجال الدراسة، ونظرا للعدد الغير المناهي لهذه القيم نقسم مجال الدراسة إلى مجالات جزئية تسمى الفئات.

3- **مصادر المعلومات الإحصائية** (جلاطو جيلالي ،2005، ص8) :

يعتمد الباحثون على مصدرين للحصول على المعلومات الإحصائية الخاصة بظاهرة معينة، الأول مباشر والثاني غير مباشر.

1- **المصدر المباشر:** يمكن الحصول على المعطيات الإحصائية بواسطة طريقتين،:

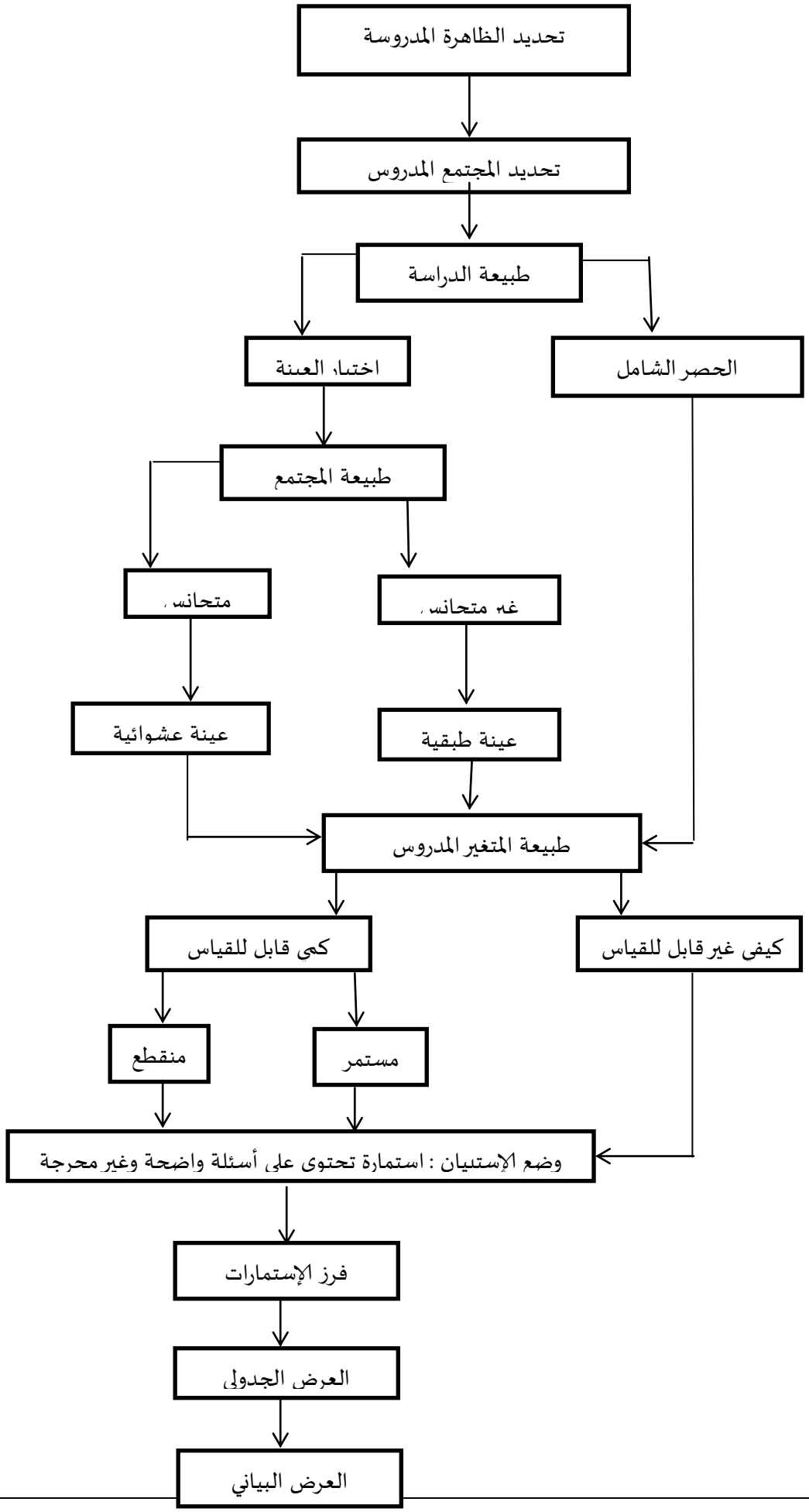
أ- **الطريقة المباشرة:** يقوم الباحث بوسائله الخاصة بإجراء دراسة ميدانية على كل الوحدات التي تشكل المجتمع المدروس أو بما يسمى الحصر الشامل، ولكن هذه الطريقة تصبح مستحيلة في كثير من الحالات.

ب- **الطريقة غير المباشرة:** وعند استحالة تطبيق الطريقة المباشرة، يقوم الباحث بإنتقاء أو إختيار عينة من المجتمع المدروس أو بما يسمى بالمعاينة ثم يعمم نتائج الدراسة على المجتمع ككل وفق طرق وأساليب إحصائية معينة.

2- **المصدر غير المباشر:** يتحصل الباحث على المعلومات الإحصائية من الدراسات والتحقيقات المنشورة على صفحات الجرائد، الكتب ، والوثائق والتقارير الخاصة...إلخ، ولكن هدف الباحث يختلف بصفة عامة عن هدف الهيئات التي قامت بهذه الدراسات، ولذلك فالبيانات المتحصل عليها من المصدر غير المباشر تكون في بعض الأحيان غير هامة.

3- **المرحلة القبلية لعرض المعلومات الإحصائية:**

تأتي عملية عرض المعلومات في مرحلة تالية لمرحلة تحديد الظاهرة المدروسة والمجتمع ونوع المتغير وطبيعة السحب (سحب العينة) وعملية تحضير الاستبيان وعملية الفرز ثم تفريغ المعطيات الأولية في جداول أولية وأخيرا توضع المعطيات الأولية في جداول نهائية تسمى بجداول المعطيات والمخطط التالي يبين مختلف المراحل الهامة القبلية لعملية العرض الجدولي:


تحديد الظاهرة المدروسة
تحديد المجتمع المدروس
طبيعة الدراسة
اختيار العينة
الحصر الشامل
طبيعة المجتمع
متجانس
غير متجانس
عينة عشوائية
عينة طبقية
طبيعة المتغير المدروس
كمي قابل للقياس
كيفي غير قابل للقياس
منقطع
مستمر
وضع الإستبيان : استمارة تحتوى على أسئلة واضحة وغير محرجة
فرز الإستمارات
العرض الجدولي
العرض البياني

المحاضرة الثانية : العرض الجدولي

- ✓ تعريف العرض الجدولي
- ✓ العرض الجدولي في حالة خاصية واحدة
  - جدول التكرارت لخاصية كمية منقطعة
  - جدول التكرارات لخاصية كمية مستمرة
  - جدول التكرارات لخاصية كيفية
- ✓ العرض الجدولي في حالة خاصيتين فاكثر " المزدوج"

الهدف منها : التعرف على انواع الجداول وكيفية تصنيف وتبويب المعلومات المتحصل عليها

6- العرض الجدولي : (نجيب حسن محمد ،2004، ص47-68)

يقوم الباحث بتفريغ النتائج المتحصل عليها في جداول أولية بعد الانتهاء من عملية فرز الإستمارات ثم يلخصها في جداول نهائية تسمى جداول المعطيات.

-تعريفه: هو عبارة عن وضع المعلومات الإحصائية في جداول نهائية، يحتوي كل منها عمودين (سطرين) يبين العمود (السطر) الأول قيم الظاهرة أو المتغير المدروس، وتكون هذه القيم على شكل قيم نقطية أو على شكل مجالات، أما العمود (السطر) الثاني فيحتوي على تكرارات هذه القيم أو المجالات (نجيب حسن محمد ،2004،ص49).

1- العرض الجدولي في حالة خاصية واحدة:

أ- **جدول التكرارات لخاصية كمية متقطعة:** الخاصية الكمية المنقطعة أو المتغير الكمي المنقطع هو ذلك المتغير الذي لا يمكن تجزئة قيمه، حيث يأخذ قيما صحيحة.

مثال : لدراسة متوسط عدد الأطفال في العائلة في مجتمع ما، أخذت عينة عشوائية حجمها n=100، فكانت النتائج كالتالي:

الجدول رقم:1.1

عدد الاطفال	0	1	2	3	4	5	6	7	∑
عدد العائلات	10	12	16	18	17	14	8	5	100

حيث أن Xi : قيم المتغير الإحصائي ، ni : تكرارات هذه القيم

ب- **حالة متغير كمي مستمر:**

عند دراسة متغير كمي مستمر، يضم مجال الدراسة مالا نهاية من القيم، ولتعذر وضع كل هذه القيم ، نقسم هذا المجال إلى مجالات جزئية تسمى الفئات ، حيث يحدد عدد هذه الفئات حسب حجم العينة وحسب توزيع الوحدات الإحصائية على مجال الدراسة، ولكن قصد تسهيل العملية وضع الإحصائي ستورج (Sturge) قاعدة تجريبية لتحديد طول الفئات وتعتمد هذه القاعدة على مجال الدراسة وحجم المجتمع أو العينة:

$$K = \frac{E}{1+1.32\ \ln(n)} \qquad K = \frac{E}{1+3.32\ \log(n)}$$

N حجم العينة أو المجتمع، E المدى العام وهو الفرق بين أكبر قيمة وأصغر قيمة: $E = Xn - X_1$، K طول الفئة ،

(1+ 3.32$\log n$) عدد الفئات.

ملاحظة: تعتبر الفئة جزء من مجال الدراسة، حيث لا يوجد طول متفق عليه لهذه الفئة، كما أنه لا يوجد عدد متفق عليه من الفئات لتوزيع إحصائي معين، ولكن يتحدد عدد الفئات وطولها حسب أهمية الدراسة وطول مجال التعريف وحجم العينة أو المجتمع، والحفاظ على توازن جدول التكرارات من الأفضل أن يتراوح عدد الفئات من 9 إلى 25 فئة. كما يستحسن أن تكون أطوال الفئات متساوية حتى يمكن مقارنة الظاهرة من مجال إلى مجال أخر.

وكما أن قاعدة (Stuege) ليست إجبارية بل على الباحث إختيار طول الفئة المناسب.

مثال: لدراسة ظاهرة تأخر العمال عن الإلتحاق بعملهم في بداية الفترة الصباحية، لوحظ أن 75 عاملا التحقوا متأخرين عن العمل في مؤسسة ما وفي يوم معين، والجدول التكراري الاتي يبين ذلك:

الجدول رقم:1.2

الفئات	التكرار
0-5	4
5-10	7
10 -15	10
15 -20	15
20 – 25	18
25 – 30	12
30 -35	6
35 – 40	3
$\sum$	75

نلاحظ أن طول الفئة يساوي K = 5

-جدول التكرارات التجميعية:Tableau des fréquences cumulées

يمكن أن نميز بين ثلاثة أنواع من التكرارات :

1- التكرار المطلق وهو التكرار العادي.

2- التكرار النسبي وهو عبارة عن نسبة مئوية.

3- التكرارات التجميعية، وهذه الأخيرة تنقسم إلى قسمين:

أ- التكرارات التجميعية الصاعدة: التكرار التجميعي الصاعد لأي فئة هو عبارة عن تكرار هذه الفئة مضافا إليه مجموع تكرارات الفئات السابقة.

ب- التكرارات التجميعية النازلة: التكرار التجميعي النازل لأي فئة هو عبارة عن مجموع التكرارت ($\sum Ni$) مطروحا منه تكرارات الفئات السابقة.

مثال: نفس المثال السابق ، المطلوب وضع التكرار التجميعي الصاعد والنازل.

الجدول رقم: 1.3

الفئات	n_i	ت.ت.ص	ت.ت.ن
0 - 5	4	4	75
5 – 10	7	11	71
10 – 15	10	21	64
15 – 20	15	36	54
20 – 25	18	54	39
25 – 30	12	66	21
30 – 35	6	72	9
35 – 40	3	75	3
$\sum n_i$	75	/	/

<u>الملاحظة الأولى:</u> القواعد الواجب اتباعها عند تشكيل الجدول الإحصائي هي كما يلي:

1- عنوان واضح في أعلى الجدول ، يعطي فكرة عن البيانات التي يحتويها هذا الجدول.

2- ذكر مصدر البيانات في أسفل الجدول ، حتى يتمكن القارئ من الرجوع إلى المصدر الأصلي للمعلومات مثلا: التقرير السنوي للدوان الوطني للإحصائيات، د.م.ج، الجزائر 1996، الصفحة 25.

3- ذكر وحدة القياس المستعملة.

4- ذكر عنوان كل عمود.

5- وضع رقما للجدول.

الملاحظة الثانية: نستعمل نفس الطريقة،

في حالة الفئات غير المتساوية ، عند حساب وتحديد التكرارات التجميعية الصاعدة والنازلة.

ج- **العرض الجدولي في حالة خاصية كيفية:**

يتكون جدول المعطيات من عمودين (سطرين) يحتوي العمود الأول على رموز كتابية للخاصية المدروسة، أما الثاني فيحتوي على تكرارات كل رمز كتابي.

مثال: أخذت عينة عشوائية من أساتذة جامعة الجزائر ، متكونة من 100 أستاذ ، حيث أن الخاصية المدروسة هي رتبة الأستاذ ، فكانت النتائج كالتالي:

الجدول رقم : 1.4

رتبة أستاذ	أستاذ	أستاذ محاضر	مكلف بالدروس	أستاذ مساعد	معيد	∑
n_i	6	15	35	30	14	100

2-الجدول المزدوج: (Tableau de contingence)

يستعمل الجدول المزدوج عند دراسة خاصيتين في نفس الوقت في مجتمع ما، و توضع المعلومات الإحصائية بالشكل التالي : نضع أفقيا الخاصية الأولى ، ونضع عموديا الخاصية الثانية.

نرمز لقيم الخاصية الأولى بالرمز Yj حيث j تتراوح من 1 إلى m ، وبالرمز Xi لقيم الخاصية الثانية ، حيث i يأخذ القيم من 1 إلى n. مثال ذلك: لدراسة مستوى معيشة العائلات يمكن أن نتطرق إلى خاصيتين، مهنة رب العائلة و الإنفاق الإستهلاكي، نلاحظ أن الخاصية الأولى كيفية أما الخاصية الثانية فإنها قابلة للقياس (كمية).

يمكن وضع الجدول المزدوج بأبعاد أربعة حسب الشكل التالي:

الجدول رقم:1.5

	y_1	y_2	y_3	y_4	n_i
X_6	n_{11}	n_{12}	n_{13}	n_{14}	n_1
X_2	n_{21}	n_{22}	n_{23}	n_{24}	n_2
X_3	n_{31}	n_{32}	n_{33}	n_{34}	n_3
X_3	n_{31}	n_{32}	N_{43}	n_{44}	n_4
n.j	n.1	n.2	n.3	n.4	$\sum n.j = \sum ni.$

مثال: سحبت عينة عشوائية من مجتمع ما ، تتكون من 183 أسرة قصد دراسة خاصيتين هما مهنة رب الأسرة و التركيبة الأسرية من حيث عدد الأطفال ، فكانت النتائج كالتالي:

الجدول رقم : 1.6

	0	1	2	3	4	5	6	n_i
إطار متوسط	10	8	6	4	3	1	0	32
مهنة حرة	15	7	3	2	1	0	0	28
عامل متخصص	3	5	8	10	15	9	7	57
عامل بسيط	2	4	9	11	13	17	10	66
n.j	30	24	26	27	32	27	17	183

تسمى التكرارات .ni : التكرارات الحدية لـ X، والتكرارات n.j: التكرارات الحدية لـ y.

المحاضرة الثالثة : العرض البياني

- ✓ العرض البياني في حالة متغير كمي منقطع
  - العرض البياني للتكرارات البسيطة
  - العرض البياني للتكرارات التجميعية الصاعدة
  - العرض البياني للتكرارات التجميعية النازلة
- ✓ العرض البياني في حالة متغير كيفي
  - العرض الدائري
  - العمود المجزأ
  - الاعمدة البسيطة

الهدف منها : كيفية استعمال العروض البيانية وفق المتغيرات المحددة سابقا

7-العرض البياني (محسن لطفي احمد ،2011،ص،36-85):

يبين العرض البياني تطور الظاهرة المدروسة في الزمان أو في المكان وبمجرد الملاحظة الأولى. سنتطرق إلى عرض البيانات الإحصائية للحالات التالية :

1- متغير كمي منقطع.

2- متغير كيفي.

3- متغير كمي مستمر

1- العرض البياني في حالة متغير كمي متقطع: نكتفي في هذه الحالة بنوعين من العروض البيانية:

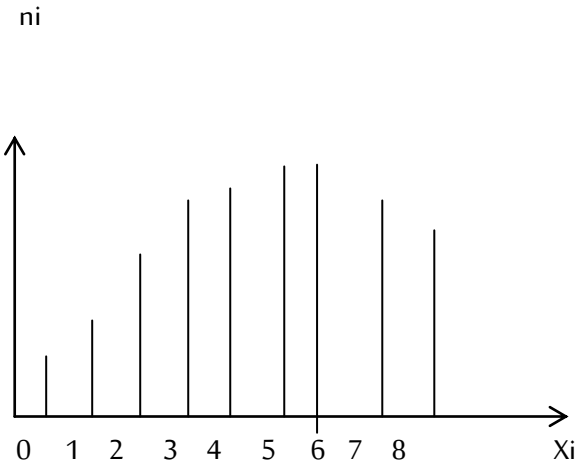
أ-العرض البياني للتكرارات البسيطة: هو عبارة عن أعمدة بسيطة تتناسب أطوالها مع التكرار المقابل لقيمة معينة للمتغير المدروس، وتسمى " الأعمدة البسيطة". (diagrammes en batons).

مثال : يبين الجدول التالي عدد الأطفال في العائلة لـ: 125 أسرة

المطلوب : العرض البياني للتكرارات البسيطة:

الجدول رقم: 1.7

Xi	Ni	ت.ت.ص	ت.ت.ن
0	6	6	125
1	9	15	119
2	10	25	110
3	14	39	100
4	16	55	86
5	20	75	70
6	25	100	50
7	15	115	25
8	10	125	10
$\sum ni$	125	/	/



الشكل رقم :1.1

ملاحظة: نلاحظ من بين الأعمدة التي تشكل العرض البياني السابق ، أن العمود الذي يقابل القيمة 6 هو أطولهم وتكراره يساوي 25 ، معنى ذلك أن أغلبية العائلات لها 6 أطفال.

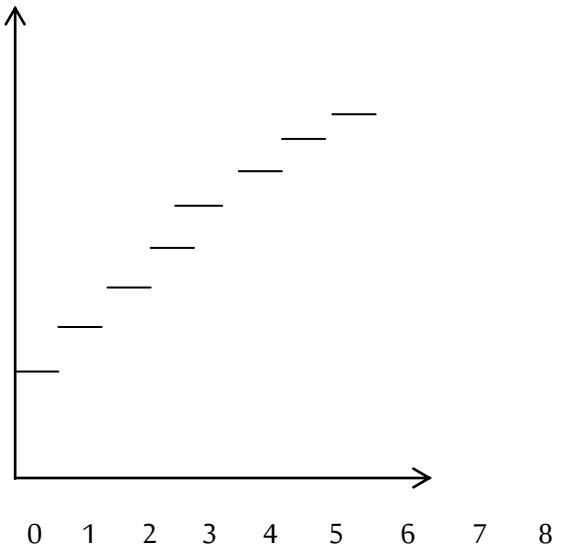
ب -العرض البياني للتكرارات التجميعية الصاعدة والنازلة:

ب-1- العرض البياني للتكرارات التجميعية الصاعدة: Effectifs cumulés croissants

هو عبارة عن قطع مستقيمة متصاعدة حسب تصاعد التكرارات التجميعية الصاعدة المقابلة لكل قيمة من قيم المتغير الإحصائي المدروس (عدد الأطفال في العائلة).

مثلا لرسم القطعة المستقيمة المقابلة للقيمة 2 والذي يساوي تكرارها الصاعد 25 ، نضع قطعة مستقيمة طولها 1 سم عند إحداثيات النقطة (2، 25) ، تبدأ من مستوى القيمة 2 لتنتهي عند مستوى القيمة 1 .

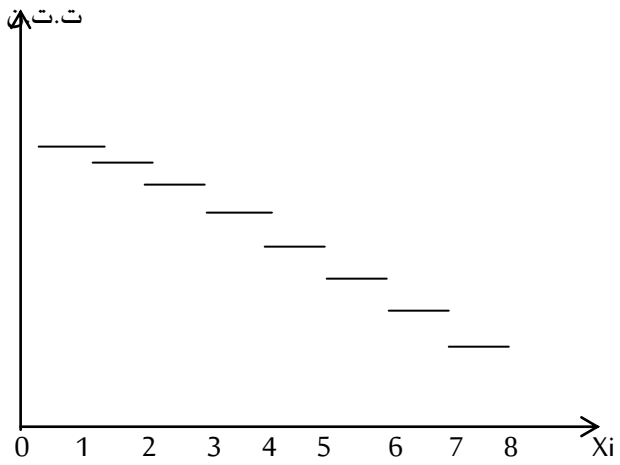
مثال: نفس المثال السابق مع إضافة عمودا إلى الجدول السابق (العمود الثالث) يحتوي على التكرارات التجميعية الصاعدة ، حيث أن العرض البياني لهذه التكرارات هو كالتالي :



الشكل رقم:1.2

ب-2--العرض البياني للتكرارات التجميعية النازلة : Effectifs cumulés décoissants_

هو عبارة عن قطع مستقيمة متنازلة حسب تنازل التكرارات التجميعية النازلة، حيث أن القطعة المستقيمة الاولى تقابل مجموع التكرارات و أصغر قيمة للمتغير المدروس والقطعة الثانية تقابل مجموع التكرارات ناقص التكرار البسيط الأول مع القيمة الثانية للمتغير الإحصائي وهكذا...



الشكل رقم:1.3

2- **العرض البياني في حالة متغير كيفي:** سنتعرض إلى أهم العروض البيانية :

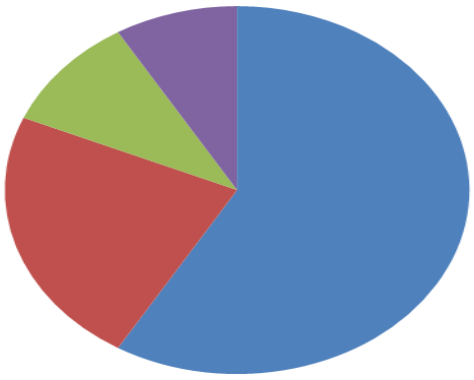
أ- **العرض الدائري : (Diagramme Circulaire)** يتمثل في دائرة مقسمة إلى عدة أجزاء كل جزء يقابل زاوية مركزية تتناسب مع التكرارات المقابلة لكل خاصية من الخصائص المدروسة ، ولتحقيق ذلك نضيف عمودا إلى جدول المعطيات يحتوي على الزوايا المركزية المقابلة لكل تكرار.

مثال: يبين الجدول التالي الإنتاج العالمي للذهب حسب القرات في سنة ما:

الجدول رقم :1.8

القارات	كمية الذهب 5 (طن)	الزوايا المركزية
أوروبا	176	57.6
آسيا	87	28.47
إفريقيا	431	141.45
أمريكا	350	114.54
أستراليا	56	18.32
المجموع	1100	/

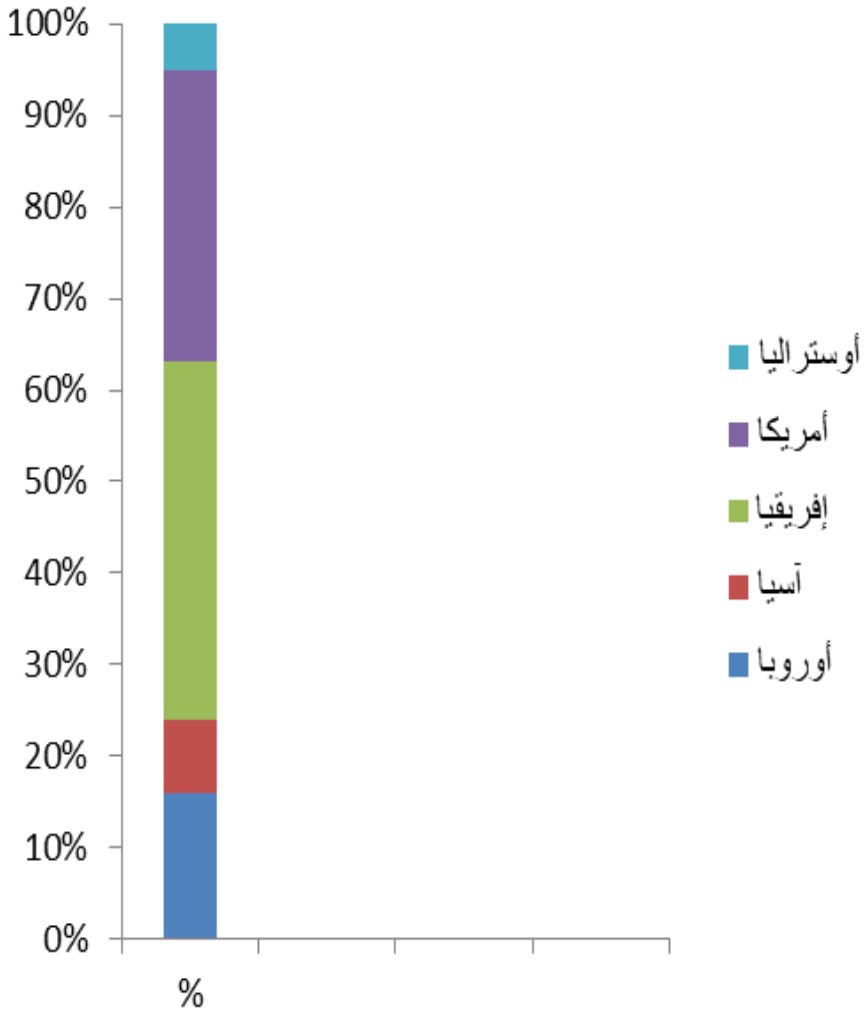
ملاحظة: يمكن إستعمال النسب المئوية لوضع العرض الدائري.



الشكل رقم: 1.4

ب- **العمود المجزأ: (Diagramme en Barres)** و هو عبارة عن مستطيل مقسم إلى عدة أجزاء، كل جزء يقابل تكرار معين للخاصية المدروسة.

مثال: نفس المثال السابق (الإنتاج العالمي للذهب) ، لوضع العمود المجزأ ، نستعمل النسب المئوية المقابلة لكل تكرار، حيث أن طول المستطيل يساوي 100%.



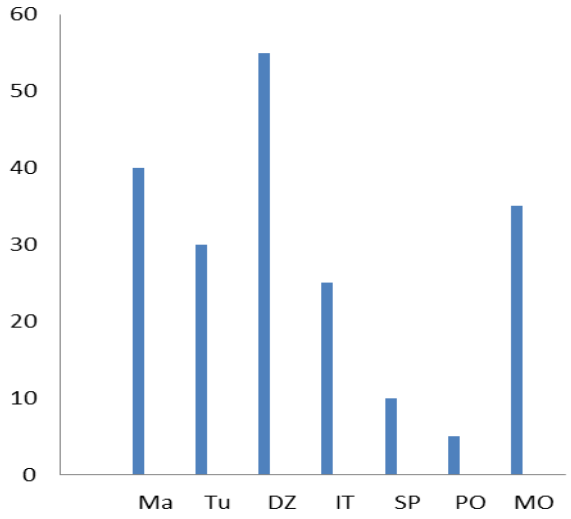
الشكل رقم:1.5

ج- **الأعمدة المستطيلة: (Diagramme en Colonnes)** و هو عبارة عن مستطيلات متباعدة بمسافات ثابتة، ولها قواعد متساوية، وتتناسب أطوالها مع التكرارات المقابلة لمكونات الخاصية المدروسة.

مثال: يبين الجدول التكراري التالي مكونات عينة عشوائية من المتغيرين في فرنسا:

الجدول رقم:1.9

الجنسية	n_i
المغرب	40
تونس	30
الجزائر	55
ايطاليا	25
إسبانيا	10
البرتغال	5
الشرق الأوسط	35
$\sum n_i$	200



الشكل رقم:1.6

المحاضرة الرابعة : تابع العرض البياني

- ✓ العرض البياني في حالة متغير كمي مستمر
  - المدرج التكراري
  - المضلع التكراري
  - العرض البياني للتكرارات التجميعية الصاعدة
  - العرض البياني للتكرارات التجميعية النازلة

الهدف منها : كيفية استعمال العروض البيانية وفق المتغيرات المحددة سابقا

3. العرض البياني في حالة متغير كمي مستمر (نجيب حسن محمد ،2004،ص 74-87) : إن العروض البيانية للمتغير الإحصائي المستمر هي أكثر العروض البيانية ، ومن أهمها ما يلي:

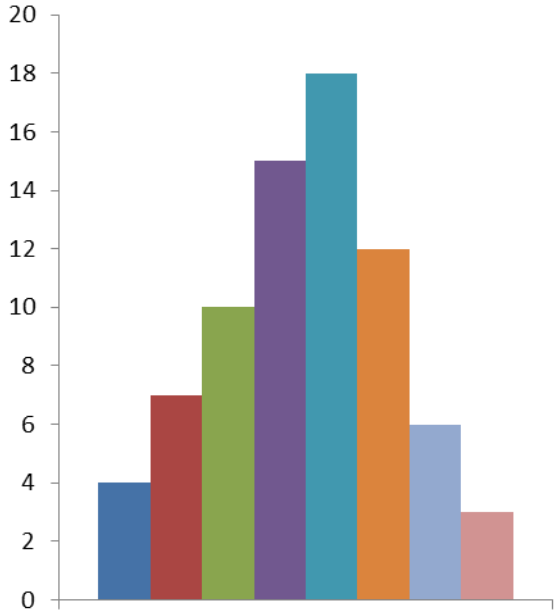
3-1- المدرج التكراري :(Histogramme) يكون على شكل مستطيلات متلاصقة، طول كل مستطيل منها يتناسب مع التكرار المقابل ، وقاعدة كل منها تساوي طول الفئة المقابلة ، حيث توضع الفئات على محور السينات وتوضع التكرارات على محور العينات، ويمكن أن نميز بين حالتين عند وضع المدرج التكراري:

الحالة الأولى : عندما تكون الفئات متساوية ، نلاحظ في هذه الحالة أن قاعدة المقارنة ثابتة ومتساوية ومن ثمة لا نجري أي تعديل على جدول المعطيات.

مثال: يبين الجدول التالي توزيع ظاهرة التأخر عن العمل على 75 عامل في إحدى المؤسسات:

الجدول رقم : 1.10

الفئات	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40	∑ni
ni	4	7	10	15	18	12	6	3	75



الشكل رقم:1.7

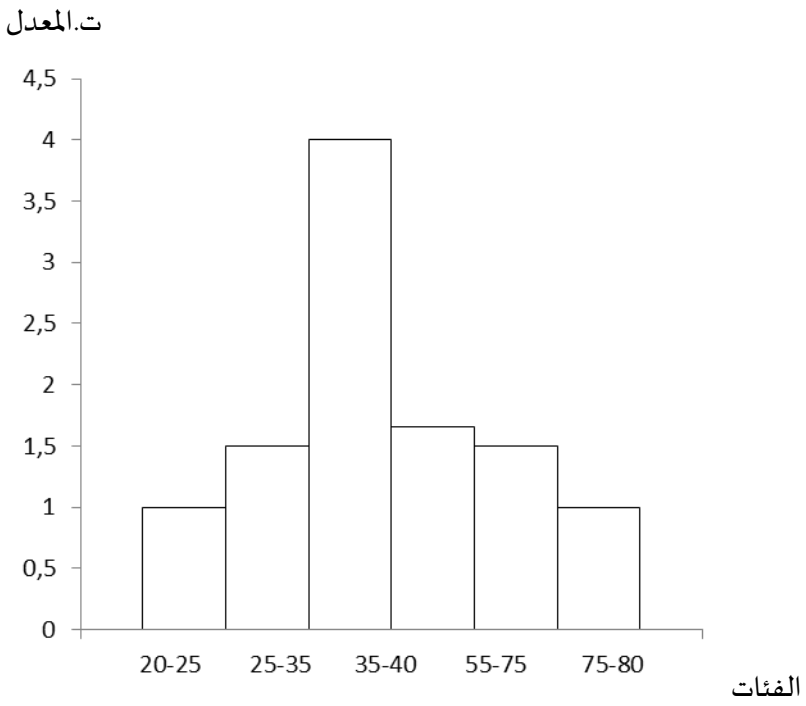
الحالة الثانية : حالة الفئات غير المتساوية: إذا كانت غير متساوية ، نقوم بتعديل التكرارات ، لأن قاعدة المقارنة غير ثابتة ، حتى يكون هناك تناسب بين طول الفئة والتكرار المقابل لها، أي إيجاد عدد الوحدات الإحصائية الموزعة على وحدة قياس معينة.

التكرار المعدل: هو عبارة عن النسبة بين التكرار البسيط وطول الفئة المقابلة .

مثال: يبين التوزيع التكراري التالي توزيع 100 عاملا حسب الأجر اليومي:

الجدول رقم :1.11

فئات الأجور	التكرار البسيط	طول الفئة	التكرار المعدل
20 -25	5	5	1
25 -35	15	10	1.5
35-40	20	5	4
40-55	25	15	1.66
55-75	30	20	1.5
75-80	5	5	1
مجموع التكرارات	100	/	/



الشكل رقم:1.8

ملاحظة : نقوم بتعديل التكرارات (في حالة فئات غير متساوية) في حالتين :

أ- عند رسم المدرج التكراري.

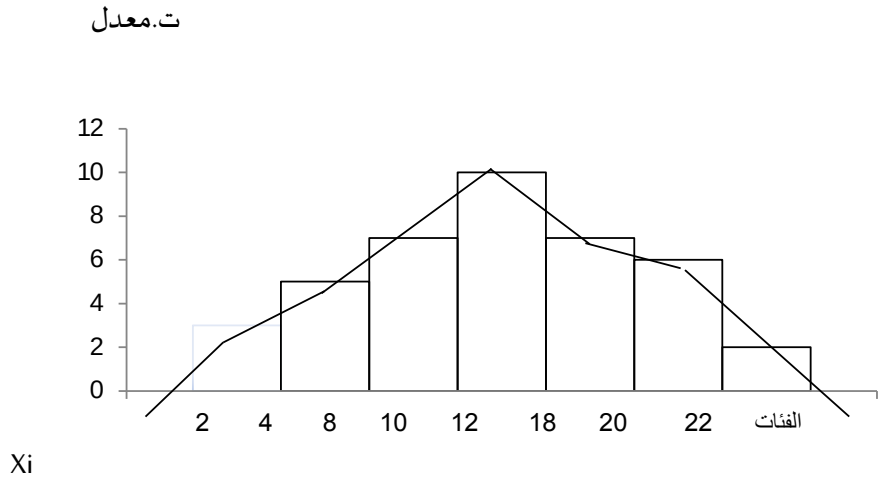
ب- عند تحديد الفئة المنوالية وحساب المنوال.

1- **المضلع التكراري:** (Polygone de Fréquence) هو مجموعة من قطع مستقيمة متصلة ومنكسرة ، تتحدد بنقاط إحداثياتها : مراكز الفئات والتكرارات المقابلة لها.

مثال: ليكن التوزيع التكراري التالي ، المطلوب وضع العرض البياني المناسب والمضلع التكراري.

الجدول رقم :1.12

الفئات	ni	Ki	التكرار المعدل
2-4	3	2	3
4 – 8	10	4	5
8 – 10	7	2	7
10 – 12	10	2	10
12 -18	21	6	7
18 – 20	6	2	6
المجموع	59	/	/



الشكل رقم:1.9

ملاحظة: الخط المنكسر يمثل المضلع التكراري ، حيث أن المساحة التي تقع تحت المضلع التكراري تساوي المساحة التي تقع تحت المدرج التكراري، وحتى نحافظ على المساحة التي تقع تحت هذا المضلع ، نفرض أن لهذا التوزيع فئتان إحداهما في بدايته و الأخرى في نهايته، تكرار كل منها يساوي الصفر ، بحيث ننطلق من مركز الفئة الإفتراضية الأولى وننتهي عند مركز الفئة الإفتراضية الأخيرة..

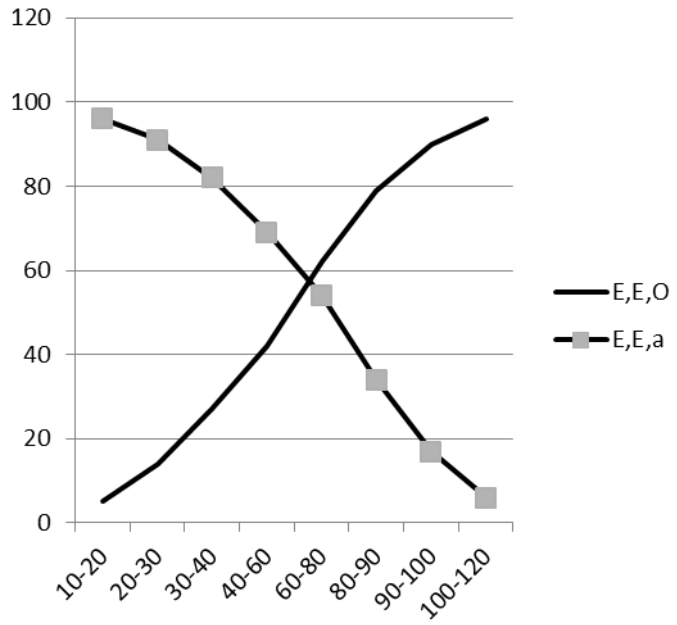
2- العرض البياني للتكرار التجميعي في حالة متغير مستمر:

- **العرض البياني للتكرار التجميعي الصاعد:** يسمى بالمنحنى التجميعي الصاعد، يرسم هذا المنحنى عن طريق إيصال مجموعة النقاط ذات الإحداثيات التالية: الحدود العليا للفئات والتكرارات التجميعية الصاعدة المقابلة لها.
- **العرض البياني للتكرار التجميعي النازل:** وهو عبارة عن المنحنى التجميعي النازل، يتم رسمه بإيصال مجموعة النقاط ذات الإحداثيات التالية : الحدود الدنيا للفئات و التكرارات التجميعية النازلة المقابلة لها.

مثال: ليكن التوزيع التكراري التالي:

الجدول رقم:1.13

الفئات	ni	ت .ت .ص ≤	ت .ت .ن
10 -20	5	5	96
20 – 30	9	14	91
30 – 40	13	27	82
40 – 60	15	42	69
60 -80	20	62	54
80 – 90	17	79	34
90 – 100	11	90	17
100 – 120	6	96	6
المجموع	96	/	/



الشكل رقم:1.10

ملاحظة: يبين كل من المنحنى التجميعي الصاعد والنازل شدة أو ضعف تطور الظاهرة المدروسة عند مستوى معين من مجال الدراسة ، إن فاصلة نقطة تقاطع المنحنى التجميعي الصاعد مع النازل يسمى " الوسيط " La Mediane، أما ترتيبها فهو $(\sum ni / 2)$.

تمارين محلولة للفصل الأول :

التمرين الأول : أجريت دراسة حول إستطلاع للرأي على 200 مستهلك لمعرفة رأيهم حول منتوج رياضي جديد عرض في السوق ، الإجابة المطلوبة من المستهلك هي " موافق أو غير موافق " . المطلوب ما يلي:

1- يتكون المجتمع من :

أ. كل المستهلكين.

ب. من المستهلكين الجزائريين.

ج. من المستهلكين المستجوبين.

د. من المجتمع الذي سحبت منه العينة.

2- ما هو المتغير المدروس ؟

الاجابة:

1- نلاحظ أن المجتمع المدروس يتكون من المستهلكين المستجوبين .

2- المتغير المدروس هو متغير كيفي ، لأنه غير قابل للقياس و يتمثل في عبارتين موافق أو غير موافق..

التمرين الثالث : يتكون مجتمع من 4 فئات إجتماعية – مهنية حجمه N = 5000 وحجم كل فئة هو كتالي: N1 = 1000 ، N2 = 1800 ، N3 = 1600 ، N4 = 600.

نريد سحب عينة حجمها n =180 : المطلوب ما يلي:

1- ماهي طبيعة المجتمع المدروس؟.

2- حدد عدد الوحدات الإحصائية التي يمكن سحبها من كل فئة؟

3- بإستعمال جدول الأعداد العشوائية ، حدد رقم الوحدات الإحصائية المسحوبة من كل فئة؟.

الاجابة:

1- المجتمع المدروس غير متجانس، لأنه يتكون من فئات إجتماعية متباينة.

2- عدد الوحدات الإحصائية التي يمكن سحبها من كل فئة هو كما يلي :

$n_1 = N_1 / N \times n = 1000/5000 \times 180 = 36$

$n_2 = N_2 / N \times n = 1800/5000 \times 180 = 65$

وبنفس الطريقة نجد أن : $n_3 = 58$ ، $n_4 = 22$

3- عند إستعمال جدول الأعداد العشوائية ، كل شخص يتحصل على أرقام تختلف عن الأرقام التي تحصل عليها شخص أخر اللهما إذا استعملا نفس الإختيار في إتجاه الأرقام.

المحاضرة الخامسة : خصائص النزعة المركزية

- ✓ مفهوم النزعة المركزية
- ✓ قياس النزعة المركزية
  - حساب الوسط الحسابي
  - خصائص الوسط الحسابي
  - حساب الوسط الهندسي
  - مجالات تطبيق الوسط الهندسي
  - حساب الوسط التوافقي

الهدف منها : كيفية حساب وتطبيق مقاييس النزعة المركزية في حالة توفر المعلومات الاحصائية

الفصل الثاني : خصائص النزعة المركزية (محمد صبحي ابو صالح ،2015، ص48- 58)

رأينا في الفصل السابق كيف يتم عرض البيانات الإحصائية جدوليا وبيانيا من أجل نقل وصف عام وسريع للظاهرة المدروسة ومن أجل وضع ترتيب معين وضروري لهذه المعلومات الإحصائية . غير أن لهذه الطريقة حدود من بينها :

1- لا يمكن إستخدامها في الأسلوب الشفهي.
2- لا يمكن إستخدامها لتحليل المعطيات .
3- لايمكن الإستفادة منها في مجال الإستقراء الإحصائي (التنبؤ و إتخاذ القرارات).

ولهذه الأسباب وضعت مقاييس عددية وصفية يمكن إستخدامها في مجالات عديدة منها التحليل والتنبؤ و إتخاذ القرار. ومن بين هذه المقاييس ، مقاييس النزعة المركزية أو مقاييس الوضع.

- **مفهوم النزعة المركزية:** تميل البيانات الإحصائية إلى التمركز حول قيمة معينة ، وكلما ابتعدنا عن هذه القيمة فإن عدد المعلومات يبدأ في التناقص ، نسمي هذه الظاهرة بالنزعة المركزية.

1- قياس النزعة المركزية : يعبر قياس النزعة المركزية عن مركز التوزيع الإحصائي ، ولقياس هذه النزعة نستعمل عدة مقاييس من بينها:

1-1- الوسط الحسابي: (Moyenne arithmétique) وهو عبارة عن مجموع القياسات الخاصة بظاهرة معينة على عدد هذه القياسات.

أ- **الوسط الحسابي البسيط:** (Moyenne airithmétique simple) يستعمل الوسط الحسابي البسيط في حالة بيانات غير مبوبة ، أي عندما يكون لقياسات المتغير المدروس نفس المستوى من الأهمية ، مثلا عندما يكون لديك 4 مواد لها نفس المعامل (نفس الأهمية) فإننا نستعمل الوسط الحسابي البسيط لحساب المعدل.

نفرض أن القياسات هي :Xn..... ; X1 ; X2 ; X3 ، حسب التعريف السابق فإن علاقة الوسط الحسابي هي: $\overline{X} = \frac{\sum Xi}{n}$

مثال : تبين السلسلة الإحصائية التالية أجور 8 عمال في مؤسسة ما ، المطلوب حساب الوسط الحسابي

700 ; 1800 ; 1300 ;1500 ; 900 ; 800 ; 1100 ; 1000

$$\overline{X} = \frac{700 + 1800 + 1300 + 1500 + 900 + 800 + 1100 + 1000}{8}$$

$$= 1137.8$$

أ- **العلاقة الثانية للوسط الحسابي:** تستعمل في هذه الحالة وسط فرضي (Moyenne peovisoire) لتحديد علاقة الوسط الحسابي حيث أن هذا الأخير، هو عبارة عن الوسط الفرضي مضافا إليه الوسط الحسابي للإنحرافات قيم المتغير الإحصائي عن الوسط الفرضي . علما أن هذه الإنحرافات هي (Xi – Xo) و Xo هو الوسط الفرضي ، إذن تعطى علاقة الوسط الحسابي بإستعمال وسط فرضي بالشكل التالي

$$\overline{X} = X0 + \sum (Xi - Xo)/n$$

مثال: نفس المثال السابق ، ونفرض أن Xo = 1000 ، فإن الوسط الحسابي هو :

$$\overline{X} = 1000 + \frac{(700 - 1000) + (1800 - 1000) + (1300 - 1000) + \cdots + (1000 - 1000)}{8} = 1137.5$$

ب- **الوسط الحسابي المرجح:** (Moyenne Arithmétique Pondérée)

تختلف أهمية قياسات المتغير الإحصائي من قيمة إلى أخرى في أغلب الأحيان ، و لهذا الغرض أدخل الترجيح في علاقة الوسط الحسابي ، والترجيح هو أهمية أو وزن قياس معين من قياسات المتغير الإحصائي . مثلا أمتحن طالب في 3 مواد : الإحصاء ، الرياضيات ، الإقتصاد الجزئي فكانت العلامات على التوالي 10 ،12 ،8 ومعاملاتها (أهميتها أو وزنها) هي على التوالي 4،3،3.

المطلوب تحديد الوسط الحسابي أو المعدل المتحصل عليه :

$$\overline{X} = \frac{12 \times 3 + +10 \times 3 + 8 \times 4}{3 + 3 + 4} = 9.8$$

ويمكن إستعمال الوسط الفرضي لحساب الوسط الحسابي ، نفرض أن Xo = 10:

$$\overline{X} = 10 + \frac{[(8 - 10) \times 4 + (10 - 10) \times 3 + (12 - 10) \times 3]}{10} = 9.8$$

ج- **الوسط الحسابي في حالة توزيع تكراري:** (distribution de fréquence)

نستعمل نفس علاقة الوسط الحسابي المرجح في حالة توزيع تكراري، غير أن الذي ينقصنا هي القيم النقطية للمتغير الإحصائي (Xi) هو عبارة عن مركز الفئة .

مثال : ليكن الجدول التكراري التالي :

الجدول رقم : 2.1

الفئات	ni	Xi	ni.Xi	Xi-Xo	ni(Xi-Xo)	$ni\frac{(Xi-Xo)}{K}$
0 -2	3	1	3	-6	-18	-9
2 - 4	6	3	18	-4	-24	-12
4 – 6	9	5	45	-2	-18	-9
6 – 8	10	7	70	0	0	0
8 – 10	8	9	72	2	16	8
10 – 12	7	11	77	4	28	14
12 – 14	2	13	26	6	12	6
المجموع	45		311		-4	-2

المطلوب : حساب الوسط الحسابي بإستعمال علاقة التعريف وعلاقة الوسط الفرضي ؟.

الإجابة:

1- نطبق علاقة التعريف : $\bar{X} = \frac{311}{45} = 6.91$

2- نطبق علاقة الوسط الحسابي بإستعمال وسط فرضي : نفرض أن Xo = 7 ، لماذا أخترنا القيمة 7 للوسط الفرضي ؟ لأنها تقع في منتصف قيم Xi ، وهذا لا يمنعنا من إختيار القيمة الوسطى للوسط الفرضي يسهل العمليات الحسابية.

$$\bar{X} = 7 + \frac{(-4)}{45} = 6.91$$

3- يمكن تطبيق علاقة ثالثة تسمى العلاقة المختصرة للوسط الحسابي:

نقسم الإنحرافات أو الفروق (Xi – X0) على طول الفئة k (في حالة طول الفئة ثابت) أ, على قاسم مشترك (في حالة طول الفئة غير ثابت).

$$\bar{X} = X0 + \frac{\sum\left(\frac{Xi-Xo}{k}\right).k}{\sum ni} \Rightarrow \bar{X} = 7 + \frac{(-2).2}{45} = 6.91$$

د- خصائص الوسط الحسابي:(Propriétés de la moyenne arithmétique) للوسط الحسابي عدة خصائص من بينها:

1- يتأثر الوسط الحسابي بالقيم المتطرفة أو الشاذة (القيم التي تقم في طرفي مجال الدراسة) :

مثال : تتكون أسرة من 5 أفراد ، تبلغ أعمارهم كما يلي : 55 ، 49 ، 13 ، 11 ، 10.

$$\bar{X}(t) = \frac{1}{30}(1500 \times x - \frac{X^2}{6} - 7.25 \times X^2) \quad / \quad {}^{30}_{1} \Rightarrow \bar{X}(t) = \frac{33975}{30} = 1132.5$$

2.1- الوسط الهندسي : (Moyenne géométrique) الوسط الهندسي لـ n قيمة من قيم المتغير الإحصائي هو عبارة عن الجذر النوني لجداء هذه القيم.

لتكن : Xn..... ; X1 ; X2 ; X3 القيم الممكنة للمتغير الإحصائي ، حيث أن صيغة الوسط الهندسي البسيط تعطى بالشكل التالي:

$$G = \sqrt[n]{X1.X2.X3 \dots\dots\dots\dots\dots\dots Xn}$$

مثال : لتكن السلسلة الإحصائية التالية: 9; 7; 16; 14; 10 ، المطلوب حساب الوسط الهندسي.

$$G = \sqrt[5]{10.14.16.7.9} = 10.71$$

لتسهيل العمليات الحسابية ، نقوم بإدخال اللوغاريتم على الصيغة السابقة ، لتصبح بالشكل التالي:

$$\log G = 1/n \log(X1.X2.X3 \dots\dots\dots\dots\dots\dots Xn)$$

$$\log G = 1/n(\log X1 + \log X2 + \log X3 + \cdots \dots\dots\dots\dots\dots \log Xn)$$

$$= 1/n \sum \log Xi$$

نقوم بتطبيق الصيغة الأخيرة على المثال السابق:

$$\log G = 1/5[\log 10 + \log 14 + \log 16 + \log 7 + \log 9] = 1.0299$$

إذن قيمة الوسط الهندسي هي : 10.71 = $G=10^{1.0299}$

مجالات تطبيق الوسط الهندسي:

تعتبر مجالات تطبيق الوسط الهندسي قليلة مقارنة بالوسط الحسابي، ومن أهم هذه المجالات هي كالتالي:

1- يستعمل في حساب الأرقام القياسية ، حيث يعتبر من أحسن المتوسطات في هذا المجال لأنه يحقق كل الخصائص الرياضية للأرقام القياسية.

2- يستعمل في حساب المعدلات (معدل الفائدة، معدل نمو السكان، ...إلخ)، مثال : نفرض أن 1000 د.ج وضعت في صندوق ا للادخار للاعب رياضي مدة 9 سنوات بمعدل فائدة وزع بالشكل التالي : 5% في 4 سنوات

الأولى ، 3% في السنتين التاليتين ، 4% في 3 سنوات الأخيرة ، المطلوب تحديد معدل سعر الفائدة المتوسط خلال 9 سنوات؟

الإجابة:

عند نهاية 9 سنوات سيصبح رأس المال ا الرياضي لأولى كالتالي، أو يحدد الرأس المال النهائي بالعلاقة التالية :

$$1000(1+r)^9 = 1000(1.05)^4(1.03)^2(1.04)^3$$

نقسم الطرفين على 1000: $(1+r)^9 = (1.05)^4(1.03)^2(1.04)^3$

ندخل اللوغاريتم على الطرفين: $9\log(1+r) = 4\log 1.05 + 2\log 1.03 + 3\log 1.04$

$$\log(1+r) = 0.0179$$

$$1 + r = 1.042 \Rightarrow r = 4.22\%$$

أ- **الوسط الهندسي المرجح:** (Moyenne géométrique pondérée)

تستعمل نفس الطريقة للحصول على علاقة الوسط الهندسي المرجح، مع الأخذ بعين اعتبار أهمية وترجيحات قيم المتغير الإحصائي المدروس ، وتعطى هذه العلاقة بالشكل التالي: $\log G = \frac{1}{\sum ni}.\sum ni.\log xi$ حيث أن **ni** هي ترجيحات أو أهمية قيم المتغير الإحصائي ، و xi هي تلك القيم.

3.1- الوسط التوافقي : (Moyenne Harmonique) يمكن تعريف الوسط التوافقي بالشكل التالي: مقلوب الوسط التوافقي هو عبارة عن الوسط الحسابي لمقلوب قيم المتغير الإحصائي.

أ- **الوسط التوافقي البسيط:** لتكن : Xn.....;X2 ;X1 قيم المتغير الإحصائي ، حيث نرمز للوسط التوافقي بالرمز H، وحسب تعريف هذا الأخير فإن:

$$\frac{1}{H} = \frac{1}{4}\sum\frac{1}{Xi} \Rightarrow H = \frac{n}{\sum\frac{1}{Xi}}$$

ب- **الوسط التوافقي المرجح:** (Moyenne Harmonique pondérée) تستعمل نفس الطريقة لوضع علاقة الوسط التوافقي المرجح ، مع إدخال أهمية وترجيحات قيم المتغير الإحصائي :: $\frac{1}{H} = \frac{\sum ni}{\sum\frac{ni}{Xi}}$

<u>ملاحظة:</u> إن مفهوم ومدلول الوسط التوافقي غير ظاهرة مقارنة مع المقاييس والمتوسطات الأخرى، والمثال التالي يبين ذلك :الأرقام التالية تبين أسعار مادة معينة خلال ثلاث فترات : 35 .65. 80 د .ج من السهل أن نقول أن متوسط السعر هو $\overline{X} = \frac{80+65+35}{3} = 60$ غير أن هذه النتيجة غير صحيحة لأن أهمية هذه السلعة أو المادة تختلف من فترة إلى فترة زمنية أخرى، فإذا كانت الميزانية المخصصة لشراء المادة تساوي 400 د.ج ، فيمكن أت نقتني 11.42 وحدة في الفترة الأولى و 6.15 وحدة في الفترة الثانية، وأخيرا 5 وحدات في الفترة الثالثة ، فعلى ضوء هذه المعلومات يصبح السعر المتوسط كالتالي

$$\overline{X} = \frac{35(11.42)+65(6.15)+80(5)}{11.42+6.15+5} = 53.13$$

نلاحظ أن النتيجة الأولى (بدون ترجيح) تختلف عن النتيجة الثانية (مع الترجيح) ومن هنا يتضح أنه لا يمكن إستعمال الوسط الحسابي في مثل هذه الحالات ، بينما عند إستعمال الصيغة البسيطة للوسط التوافقي سنلاحظ أن نتيجتها تساوي النتيجة الثانية لصيغة الوسط الحسابي، أي أن نتيجة الوسط التوافقي بدون ترجيح تساوي نتيجة الوسط الحسابي مع الترجيح، ماذا نستنتج من هذه الحالة؟ قبل أن نجيب على هذا التساؤل ، نحدد قيمة الوسط التوافقي البسيط: $H = \frac{3}{\frac{1}{35}+\frac{1}{65}+\frac{1}{80}} = 53.13$

نستنتج من هذه الحالة أن مفهوم الوسط التوافقي هنا، مرتبط بالقدرة الشرائية ، فكلما ارتفع السعر كلما انخفضت القدرة الشرائية والعكس صحيح.

عند استعمال صيغة الوسط الحسابي المرجح، سنجد نفس نتيجة الوسط التوافقي البسط.

إذن حسب ما سبق نضع النتيجة التالية :**نستعمل الوسط التوافقي في حالة وجود علاقة عكسية بين ظاهرتين.**

4.1- الوسط التربيعي: (Moyenne quadratique) *:* هو الجذر التربيعي لمتوسط مربعات قيم المتغير الإحصائي .

أ- **الوسط التربيعي البسيط:** $Q = \sqrt{\frac{\sum Xi^2}{n}}$

ب-**الوسط التربيعي المرجح:** $Q = \sqrt{\frac{\sum ni \ .Xi^2}{\sum ni}}$

ملاحظة: نستعمل أحد المتوسطات المذكورة سابقا حسب نوع العلاقة الموجودة بين الخصائص المدروسة ، وهذا وفق العلاقة العامة التالية : $f(x) = X^{\alpha}$، مثلا عندما تكون $\alpha = 1$ فالعلاقة تكون خطية بين الخصائص المدروسة ، ونستعمل في هذه الحالة الوسط الحسابي . وعندما تكون$\alpha = -1$ فالعلاقة الموجودة بين الخصائص تكون عكسية، ونستعمل الوسط التوافقي.

الجدول رقم :2.2

α	1-	0	1	2
نوع الوسيط	H	G	$\overline{X}$	Q

من الجدول السابق، نضع العلاقة التالية: $H < G < \overline{X} < Q$

مثال : ليكن التوزيع التكراري التالي، المطلوب تحديد كل من الوسط الحسابي ، الوسط الهندسي ، الوسط التوافقي، والوسط التربيعي ، ثم تحديد العلاقة بينهم:

الجدول رقم:2.3

الفئات	ni	Xi	Xi –xo	ni (xi-xo)	ni. Xi^2	$\frac{ni}{xi}$	ni log xi
0 -10	2	5	-25	-50	*50*	*0.4*	*1.39*
10 -15	5	12.5	-17.5	-87.5	781.25	0.4	5.48
15 -25	13	20	-10	-130	5200	0.65	16.91
25 -35	15	30	0	0	13500	0.5	22.15
35 -40	16	37.5	7.5	120	22500	0.42	25.18
40 -50	22	45	15	330	44550	0.48	36.37
50 -60	14	55	25	350	42350	0.25	24.36
60 -70	6	65	35	210	25350	0.09	10.87
$\sum ni$	93	/	/	742.5	154281.25	3.212	142.75

الحل:

1- تحديد الوسط الحسابي بإستعمال وسط فرضي : يتم إختيار الوسط الفرضي من قيم Xi (مراكز الفئات) ، ومن الأفضل أن تقع هذه القيمة في متصف قيم Xi ، حيث أن أحسن قيمة للوسط الفرضي في هذا المثال هي *X0 =30* (الجدول السابق يبين مختلف العمليات الحسابية) إذن :

$$\overline{X} = X0 + \frac{\sum ni(Xi - X0)}{\sum ni} \Rightarrow 30 + \frac{742.5}{93} = 37.98$$

2- **تحديد الوسط التربيعي:** $Q = \sqrt{\frac{\sum ni \,.Xi^2}{\sum ni}} = \sqrt{\frac{154281.25}{93}} = 40.73$

3- **تحديد الوسط التوافقي:** $\frac{1}{H} = \frac{\sum ni}{\sum \frac{ni}{Xi}} = \frac{93}{3.212} = 28.95$

4- **تحديد الوسط الهندسي:** $\log G = \frac{\sum .\log xi}{\sum ni}$

$$\Rightarrow G = 10^{1.5349} = 34.27$$

من النتائج المتحصل عليها نلاحظ أن : 40.73> 37.27> 34.27 > 28.95 إذن: $H < G < \overline{X} < Q$

المحاضرة السادسة : تابع مقاييس النزعة المركزي

- ✓ حساب الوسيط
  - في حالة البيانات غير مبوبة
  - في حالة البيانات مبوبة
  - خصائص الوسيط
- ✓ حساب الربيعيات
- ✓ حساب العشريات
- ✓ حساب الميئينيات
- ✓ حساب المنوال
- ✓ العلاقة بين الوسط الحسابي والوسيط والمنوال

الهدف منها : كيفية حساب وتطبيق مقاييس النزعة في حالة توفر المعلومات الاحصائية

4.1 لوسيط: (Médiane) هي تلك القيمة التي تقسم المجتمع الإحصائي إلى قسمين متساويين ، بحيث تكون قيم المتغير الإحصائي مرتبة ترتيبا تصاعديا أو تنازليا (جلاطو جيلالي ،2005،ص41).

مثال: إذا كان وسيط أجور عمال مؤسسة ما يساوي 1000 دج ، معنى ذلك أن 50% من العمال يتقاضون أجرا أقل من 10000 د.ج و 50% الآخرون لهم أجرا أكبر من 10000 د.ج.

أ- **الوسيط في حالة بيانات غير مبوبة :** يتم تحديد الوسيط في حالة بيانات غير مبوبة حسب الخطوات التالية:

1- نرتب قيم السلسلة الإحصائية ترتيبا تصاعديا أو تنازليا.

2- نحدد ترتيب الوسيط، وهنا لابد أن نميز بين حالتين:

- عندما يكون عدد القيم N فردي، في هذه الحالة ترتيب الوسيط هو عبارة عن : $\frac{N+1}{2}$.
- أما عندما يكون عدد القيم N زوجي ، فلا توجد قيمة وسيطيه ، وفي هذه الحالة فإن قيمة الوسيط هي عبارة عن الوسط الحسابي للقيمتين ذات الترتيب $\frac{N}{2}$ و $\frac{N}{2}+1$ على التوالي.

3- نضع قيمة الوسيط ، ونرمز له بالرمز Me .

مثال: في حالة N فردي: تبين السلسلة الإحصائية التالية علامات 9 طلبة في مقياس الإحصاء ، المطلوب تحديد الوسيط ، 13 ; 14 ; 7 ; 17 ; 15 ; 9 ; 8 ; 11 ; 10.

أولا: نرتب العلامات ترتيبا تصاعديا : 17 ; 15 , 14 ; 13 ; 11 ; 11 ; 10 ; 9 ; 8 ; 7.

ثانيا : نحدد ترتيب الوسيط $\frac{9+1}{2}=5$: مرتبة الوسيط هي المرتبة الخامسة إذن : Me = 11.

مثال : في حالة N زوجي : تبين السلسلة الإحصائية التالية علامات 10 طلبة في مقياس الرياضيات ، 6 ; 7 ; 9 ; 10 ;11 ; 12 ; 12 ; 13 ; 15 ; 16.

أولا : مرتبة القيم التي تدخل في حساب الوسيط: $\frac{n}{2}+1=\frac{10}{2}+1=6$ ، $\frac{n}{2}=\frac{10}{2}=5$.

ثانيا: تحديد قيمة الوسيط : $\frac{11+12}{2}=11.5$ = Me.

ب- **الوسيط في حالة بيانات مبوبة أو في حالة توزيع تكراري :**

تختلف طريقة حساب الوسيط في هذه الحالة عن الحالة الأولى وهذا حسب الخطوات التالية:

1- تحديد التكرار التجميعي الصاعد أو النازل.

2- تحديد ترتيب الوسيط وهو عبارة عن نصف مجموع التكرارت أي $\frac{\sum ni}{2}$.

3- تحديد فئة الوسيط أي الفئة التي يقع فيها الوسيط ، وهي الفئة التي تقابل التكرار التجميعي الصاعد (النازل) الذي يساوي ترتيب الوسيط أو الأكبر (الأصغر) منه مباشرة.

4- تحديد وحساب الوسيط بتطبيق العلاقة الإحصائية للوسيط.

مثال: يبين التوزيع التكراري التالي توزيع 80 فريق رياضي حسب قيمة عائداتها :

الجدول رقم:2.4

الفئات	Ni	ت.ت.ص
90 – 100	5	5
100 – 110	9	14
110 – 120	16	30
120 – 130	25	55
130 -140	13	68
140 – 150	7	75
150 – 160	3	78
160 – 170	2	80
$\sum$ni	80	/

المطلوب : تحديد الوسيط؟

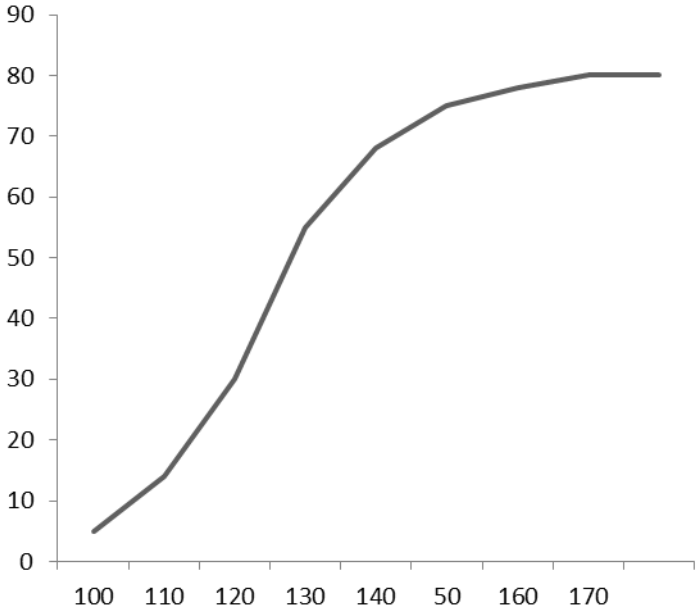
الحل:

1- تحديد التكرار التجميعي الصاعد في العمود الثالث من الجدول .

2- ترتيب الوسيط : $\frac{\sum ni}{2} = \frac{80}{2} = 40$.

3- تحديد الفئة الوسيطية هي الفئة التي تقابل التكرار التجميعي الصاعد الذي يساوي 40 أو أكبر من 40 مباشرة ، نلاحظ أن العدد 40 غير موجود من بين ت.ت.ص ولكن العدد الأكبر منه مباشرة هو 55 إذن الفئة التي تقابل ت.ت.ص 55 هي الفئة الوسيطية [120-130].

4- رسم المنحنى التجميعي الصاعد لتحديد العلاقة الإحصائية للوسيط بإستعمال طريقة ميل معادلة مستقيم ، ولتحقيق ذلك، نفرض أن المنحنى التجميعي الصاعد هو عبارة عن خط مستقيم في المجال أو الفئة [120- 130]، وبالتالي فإن الميل يكون ثابتا ، ثم نستخرج علاقة الوسيط من النسب المتساوية التالية :



الشكل رقم :2.1

مادام أن هذه المقادير متساوية ، نختار مقدارين يحتويان على مجهول واحد ألا وهو الوسيط:

من المقدارين من الأول والثاني نستخرج قيمة الوسيط: $\frac{130-120}{55-30} = \frac{130-Me}{55-30} = \frac{Me-120}{55-30}$

إنطلاقا من هذه العلاقة نضع $Me = 120 + \frac{(40-30)(130-120)}{55-30} = 120 + \frac{(40-30)10}{25} = 124$

الشكل النهائي للعلاقة الإحصائية للوسيط:

$$M_e = A + \frac{\left(\frac{\sum ni}{2} - M_{n-1}\right).k}{E_{SMe}}$$

حيث أن : A: الحد الأدنى للفئة الوسيطية ، K طول الفئة الوسيطية ، M_{n-1} التكرار التجميعي الصاعد السابق للفئة الوسيطية ، E_{Sme} التكرار الأصلي للفئة الوسيطية .

<u>ملاحظة:</u> إن قيمة الوسيط هي قيمة تقريبية وتكون كذلك عند توفر الشروط التالية:

1- يجب ألا يكون طول الفئة كبيرا بشكل ملحوظ.

2- يجب أن يكون حجم المجتمع أو العينة هاما.

3- من الأفضل أن يكون التوزيع التكراري المدروس أقرب إلى التوزيع المتناظر.

ج- **خصائص الوسيط:**

1- يعتبر الوسيط كلما غيرنا أطوال الفئات بالنسبة لنفس التوزيع التكراري إذن يتميز الوسيط بعدم الثبات.

2- لا يتأثر الوسيط بالقيم المتطرفة أو الشاذة.

3- يقسم المدرج التكراري إلى مساحتين متساويتين.

4- يستعمل الوسيط في عدة مجالات منها على وجه الخصوص :دراسة الاجور و الأسعار ، في نظرية أخطاء القياس ، في دراسة الوفيات ، في دراسة المدة المتوسطة للحياة...إلخ.

5.الربيعيات: Les Quartiles : نستعمل نفس الطريقة المتبعة لإيجاد الوسيط، غير أن الذي يتغير هو الترتيب وما يترتب عنه (نجيب حسن محمد ،2004،ص143).

تنقسم الربيعيات إلى ثلاثة أقسام : الربيعي الأول ، الربيعي الثاني (الوسيط) ، الربيعي الثالث.

أ- **الربيعي الأول:**(Premier quantile) هي قيمة المتغير الإحصائي التي تقسم المجتمع الإحصائي إلى قسمين، يحتوي القسم الأول على 25% من الوحدات الإحصائية، أما القسم الثاني فيحتوي على 75% من هذه الوحدات . تقع قيمة الربيعي الأول في نهاية الربع الأول من التوزيع الإحصائي ، وفي هذا الموضع تكون مرتبته تساوي 25% أو $\frac{\sum ni}{4}$ حسب الترتيب التصاعدي لقيم المتغير الإحصائي ، ويرمز له بالرمز Q_1.

العلاقة الإحصائية للربيعي الأول: $$Q_1 = A + \frac{\left(\frac{\sum ni}{4} - M_{n-1}\right).K_{Q1}}{E_{SQ1}}$$

خصائص الربيعي الأول: له نفس خصائص الوسيط.

مثال : نفس المثال السابق ، مرتبة الربيعي الأول هي : $\frac{\sum ni}{4} = \frac{80}{4} = 20$ ، فئته هي الفئة التي تقابل المرتبة 20 أو أكبر منها مباشرة ضمن ت.ت.ص وهي : [120 – 110]، أما قيمته ، فتعطى بالعلاقة السابقة :

$$Q_1 = 110 + \frac{(20-14).10}{16} = 113.75$$

ب-الربيعي الثالث:(Troisième quantile) هو عبارة عن قيمة المتغير الإحصائي التي تقسم المجتمع إلى قيمين ، يحتوي القسم الأول على 75% من المجتمع المدروس ، أما القسم الثاني فيحتوي على 25 % المتبقية منه حسب الترتيب التصاعدي للمتغير المدروس، ويرمز له بالرمز Q_3 ، ومرتبته 75% أو $\frac{3\sum ni}{4}$.

علاقته الإحصائية: هي نفس علاقة الوسيط، طبعا مع تغيير في الفئة والتكرار التجميعي الصاعد السابق وكذلك التكرار الأصلي لفئته Q_3.

$$Q_1 = A + \frac{\left(\frac{3\sum ni}{4} - M_{n-1}\right).K_{Q1}}{E_{SQ3}}$$

خصائصه: له نفس خصائص الوسيط.

مثال :نفس المثال السابق الخاص بالوسيط:

1- ترتيب Q_3: $\frac{3\sum ni}{4} = \frac{3(80)}{4} = 60$.

2- فئة Q_3 [130 – 140].

3- قيمة Q_3: $Q_3 = \frac{[60-55]}{13}.10 = 133.84$.

6.**العشريات:**(les Déciles): يوجد على طول مجال الدراسة 9 عشريات هي : D1 ; D2 ; D3 ; D4 ;.....D9.D10، مرتبة كل منها هي على التوالي:100%..... . 10% . 20% . 30% أو $\frac{\sum ni}{10} ; \frac{\sum ni}{5} ; \dots\dots\dots\dots \sum ni$

العلاقة الإحصائية العامة للعشريات : نستعمل نفس فكرة الوسيط هذه العلاقة وهي كالتالي :

$$D_1 = A + \frac{\left(i\frac{\sum ni}{10} - M_{n-1} \right).k}{E_{SDI}}$$

مثال : ليكن التوزيع التكراري التالي :

الجدول رقم:2.5

الفئات	Ni	ت.ت.ص <	Xi	xi-xo	ni(xi-xo)
0 -4	3	3	2	-12	-36
4 -8	6	9	6	-8	-48
8 -10	7	16	9	-5	-35
10 -12	9	25	11	-3	-27
12 -16	16	41	14	0	0
16 -20	13	54	18	4	52
20 -26	8	62	23	9	72
26 -30	5	67	28	14	70
30 – 40	3	70	35	21	63
∑ni	70	/	/	/	111

المطلوب ما يلي:

1- حدد الوسيط الحسابي بطريقة الوسط الفرضي؟

2- حدد قيمة المتغير الإحصائي التي تقسم المجتمع الإحصائي إلى قسمين متساويين؟

3- حدد الربيعي الأول ز الثالث؟

4- حدد الشعيري الرابع والسادس؟

الحل:

1- تحديد الوسط الحسابي بواسطة الوسط الفرضي : إنطلاقا من العمليات الحسابية الموجودة في الجدول السابق، نقوم بتحديد الوسط الحسابي .

$$\bar{X} = X0 + \frac{\sum ni(Xi - X0)}{\sum ni} \Rightarrow 14 + \frac{111}{70} = 15.58$$

2- قيمة المتغير الإحصائي التي تقسم المجتمع إلى قسمين متساويين هي الوسيط:

$M_e = A + \frac{\left(\frac{\sum ni}{2} - M_{n-1}\right).k}{E_{SMe}}$ ، حيث أن ترتيب Me هو: $\frac{\sum ni}{2} = \frac{70}{2} = 35$ ، إذن فئة Me هي : [12-16] ، أما قيمة Me فهي : $M_e = 12 + \frac{(35-25).4}{16} = 14.5$.

3- تحديد الربيعي الأول والربيعي الثالث:

حساب Q_1 : ترتيب Q_1 هو $\frac{\sum ni}{4} = \frac{70}{4} = 17.5$ ، و فئة Q_1

5-المئينيات : (les Percentiles): يمكن تعميم طريقة حساب الوسيط والعشريات إلى بقية مراتب قيم المتغير الإحصائي، حيث يمكن تحديد وحساب قيمة المتغير الإحصائي إذا عرفت مرتبته، ومن بين هذه القيم المتينات ، ومراتبها على التوالي :1%; 2% ; 3% ;.....100%، أما العلاقة العامة لها فهي (نجيب حسن محمد ،2004،ص153) :

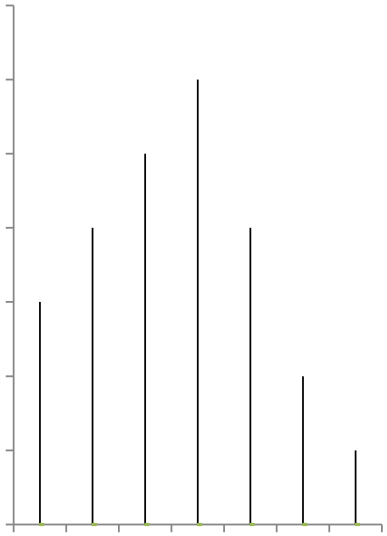
حيث أن i هة مرتبة المتغير الإحصائي.

7. المنوال : (le Mode): هو قيمة المتغير الإحصائي الأكثر إنتشارا أو تكرارا، ويرمز له بالرمز Mo.

الحالات الممكنة لقيم المنوال: الأشكال التالية تبين مختلف الحالات الممكنة لقيم المنوال (محمد بوعلاق ،2009،ص45):

الحالة الأولى: توزيع تكراري وحيد المنوال، أي التوزيع الذي يحتوي على قيمة شائعة واحدة.

ni

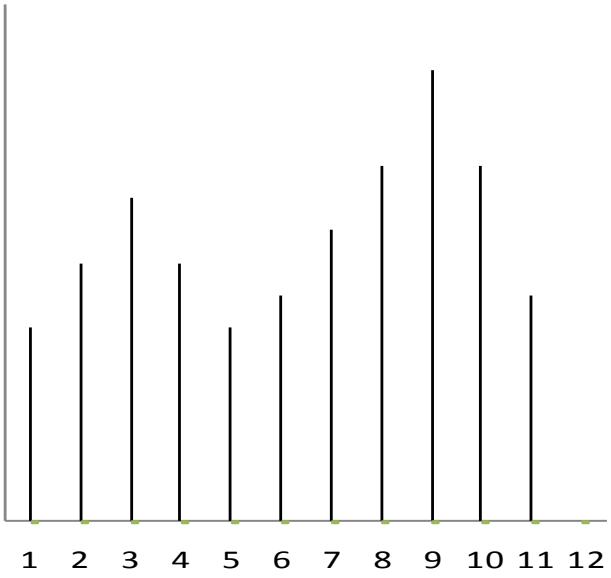


Xi

الشكل رقم:2.2

<u>الحالة الثانية</u>: توزيع تكراري ثنائي المنوال، أي التوزيع الذي يحتوي على قيمتين شائعتين (أكثر تكرارا).

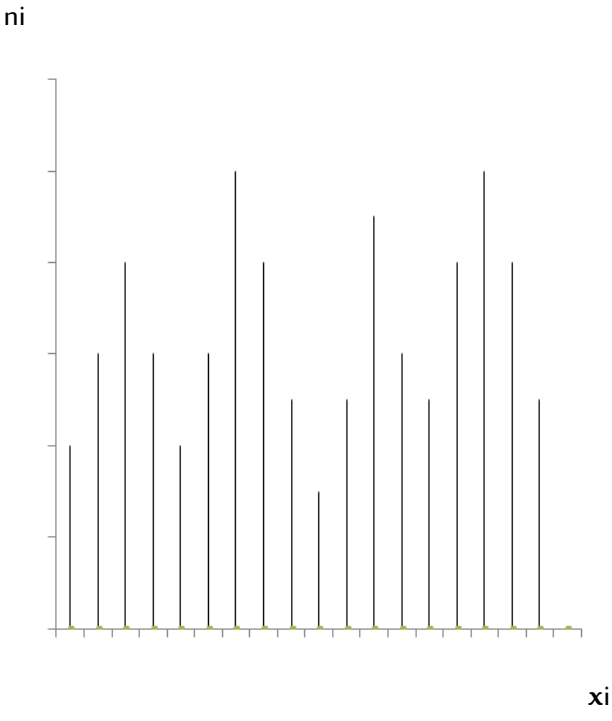
ni



xi

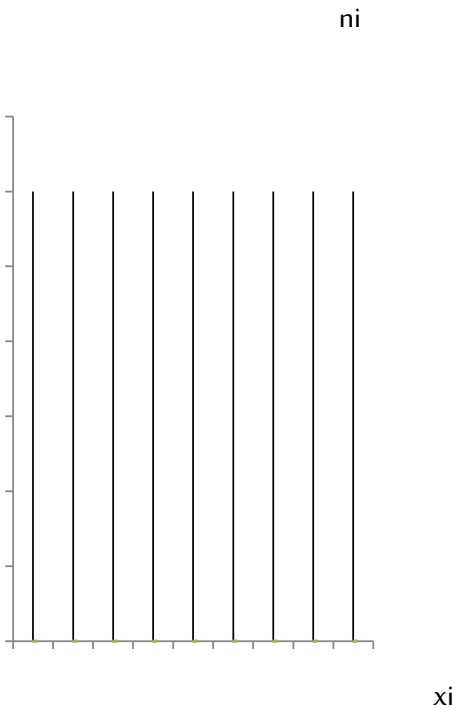
الشكل رقم:2.3

الحالة الثالثة: توزيع تكراري متعدد المنوال، أي التوزيع الذي يضم على أكثر من منوالين، في هذه الحالة لا يكون للمنوال أي مدلول إحصائي ، ولا يستعمل كمقياس من مقاييس النزعة المركزية إلا في حالة إستثنائية مثلا: عند تحديد مختلف الأغلبيات والتكتلات في البرلمنات المنتخبة.



الشكل رقم: 2.4

الحالة الرابعة: توزيع تكراري عديم المنوال، في هذه الحالة كل قيم المتغير الإحصائي لها نفس المستوى من الأهمية أي نفس التكرار، وهذه الحالة لا تحدث في الواقع إلا نادرا.



الشكل رقم: 2.5

1- **المنوال في حالة بيانات غير مبوبة:**

مثال: السلسلة الإحصائية التالية تبين علامات 10 طلبة في الإقتصاد الجزئي 6; 16 ; 15 ; 10 ; 9 ; 11; 10 ; 9 ; 10 ; 7

نلاحظ أن القيمة الأكثر تكرارا و إنتشارا هي القيمة 10 إذن : Mo=10، وهي سلسلة وحيدة المنوال.

2- **المنوال في حالة بيانات مبوبة:**

إن طريقة حساب المنوال في هذه الحالة تختلف عن الحالة الأولى ، ولحسابه نتبع الخطوات التالية :

أ- تحديد الفئة المنوالية : وهي التي تقابل أكبر تكرارا عندما يكون طول الفئة ثابت (K= consrante)، أو الفئة التي تقابل أكبر تكرارا معدل عندما تكون أطوال الفئات غير متساوية.

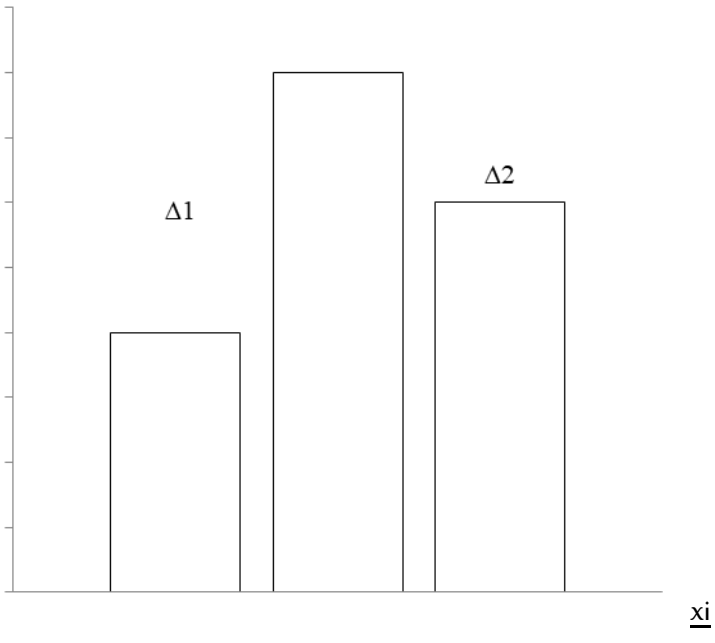
ب- وضع المدرج التكراري قصد تحديد علاقة المنوال ، والعلاقة الأكثر إنتشارا هي علاقة الفروق أو علاقة (Pearson).

مثال : يبين الجدول التالي توزيع مداخيل 50 عاملا في مؤسسة ما : المطلوب تحديد المنوال؟

نحدد علاقة المنوال حسب طريقة (Pearson) إنطلاقا من المدرج التكراري التالي:

الجدول رقم :2.6

الفئات	30- 40	40- 50	50 – 60	60-70	70-80	80 -90	90 -100	∑ni
ni	4	6	8	12	9	7	4	50



الشكل رقم:2.6

ملاحظة : أكتفينا بالفئة الثالثة والخامسة لتحديد المنوال ، لأن هذا الأخير يعتمد أساسا على الفئة المنوالية والفئة السابقة واللاحقة لها.

- تحديد الفئة المنوالية: وهي الفئة التي تقابل التكرار 12 (أكبر تكرار) لأن أطول الفئات متساوية، وهذه الفئة هي [60-70] ، نفرض أن Δ_1 هو الفرق بين تكرار الفئة المنوالية وتكرار السابقة لها، Δ_2 هو الفرق بين تكرار الفئة المنوالية وتكرار الفئة اللاحقة لها، X هو الفرق بين قيمة المنوال والحد الأدنى للفئة المنوالية ، K طول الفئة المنوالية . من الشكل السابق ، نلاحظ أن المقادرين التاليين متساويين:

علما أن Mo = A+X نعوض بقيمة X حسب العلاقة السابقة :

ويمكن تعميم هذه العلاقة بالشكل التالي :

3- **كيفية إختيار خاصية من خصائص النزعة المركزية:** يرتبط إختيار مقياس من مقاييس النزعة المركزية بطبيعة وهدف الدراسة وبنوع العلاقة الموجودة بين الخصائص المدروسة، فالأمثلة التالية تبين بعض هذه الحالات (جلاطو جيلالي ،2005،ص54):

أ- حالة استراد مادة إستهلاكية أساسية : في هذه الحالة نبحث الوسط الحسابي لتحديد الكمية المستوردة، فهذه الكمية هي عبارة عن الوسط الحسابي في حجم المجتمع $N, \bar{X}$.

ب- يتطلع المهتمون إلى معرفة الكتلة الرلمانية ذات الأغلبية المطلقة أو النسبية عند دراسة نتائج الإنتخابات البرلمانية وذلك بإستعمال المنوال أي القيمة الأكثر إنتشارا، أو عند إنتاج نوع من الملابس والأحذية ، فالمنتج يقوم بإنتاج القياسات الأكثر إستعمالا (المنوال).

ج- نبحث عن الأجر أو الدخل الذي يقسم المجتمع إلى قسمين متساويين أو غير متساويين ، عند دراسة الأجور والمداخيل (الوسيط أو الربيعيات).

د- يستعمل الوسيط كمقياس من مقاييس النزعة المركزية في حالة التوزيعات التكرارية المفتوحة لأنه لا يتأثر بالقيم المتطرفة ... إلخ.

4- **العلاقة بين الوسط الحسابي ، الوسيط، المنوال:** يقع الوسيط في كل الحالات بين الوسط الحسابي والمنوال، وذلك حسب الحالات التالية:

<u>الحالة الأولى:</u> تكون قيم المقاييس الثلاثة متساوية $\bar{X} = Me = Mo$ ، في هذه الحالة يكون التوزيع التكراري المدروس متماثل أو متناظر.

<u>الحالة الثانية :</u> عندما يكون التوزيع التكراري المدروس غير متناظر من اليمين تكون قيم المقاييس الثلاثة بالشكل التالي: $\bar{X} > Me > Mo$.

<u>الحالة الثالثة :</u> عندما يكون التوزيع التكراري غير متناظر من اليسار، فتصبح قيم المقاييس الثلاثة بالشكل التالي: $\bar{X} < Me < Mo$.

تمارين محلولة خاصة بالفصل الثاني :

التمرين الأول: يبين الجدول التكراري التالي توزيع أجور الرياضيين في فريقين مختلفتين :

الجدول رقم: 2.7

الفئات	Pi(1)	Pi (2)	*Xi*	Pi .Xi (1)	Pi .Xi (2)	*ت.ن.ت..ص*
18 -28	*0.042*	*0.0279*	*23*	*0.966*	*0.64*	*0.0279*
28-36	0.57	0.335	32	11.42	10.72	0.3629
36-48	0.378	0.223	42	15.87	9.36	0.5859
48-50	0.126	0.167	49	6.174	8.18	0.7525
50-66	0.04	0.1117	58	2.32	6.47	0.8646
66-78	0.029	0.078	72	2.088	5.61	0.9426
78-88	0.0168	0.033	83	1.394	2.739	0.9756
88-98	0.0084	0.0223	93	0.781	2.07	1
$\sum$Pi	1	1		41.023		

المطلوب:

1- حدد الوسط الحسابي بإستعمال علاقة التعريف في الفريقين ، ماذا تلاحظ؟

2- حدد كل من الوسيط، العشير الثالث (*D3*)، والمئيني الأربعين في الفريق الثاني ؟

التمرين الثاني: تبين السلسلة الإحصائية التالية مراكز الفئات لتوزيع تكراري:

3.42 ; 3.39 ; 3.36 ; 3.33 ; 3.27 ; 3.24 ; 3.21 .

المطلوب تحديد ما يلي :

1- طول الفئة ؟ 2-المدى العام؟ 3- حدود الفئة؟

التمرين الثالث: يبين التوزيع التكراري التالي توزيع النفقات الاستهلاكية الشهرية (بآلاف الدينارات) لعينة من الرياضيين سحبت بطريقة عشوائية من مجتمع معين:

الجدول رقم:2.8

الفئات	ni	Xi	≤	Er
8 -10	4	9	4	2
10 -12	n5	11	26-n5	3
12 -15	12	13.5	38-n5	4
15 – 18	15	16.5	53-n5	5
18 -20	n5	19	53	8
20 – 24	28	22	81	7
24 – B	24	26	105	6
B – 30	10	29	115	5
30 – 33	12	31.5	127	4
33 – 36	9	34.5	136	3
36 – 40	4	38	140	1
/	140	/	/	/

المطلوب ما يلي:

1- أوجد قيم $n2$،$n5$ علما أن العشير الثالث $D3 = 18.625$ ؟

2- أوجد الحد الأدنى للفئة السابعة علما أن الوسط الحسابي هو $\bar{} = 22.757$ ؟

3- ما هو العرض البياني المناسب؟

4- حدد الإنفاق الإستهلاكي للرياضيين ذات المرتبة 75% ؟

5- أدرس شكل هذا التوزيع بإستعمال المقاييس الثلاثة للنزعة المركزية؟

التمرين الرابع : قام مصححان بتصحيح ، كل منهما ، مجموعة من أوراق إمتحان إحدى المسابقات فكانت النتائج كالتالي:

الجدول رقم: 2.9

الفئات	n1i	n2i
0 – 3	15	20
3 – 8	35	25
8 -10	45	40
10 -12	28	35
12 -14	20	18
14- 16	12	3
16 -20	4	2
$\sum$ni	159	143

المطلوب : تحديد ما يلي :

1- العرض البياني المناسب؟

2- الوسط الحسابي بطريقة الوسط الفرضي للتوزيعين؟

3- المعدل العام للطلبة؟

4- العلامة المتحصل عليها من طرف أغلبية الطلبة لكل توزيع؟

5- شكل التوزيعين بإستعمال الخصائص الثلاثة للنزعة المركزية ؟

6- علامة الطالب ذو المرتبة 20 %؟

7- إذا لوحظ وجود فرق بين علامات المصححين ، كيف يتم تقليص هذا الفرق؟

التمرين الخامس : يبين التوزيع التكراري التالي تطور عدد الطلبة في السنة الثالثة ليسانس لمعهد التربية البدنية والرياضية حسب التخصصات التالية خلال 10 سنوات :

الجدول رقم: 2.10

التخصصات	85/86	95/96
كرة السلة	300	450
كرة اليد	160	300
كرة القدم	70	100
السباحة	40	90
الجمباز	50	80
الالعاب القوى	40	60
$\sum ni$	660	1080

المطلوب : ما يلي :

1- ضع العرض البياني المناسب؟

2- حدد المعدل السنوي المتوسط لتطور عدد الطلبة في كل فرع؟

3- حدد الوسط الحسابي لتطور عدد الطلبة في كل فرع؟

4- حدد الوسط الحسابي المرجح للمعدلات السنوية المتوسطة (نأخذ بعين الاعتبار ترجيحات سنة 85/86)؟

5- حدد المعدل السنوي المتوسط لمجموع الطلبة في الفترتين؟

6- ماذا تلاحظ من نتيجة السؤال الرابع و الخامس؟

التمرين السادس : بين صحة العلاقة التالية ، علما أن :

$$\bar{X} = Me = Mo \qquad 2\sum ni.Xi = 2A\sum ni + k.\sum ni$$

حل تمارين الفصل الثاني :

التمرين الأول:

1- تحديد الوسط الحسابي :

أ- في الفريق الاول : $\overline{X}_1 = \frac{\sum ni.xi}{\sum ni} = \sum Pi.xi = 41.123$ نفرض أن $\sum ni = \overline{\quad}$ لأن التكرارات المعطاة هي تكرارات نسبية .

ب- في الفريق الثاني : $\overline{X}_2 = \frac{\sum ni.xi}{\sum ni} = \sum Pi.xi = 45.81$ نلاحظ أن متوسط الأجر في الفريق الثاني أكبر من متوسط الأجر في الفريق الأولى رغم وجود نفس مجال الدراسة أي نفس المدى العام.

2- تحديد الوسيط والعشير الثالث ولمئيني الأربعين في الفريق الثاني :

أ- انطلاقا من العمود السابع لجدول المعطيات فإن الفئة الوسيطية هي [48 – 36] لأن ترتيب الوسيط هنا هو 50% أننا في حالة تكرارات نسبية، إذن قيمة الوسيط هي $M_e = 36 + \frac{[50-36.29]12}{22.3} = 43.377$.

ب- حساب العشير الثالث (D3) : مرتبته هي 30 %، والفئة التي ينتمي إليها هي [28-36] إذن قيمته هي $D3 = 28 + \frac{[30-2.79].8}{33.5} = 34.4979$.

ج- حساب المئين الأربعين (C40) : مرتبته هي 40 % وفئته هي [36-48] إذن تعطى قيمته بالشكل التالي $C40 = 36 + \frac{[40-36.29].12}{22.3} = 37.996$.

التمرين الثاني:

أ- تحديد طول الفئة: نلاحظ أن الفرق بين مراكز الفئات ثابت، معنى ذلك أ، طول الفئات متساو وثابت إذن طول الفئة K= 0.03 .

ب- تحديد المدى العام : أولا نحدد الحد الأدنى للمجال، الحد الأدنى = مركز الفئة الأولى – نصف الفئة أي أن A=3.21-0.03/2=3.195، بنفس الطريقة نحدد الحد الأعلى للمجال، ونرمز له بالرمز B حيث B=3.42+0.03/2=3.435، إذن المدى العام هو 3.435-3.195=0.24 ، أو أن المدى العام = طول الفئة x عدد الفئات = 0.03 x8 = 0.24.

ج- لتحديد حدود الفئات نضيف طول الفئة إلى الحد الأدنى للمجال للحصول على الفئة الأولى، ثم نضيف طول الفئة للحد الأعلى للفئة الأولى للحصول على الفئة الثانية وهكذا حتى تنتهي العملية، إذن الفئات على التوالي هي: [3.195-3.225] ؛ [3.225-3.255] ؛ [3.225-3.285] ؛ [3.285-3.315] ؛ [3.315-3.345]؛ [3.3753.345]؛ [3.405-3.375] ؛ [3.405-3.435]

التمرين الثالث :

1- تحديد قيم كل من n5.n2 لتحديد هذه القيم نضع المعادلتين التاليتين

$$\sum ni = 4 + n2 + 12 + 15 + n5 + 28 + 24 + 10 + 12 + 9 + 4 = 140$$

$$\Rightarrow n2 + n5 + 118 = 140$$

n2=22-n2.....(1)

ولدينا كذلك D_3= 18.62 لوضع العلامة كاملة ، نبحث أولا عن ترتيب D_3، ثم نحدد فئته:

الترتيب هو : 42 = 0.3x140 ، أما فئته فهي [18-20] حسب التكرار التجميعي الصاعد ، بالتعويض في علاقة D_3 نتحصل على مايلي: $18.625 = 18 + \frac{(42-(31+n_2)).2}{n_5}$ ، أي أن المعادلة الثانية تعطى بالشكل التالي :

n2= 11-0.32=125 n5.....(2)

إذن بناءاه على المعادلتين الأولى والثانية فإن قيم كل من n2 ; n5 هما كالتالي : n2=6 ; n5= 16.

2- تحديد الحد الأعلى للفئة السابعة علما أن الوسط الحسابي يساوي : $\bar{X}$= 22.757 ، بالتعويض في المعادلة نتحصل على ما يلي: $\sum ni\ 22.757 = \sum Xi.ni \Rightarrow 2.757\ .140 = \sum Xi.ni$

$$3185.98 = 36 + 66 + 162 + 247.5 + 304 + 616 + 24\frac{(24+B)}{2} + 10\frac{(B+30)}{2} + 378 + 310.5 + 152$$

$$3185.98 = 2272 + 12(24+B) + 5(B+30) \Rightarrow B = \frac{476}{17} = 28$$

3- العرض البياني المناسب هو المدرج التكراري بعد تعديل التكرارات لأن أطوال الفئات غير متساوية.

4- تحديد قيمة الإنفاق الإستهلاكي الذي تبلغ مرتبته 75%: إن قيمة المتغير الإحصائي الذي تبلغ مرتبته 75% هو الربيعي الثالث، ولحسابه، نحدد أولا التكرار التجميعي الصاعد ، ونلاحظ من جدول المعطيات أن فئة الربيعي الثالث هي: [24-28] وحدها الأعلى يساوي مرتبته Q3 وفي هذه الحالة نستنتج مباشرة أن قيمة الإنفاق الإستهلاكي تساوي الحد الأعلى أي القيمة Q3=28 ، وعند تطبيق علاقة Q3 سنجد نفس القيمة ونفس النتيجة.

5- دراسة شكل التوزيع باستعمال Mo ; Me ; $\bar{X}$: أولا نقوم بتحديد قيم هذه المقاييس ثم نقارن بينها علما أن الوسط الحسابي معطى ، أما القيم الأخرى فهي كالتالي:

أ- حساب Mo : الفئة المنوالية هي الفئة التي تقابل أكبر تكرار معدل، ومن الجدول نلاحظ أن هذه الفئة هي [18-20] أما المنوال فيساوي : $M_0 = 18 + \frac{(8-5)\times 2}{(8-5)+(8-7)}$=19.5.

ب- حساب Me : ترتيبه هو 70 = 0.5×140، وفئته هي [20-24] أما قيمته فهي :

$M_e = 20 + \frac{(70-53)\times 4}{28}$=22.428.

حسب قيم الخصائص الثلاثة للنزعة المركزية ، نستنتج أن التوزيع المدروس غير متناظر من اليسار لأن $\bar{X} > M_e > M_o$.

التمرين الرابع : يحتوي الجدول التالي الحسابات الضرورية للإجابة على الأسئلة المطروحة:

الجدول رقم: 2.12

E_{i1r}	E_{i2r}	Xi	$N_{i1}(x_i-x_0)$	$n_{i2}(x_i-x_0)$	ت.ت.ص	ت.ت.ن
5	6.66	1.5	-142.5	-190	15	20
7	5	5.5	-192.5	-137.5	50	45
22.5	20	9	-90	-80	95	85
14	17.5	11	0	0	123	120
10	9	13	40	36	143	138
6	1.5	15	48	12	155	141
1	.5	18	28	14	159	143
/	/	/	-309	-345.5	/	/

1- العرض البياني المناسب في كلا الحالتين هو المدرج التكراري ، وذلك بعد تعديل التكرارات.

2- تحديد الوسط الحسابي بطريقة الوسط الفرضي: الوسط الفرضي في كلا الحالتين: $X_0 = 11$.

أ- الوسط الحسابي للتوزيع الأول :

$$\bar{X} = \frac{\sum n_{i1}(X_i - X_0)}{\sum n_{i1}} \Rightarrow \bar{X} = 11 + \frac{-309}{159} = 9.056$$

ب- الوسط الحسابي للتوزيع الثاني : $\bar{X} = 11 + \frac{-345.5}{143} = 8.5839$

3- حساب المعدل العام: $\bar{X} = \frac{9.056 \times 159 + 8.5839 \times 143}{159+143} = 8.832$

4- تحديد العلامة المتحصل عليها من طرف أغلبية الطلبة :

أ- بالنسبة للتوزيع الأول :

الفئة المنوالية هي الفئة التي تقابل التكرار 22.5 (أكبر تكرار معدل) وهي [8-10]، مع العلم أننا نستعمل التكرار المعدل لحساب المنوال، وقيمته كالتالي: $M_e = 8 + \frac{(22.5-7)\times 2}{(22.5-7)+(22.5-14)} = 9.29$

ب- بالنسبة للتوزيع الأول :

الفئة المنوالية هي [8-10]، أما قيمة المنوال فهي: $M_o = 8 + \frac{(20-5)\times 2}{(20-5)+(20-17.5)} = 9.714$

5- تحديد شكل التوزيع بإستعمال المقاييس الثلاثة للنزعة المركزية:

أ- حساب هذه المقاييس : لقد تم حساب كل من الوسط الحسابي والمنوال للتوزيعين، ونقوم بحساب الوسيط:

1- وسيط التوزيع الأول : مرتبته هي 159/2 =79.5 أما فئته فهي [8-10] إذن قيمة الوسيط هي كالتالي :

$$M_e = 8 + \frac{(79.5-50)\times 2}{45} = 9.31$$

2- وسيط التوزيع الثاني : مرتبته هي 143/2 =71.5 أما فئته فهي [8-10] إذن قيمته هي :

$$M_e = 8 + \frac{(71.5-45)\times 2}{40} = 9.325$$

إن شكل التوزيع الأول حسب قيم الوسط الحسابي والوسيط والمنوال هو كالتالي : $\bar{X} > M_e > M_o$ معنى ذلك أن التوزيع غير متماثل من اليمين.

أما شكل التوزيع الثاني فهو كذلك غير متماثل من اليمين لأن $M_o > M_e > \bar{X}$.

6- تحديد علامة الطالب ذو المرتبة 20% : نكتفي بحساب هذه العلامة بالنسبة للتوزيع الأول ، فمرتبته هي 159×0.2=31.8 ، وحسب هذه المرتبة فإن فئة العشير الثاني هي [3-8] أما قيمته فهي :

$$D_2 = 3 + \frac{[31.5-15].5}{35} = 2.40$$

7- نلاحظ أن هناك فرق بين المعدلين ، ولتوحيد علامات التوزيعين نستعمل النسبة بين المعدل العام والوسط الحسابي:

أ- بالنسبة للتوزيع الأول: 0.975 = 8.832/9.056 ، لتوحيد علامات التوزيع الأول يجب ضرب علامات المصحح الأول × 0.975.

ب- بالنسبة للتوزيع الثاني : 1.029 =8.583 / 8.832 ، يجب ضرب علامات المصحح الثاني ×1.029، حتة نتحصل عل علامات منسجمة.

التمرين الخامس :

1- العرض البياني المناسب هو الأعمدة المتلاصقة ، لأن توزيع الطلبة في السنة الثالثة يتم حسب الفروع وحسب الزمن.

للإجابة على بقية الأسئلة نضع الجدول التالي:

الجدول رقم:2.13

الفروع	85/86	95/96	المعدل خلال 11 سنة	معدل سنوي متوسط	متوسط سنوي Xi	$n_i.X_i$
كرة السلة	300	450	0.5	0.037	0.0454	13.62
كرة اليد	160	300	0.875	0.058	0.0795	12.72
كرة القدم	40	90	1.25	0.076	0.1136	4.54
السباحة	70	100	0.4281	0.0329	0.0389	2.723
الجمباز	50	80	0.6	0.0436	0.0454	2.725
الالعاب القوى	40	60	0.5	0.037	0.0454	1.816
$\sum ni$	660	1080	/	0.0457	0.0578	38.148

2- تحديد المعدل السنوي المتوسط لكل فرع:

أ- فرع العلوم المالية: $(1+I)^{11} = 1.5 \Rightarrow i = 1.5^{0.09} - 1 \Rightarrow i = 3.754\%$

ونتبع نفس الطريقة بالنسبة للفروع الأخرى ، كما هو مبين في الجدول السابق.

3- تحديد الوسط الحسابي لتطور الطلبة في كل الفروع:

بالنسبة لفرع العلوم المالية: $\bar{X}$= 0.5/11=0.0454.

ونفس الشيء يمكن فعله بالنسبة لبقية الفروع، كما هو مبين في العمود السادس من الجدول السابق.

4- تحديد الوسط الحسابي المرجح (نأخذ بعين الاعتبار ترجيحات 85/86) :

$$\bar{X} = \frac{\sum ni.Xi}{\sum ni} = \frac{38.148}{660} = 5.78\%$$

5- تحديد المعدل السنوي العام المتوسط :

$$(1+I)^{11} = 1080/660 = 1.636 \Rightarrow i = 4.57\%$$

6- نلاحظ من نتيجة السؤالين 4 و 5 أنهما مختلفين إختلاف كبيرا، لأننا في (4) اعتمدنا على ترجيحات سنة 85/86، بينما (5) يبين تطور الظاهرة في كل سنة ، أي لابد أن نفرق ونميز بين نتيجة الوسط الحسابي والوسط الهندسي.

التمرين السادس :

نبين صحة العلاقة $\bar{X} = M_e = M_o$ علما أن $2\sum ni.Xi = 2A\sum ni + k.\sum ni$ أي أن التوزيع متناظر ،

نلاحظ من صيغة العلاقة السابقة أن الأمر يتعلق بالوسط الحسابي والمنوال، أي أن :

$\frac{\sum n_i \times X_i}{\sum n_i} = A + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times K$ ، نضرب الطرفين في ∑ni، حيث أن : $\Delta_1 = \Delta_2$ لأن التوزيع متناظر ، ووفق هذه المعلومات تصبح العلاقة بالشكل التالي: $\sum ni \times Xi = A\sum ni + 1/2K\sum ni$ ثم نضرب الطرفين في 2 لنتحصل على المطلوب.

المحاضرة السادسة : مقاييس التشتت

- ✓ المدى العام
- ✓ المدى الربيعي
  - نصف المدى الربيعي
  - علاقة بين المدى الربيعي والمدى العام
- ✓ الانحراف المتوسط
- ✓ التباين
- ✓ الانحراف المعياري
- ✓ معامل الاختلاف

الهدف منها : تعرف على مقاييس التشتت وكيفية تطبيقها

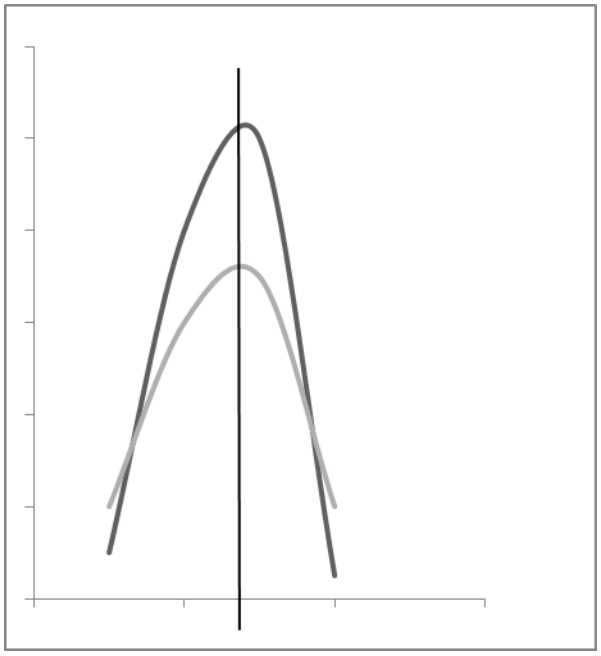
الفصل الثالث : مقاييس التشتت (محسن لطفي احمد، 2011، ص131- 145)

تمهيد : تظهر مقاييس النزعة المركزية القيمة المركزية للتوزيع الإحصائي دون أن تبين كيفية توزيع وانتشار قيم المتغير الإحصائي حول هذه القيمة. نلاحظ مثلا أن التوزيعين التاليين لهما نفس القيمة المركزية ، غير أنهما يختلفان اختلافا كبيرا من حيث انتشار وتوزيع قيمهما على مجال الدراسة:

التوزيع الأول : 68،69،70،70،70،71،71،72.

التوزيع الثاني: 30،50 ،110،90،90،70،70،70،50.

الخصائص الثلاثة للنزعة المركزية متساوية في التوزيعين $\bar{} = M_e = M_o$ ، غير أن طول مجال الدراسة في التوزيعين (المدى العام) يختلف من التوزيع الأول إلى التوزيع الثاني. كما أنه يمكن أن نبين هذه الظاهرة بيانيا:



الشكل رقم:2.7

إن مجال التوزيع الثاني أكبر من مجال التوزيع الأول، معنى ذلك أن التوزيع الثاني أكثر تشتتا من التوزيع الأول بالنسبة للقيمة المركزية. ولقياس هذا التشتت ، وحتى تستكمل دراسة التوزيع الإحصائي، نتطرق في هذا الفصل إلى مقاييس عددية تبين لنا كيفية توزيع وانتشار قيم المتغير الإحصائي حول القيمة المركزية، وسنكتفي بالمقاييس الأكثر إستعمالا. كما سنتطرق إلى مقاييس أخرى، تبين لنا تناظر أو عدم تناظر التوزيع الإحصائي بالنسبة للقيمة المركزية (مقاييس الشكل).

مقاييس التشتت : سنتطرق إلى مقاييس التشتت حسب درجة أهميتها، فمن الأقل أهمية إلى الأكثر أهمية.

1- **المدى العام:**(Etendue ou Range) هو الفرق بين أكبر وأصغر قيمة للتوزيع الإحصائي أو هو عبارة عن الفرق بين الحد الأعلى للفئة الأخيرة والحد الأدنى للفئة الأولى. وحسب تعريف هذا المقياس، فإنه لا يأخذ بعين اعتبار إلا القيميتين المتطرفتين (الأولى والأخيرة) ، وبالتالي فإنه يتأثر بالقيم المتطرفة. يتميز المدى العام بأنه يستعمل في المقارنة بين توزيعين إحصائيين أو أكثر وأنه يضم كل الوحدات الإحصائية التي تشكل المجتمع الإحصائي.

مثال: تبين السلستان الإحصائيتان التاليتان توزيع مجموعتين من العمال في مؤسستين مختلفتين حسب أجورهم اليومية:

التوزيع الأول:

الجدول رقم:3.1

رقم العامل	1	2	3	4	5
الأجر اليومي	120	135	140	150	165

التوزيع الثاني:

الجدول رقم:3.2

رقم العامل	1	2	3	4	5
الأجر اليومي	80	100	140	180	200

نلاحظ أن الوسط الحسابي للتوزيعين متساوي $\overline{X_1} = \overline{X_2} = 140$ فهذا لا يعني أنهما متشابهان ، سنرى ذلك من خلال المدى العام ونرمز له بالرمز E: حيث أن : $E_1 = 165 - 120 = 45$ أما $E_2 = 200 - 80 = 120$

وبالتالي فإن $E_1 < {}_2$ ، إذن التوزيع الثاني أكثر تشتتا من التوزيع الأول ، معنى ذلك أن أجور المجموعة الثانية أكثر تباعدا فيما بينها مقارنة مع المجموعة الأولى، ونستنتج هنا أن المجموعة الأولى أكثر تجانسا وأكثر عدالة في توزيع الأجور من المجموعة الثانية.

2- **المدى الربيعي:**(Intervalle interquartile) وهو الفرق بين الربيعي الثالث والربيعي الأول (Q3-Q1) ، ويرمز له بالرمزIQ . ويتميز المدى الربيعي بالخصائص التالية :

أ- يضم 50% من المجتمع مهما كان التوزيع الإحصائي.

ب- يتغير طوله مقارنة بالمدى العام حسب طبيعة التوزيع.

ج- استعمالاته محدودة نظرا لبساطته، غير أنه أحسن من المدى العام.

د- يستعمل في المقارنة بين توزيعين إحصائيين أو أكثر.

3- **نصف المدى الربيعي:**(Le Semi-interquartile) وهو عبارة عن $\frac{IQ}{2} = \frac{Q_3 - Q_1}{2}$ ، إن الذي يهمنا في هذا المقياس هو مدلوله الإحصائي ن حيث نبينه من خلال المثال:

مثال: السلسلة الإحصائية التالية تبين مداخيل 11 وحدة إحصائية ، المطلوب حساب المدى الربيعي:

700،800،900،900،1000،1100،1100،1200،1300،1300،1400.

أولا: نحدد قيمة Q_1، ولحساب هذه القيمة ، نحدد مرتبتها، نلاحظ أن عدد القيم فردي إذن المرتبة هي : $\frac{11+1}{4} = 3$. وبعد ترتيب القيم ترتيبا تصاعديا نجد أن $Q_1 = 900$. وبنفس الطريقة نجد أن $Q_3 = 300$ ومن ثمة فإن $\frac{Q}{2} = \frac{1300-900}{2}$. السؤال المطروح الآن : ما هو مدلول هذه القيمة؟ يمكن أن نقول أن 50% من المداخيل تبعد في المتوسط عن الوسيط بأقل من 200.

4- **النسبة بين المدى الربيعي والمدى العام:** يبين هذا المقياس تشتت 50% من الوحدات الإحصائية المركزية (أي التي تقع في مركز التوزيع) حول الوسيط مقارنة بالمدى العام. وتكتب علاقة هذا المقياس بالشكل التالي: $R = \frac{Q_3 - Q_1}{E} \times 100$، ومن مميزاته ما يلي:

أ- يستعمل لقياس 50% من الوحدات الإحصائية التي تقع حول القيمة المركزية بالنسبة لنفس التوزيع.

ب- إذا كان R=50% يكون التوزيع الإحصائي متماثل أو متناظر، أما إذا كان هذا المقياس أكبر من 50% فإن التوزيع يكون قوي التشتت بالنسبة للقيمة المركزية ، و إذا كان عكس ذلك فإن التوزيع يكون قليل التشتت بالنسبة للقيمة المركزية (الوسيط).

مثال: نفس المثال السابق ، ماهي درجة تشتت مداخيل 11 وحدة إحصائية ؟

$R = \frac{1300-900}{1400-700} \times 100 = 57.14\%$، نلاحظ أن المدى الربيعي يمثل 57.14% من المدى العام ويحتوي على 50% من الوحدات الإحصائية إذن يوجد تشتت نسبي و أقرب إلى التماثل.

5- **الإنحراف المتوسط :** (Ecart moyen):هو البعد المتوسط لقيم المتغير الإحصائي عن قيمة مركزية (الوسط الحسابي) ، و الإنحراف هو الفرق بين قيم المتغير الإحصائي والقيمة المركزية ($\bar{X}; M_O; M_e$) ويمكن كتابته بالشكل التالي $X1 - \bar{X}; X2 - \bar{X}; X3 - \bar{X}; \dots\dots\dots Xi - \bar{X}$

والانحراف المتوسط هو عبارة عن الوسط الحساب لهذه الفروق، وحسب خاصية من خصائص الوسط الحسابي فإن مجموع هذه الفروق يساوي الصفر، وحتى نتفادى هذا الإشكال ، نستعمل هذه الفروق بالقيم المطلقة ، إذن تكتب علاقة الإنحراف المتوسط بالشكل التالي: $E_{\bar{X}} = \frac{\sum ni|Xi - \bar{X}|}{\sum ni}$ وهي العلاقة البسيطة، أما العلاقة المرجحة أو في حالة توزيع تكراري فهي: $E_{\bar{X}} = \frac{\sum ni|Xi - \bar{X}|}{\sum ni}$

المثال الأول: لتكن السلسلة الإحصائية التالية:

13; 11; 10; 9; 8; 7; 5; 4; 2; 1، المطلوب تحديد الإنحراف المتوسط؟

أولا يتم تحديد الوسيط الحسابي: $\bar{X} = \frac{70}{10} = 7$ ثم نحدد الفروق:

$\sum|X_i - \bar{X}| = 6+5+3+2+0+1+2+3+4+6 = 32$، إذن: $E_{\bar{X}} = \frac{32}{10} = 3.2$.

<u>المثال الثاني:</u> ليكن التوزيع التكراري الآتي، المطلوب تحديد الإنحراف المتوسط؟

الجدول رقم: 3.3

الفئات	ni	Xi	$(Xi - Xo)$	$ni(Xi - Xo)$	$\|Xi - \bar{X}\|$	$ni\|Xi - \bar{X}\|$
89.5 - 99.5	5	94.5	-30	-150	29.4	147
99.5 - 109.5	9	104.5	-20	-80	19.4	174.6
109.5 - 119.5	16	114.5	-10	-160	9.4	150.4
119.5 - 129.5	25	124.5	0	0	0.6	15
129.5 - 139.5	13	134.5	10	130	10.6	137.8
139.5 - 149.5	7	144.5	20	140	20.6	144.2
149.5 - 159.5	3	154.5	30	90	30.6	91.8
159.5 - 169.5	2	164.5	40	80	40.6	81.2
$\sum ni$	80	/	/	-50	/	942

الحل:

1- تحديد الوسط الحسابي: نفرض أن Xo=124.5.

$$\bar{X} = X_0 + \frac{\sum ni(X_1 - X_0)}{\sum ni}$$

$$\bar{X} = 124.5 + \frac{(-50)}{80} = 12.87$$

2- حساب الإنحراف المتوسط: $E_{\bar{X}} = \frac{\sum ni|Xi - \bar{X}|}{\sum ni} = \frac{942}{80} = 11.8$.

ملاحظة: يعتبر الإنحراف المتوسط أحسن من سابقيه (المدى العام، الدى الربيعي) ، غير أنه لا يستعمل بشكل واسع نظرا لوجود القيمة المطلقة.

6- **التباين و الإنحراف المعياري:** (la variance et l'ecart-type)

أ-**التباين:** هو عبارة عن الوسط الحسابي لمربعات الفروق بين قيم المتغير الإحصائي و الوسط الحسابي. ويرمز له بالرمز V(X)، ومن خلال هذا التعريف يمكن أن نستخرج علاقتين:

1- **علاقة التعريف:**

- في حالة بيانات غير مبوبة: $V(X) = \frac{\sum(X_i - \bar{X})^2}{N}$.
- في حالة توزيع تكراري: $V(X) = \frac{\sum ni(X_i - \bar{X})^2}{\sum ni}$.

2- **العلاقة الثانية للتباين:** إنطلاقا من علاقة التعريف، نتحصل على العلاقة التالية:

$$V(X) = \frac{\sum(X_i - \bar{X})^2}{n} = \frac{\sum({X_i}^2 - 2X_i.\bar{X} + \bar{X}^2)}{n}$$

$$= \frac{\sum {X_i}^2}{n} - 2\bar{X}.\frac{\sum X_i}{n} + \frac{\sum \bar{X}^2}{n} = \frac{\sum {X_i}^2}{n} - 2X_i.\bar{X} + \bar{X}^2$$

- في حالة توزيع بيانات غير مبوبة : $V(X) = \frac{\sum {X_i}^2}{N} - \bar{X}^2$
- في حالة توزيع تكراري : $V(X) = \frac{\sum ni{X_i}^2}{\sum ni} - \bar{X}^2$

ملاحظة : من الأفضل إستعمال العلاقة الثانية للتباين لسهولة العمليات الحسابية، وهي عبارة عن الفرق بين مربع الوسط التربيعي ومربع الوسط الحسابي.

ب- **الإنحراف المعياري :** وهو الجذر التربيعي للتباين ، ويرمز له بالرمز SD.

- في حالة بيانات غير مبوبة:

$$SD = \sqrt{\frac{\sum {X_i}^2}{n} - \bar{X}^2} = \sqrt{\frac{\sum(X_i - \bar{X})^2}{n}}$$

- في حالة توزيع تكراري

$$SD = \sqrt{\frac{\sum ni(X_i - \bar{X})^2}{\sum n_i}} = \sqrt{\frac{\sum ni.{X_i}^2}{\sum n_i} - \bar{X}^2}$$

7- **خصائص الإنحراف المعياري:** (Peopriétés de lécart-type)

أ- يعتبر الإنحراف المعياري من أهم مقاييس التشتت لأنه يستعمل في حساب عدة مؤشرات نذكر منها على وجه الخصوص: معامل الإرتباط ، في تحديد أشكال التوزيعات الإحصائية والاحتمالية ...إلخ.

ويستعمل في تحديد نسب عدد الوحدات الإحصائية بالنسبة لتوزيع إحصائي معين بكون قريب إلى التماثل أو التناظر حسب الحالات التالية:

المجال الأول يحتوي على 50% من المجتمع $[\bar{X} \pm 0.67.SD]$

المجال الثاني يحتوي على 68% من المجتمع $[\bar{X} \pm SD]$

المجال الثالث يحتوي على 95% من المجتمع $[\bar{X} \pm 2.SD]$

المجال الرابع يحتوي على 99% من المجتمع $[\bar{X} \pm 3.SD]$

ب- يتأثر الإنحراف المعياري بالقيم المتطرفة كما هو الحال بالنسبة للوسط الحسابي .

ج- الإنحراف المعياري لقيمة ثابتة يساوي الصفر $V(a)=0$.

د- $V(aX) = a^2.V(X)$.

هـ- $V(a+X) = V(a) + V(X) = V(X)$.

و- العلاقة بين الإنحراف المعياري ، الإنحراف المتوسط والمدى الربيعي :

لتحديد هذه العلاقة نفرض أن التوزيع الإحصائي المدروس يكون قريب إلى التماثل أو التناظر بالنسبة للقيمة المركزية.

8- **العلاقة بين الإنحراف المعياري والمدى الربيعي:** نستخرج العلاقة بين هاذين المقياسيين إنطلاقا من المجالين التاليين $\bar{X} \pm 0.67.SD$ [Q1; Q2] الذي يضم كل منهما على 50% من الوحدات الإحصائية المركزية، وبالتالي فإن الحد الأدنى للمجال الأول يساوي الحد الأدنى للمجال الثاني، الحد الأعلى للمجال الأول يساوي الحد الأعلى للمجال الثاني؟، إذن $Q1 = \bar{X} - 0.67.SD$ و $Q3 = \bar{X} + 0.67.SD$، و من هاتين العلاقتين نستنتج أن Q3-Q1=1.34.SD.

9- **العلاقة بين الإنحراف المعياري والانحراف المتوسط:** تعتبر هذه العلاقة تجريبية فكلما كان التوزيع الإحصائي أقرب إلى التناظر والتماثل كلما كانت العلاقة بين المقاييس صحيحة $E_{\bar{X}} = \frac{4}{5}SD$.

10- **معامل الاختلاف:** (Coéfficie,t de variation) هو عبارة النسبة بين الإنحراف المعياري والوسط الحسابي ، ويعتبر من مقاييس التشتت النسبية. يستعمل خاصة في المقارنة بين توزيعات إحصائية غير متجانسة (وحدات قياس مختلفة)، ويرمز لهذا المقياس بالرمز CV، ونكتب علاقته بالشكل التالي: $CV = \frac{SD}{X}$.

المحاضرة السابعة : مقاييس الشكل

- ✓ العزوم البسيطة
- ✓ العزوم المركزية
- ✓ العلاقة بين العزوم المركزية والبسيطة
- ✓ تحديد شكل التوزيع
- ✓ مقاييس تحديد شكل التوزيع الاحصائي
  - معامل فيشر
  - معامل بارسون
  - معامل بول وكاندال

الهدف منها : التعرف على مقاييس الشكل وكيفية التعامل معها وفق المعطيات المتوفرة

مقاييس الشكل (محمد صبحي ابو صالح،2007، ص135- 145):

إضافة إلى مقاييس الوضع ومقاييس التشتت ، سنتطرق إلى مقاييس تبين شكل التوزيع الإحصائي (الإلتواء، التطاول و التقاطع) مقارنة بتوزيع مرجعي (التوزيع المتناظر بالنسبة للإلتواء والتوزيع الطبيعي بالنسبة للتطاول والتفلطح)، تسمى هذه المقاييس مقاييس الشكل. تعتمد هذه الأخيرة على العزوم البسيطة والعزوم المركزية.

1- **العزوم البسيطة:** (Les Moments simples)

لتكن $X^k{}_1 ; X^k{}_2 ; X^k{}_3 \dots\dots\dots\dots\dots\dots\dots\dots X^k{}_n$ ، العزم البسيط من المرتبة K هو عبارة عن الوسط الحسابي لقيم المتغير الإحصائي مرفوعة إلى القوة K.

ويكتب بالشكل التالي:

- في حالة بيانات غير مبوبة: $m_k = \frac{\sum {X_i}^k}{n}$
- في حالة بيانات مبوبة أو توزيع تكراري: $m_k = \frac{\sum ni.\ {X_i}^k}{\sum ni}$

إن مراتب العزوم البسيطة تتراوح من 0 إلى K ، وهذه بعض الحالات الخاصة:

$$m_0 = 1; m_1 = \bar{X}. m_2 = Q^2. m_3 = \frac{\sum {X_i}^3}{n} \dots\dots\dots\dots$$

2- **العزوم المركزية:** (Les Moments centrés)

تكتب العلاقة العامة للعزوم بالشكل التالي: $m_k = \frac{\sum ni\ .(X_i - X_0)^2}{\sum ni}$

عندما $X_0 = 0$ فإننا نكون في حالة عزوم بسيطة، أما عندما $X_0 = \bar{X}$ فهذه الحالة تنطبق على العزوم المركزية ، ويرمز لها بالرمز μ :

- في حالة بيانات غير مبوبة: $\mu_k = \frac{\sum (X_i - \bar{X})^k}{n}$
- في حالة بيانات مبوبة أو توزيع تكراري: $\mu_k = \frac{\sum .(X_i - \bar{X})^k}{\sum ni}$

تتراوح مراتب العزوم المركزية كم 0 إلى K ، وهذه بعض الحالات الخاصة: $\mu_0 = 1; \mu_1 = 0; \mu_2 = v(0)$

3- **العلاقة بين العزوم المركزية و البسيطة:**

لإيجاد العلاقة بين العزوم المركزية والعزوم البسيطة نقوم بنشر ثنائي نيوتن للمقدار $\mu_k = \frac{1}{N}\sum (X_i - \bar{X})^K$

$$\mu_k = \sum \{(-1)^k . C_n^k . X_i^{n-k} . \bar{X}^k\}$$

بالعزوم المركزية الخمسة الأولى بدلالة العزوم البسيطة:

$$k = 0 \Rightarrow \mu_0 = 1$$

$$k = 1 \Rightarrow \mu_1 = 0$$

$$k = 2 \Rightarrow \mu_2 = \frac{\sum X_i^2}{N} - \bar{X}^2 = m_2 - m_1^2$$

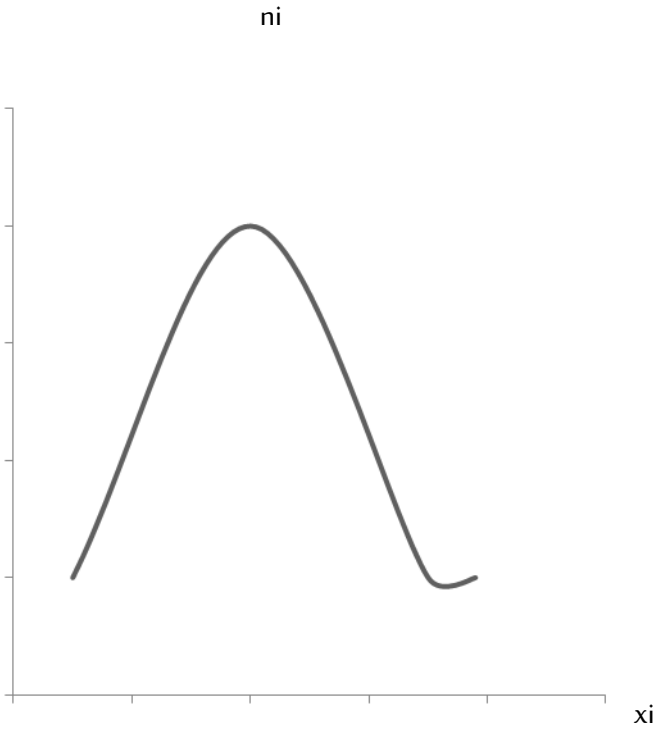
$$k = 3 \Rightarrow \mu_3 = \frac{\sum (X_i - \bar{X})^3}{N} = m_3 - 3.m_1.m_2 + 2.m_1^2$$

$$k = 4 \Rightarrow \mu_4 = m_4 - 4.m_1.m_3 + 6.m_1^2.m_2 - 3m_1^4$$

4- **تحديد شكل التوزيع:** يحدد شكل التوزيع التكراري بالنسبة للقيمة المركزية (الوسط الحسابي أو الوسيط)، ويمكن أن نميز بين نوعين من أشكال التوزيعات الإحصائية:

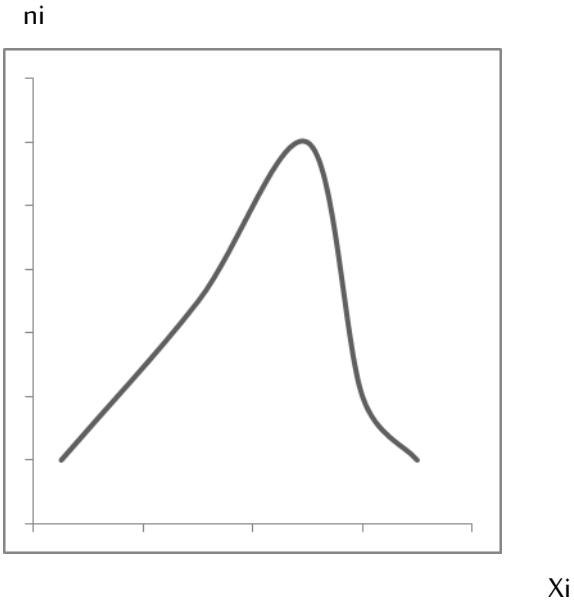
النوع الأول: يخص هذا النوع من الأشكال التوزيعات الإحصائية غير المتناظرة من اليسار أو من اليمين (dissymétrie à gauche ou à droite) مقارنة بتوزيع متناظر بالنسبة للقيمة المركزية:

أ- عدم تناظر من اليسار:



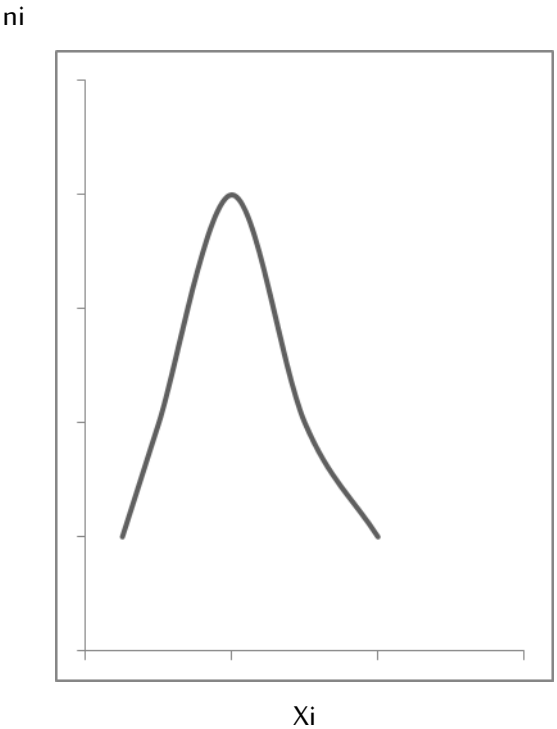
الشكل رقم:3.2

ب- عدم تناظر من اليمين:



الشكل رقم:3.3

ج- توزيع متناظر

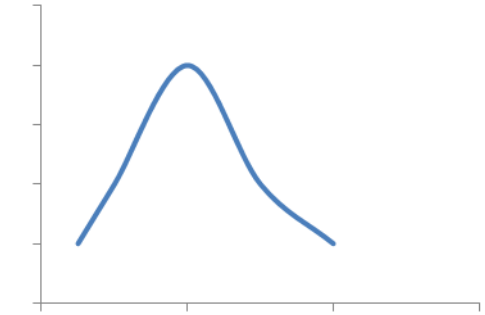


الشكل رقم:3.4

النوع الثاني: يخص هذا النوع من الأشكال التوزيعات الإحصائية المتناظرة والتي تكون في نفس الوقت متطاولة (Leptocurtique) أو متفلطحة (platicurtique) مقارنة بالشكل الجرسي أو بما يسمى بالتوزيع الطبيعي (mésocurtique)، وذلك حسب الأشكال التالية :

أ- توزيع متطاول (ذو تشتت ضعيف):

ni

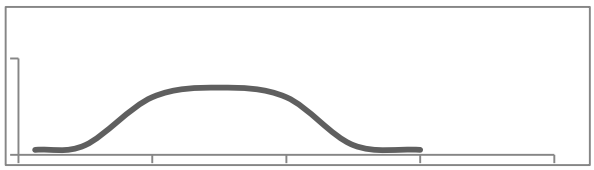


. xi

الشكل رقم:3.5

ب- توزيع متفلطح (ذو تشتت قوي):

ni



xi

الشكل رقم:3.6

أن الفرق الأساسي بين هذه التوزيعات يتمثل في درجة قوة التشتت، ففي التوزيع المتقاطع يكون التشتت قوي بالنسبة للقيمة المركزية (الوسط الحسابي)، أما بالنسبة للتوزيع المتطاول فالتشتت يكون ضعيفا بالنسبة للوسط الحسابي.

5- **مقاييس تحديد شكل التوزيع الإحصائي :**

أ- **النوع الأول:**

1- **معامل فيشر:** (Ronald Aylmer Fisher (1890-1962), England)

يقيس هذا المعامل درجة إلواء شكل التوزيع الإحصائي، ونعتمد في ذلك على قيمة العزم المركزي من المرتبة الثالثة، ولاستبعاد وحدة القياس، نقسمه على الإنحراف المعياري من نفس المرتبة.

يرمز لمعامل فير للنوع الأول بالرمز F1: $F_1 = \frac{\mu_3}{SD^3}$

لماذا نعتمد على العزم المركزي من المرتبة الثالث؟

لأن قيمته في حالة توزيع متناظر تساوي إلى الصفر ، إذن بناءه على هذه القيمة، نضع الحالات الممكنة التالية:

الحالة الأولى : توزيع إحصائي متناظر $F_1 = 0 \Rightarrow$.

الحالة الثانية : عدم تناظر من اليمين $F_1 > 0 \Rightarrow$ ، قيم Xi التي تكون أكبر من $\bar{X}$ يكون عندها أكبر عدد من القيم التي تكون أصغر من $\bar{X}$، معنى ذلك أن العلاقة بين المقاييس الثلاثة للنزعة المركزية تكون بالشكل التالي $\bar{X} > M_e > M_o$.

الحالة الثالثة : عدم تناظر من اليسار $F_1 < 0 \Rightarrow$ ، أي أن قيم Xi التي تتجاوز $\bar{X}$ تكون أقل من عدد القيم التي تكون أصغر من الوسط الحسابي.

ملاحظة:

- لا يوجد حد أدنى أو أعلى بقيم F_i ، غير لأنه من أجل توزيع إحصائي ضعيف الالتواء تكون $F_1 \in [-2\ ;\ +2]$
- يستعمل في بعض الحالات العلاقة أو المقدار $\frac{F_i}{2}$ لقياس الالتواء وذلك لأن $\frac{F_1}{2} = \frac{\bar{X} - M_0}{SD}$
- تكون العلاقة التالية صحيحة $\bar{X} - M_0 = 3(\bar{X} - M_e)$ في حالة توزيع ضعيف الالتواء.

2- **معامل بارسون:** (Kari Pearson ظهرت أعماله بين سنة 1890 و 1920 في بريطانيا): وهو عبارة عن النسبة بين مربع العزم المركزي من المرتبة الثالثة ومكعب العزم المركزي من المرتبة الثانية، ويرمز له بالرمز $P_1 : P_1 = \frac{\mu_3^2}{\mu_2^3}$، ولقد تم إختيار العزم المركزي من المرتبة الثالثة لقياس الالتواء لأن قيمته تساوي الصفر في حالة توزيع متناظر، أما بالنسبة للحالات الممكنة لهذا المعامل فإنها نفس الحالات الثلاثة لمعامل فيشر.

3- **معامل بول وكاندال:** (Yule & kendall)

مل هذا المعامل في حالة جداول مفتوحة، وهو عبارة عن النسبة[(Q1 - Q2) - (Q2 - Q3)] والمدى الربيعي، ويرمز له بالرمز Cyk ويكتب بالشكل التالي:

$$C_{YK} = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1}$$

الحالات الممكنة :

1- توزيع متناظر $C_{YK} = 0$.

2- التواء سالب $C_{YK} < 0$

3- التواء موجب $C_{YK} > 0$

ب- **مقاييس تحديد شكل النوع الثاني:** يستعمل هذه المقاييس في حالة توزيعات متناظرة، وذلك لتحديد مدى تطاولها أو تفلطحها مقارنة مع التوزيع الطبيعي، حيث يعتمد كل من فيشر وبارسون على المركزي من المرتبة الرابعة.

1- **معامل بارسون:** (Karl Pearson): وهو عبارة عن النسبة بين العزم المركزي من المرتبة الرابعة ومربع العزم المركزي من المرتبة الثانية، لماذا يعتمد كل من فيشر وبارسون على العزم المركزي من المرتبة الرابعة؟ لأن العزم

المركزي من المرتبة الرابعة يساوي 3 في حالة توزيع طبيعي أي توزيع على شكل جرسي، ويرمز لهذا المعامل بالرمز P_2 ، والعلاقة الإحصائية له هي كالتالي: $P_2 = \frac{\mu_4}{SD^4}$

ويمكن دراسة هذا المعامل حسب الحالات التالية:

الحالة الأولى : توزيع طبيعي أو على شكل جرسي $\Rightarrow P_2 = 3$.

الحالة الثانية : توزيع متطاول (تشتت ضعيف) $\Rightarrow P_2 > 3$.

الحالة الثالثة :توزيع متفلطح (تشتت قوي بالنسبة لمركز التوزيع) $\Rightarrow P_2 < 3$.

2- **معامل Fisher** : وهو عبارة عن معامل Pearson مطروحا منه 3، ويرمز له بالرمز F_2، وتكتب علاقته بالشكل التالي: $F_2 = P_2 - 3$، حيث يدرس حسب الحالات السابقة.

تمارين تطبيقية محلولة للفصل الثالث :

التمرين الأول: يبين التوزيع التكراري التالي نتائج الدراسة ما على عينة حجمها 50 وحدة حسب الاستعمال :

الجدول رقم: 3.4

الفئات	Ni	XI
100 -500	2	300
500 – 800	4	650
800 -900	6	850
900 -1000	18	950
1000 -1100	15	1050
1100 -1500	5	1300
$\sum$ni	50	/

المطلوب ما يلي :

. حدد شكل التوزيع بإستعمال مقاييس النزعة المركزية الثلاثة؟

. أحسب الإنحراف المتوسط والمعياري؟

. حدد النسبة التي يحتويها المجال $[\bar{X} \pm SD]$؟

التمرين الثاني : أرادت مؤسسة ما استراد نوعا معينا من الآلات الرياضية ، وأن لديها الإختيار بين ثلاثة أنواع، معتمدة في ذلك على الجودة.

وبعد إجراء الاختبارات الضرورية على عينة من كل نوع، تحصلت على النتائج التالية : (الفئات = مدة الاستعمال)

النوع الأول :

الجدول رقم:3.5

الفئات	0 -3	3 – 6	6 -9	9 -12	12 – 15	15 -18	18 -21
ni	2	4	8	12	17	6	1

النوع الثاني :

الجدول رقم:3.6

الفئات	0- 3	3 – 5	5 – 10	10- 15	15 -18	18 -21
ni	1	2	7	20	15	5

النوع الثالث :

الجدول رقم:3.7

الفئات	0 – 5	5 – 10	10 – 12	12 – 16	16 – 18	18 – 21
Ni	3	5	6	23	10	3

المطلوب ما يلي:

1- ما هو عدد الآلات التي لديها مدة إستعمال أقل من16، لكا نوع من الأنواع الثلاثة؟
2- ماهي مدة الاستعمال الأكثر تكرارا لكل نوع من الأنواع الثلاثة؟
3- ماهي مدة الاستعمال الوسيطية ، لكل نوع من الأنواع الثلاثة؟
4- ما هو النوع الذي يجب استرادة، باستخدام مقياس من مقاييس التشتت؟

التمرين الثالث : يبين التوزيع التكراري التالي المداخيل النسبية للرياضيين في فترة ما .

الجدول رقم:3.8

الفئات	%
$X_i < 10$	17.2
10 - 20	11.7
20 – 30	12.1
30 – 40	14.8
40 – 50	15.9
50 – 60	11.9
60 – 70	5
70 – 80	4
80 – 90	3.7
$X_i > 90$	3.6

المطلوب ما يلي:

1- هل يمكن تحديد SD لهذا التوزيع ولماذا؟

2- ما هو مقياس التشتت المناسب؟ أحسبه؟

3- حدد حدود المجال التالي: $[P_{19;}P_{80}]$؟

4- حدد حدود ونسبة المجال التالي $[\overline{X} \pm 0.67.\ SD]$؟

5- أدرس شكل هذا التوزيع بإستعمال مقياس فيشر؟

التمرين الرابع : لتكن السلسلتان الإحصائيتان التاليتان A ;B :

A : 5 ; 18 ; 10 ; 15 ; 3 ; 7 ; 6 ; 12

B : 18 ; 9 ; 8 ; 8 ; 3 ; 9

المطلوب ما يلي:

1- حدد الوسيط والمنوال لهاتين السلسلتين؟

2- مهل يمكن إستعمال المدى العام للمقارنة بين السلسلتين؟

3- ما هو مقياس التشتت المناسب في المقارنة بين السلستين؟

التمرين الخامس : يبين الجدول التكراري التالي توزيع 100 طالب حسب علاماتهم في مقياس الإحصاء والمنهجية

Xi	2	4	5	7	9	10	11	13	14	15	16	∑ni
ni	1	3	6	8	14	20	25	18	2	2	1	100
ni2	2	3	5	7	10	15	20	22	10	5	1	100

المطلوب ما يلي:

1- أحسب معامل الاختلاف في الحالتين ماذا تلاحظ؟

2- أحسب معامل التغير الربيعي في الحالتين؟

3- قارن تشتت العلامات في الحالتين بناءاه على السؤالين 1 و 2؟

4- قارن شكل التوزيعين بإستعمال معامل بارسون؟

التمرين السادس : أجب على الأسئلة النظرية التالية باختصار:

. متى تستخدم مقاييس التشتت المطلقة والنسبية؟

. ما هو الفرق بين المدى العام والمدى الربيعي؟

. إذا كان $\bar{} = Mo = Me$ لتوزيعين تكراريين فهل هذا يعني أنهما متشابهان بين ذلك

. متى يستعمل كل من الوسط الحسابي، الوسط الهندسي والوسط التوافقي؟

. ما هو دور كل من مقاييس النزعة المركزية والتشتت والشكل ، وما هو الفرق بينهم؟

التمرين السابع : يبين الجدول التكراري التالي توزيع 65 رياضي حسب أجورهم اليومية: المطلوب ما يلي:

. ما هو الأجر المتوسط؟

. قس تشتت هذا التوزيع بإستعمال المقياس المناسب؟

. قس التواء هذا التوزيع بإستعمال مقياس ملائم؟

. وضح لماذا لا يعد الإنحراف المعياري مقياسا ملائما للتشتت؟

$X_i > 110$	100 – 110	90 – 100	80 – 90	70 – 80	60 – 70	$X_i < 60$	%
3.07	7.69	15.38	21.53	24.61	15.38	12.3	الفئات

. إذا فرضنا أن طول كل من الفئة الأولى و الأخيرة يساوي 10 فقارن بين النسبة النظرية والنسبة التجريبية للمجال التالي $[\bar{X} \pm SD]$؟

حل تمارين الفصل الثالث :

التمرين الأول:

1- نستعمل مقاييس النزعة المركزية الثلاثة لتحديد شكل التوزيع الإحصائي المركزية ونحدد أولا هذه المقاييس إنطلاقا من جدول الحسابات التالي:

الجدول رقم:3.11

ت.ت.ص	E_r	$ni(Xi - Xo)$	X_i^2	$ni.X_i^2$	$\|Xi - \bar{X}\|$	$\|Xi - \bar{X}\|ni$
2	2/4	-26	90	18	653	1306
6	4/3	-24	422.5	169	303	1212
12	6	-12	722.5	433.5	103	618
30	18	0	902.5	1624.5	3	54
45	15	30	1102.5	1653.75	97	1455
50	5/4	35	1690	845	347	1735
/	/	/	3	4743.75	/	6380

أ- تحديد الوسط الحسابي بإستعمال الوسط الفرضي:

$$\bar{X} = X_0 + \frac{\sum n_i \frac{X_i - X_0}{K})}{\sum n_i}.K = 950 + \frac{3}{50}.50 = 953$$

ب- تحديد المنوال : لحساب المنوال ، نعدل التكرارات لأن الفئات غير متساوية :

$$M_0 = 900 + \frac{12}{12+3}.100 = 980$$

ج- تحديد الوسيط: بإتباع خطوات حساب الوسيط نتحصل على ما يلي:

$$M_0 = 900 + \frac{(25-12)}{18}.100 = 972.22$$

نلاحظ أن $M_o > M_e > \bar{X}$: إلتواء سالب.

2- تحديد الإنحراف المتوسط: $E_{\bar{X}} = \frac{\sum n_i|X_i - \bar{X}|}{\sum n_i} = \frac{6380}{50} = 127.6$

3- تحديد الإنحراف المعياري:

$$V(X) = \frac{\sum niX_i^{\,2}}{\sum ni} - \bar{X}^2 = \frac{4743.75 \times 10^4}{50} - 908209$$

$$V(X) = 40541 \Rightarrow SD = 201.34$$

4- تحديد النسبة التي يحتويها لمجال $[\bar{X} \pm SD]$:

[953 -201.34 ; 953+201.34] = [752 ; 1154]

لتحديد النسبة التي يحتويها هذا المجال، ننطلق من الفئة الثانية حيث أن 752 تقع بين 500 و800 ، ونلاحظ أنه من 500 إلى 800 لدينا تكرار يساوي 4 أما من 752 إلى 800 فيوجد X بإستعمال القاعدة الثلاثية ونفرض أن التكرارات تتوزع بشكل منتظم داخل المجال، نجد أن عدد التكرارات بين القيميتين 752 و 800 تساوي إلى 064 حيث نقرب هذه القيمة إلى 1، أما تكرارات الفئات الثالثة والرابعة والخامسة فتقع كلها في المجال المدروس ، أما المجال الجزئي من الفئة السادسة، نطبق عليه القاعدة الثلاثية بنفس طريقة المجال الجزئي للفئة الثانية، و نتحصل على التكرارات التالية للمجال المدروس : 41 = 1 + 15 + 18 +6+ 1 ، أما النسبة فهي : 82% = 100(41/50) ، نلاحظ أن هذه النسبة بعيدة بشكل كبير عن النسبة النظرية والتي تساوي 68% ، نستنتج أن التوزيع غير متناظر.

التمرين الثاني:

1- عدد الآلات التي لها مدة إستعمال أقل من 16 وحدة زمنية:

- النوع الأول : تكرارات الفئات الخمسة الأولى + عدد الوحدات المحصورة بين 15 و 16 ، وهذا بإستعمال القاعدة الثلاثية كما رأينا في التمرين السابق، وبالتالي فإن عدد هذه الآلات هو :

2 + 4 + 8 +12 + 17 + 2 = 45

- النوع الثاني : بنفس الطريقة ، نجد أن عدد هذه الآلات هو 35 = 5 + 20+7+2+1.
- النوع الثالث: عدد الآلات هو 37 = 23+6+5+3.

2- مدة الاستعمال الأكثر إنتشارا وتكرارا: هي المنوال.

- النوع الأول نلاحظ أن أطوال الفئات متساوية، وبالتالي فإن الفئة المنوالية : [12-15]، والمنوال هو :

$$M_e = 12 + \frac{5}{5+11}.3 = 12.937$$

- النوع الثاني : نلاحظ أن أطوال الفئات غير متساوية ، إذن يجب تعديل التكرارات من أجل حساب المنوال:

الجدول رقم: 3.12

Ki	3	2	5	5	3	3
Er	1/3	1	1.4	4	5	5/3

بعد تعديل التكرارات ، أصبحت الفئة المنوالية للفئة الخامسة بدلا من الفئة الرابعة :

[18 – 15] ، والمنوال هو $M_e = 15 + \frac{15}{1+10/3}.3 = 15.69$

- النوع الثالث: نلاحظ أن أطوال الفئات غير متساوية كذلك، إذن نعدل التكرارت:

الجدول رقم: 3.13

Ki	5	5	2	4	2	3
Er	3/5	1	3	5.75	5	1

الفئة المنوالية هي : [16 – 12] (أي الفئة التي تقابل أكبر تكرار معدل) ، إذن قيمة المنوال هي كالتالي:

$$M_e = 12 + \frac{2.75}{3.5}.4 = 15.14$$

3- مدة الاستعمال الوسيطية : يتعلق الأمر هذا بتحديد الوسيط:

- النوع الأول : أولا- نحدد التكرار التجميعي الصاعد:

الجدول رقم: 3.14

ت.ت.ص	2	6	14	26	43	49	50

ثانيا: ترتيب الوسيط: $\frac{\sum ni}{2} = \frac{50}{2} = 25$

ثالثا: الفئة الوسيطية: هي الفئة التي تقابل التكرار التجميعي الصاعد الذي يساوي ترتيب الوسيط أو الأكبر منه مباشرة، وهي الفئة الرابعة [12 – 9].

رابعا: قيمة الوسيط: $M_e = 9 + \frac{[25-14]}{3.125}.3 = 11.75$

بنفس الطريقة نحدد قيمة الوسيط للنوع الثاني والثالث:

النوع الثاني: $M_e = 10 + (15/20).5 = 13.75$

النوع الثالث $M_e = 12 + (11/23).4 = 13.91$

4- تحديد عدد الآلات التي تقع في المجال [19 – 11] بالنسبة لكل فرع:

- النوع الأول : لتحديد هذا العدد ، نحدد موضع طرفي المجال المدروس بين فئات التوزيع التكراري، فالفئات التي تقع بكاملها داخل المجال تحسب تكراراتها كاملة أما الفئات التي يقع جزء منها داخل هذا المجال، نطبق عليها القاعدة الثلاثية (كما رأينا في التمرين السابق) إذن عدد الآلات في المجال السابق هي كالتالي:
- النوع الأول : 27 = 4 +6+17.
- النوع الثاني : بنفس الطريقة نجد أن عدد الآلات هو : $23 + 10 + 1 + 3 \cong 33$.
- النوع الثالث : عدد الآلات هو :37 = 3+ 1+10+23.

5- نوع الآلات التي يجب سأترادها من طرف المؤسسة هي تلك التي يكون تشتتها ضعيفا مقارنة بالقيمة المركزية (الوسط الحسابي)، ولتحديد ذلك نستعمل معامل الاختلاف مثلا، ثم مقارنته بالنسبة لكل فرع، فأصغر معامل إختلاف يقابل أحسن نوع، أي كلما كان الإنحراف المعياري صغيرا بالنسبة للوسط الحسابي كلما كان النوع جيدا ، وهو النوع الذي يجب سترادة.

إذن نقوم بتحديد معامل الاختلاف لكل نوع :

- النوع الأول : معامل الاختلاف هو عبارة عن النسبة بين الإنحراف المعياري والوسط الحسابي، وجدول الحسابات التالي يحتوي على مختلف العمليات الحسابية :

الجدول رقم: 3.15

X_i	$\frac{X_i - X_0}{3}$	$n_i(\frac{X_i - X_0}{3})$	X_i^2	$n_i.X_i^2$
1.5	-3	-6	2.25	4.5
4.5	-2	-8	20.25	81
7.5	-1	-8	56.25	450
10.5	0	0	110.25	1323
13.5	1	17	182.25	3098.25
16.5	2	12	272.25	1633.5
19.5	3	3	380.5	380.5
/	/	10	/	6970.5

أ- تحديد الوسط الحسابي: نستعمل الطريقة المختصرة، حيث نقسم الفروق (Xi-Xo) على طول الفئة ونفرض أن الوسط الفرضي يساوي 10.5 إذن :

$$\bar{X} = 10.5 + \frac{10}{50}.3 = 11.1$$

ب- الإنحراف المعياري: $SD = \sqrt{\frac{\sum ni.{X_i}^2}{\sum ni} - \bar{X}^2} = \sqrt{\frac{6970.5}{50} - 123.21}$

$$= \sqrt{16.2} = 4.0249$$

ج- تحديد معامل الاختلاف: $C.V = \frac{SD}{\bar{X}} = \frac{4.0249}{11.1} = 0.3626$

- تحديد معامل الاختلاف للنوع الثاني : بنفس الطريقة السابقة نحدد هذا المعامل ، اعتمادا على جدول الحسابات التالي :

الجدول رقم:3.16

X_i	$X_i - X_0$	$n_i(X_i - X_0)$	${X_1}^2$	$n_i.{X_1}^2$
1.5	-11	-11	2.25	2.25
4	-8.5	-17	16	32
7.5	-5	-35	56.25	393.75
12.5	0	0	156.25	3125
16.5	4	60	272.25	4083.75
19.5	7	35	380.25	1901.25
/	/	32	/	9538

$$\bar{X} = 12.5 + \frac{32}{50} = 13.14$$

$$SD = \sqrt{\frac{9538}{50} - 172.659} = \sqrt{18.101} = 4.2545$$

$$C.V = \frac{SD}{\bar{X}} = \frac{4.2545}{13.14} = 0.3237$$

النوع الثالث: بنفس الطريقة ، نحدد معامل ا لاختلاف حسب الجدول التالي :

الجدول رقم:3.16

X_i	$X_i - X_0$	$n_i(X_i - X_0)$	$X_1^{\ 2}$	$n_i . X_1^{\ 2}$
2.5	-11.5	-34.5	6.25	18.75
7.5	-6.5	32.5	56.25	281.25
11	-3	-18	121	726
14	0	0	196	4508
17	3	30	289	2890
19.5	5.5	16.5	380.25	1140.75
		-38.5		9564.75

$$\bar{X} = 14 + \frac{(-38.5)}{50} = 13.23$$

- معامل الاختلاف :

$$SD = \sqrt{\frac{9564.75}{50} - 175.03} = \sqrt{16.26} = 4.032$$

$$SD = \frac{SD}{\bar{X}} = \frac{4.032}{13.23} = 0.3047$$

النوع الواجب استرادة: نقارن بين معاملات الاختلاف:

إذن نلاحظ أن أصغر معامل إختلاف هو معامل الاختلاف للنوع الثالث، وبالتالي فالنوع $CV_3 < CV_2 < CV_1$ الثالث هو الذي يجب ستراده.

التمرين الرابع:

1- تحديد الوسيط والمنوال للسلسلتين A & B .

أ- تحديد الوسيط:

- السلسلة A : أولا ترتيب قيم المتغير الإحصائي ترتيبا تصاعديا.

3 ; 5 ; 6 ; 7 ; 10 12 ; 15 ; 18

الوسيط هو عبارة عن الوسط الحسابي للقيمة الرابعة والخامسة:

$M_{eA}=17/2=8.5$

- السلسلة B: بنفس الطريقة السابقة، يحدد الوسيط لهذه السلسلة :

$M_{eB}=17/2=8.5$

ب- تحديد المنوال :

- السلسلة A : لا يوجد منوال لأن كل القيم لها نفس التكرار

- السلسلة B: سلسلة ثنائية المنوال: $M_{01}=8\ ;\ M_{02}=9$، لأن القيمتين مكررتين بنفس التكرار.

2- لايمكن إستعمال المدى العام للمقارنة بين السلسلتين، ويمكن إستعمال المدى الربيعي مثلا ، لأن المدى العام للسلسلة A يساوي المدى العام للسلسة B .

3- مقياس التشتت المناسب هو المدى الربيعي :

- المدى الربيعي لـA: Q3-Q1=12-5=7

- المدى الربيعي لـB: Q3-Q1=9-8=1 . نلاحظ أن التوزيع *A* أكثر تشتتا من التوزيع *B*.

التمرين الخامس:

1- حساب معامل الاختلاف :

أ- حساب الوسط الحسابي للتوزيعين (الإحصاء والمنهجية) : الجدول التالي يبين مختلف العمليات الحسابية:

$$Z_1=X_i-X_0\ ;Z_2=ni_1(X_i-X_0);\ Z_3=ni_2(X_i-X_0)$$

الجدول رقم:3.18

X_i	X_i-X_0	$ni_1(X_i-X_0)$	$ni_2(X_i-X_0)$
2	-8	-8	-16
4	-6	-18	-18
5	-5	-30	-25
7	-3	-24	-21
9	-1	-14	-10
10	0	0	0
11	1	25	20
13	3	54	66
14	4	8	40
15	5	10	25
16	6	6	6
$\sum$	/	9	67

$$\overline{X_1}=10+\frac{9}{100}=10.09\ ;\overline{X_2}=10+\frac{67}{100}=10.67$$

ب- حساب الإنحراف المعياري للتوزيعين:

$$SD_1 = \sqrt{108.93 - 101.808} = 2.668$$

$$SD_1 = \sqrt{123.13 - 113.84} = 3.0464$$

ج- حساب معامل الاختلاف للتوزيعين:

$$CV_2 = \frac{3.0464}{10.67} \times 100 = 28.55\% \qquad CV_1 = \frac{2.668}{10.09} \times 100 = 26.44\%$$

التوزيع الأول أقل تشتتا من الثاني .

2- حساب معامل التغير الربيعي: لحساب ذلك نضع التكرار التجميعي الصاعد للتوزيعين:

الجدول رقم: 3.19

ت.ت.ص	1	4	10	18	32	52	77	95	97	99	100
ت.ت.ص	2	5	10	17	27	42	62	84	94	99	100

أ- معامل التغير الربيعي الأول : هو عبارة عن النسبة بين المدى الربيعي والربيعي الثاني، نحدد كل من الربيعي الأول والثاني والثالث:

Q1=9 ; Q2=10 ; Q3=13 ، إذن :

(Q3-Q1)/Q2=4/10=0.4

ب- معامل التغير الربيعي للتوزيع الثاني:

$$\frac{Q_3 - Q_1}{Q_2} = \frac{13 - 9}{11} = 0.363$$

3- مقارنة تشتت العلامات في التوزيعين بناء على نتائج السؤال الأول والثاني: عند إستعمال معامل الإختلاف وجدنا أن التوزيع الأول أقل تشتتا من التوزيع الثاني. أما عند قياس التشتت بالنسبة لمركز التوزيع فقط أي بالنسبة لـ: 50% المركزية من الوحدات الإحصائية وجدنا أن التوزيع الثاني أقل تشتتا من التوزيع الأول.

4- مقارنة شكل التوزيعين بإستعمال معامل بيرسون:

أ- قياس الاتواء التوزيع الأول: $P_1 = \frac{59.69}{360.67} = 0.165$ إلتواء موجب.

ب- قياس الاتواء التوزيع الثاني: $P_1 = (-22.96)^2/799.32 = 0.659$

ج- قياس تطاول التوزيع الأول:

P2 < 3 : وجود تفلطح في التوزيع الأول

$$m_2 = 108.93 \; ; \; m_3 = 1250.55 \; ; \; m_4 = 14393.37$$

د- قياس تطاول التوزيع الثاني:

$$P_2 = \frac{\mu_4}{SD^4}$$

$$\mu_4 = m_4{}^{-4}.m_1.m_3{}^{+6}.m_1^2.m_2{}^{-3}.m_1^4$$

=18603.25-63545.82+184109.29-38884.716=100282.0039

. التوزيع الثاني ذو تطاول شديد $P_2 = \frac{100282.0039}{9.28} = 1164.32\,;\, P_2 > 3 \Rightarrow$

التمرين السادس :

الإجابة عن الأسئلة النظرية :

1- نستخدم مقاييس التشتت المطلقة في المقارنة بين عدة توزيعات إحصائية بوحدات قياس متماثلة، كما تستخدم في تحديد التوزيعات النظرية الموافقة لتوزيعات تجريبية معينة. أما مقاييس التشتت النسبية ، فتستخدم في مقارنة تشتت عدة توزيعات إحصائية بوحدات قياس مختلفة، كما تستخدم في تحديد جودة التوزيعات الإحصائية.

2- الفرق بين المدى العام والمدى الربيعي:

المدى الربيعي هو جزء من المدى العام. ويحتوي المدى العام على كل الوحدات الإحصائية التي تشكل العينة أو المجتمع ، بينما يحتوي المدى الربيعي على 50% من المجتمع أو العينة.

يمثل المدى الربيعي 50 % من المدى العام في حالة توزيع متماثل، ويمثل أقل من هذه النسبة في حالة توزيع ذو تشتت ضعيف ، كما يمثل أكثر من النسبة السابقة في المدى العام في حالة تشتت قوي نسبيا.

3- إذا كانت الخصائص الثلاثة للنزعة المركزية متساوية $(\bar{X} = M_o = M_e)$ لتوزيعين تكراريين، فهذا لا يعني أنهما متشابهان ، فيمكن أن يكون إحداهما متطاول والآخر متفلطح، وهذا حسب قيمة مقياس الشكل المستعمل.

4- يستعمل الوسط الحسابي في حالة وجود علاقة طردية بين خاصيتين وفي كل الحالات الأخرى ماعدا الاستعمالات الخاصة بالوسط الهندسي والوسط التوافقي.

ويستعمل الوسط الهندسي في حالات خاصة منها : في حساب متوسط الأرقام القياسية وفي حساب المعدلات (المعدل المتوسط لسعر الفائدة، المعدل المتوسط للنمو ...إلخ).

أما الوسط التوافقي فيستعمل في حالات إستثنائية ، كوجود علاقة عكسية بين خاصيتين (السعر المتوسط(علاقة عكسية بين السعر والقدرة الشرائية)، السرعة المتوسطة (ة علاقة عكسية بين الزمن والسرعة) ، ...إلخ).

التمرين السابع :

عندما تكون الخصائص الثلاثة للنزعة المركزية (الوسط الحسابي، الوسيط، المنوال) متساوية فإن التوزيع التكراري يكون متناظر بالنسبة لهذه الخصائص، وتقع هذه الأخيرة في مركز الفئة المنوالية أو الوسيطية، وبالتالي فإن هذه الخصائص تساوي ما يلي: ، إذن حسب العلاقة المعطاة في السؤال، يمكن كتابة ما يلي:

$$M_0 = A + \frac{\Delta_1}{\Delta_1+\Delta_2}.k + += A + \frac{k}{2} \Rightarrow \frac{\Delta_1}{\Delta_1+\Delta_2} = \frac{1}{2}$$

المحاضرة الثامنة : الانحدار الخطي البسيط

- ✓ نوع العلاقات الموجودة بين المتغير التابع والمستقل
- شكل العلاقة الموجودة بين المتغير التابع والمستقل
- طريقة تحديد نوع العلاقة بين المتغيرين
- شروط دراسة العلاقة بين متغيرين

الهدف منها : التعرف على الادوات الضرورية لدراسة العلاقة بين متغيرين

الفصل الرابع : الانحدار والارتباط: Regression & Coeeelation

يحتوي هذا لفصل على مدخل في الانحدار والارتباط الخطي البسيط، ويتضمن أساسا على الأدوات الضرورية لدراسة العلاقة بين متغيرين إحداهما تابع والآخر مستقل، مع إمكانية التنبؤ بقيم المتغير التابع وفق قيم المتغير المستقل (إذا بقت نفس الظروف على حالها) . ظهرت أول دراسة في مجال الانحدار الخطي من طرف :

(Sir Francis Galton : 1822-1911)، وكانت تخص العلاقات الوراثية (العلاقة بين طول قامة الآباء و الأبناء) ، ثم من طرف (Karl Pearson : 1857-1936) في نفس الميدان..

أولا : الانحدار الخطي البسيط (حسن ياسين طعمة،2009، ص261 - 315) :

تقتصر دراسة الانحدار الخطي البسيط على العلاقة الخطية بين متغيرين فقط، إن كلمة خطي تعني أن نسبة الزيادة في المتغير المستقل (المفسر: بكسر السين) تساوي بالتقريب نسبة الزيادة في المتغير التابع (المفسر: بفتح السين) ، أما كلمة بسيط فيقصد بها العلاقة بين متغيرين فقط.

تعتبر هذه العلاقة عن تطور ظاهرة معينة (اقتصادية ، اجتماعية ، بيولوجية ...إلخ)، في مكان أو زمن معين. إن دراسة هذه العلاقة تؤدي إلى صياغة إحصائية على شكل نموذج لهذه الظاهرة ، وهذا قصد التنبؤ بقيمها مستقبلا إذا بقيت نفس الظروف على حالها، بحيث أن الظواهر الاقتصادية والاجتماعية ...إلخ، لا تتطور بصفة عفوية وعشوائية بل هناك مسببات تؤدي بها إلى التغير عبر مراحل أوضاع مختلفة وهذا حسب قوة التأثير. إن لدراسة أي ظاهرة من الظواهر مراحل محددة ، تلخص بالشكل التالي:

- تحديد المتغير التابع المدروس .
- تحديد العوامل التي لها علاقة بهذه الظاهرة وترتيبها حسب درجة التأثير.
- اختيار العامل الأكثر تأثيرا وتفسيرا.
- تحديد نوع العلاقة الموجودة بين المتغيرين.
- وضع الصياغة الرياضية لشكل العلاقة (خطية أو غير خطية).
- وضع الشكل النهائي للنموذج بناءا على النماذج الرياضية المعروفة.

حددت النظرية الاقتصادية العوامل المؤثرة في عدة ظواهر نذكر منها:

أ- الدخل هو العامل الرئيسي في تحديد استهلاك الأفراد والأسر ، ووضعت الصياغة الإحصائية بالشكل التالي: $C = \alpha + \beta Y + u$، حيث أن C الاستهلاك ; α β معلمات يمكن تقديرها وهي تمثل على التوالي نسبة الدخل المخصصة للاستهلاك و الحد الأدنى للاستهلاك ، أما u فيمثل بقية العوامل الثانوية المؤثرة في الاستهلاك (الأسعار ، حجم الأسرة ....إلخ) و العوامل غير القابلة للقياس (المتغيرات الكيفية : العادات والسلوك الاستهلاكي ...إلخ).

ب- معدل الفائدة هو العامل الرئيسي في تحديد حجم الاستثمار: $I = I_0 + \beta_i$، يمثل i معدل الفائدة .

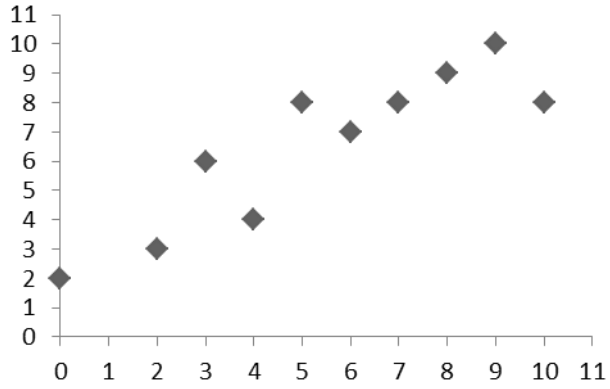
تعتبر قيم المعلمات $\alpha; \beta; I_0$ مجهولة ، هدفنا هو تحديدها اعتمادا على معطيات سابقة للظاهرة المدروسة والعامل الرئيسي المؤثر فيها (المتغير التابع و المتغير المستقل). سنكتفي باستعمال طريقتين لتحديد هذه المعلمات: الطريقة البيانية والطريقة التحليلية وسنركز على هذه الأخيرة لأنها أكثر عملية من الأولى. وقبل التطرق إلى هاتين الطريقتين ، نحاول أن نبين نوع العلاقات التي يمكن أن تكون بين المتغيرين المدروسين . وسنركز على الظواهر الإقتصادية-الاجتماعية لدرجة الأولى.

1- نوع العلاقات الموجودة بين المتغير التابع والمستقل:

إن العلاقة التي يمكن أن تكون بين المتغير التابع والمستقل هي إما علاقة خطية أو غير خطية، سنكتفي في هذا الفصل بالعلاقة الخطية فقط، ونحاول عن طريق أمثلة الانتقال من بعض العلاقات غير الخطية إلى العلاقة الخطية. ويقصد بالعلاقة الخطية أن نسبة تغير المتغير التابع تساوي بالتقريب نسبة تغير المتغير المستقل، أي أن سحابة النقاط الممثلة للمتغيرين تتبع خط مستقيم بالتقريب.

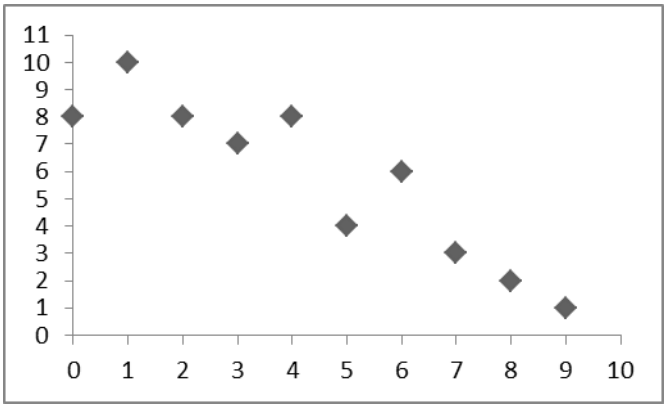
أ- **شكل العلاقة الموجودة بين المتغير التابع والمستقل:** يمكن أن نميز بين ثلاثة حالات:

- **الحالة الأولى:** علاقة خطية طردية: الزيادة في المتغير المستقل تؤدي إلى زيادة في المتغير التابع والعكس صحيح.



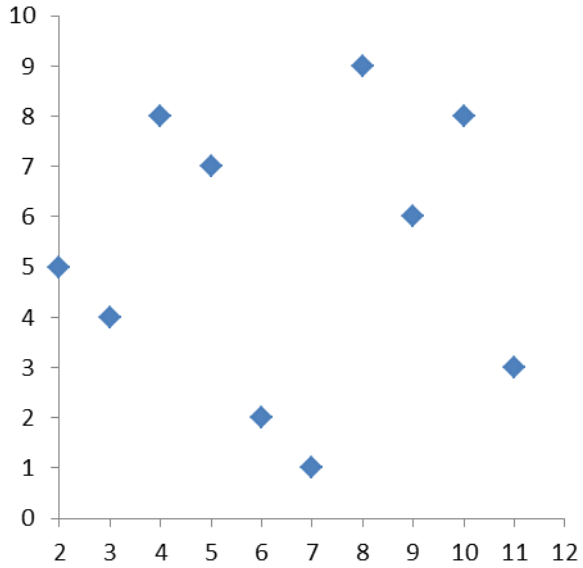
الشكل رقم :5.1

- **الحالة الثانية:** علاقة خطية عكسية: الزيادة في المتغير المستقل تؤدي إلى الانخفاض في المتغير التابع والعكس صحيح.



الجدول رقم:4.2

- **الحالة الثالثة:** لا توجد علاقة بين المتغيرين، لأن النقاط مبعثرة ولا يوجد اتجاه واضح لها.



الجدول رقم:4.3

ملاحظة : يمكن أن نحدد نوع وشكل العلاقة بين متغيرين انطلاقا من العرض البياني.

ب- طريقة تحديد نوع العلاقة بين المتغيرين:

نفرض أن قيم المتغير التابع و ، قيم المتغير المستقل:

$$y_1 \; ; \; y_2 \; ; y_3 ; y_4 ; \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots y_n$$

$$X_1 \; ; \; X_2 \; ; X_3 ; X_4 ; \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots X_n$$

نستعمل النسب التالية لتحديد نوع العلاقة بين المتغيرين:

$$\frac{Y_i - Y_{i-1}}{X_i - X_{i-1}} = \frac{Y_{i-1} - Y_{i-2}}{X_{i-1} - X_{i-2}} = \dots\dots\dots\dots\dots\dots = \frac{Y_2 - Y_1}{X_2 - X_1} \text{ ثابت}$$

كانت هذه النسب متساوية أو متقاربة، نقول أن العلاقة بين المتغيرين خطية(طردية أو عكسية) ، أما إذا كانت غير ذلك أي متباعدة فنقول أن العلاقة بينها غير خطية.

ج- شروط دراسة العلاقة بين متغيرين:

لدراسة العلاقة بين متغيرين يجب توفر الشروط التالية:

- أن تكوت بين الظاهرتين المدروستين علاقة جدلية واضحة.
- أن تكون إحداهما ناتجة والأخرى مسببة.
- أن تكون الظاهرتان قابلتان للقياس.
- أن تكوت كل القياسات المتحصل عليها متقابلة بالنسبة للزمن والمكان أو كلاهما معا.

المحاضرة التاسعة : تابع الانحدار والارتباط

- ✓ طرق تحديد العلاقة الخطية البسيطة
  - الطريقة البيانية
  - الطريقة التحليلية
- ✓ اهم خصوصيات معامل الارتباط
- ✓ معامل الارتباط الرتبي
  - قياس الارتباط الرتبي
  - دراسة صلاحية النموذج

الهدف منها : التعرف على الادوات الضرورية لدراسة العلاقة بين متغيرين

2- الإنحدار Regression (حسن ياسين طعمة ،2009، ص 325- 356) :

وهو إيجاد علاقة أو ترابط أحد المتغيرات بالمتغيرات الأخرى، تسمى الدالة التي تربط بينهم بدالة الانحدار أو التراجع الخطي البسيط (Regression Lineaire Simple) . نكتفي في هذا الفصل بدراسة الشكل الخطي البسيط لهذه العلاقة.

مثال عن مفهوم الانحدار: لاحظ البيولوجيون ، عند دراسة طول قامة الآباء والأولاد أنه عندما يكون طول قامة الأب أكبر من المتوسط، فإن طول قامة الابن تكون أقل بقليل من المتوسط، معناه وجود نزعة من طرف الأبناء على الابتعاد عن طول قامة الآباء أو التراجع أو النزول إلى متوسط أقل.

أ-**طرق تحديد العلاقة الخطية البسيطة:** نتبع خطوتين هامتين لصياغة العلاقة النهائية بين متغيرين، تتعلق الأولى بوصف وتحديد نوع المتغيرات التي تدخل في النموذج، وهذا باستعمال الإحصائيات المتوفرة لدينا(مثلا: إذا أرادت إحدى المؤسسات دراسة العوامل التي تدفع المستهلكين إلى اقتناء منتوجها، تحدد المتغير الأكثر تأثيرا على سلوكهم و أذواقهم)، وتتمثل الخطوة الثانية في تحديد فيما إذا كان التغير في المتغير المستقل يؤدي إلى التغير في المتغير التابع وبنفس الاتجاه أو بالاتجاه المعاكس. ثم ننتقل إلى وضع شكل النموذج مع تقدير معلماته، وفي هذا الصدد نكتفي بطريقتين : البيانية والتحليلية.

- **الطريقة البيانية:** (Méthode graphique) نقوم بتمثيل وعرض البيانات الإحصائية المتعلقة بالمتغيرات المدروسة بيانيا، فنلاحظ في الأخير أننا تحصلنا على كوكبة أو سحابة من النقاط، فشكل هذه الأخيرة هو الذي يمكننا من تقريبه إلى أحد الأشكال الرياضية المعروفة (بناء على معرفة الباحث بالأشكال والنماذج الرياضية). نتحصل في أغلب الأحيان على كوكبة مستطيلة من النقاط، إن هذه الاستطالة تعني وجود علاقة خطية (طردية أو عكسية)، ونتحصل في حالات أخرى على أشكال مختلفة أخرى.

يهمنا في هذه الفصل الشكل المستطيل من النقاط والذي يقع بالتقريب على استقامة واحدة ففي هذه الحالة، نلاحظ أن الشكل الرياضي المناسب هو الخط المستقيم، ولوضع هذا الخط لابد من توفر شرطان أساسيان:

أولا: يجب أن يمر هذا المستقيم على النقطة المركزي $(\bar{X}\ ;\ \bar{Y})$.

ثانيا : يجب أن تكون المسافات التي تفصل بين نقاط الكوكبة ونقاط الخط المستقيم أصغر ما يمكن.

تتطلب الطريقة البيانية دقة كبيرة يصعب علينا بلوغها، لتحديد المسافات السابقة و لتقدير معلمات النموذج، وبالتالي ننتقل إلى طريقة عملية تسمى بالطريقة التحليلية.

- **الطريقة التحليلية :** (Méthode Analytique) تسمى هذه الطريقة بطريقة المربعات الصغرى: Lé méthode) des moindes carrés) نتبع نفس خطوات وشروط الطريقة البيانية، لنصل إلى الشرط المتعلق بالمسافات التي تفصل بين نقاط الكوكبة والنقاط التي تقع على الخط المستقيم، فمجموع هذه المسافات أو الفروق يكون في بعض الأحيان مساويا إلى الصفر ، وتفاديا لهذه النتيجة ، نستعمل مربعات هذه المسافات أو الفروق.

تنص طريقة المربعات الصغرى بأن يكون مجموع مربعات هذه الفروق أصغر ما يمكن، ومن هنا جاءت تسمية هذه الطريقة ، وتكون هذه الطريقة صالحة في حالة النماذج الخطية البسيطة.

نفرض أن النموذج الخطي البسيط يكون بالشكل التالي: $Y_i = a + bX_i + U_i$

Y_i : قيم المتغير التابع أو القيم الحقيقة للظاهرة المدروسة.

X_i قيم المتغير المفسر أو المستقل.

a مقدر ثابت . b معامل الانحدار.

U_i مقدار الخطأ أو الفرق بين القيم الحقيقة التي تشكل نقاط الكوكبة والقيم التي تقع على الخط المستقيم، ويسمى هذا المتغير بالمتغير العشوائي والذي يمثل جميع العوامل الأخرى التي تؤثر في المتغير التابع والتي لم تؤخذ بعين الاعتبار أو التي لا يمكن قياسها.

نرمز لقيم المتغير التابع التي تقع على الخط المستقيم التابع والتي تسمى بالقيم المقدرة بالرمز * .

ويمكن كتابة الفرق بين القيم الحقيقية والقيم المقدرة بالشكل التالي: $U_i = {}_i - Y_i^*$

يتمثل مبدأ طريقة المربعات الصغرى البسيطة في تقدير المعلمات a ; b بشرط أن يكون مجموع مربعات الفروق أصغر ما يمكن ، ويكتب هذا الشرط بالشكل التالي: $\sum(Y_i - Y_i^*)^2$

يمر التابع f (a ; b) بنهايته الصغرى عند تساوي المشتقتان الجزئيتان لهذا التابع بالنسبة للمعلمتين a ; b إلى الصفر

$$\frac{\delta f(a;b)}{\delta a} = 0 \quad ; \frac{\delta f(a;b)}{\delta b} = 0$$

نشتق المقدارين الأول والثاني بالشكل التالي:

$$\frac{\delta f(a;b)}{\delta a} = \frac{\delta \sum(Y-a-bX)^2}{\delta a} = -2\sum(Y - a - bX) = 0$$

$$\Rightarrow \sum(Y - a - bX) = 0 \dots\dots\dots\dots (1)$$

$$\frac{\delta f(a;b)}{\delta b} = \frac{\delta \sum(Y-a-bX)^2}{\delta b}$$

$$\Rightarrow \sum X(Y - a - bX) = 0 \dots\dots\dots\dots (2)$$

من هاتين المعادلتين نستخرج قيمة كل من a ; b :

- من المعادلة الأولى نستخرج قيمة a بدلالة b :

$\sum Y - \sum a - \sum bX = 0 \Rightarrow \sum Y - na - b\sum X = 0$ نقسم الطرفين على n لنجد a بدلالة b:

$$a = \frac{\sum Y}{n} - b\frac{\sum X}{n} \Rightarrow a = \bar{Y} - B\bar{X} \dots\dots\dots\dots (3)$$

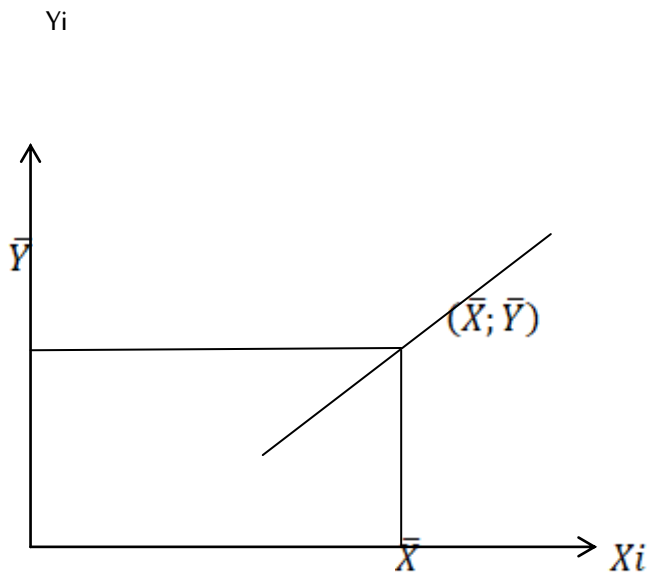
أما من المعادلة الثانية نستخرج معامل الانحدار b :

$\sum XY - \sum a - b\sum X_i^2 = 0 \Rightarrow \frac{\sum XY}{n} - a\frac{\sum X}{n} - b\frac{\sum X^2}{n} = 0$ ، نعوض بقيمة a في هذه المعادلة:

$$\frac{\sum X_I Y_i}{n} - \bar{Y}\bar{X} + b\bar{X}^2 - b\frac{\sum X_i^2}{n} = 0$$

$$\Rightarrow \frac{\frac{\sum X_i Y_i}{n} - \bar{Y}\bar{X}}{\frac{\sum X_i^2}{n} - \bar{X}^2} \dots\dots\dots\dots\dots\dots (4)$$

ملاحظة : تبين المعادلة الثالثة أن المستقيم الذي يمثل كوكبة النقاط (القيم الحقيقة للمتغيرين) لابد أن يمر من النقطة $(\bar{X}\,;\bar{Y})$حسب الشكل التالي:



الشكل رقم : 4.4

بعض فرضيات النموذج الخطي البسيط:

1- يجب أن يكون قياس المتغيرين بدقة وبدون خطأ.

2- متوسط الخطأ يساوي إلى الصفر $\bar{U}_i = 0$.

3- تباين الخطأ مستقل عن الزمن $V_{(u_t)} = E(\ \ ^2_t) = \sigma^2 \forall t$ أو بما يسمى بتجانس تباين الخطأ (Homoscédasticité).

4- عدم وجود ارتباط ذاتي (Autocorrélation) بين الأخطاء، معناه أن التبيان المشترك بين الأخطاء معدوم.

$$COV(u_t; u_t) = 0;\ \forall t, s: t \neq s$$

مثال : يبين التوزيع التكراري التالي العلاقة بين استهلاك الكهرباء (X) ومردودية الرياضي (Y) في 10 ملاعب خلال سنة معينة:

الجدول رقم :4.1

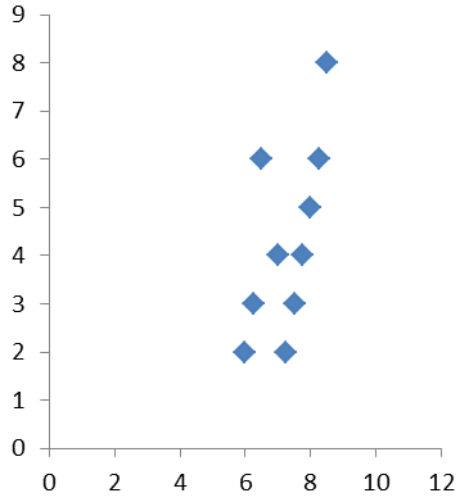
رقم الرياضي	Y	X
1	2	6
2	3	6.1
3	6	6.8
4	4	7.2
5	2	7.4
6	3	7.9
7	4	8.2
8	5	8.5
9	6	8.9
10	8	9.1

المطلوب ما يلي:

1- ما هو شكل العلاقة بين مردودية الرياضي و استهلاك الكهرباء؟
2- نفرض أن العلاقة خطية بين المتغيرين: قدر معلمات النموذج الخطي البسيط ؟
3- ما هي قيمة المردودية عندما يبلغ استهلاك الكهرباء 11 وحدة؟

الحل:

1- لتحديد العلاقة بين مردودية الرياضي واستهلاك الكهرباء، نقوم بوضع العرض البياني المناسب :



الشكل رقم :4.5

نلاحظ أن كوكبة النقاط تأخذ مسار خطي متزايد، بناءا على هذا المسار، يمكن تقريب العلاقة بين المتغيرين إلى علاقة خطية وطردية.

2- تقدير معلمات النموذج فرضا أن العلاقة خطية بين المتغيرين:

$$Y_i = a + bX_i + u_i$$

- أولا: نحدد الوسط الحسابي للمتغيرين:

$$\bar{X} = \frac{\sum X_i}{n} = \frac{76.1}{10} = 7.61$$

$$\bar{Y} = \frac{\sum Y_i}{n} = \frac{43}{10} = 4.3$$

- ثانيا : تحديد معامل الانحدار:

$$b = \frac{\frac{\sum x_i.Y_i}{n} - \bar{x}.\bar{y}}{\frac{\sum x_i^2}{n} - \bar{X}^2} = \frac{\frac{339.9}{10} - 4.3 \times 7.61}{\frac{589.97}{10} - 57.91} = \frac{12.67}{10.849} = 1.1678$$

- ثالثا: تحديد الثابت a

$$a = \bar{Y} - b\bar{X} = 4.3 - 1.1678 \times X = -4.587:$$

- رابعا: كتابة معادلة الانحدار:

$$Y = -4.587 + 1.1678 \times X$$

3- المردودية عندما يبلغ مستوى استهلاك الكهرباء 11 وحدة

$$Y = -4.587 + 1.1678 \times 11 = 8.25:$$

ملاحظة: يستعمل النموذج الخطي البسيط المتحصل عليه في عملية التنبؤ بقيم المتغير التابع، بشرط أن تبقى الظروف على حالها . إذن يمكن أن نحدد مردودية الرياضي عند أي مستوى من استهلاك الكهرباء.

التشفير الاقتصادي لهذا النموذج هو كالتالي : كلما أزداد X بوحدة واحدة فإن Y يزداد بـ1.1678 وحدة، إذن كلما كان معامل الانحدار كبيرا كلما أصبحت العلاقة بين المتغيرين قوية.

توضيح: لا يمكن تحديد مردودية الرياضي عند أي مستوى من استهلاك الكهرباء بل عند مستويات محدودة ، لأن هناك قيود تحد من ذلك، مثلا قيد الإشباع، فعندما يصل استهلاك الكهرباء إلى مستوى معين فإن المردودية لا يمكن أن تزداد مهما ارتفع الاستهلاك. إن إختيار المتغير المستقل الأكثر أهمية له دور كبير في تحديد و إعطاء قيم أكثر دقة للتنبؤات.

ثانيا : الانتقال من الانحدار غير الخطية إلى الانحدار الخطي (جلاطو جيلالي،2005، ص155):

يمكن تطبيق طريقة المربعات الصغرى على النماذج غير الخطية بعد الانتقال من العلاقة غير الخطية إلى العلاقة الخطية بين المتغيرين بواسطة النماذج اللوغاريتمية أو نصف اللوغاريتمية، وسنتطرق إلى بعض الحالات:

الحالة الأولى النموذج نصف اللوغاريتمي

ننتقل في هذه الحالة من أحد الأشكال الأسية إلى الشكل النصف اللوغاريتمي ، ولكن مثلا لتقدير معلمات هذا النموذج ننتقل أولا غلى الشكل الخطي بضرب الطرفين في اللوغاريتم: $\log Y = \log \beta + X \log \alpha$

نلاحظ أن المعطيات الأصلية تتحول إلى قيم لوغاريتمية قصد تقدير لوغاريتمي المعلمات، وبعد الإنتهاء من هذه العملية ، نعود إلى النموذج الأصلي بوضع القيم الأصلية للمعلمات المقدرة.

مثال: الوحدات الإحصائية التي تشكل مجتمعا معينا معرفة بالخاصيتين المبينتين في الجدول التالي:

الجدول رقم :4.2

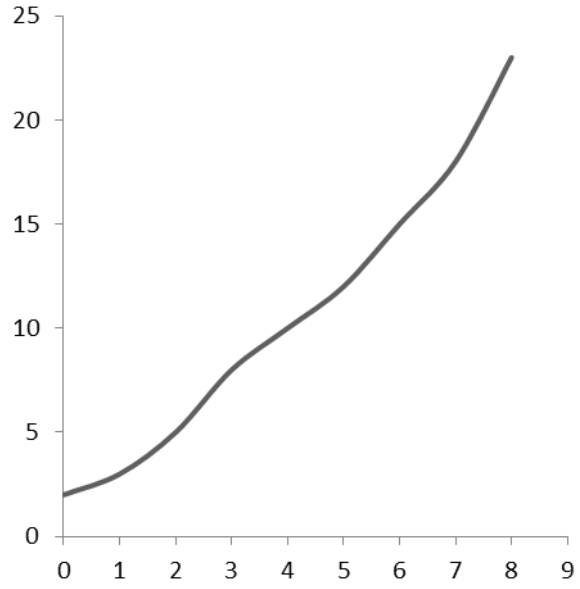
X	1	3	4	6	7	8	9
Y	2.6	4.6	5.4	9.8	12.3	16.3	21.2

المطلوب ما يلي:

1- تحديد شكل العلاقة الموجودة بين X وY ؟
2- تقدير معلمات النموذج المتحصل عليه؟

الحل:

1- تحديد شكل العلاقة بين المتغيرين من خلال العرض البياني :



الشكل رقم :4.6

نلاحظ من خلال العرض البياني أن الشكل الرياضي يمكن إعطاؤه للعلاقة الموجودة بين المتغيرين هو كالتالي : $Y = \beta . \alpha^X$ وهو عبارة عن نموذج أسي يمكن تحويله إلى نموذج نصف لوغاريتمي .

2- تقدير α و β :

$$r = \sqrt{b_1 . b_2} = \sqrt{\frac{(\sum X_i Y_i - n.\bar{X}\bar{Y})(\sum X_i Y_i - n.\bar{X}\bar{Y})}{(\sum X_i^2 - n.\bar{X}^2)(\sum y_i^2 - n.\bar{y}^2}}$$

$$\frac{\sum X_i Y_i - n.\bar{X}\bar{Y}}{\sqrt{(\sum X_i^2 - n.\bar{X}^2)(\sum y_i^2 - n.\bar{y}^2)}}$$

ويكتب الشكل النهائي لعلاقة معامل الإرتباط بالشكل التالي:

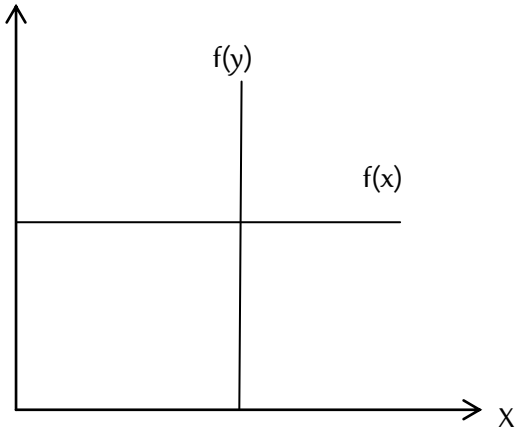
$$r = \sqrt{b_1 . b_2} = \frac{\sum X_i Y_i - n.\bar{X}\bar{Y}}{\sqrt{(\sum X_i^2 - n.\bar{X}^2)(\sum y_i^2 - n.\bar{y}^2)}} = \frac{COV(X,Y)}{SD_X . SD_Y}$$

من العلاقة الأخيرة يمكن إعطاء تعريف آخر لمعامل الإرتباط : وهو عبارة عن النسبة بين التباين المشترك وجداء الإنحراف المعياري لكل من المتغير الأول والثاني .

أهم خصوصيات معامل الإرتباط :

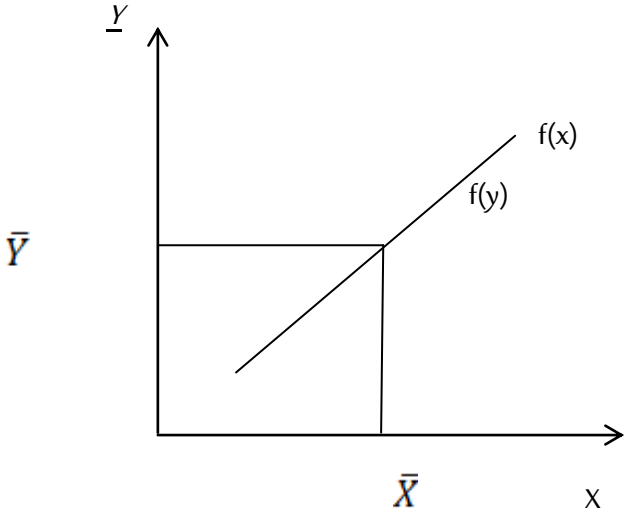
نرتكز في قياس وتفسير معامل الإرتباط على معادلتي خطي الانحدار ، حسب الحالات التالية:

- **الحالة الأولى :** عندما يكون خطي الانحدار f(y) ; f(x) متعامدين ، فإن سحابة النقاط الممثلة للقيم الحقيقة تكون مبعثرة جدا ولا يوجد أي إتجاه لها، ففي هذه الحالة لا توجد أيه علاقة بين المتغيرين، معناه وجود ارتباط معدوم r = 0 ، وذلك وفق العرض البياني التالي:



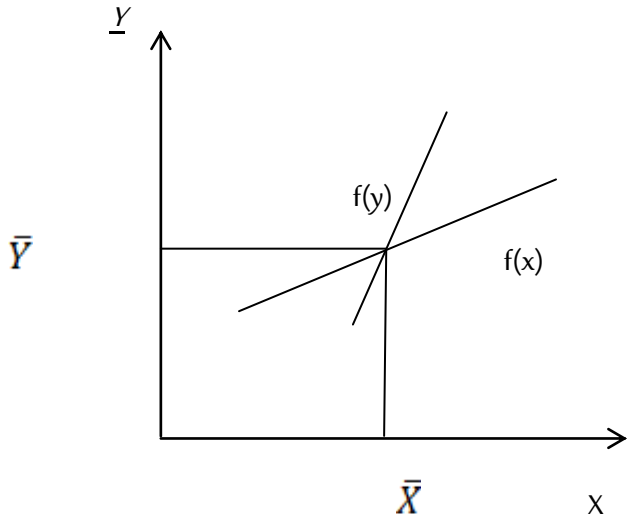
الشكل رقم :4.8

- **الحالة الثانية :** وجود تطابق بين خطي معادلتي الانحدار في هذه الحالة تكون النقاط الحقيقية لقيم f(y) ; f(x) على استقامة واحدة ، معناه توجد علاقة قوية _ ارتباط تام) بين المتغيرين.



الشكل رقم : 4.9

- **الحالة الثالثة :** يكون تقاطع خطي معادلتي الانحدار f(y) ; f(x) زاوية حادة، وتقع هذه الحالة بين الحاليتين السابقتين ، وتعتبر من الحالات العادية و الأكثر واقعية. و تنتمي قيم معامل الإرتباط إلى المجال :] 1+ ; 1- [$r \in$ ، وذلك حسب الشكل التالي:



الشكل رقم :4.10

النتيجة:

- عندما يكون r=0 : ارتباط معدوم (وجود علاقة بين المتغيرين).
- عندما يكون $r \in]-1;+1[$: وجود ارتباط نسبي (وجود علاقة نسبية عكسية أو طردية).
- عندما يكون $r = \pm 1$: وجود ارتباط تام (وجود علاقة تامة عكسية أو طردية).

مثال : يبين الجدول التالي قيمة الأصول والأرباح السنوية لخمسة مؤسسات بملايين الدينارات:

الجدول رقم :4.6

المؤسسات	1	2	3	4	5
الأصول X	10	20	30	40	50
الأرباح Y	1	3	2	5	4

المطلوب ما يلي :

1- حدد معلمات النموذج الخطي البسيط ؟

2- حدد معامل الإرتباط؟ ماذا تلاحظ؟

الحل:

1- تحديد معلمات النموذج الخطي البسيط: $Y_i = a + bX + U_i$

$$\bar{Y} = 3\ ;\ \bar{X} = 30\ ;V(Y) = 2;\ V(Y) = 200;COV(X,Y) = 16$$

$$b = \frac{cov(X,Y)}{V(X)} = \frac{16}{200} = 0.08\ ;a = 3 - 0.08 \times 30 = 0.6$$

$$Y = 0.6 + 0.08X$$

2- تحديد معامل الإرتباط :

$r = \frac{16}{\sqrt{200}\times\sqrt{2}} = 0.8$ نلاحظ أن العلاقة بين الأصول و الأرباح أي أن الأرباح متعلقة بالأصول نسبيا.

ملاحظة: لابد نميز بين العلاقة الانحدار و العلاقة الإرتباط:

- العلاقة الانحدار : نفرض أن لدينا العلاقة التالية $Y_i = a + bX_i + U_i$ ، نقول أن العلاقة بين Y_i و X_i هي علاقة انحداريه إذا كان المتغير المستقل يفسر المتغير التابع أو إذا كان تغير قيم المتغير المفسر يؤدي إلى تغير قيم المتغير التابع (Problème de regression) .

- أما عندما لا يمكن أن نميز بين متغيرين ولا يمكن تحديد بصفة واضحة المتغير التابع والمتغير المستقل لكنه في حالة مسألة ارتباطية (Problème de corrélation) .

توضيح: يمكن إستعمال القيم المعيارية للمتغيرات المدروسة لقياس الإرتباط الخطي، والقيمة المعيارية هي عبارة عن النسبة بين أي فرق لقيم المتغير الإحصائي بالنسبة للوسط الحسابي ومتوسط هذا الفرق (الإنحراف المعياري) وتكتب بالشكل التالي : $\frac{Y_i-\bar{Y}}{SD_Y}$; $\frac{X_i-\bar{X}}{SD_X}$، ولقياس درجة الإرتباط بينهما نستعمل متوسط جداء هذه القيم .

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{nSD_XSD_Y}$$

النوع الثاني : معامل الإرتباط الرتبي: (Coefficient de Corrélation de range)

يستخدم معامل الإرتباط الرتبي لدراسة الإرتباط في حالة متغيرات كيفية، حيث تستعمل رتبا تصاعديا أو تنازلية عوضا عن القيم العددية للمتغيرات المدروسة. فإذا كان الترتيب التصاعدي للمتغير المستقل يرافقه **ترتيب** تصاعدي للمتغير التابع نقول أن الإرتباط موجب (علاقة طردية)، أما إذا كان الترتيب التصاعدي للمتغير المستقل يرافقه ترتيب تنازلي للمتغير التابع نقول أن الإرتباط بين المتغيرين سالب(علاقة عكسية).

- **قياس الإرتباط الرتبي :** لقياس الإرتباط الرتبي بين مفردات المتغير التابع و المستقل نرتب كلاهما حسب أهمية خصائصهما ثم نحسب مجموع مربعات الفروق بين كل الرتب المتقابلة، لتصبح قياسات المتغيرين عبارة عن متوالية حسابية. ولاستخراج العلاقة الإحصائية لمعامل الإرتباط الرتبي نعوض في علاقة الإرتباط الخطي البسيط بقيم هاتين المتواليتين حسب التغيرات التالية:

1- مجموع n عددا مرتبا ترتيبا تصاعديا هو: $\sum X_i = \sum Y_i = \frac{n(n+1)}{2}$

2- مجموع مربعات هذه الأعداد هو: $\sum X_i^2 = \sum Y_i^2 = \frac{n(n+1)(2n+2)}{6}$

3- متوسط n عددا مرتبا ترتيبا تصاعديا هو: $\bar{X} = \bar{Y} = \frac{n+1}{2}$

4- أما تباين هذه القيم فهو :

$$V(X) = V(Y) = \frac{n(n+1)(2n+1)}{6n} - \left(\frac{n+1}{2}\right)^2 = \frac{n^2-1}{12}$$

5- مجموع مربعات الفروق بين رتب المتغيرين هو: $\sum d_1^2 = \sum (X_i - Y_i)^2$

حسب هذه المعطيات تصبح علاقة معامل الإرتباط الرتبي بالشكل التالي:

$$r = \frac{\frac{n(n+1)(2n+1)}{6} - \frac{\sum d_i^2}{2n} - (\frac{n+1}{2})^2}{\frac{n^2-1}{12}}$$

$$r = 1 - \frac{\frac{\sum d_i^2}{2n}}{\frac{n^2-1}{12}} = 1 - \frac{\sum d_i^2 . 12}{2n(n^2-1)}$$

إذن: $r = 1 - \frac{6.\sum d_i^2}{n(n^2-1)}$ ، تسمى العلاقة بعلاقة : Spearman

مثال: يبين الجدول التالي الدرجة التقديرية لستة طلبة في امتحاني الإحصاء و الإقتصاد .

المطلوب حساب معامل الإرتباط بين الدرجات التقديرية؟

الجدول رقم :4.7

رقم الطالب	1	2	3	4	5	6
الدرجة التقديرية في الإقتصاد	ضعيف	ممتاز	جيد	ضعيف جدا	مقبول	جيد جدا
الدرجة التقديرية في الإحصاء	مقبول	جيد جدا	جيد	ضعيف	ضعيف جدا	ممتاز

الحل :

أولا: نرتب الدرجات التقديرية ترتيبا تصاعديا.

ثانيا : نحسب مربع الفروق بين الدرجات التقديرية للمقياسين حسب الجدول التالي:

الجدول رقم: 4.8

الدرجة التقديرية في الإقتصاد	الدرجة التقديرية في الإحصاء	رتب الإقتصاد	رتب الإحصاء	d_i	d_i^2
ضعيف	مقبول	5	4	1	1
ممتاز	جيد جدا	1	2	-1	1
جيد	جيد	3	3	0	0
ضعيف جدا	ضعيف	6	5	1	1
مقبول	ضعيف جدا	4	6	-2	4
جيد جدا	ممتاز	2	1	1	1
/	/	/	/	/	8

نلاحظ أن العلاقة قوية نسبيا بين الدرجات التقديرية للمقاييس . $r = 1 - \frac{6\times 8}{6(36-1)} = 0.772$

رابعا : دراسة صلاحية النموذج:(Etude de la qualité du modèle)

تعتمد دراسة صلاحية النموذج الخطي البسيط على إختيار المتغير المستقل الذي يعطي تفسيرا معتبرا مقارنة مع بقية المتغيرات التفسيرية، ويرتكز هذا الإختيار على النظريات الموجودة في تخصص معين أو وفق دراسات تجريبية معينة. وتعتمد الدراسات الإحصائية على معامل الإرتباط و الإنحرافات المعيارية للمعلمات المقدرة، فكلما كانت قيم هذه الأخيرة صغيرة مقارنة مع قيم المعلمات المقدرة كلما كان النموذج أحسن والعكس صحيح.

أ- الشكل النهائي للنموذج الخطي البسيط: يتضمن الشكل النهائي لهذا النموذج على العناصر التالية :

1- تقدير معلماته.

2- تحديد معامل الإرتباط الخطي البسيط.

3- تحديد الإنحرافات المعيارية للمعلمات المقدرة.

تطرقنا في الفقرات السابقة إلى العنصر الأول والثاني ، أما الفقرة التالية فسنخصصها للعنصر الثالث.

ب- تحديد الإنحرافات المعيارية للمعلمات المقدرة:

1- الإنحراف المعياري لمعامل الانحدار: يعتمد أساسا على الفروق بين القيم الحقيقة والقيم المقدرة للمتغير التابع وعلى فروق قيم المتغير المستقل بالنسبة لوسطه الحسابي. وهو عبارة عن النسبة بين الإنحراف المعياري للمتغير

العشوائي (الفرق بين القيم الحقيقة والقيم المقدرة للمتغير التابع) والجذر التربيعي لمجموع مربعات فروق قيم المتغير المستقل لوسطه الحسابي ، وتكتب علاقته بالشكل التالي: $SD_b = \frac{SD}{\sqrt{\sum(X_i - \bar{X})^2}}$

t.q

$$SD_u = \sqrt{\frac{\sum(Y_i - Y_i^*)^2}{n-2}}$$

2- الإنحراف المعياري للمعامل a : يعطى شكل العلاقة الإحصائية كما يلي :

$$SD_u = \bar{X}.SD_b + \frac{SD_u}{n}$$

تمارين تطبيقية محلولة للفصل الرابع:

التمرين الأول: يبين الجدول التالي الوقت المحقق من طرف الحائزين على ميدالية ذهبية في سباق 100 م عند الرجال و النساء في الألعاب الأولمبية من سنة 1896 إلى سنة 1988.

الجدول رقم : 4.9

السنة	زمن : الرجال	زمن : النساء
1896	12	-
1900	11	-
1904	11	-
1908	10.8	-
1912	10.8	-
1920	10.8	-
1924	10.6	-
1928	10.8	12.2
1932	10.3	11.9
1936	10.3	11.5

1948	10.3	11.9
1952	10.4	11.5
1956	10.5	11.5
1960	10.2	11
1964	10	11.4
1968	9.95	11
1972	10.14	11.07
1976	10.06	11.08
1980	10.25	11.06
1984	9.99	10.97
198	9.92	10.54

Source : the Canadian World Almanac and Book of Facts; 1990, Toronto , Global Press , Canada.

المطلوب ما يلي:

1- حدد إتجاه هذه الظاهرة عند الرجال والنساء بواسطة معادلة خط الانحدار؟
2- ماذا تفكرون في المقولة التي تقول سيتحصل النساء في يوم ما على نتائج أحسن من الرجال في سباق 100 متر.

التمرين الثاني : أجب بنعم أو لا

1- يكون معامل الإرتباط بين متغيرين $\simeq 0$ إذا كانت توجد بينهما استقلالية؟
2- عندما يكون معامل الإرتباط الخطي سالبا معناه يوجد ارتباط ضعيف؟
3- إذا كان متغيران مرتبطان خطيا معنى ذلك أن تغير أحدهما هو نتيجة تغير الآخر؟
4- إذا كان لدينا n زوج من النقاط $(X_i; Y_i)$ وأضفنا لهما m نقطة كل منها يساوي $(\bar{X}; \bar{Y})$، فإن معامل الإرتباط لـ: n زوج من النقاط هو نفسه لـ: n+m ؟

التمرين الثالث : يبين الجدول التالي تطور الدخل والاستهلاك الشهري للأساتذة التربية البدنية والرياضية في زمن معين :

الجدول رقم: 4.10

الزمن	t_0	t_1	t_2	t_3	t_4	t_5
الرقم القياسي لأسعار الاستهلاك (%)	100	90	105	100	120	130
الدخل الشهري	80	90	80	95	100	105

إذا كانت الكمية المستهلكة في فترة الأساس t_0 هي 15 كلغ وثمن كلغ ويساوي 10د.د .

المطلوب ما يلي:

1- أدرس العلاقة بين الدخل و الاستهلاك والدخل إن وجدت؟
2- ماهي نسبة تفسر الدخل الشهري للاستهلاك؟
3- قدر قيمة استهلاك الفترة t_7 إذا توقعنا أن الدخل الشهري سيصبح 110 د.ج؟

التمرين الرابع : يبين التوزيع التكراري التالي علامات 10 طلبة في امتحانين لمقياس الإحصاء الوصفي:

الجدول رقم: 4.11

البيان	1	2	3	4	5	6	7	8	9	10
X(الامتحان الأول)	6	5	8	8	7	6	10	4	9	7
Y(الامتحان الثاني)	8	7	7	10	5	8	10	6	8	6

المطلوب ما يلي:

1- ضع شكل الانتشار ؟
2- أوجد معادلة خط الانحدار : $f(y)$; $f(x)$؟
3- أرسم خطي الانحدار ؟ ماذا تلاحظ؟
4- قس قوة الإرتباط؟
5- حدد الشكل النهائي للنموذج؟

التمرين الخامس : تظهر البيانات التالية العلاقة الموجودة بين السن(X) وضغط الدم (Y)لإثني عشرة لاعب :

الجدول رقم:4.12

ضغط الدم Y	147	125	160	118	142	128	150	145	125	140	152	155
السن X	56	42	72	36	63	47	55	49	38	42	68	60

المطلوب ما يلي:

1- أوجد معادلة خط الانحدار f(X) باستخدام طريقة المربعات الصغرى؟
2- قس قوة الإرتباط الخطي بين المتغيرين؟
3- قدر ضغط الدم للاعب و عمره 45 سنة؟

حل تمارين للفصل الرابع:

التمرين الثاني :

1- نعم ، 2- لا ، 3- لا، 4- نعم.

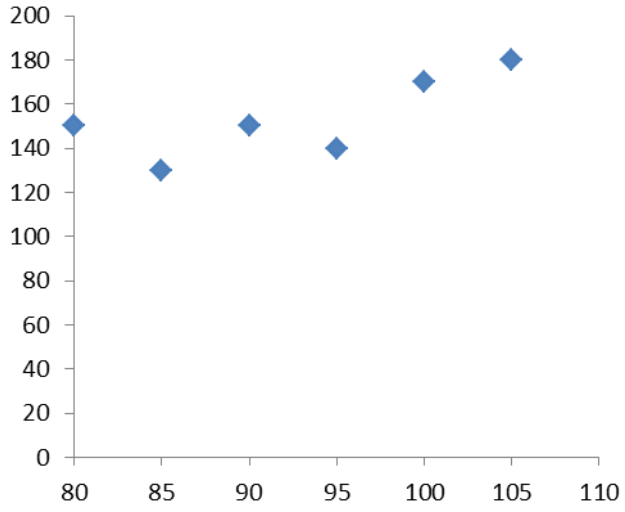
التمرين الثالث :

تتطلب الإجابة عن السؤال الأول تحديد الاستهلاك الشهري لكل الفترات :

ج:4.14

t_0	t_0	t_1	t_2	t_3	t_4	t_5
C_t	150	135	157.5	150	180	195

لدراسة العلاقة بين المتغيرين، نحدد أولا شكل العلاقة بينهما بيانيا:



الشكل رقم :4.11

. من العرض البياني ، يمكن اعتبار أن العلاقة طردية خطية بين المتغيرين ، لأن إتجاه كوكبة النقاط هو إتجاه خطي :

$$C_t = a + bR_t + U_t$$

. تحديد معلمات النموذج :

$$\bar{C} = 161.25 \; ; \; \bar{R} = 91.666 \; ; \sum C_t . R_t = 89475 \; ; \sum C_t^2 = 158456.25$$

أولا : معامل الانحدار: $b = \frac{\frac{89475}{6} - 161.25 \times 91.666}{\frac{50950}{6} - 8401.55} = 1.476$

ثانيا: الثابت a: $a = 161.25 - 1.476 \times 91.66 = 25.97$

ثالثا: معادلة خط الانحدار: $C_t = 25.97 + 1.476R_t$

. تحديد نسبة تفسير الدخل الشهري للاستهلاك ، إيجاد معامل الإرتباط

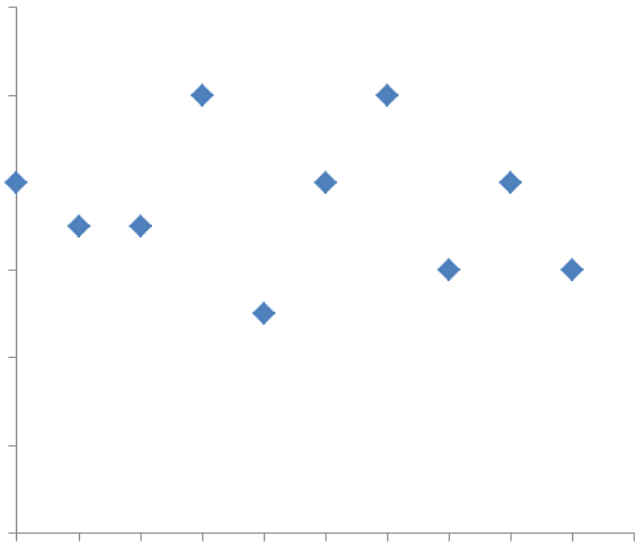
$$r = \frac{\frac{89475}{6} - 161.25 \times 91.666}{9.434 \times 20.2} = 0.689 \cong 69\%$$

. تحديد أو تقدير قيمة الاستهلاك للفترة t_7 : .

$$C_7 = 25.97 + 1.476 \times 110 = 188.3$$

التمرين الرابع :

1- تحديد شكل كوكبة النقاط:



الشكل رقم :4.12

نلاحظ من العرض البياني أن العلاقة بين علامات الامتحانين ضعيفة نسبيا لأن شكل كوكبة النقاط موازيا تقريبا لمحور السينات.

2- تحديد معادلتي خطي f(x) ; f(y):

أولا : معادلة خط الانحدار f(x) :

$$\bar{Y} = 7.5 \; ; \; \bar{X} = 7 \; ; \; \sum Y_i X_i = 540 \; ; \; \sum X_i^2 = 520 \; ; \; \sum Y_i^2 = 587$$

$$a_1 = 7.5 - 0.5 \times 7 = 4\ , b_1 = \frac{\frac{540}{10} - 7 \times 7.5}{\frac{420}{10} - 49} = 0.5$$

$$f(x) = Y = 4 + 0.5 \times X$$

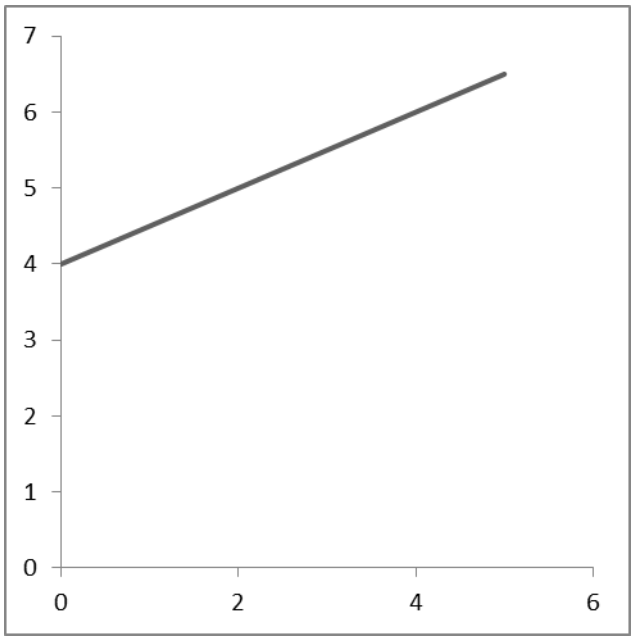
ثانيا: معادلة خط الانحدار f(y):

$$a_2 = 7 - 0.62 \times 7.5 = 2.41\ , b_2 = \frac{54 - 52.5}{\frac{587}{10} - 56.25} = 0.612$$

$$f(x) = X = 2.41 + 0.612 \times Y$$

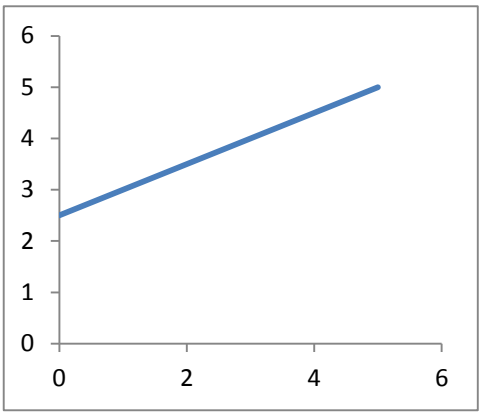
3- رسم خطي الانحدار:

. خط الانحدار f(x):



الشكل رقم: 4.13

- خط الانحدار f(y):



الشكل رقم: 4.14

4- تحديد معامل الإرتباط: $r = \sqrt{b_1 \times b_2} = \sqrt{0.5 \times 0.612} = 0.55$

5- تحديد الشكل النهائي للنموذج f(y):

أ- تحديد قيمة المقدرة لعلامات الامتحان الثاني (Y^*) ، وحساب الفرق بين القيم الحقيقية والقيم المقدرة:

الجدول رقم : 4.15

Y^*	U_i	U_i^2	$(X_i - \bar{X})^2$
7	-1	1	1
6.5	-0.5	0.25	4
8	1	1	1
8	-2	4	1
7.5	2.5	6.25	0
7	-1	1	1
7	-1	1	9
6	0	0	9
8.5	0.5	0.25	4
7.5	1.5	2.25	0
$\sum$		17	30

ب- تحديد القيمة المقدرة للانحراف المعياري للمتغير العشوائي U :

$$SD_u = \sqrt{\frac{\sum(Y_i - Y^*)^2}{n-2}} = \sqrt{\frac{17}{10-2}} = 1.46$$

ج- تحديد القيمة المقدرة للانحراف المعياري لمعامل الانحدار:

$$SD_u = \frac{SD_u}{\sqrt{\sum(X_i - \bar{X})^2}} = \frac{1.46}{\sqrt{30}} = 0.266$$

د- تحديد القيمة المقدرة للانحراف المعياري للمعامل a :

$$SD_u = \bar{X}.SD_b + \frac{SD_u}{\sqrt{n}} = 7 \times 0.266 + \frac{1.46}{3.16} = 1.92$$

إذن الشكل النهائي للنموذج هو: $r = 0.55$; $Y = 4 + 0.5 \times X$

(1.92) (0.266)

ملاحظة: إن معامل الإرتباط بين درجات الامتحان الثاني و الامتحان الأول ضعيف نسبيا، وه1ا راجع لوجود عوامل أخرى تؤثر في علامات الامتحان الثاني.

كما أن الإنحرافات المعيارية للمعلمات المقدرة كبيرة نسبيا مقارنة مع قيم هذه المعاملات إذن يعتبر تفسير X لـ Y تفسيرا ضعيفا نسبيا.

التمرين الخامس :

1- إيجاد معادلة خط الانحدار f(X):

$$b = \frac{129.53}{129.57} = 0.99 \; ; a \; 140.58 - 0.99 \times 52.33 = 88.32$$

$$y = 88.32 + 0.99 \times X$$

2- تحديد معامل الانحدار:

$$r = \frac{\frac{89833}{12} - 52.33 \times 140.58}{\sqrt{21942.42}} = 0.874$$

3- تقدير ضغط الدم لاعب عمره 45 سنة:

$$Y = 88.302 + 0.99 \times 45 = 133.25$$

4- تحديد الشكل النهائي للنموذج: نستعمل نفس طريقة التمرين السابق.

$$SD_U = 6.93; \; SD_b = 0.176 \; ; SD_a = 9.42$$

$$Y = 88.302 + 0.99 \times X \; ; \; (0.874)$$

(9.42) (0.176)

ملاحظة: يفسر المتغير التابع بشكل مقبول للأسباب التالية:

- الإرتباط قوي نسبيا.
- الإنحرافات المعيارية للمعلمات المقدرة صغيرة مقارنة مع قيم المعلمات نفسها .

المصادر والمراجع :

1- احمد سليمان عودة وخليل يوسف الخليلي : الاحصاء للباحث في التربية والعلوم الانسانية ،ط2،دار الامل ،الاردن ، 2000.

2- احمد فاروق عبد العظيم: الاحصاء ،المكتب الجامعي الحديث ،الاسكندرية ،1984.

3- اسماعيل محمد بن قانة : الاحصاء الوصفي والحيوي ،دار اسامة للنشر والتوزيع ،الاردن ،2011.

4- انيس كانجو : الاحصاء وطرق تطبيقه في ميادين البحث العلمي ،مؤسسة الرسالة ،بيروت ،1980.

5- بوعلاق محمد : الموجه في الاحصاء الوصفي والاستدلالي ،دار الامل للطباعة والنشر والتوزيع ،الجزائر ،2009.

6- جلاطو جيلالي : الاحصاء مع تمارين ومسائل محلولة ،ط5،ديوان المطبوعات الجامعية ،الجزائر ،2005.

7- حسن ياسين طعمة واخرون : طرق الاحصاء الوصفي ،دار صفاء للنشر والتوزيع ،الاردن ،2009.

8- عبد القادر حلمي : مدخل الى الاحصاء ،ط2،ديوان المطبوعات الجامعية ،الجزائر ،2005.

9- عبد الله فلاح المنيزل واخرون : الاحصاء التربوي ،ط4، دار الميسرة للنشر والتوزيع والطباعة ،الاردن ،2010.

10- محسن لطفي احمد : مقدمة في الاحصاء الاجتماعي ،النشر العلمي والمطابع ، المملكة العربية السعودية ،2011.

11- محمد خير و سليم ابوزيد: التحليل الاحصائي للبيانات باستخدام برمجة ،دار صفاء للنشر والتوزيع ،عمان ،2018.

12- محمد صبحي ابو صالح : مبادئ الاحصاء ،ط7،دار اليازوري العلمية للنشر والتوزيع ،الاردن ،2007.

13- محمد صبحي ابو صالح واخرون : مقدمة في الاحصاء ،ط9، دار الميسرة للنشر والتوزيع والطباعة ، الاردن ،2015.

14- مصطفى زايد : الاحصاء ووصف البيانات ،دار الكتب ،مصر ،1998

15- نجيب حسن محمد واخرون: المدخل الى علم الاحصاء ،دار شموع الثقافة للطباعة والنشر والتوزيع ،ليبيا ،2004.