الجامعـة الافتراضيّة السوريّة

SYRIAN VIRTUAL UNIVERSITY

الإحصاء و الاحتمالات

د.منذر عواد - د.حسام كمرجي

ISSN: 2617-989X

Books & Refrences

الإحصاء و الاحتمالات

د.منذر عواد - د.حسام كمرجي

من منشورات الجامعة الافتراضية السورية

الجمهورية العربية السورية 2020

Probability &Statistics

Monzer Awad – Housam Kamarji

Publications of the Syrian Virtual University (SVU)

Syrian Arab Republic, 2020

جدول المصطلحات والكلمات المفتاحية:

الفهرس

تمهيد (الاحصاء واهميته):

تعتبر عملية اتخاذ القرار على كافة المستويات و المجالات ، سواء كانت سياسية او اقتصادية او ادارية او عسكرية او تربوية ، من العمليات المتكررة الحدوث كثيرا . في اغلب الاحيان تكون هذه القرارت هامة جدا سواء كانت على مستوى فرد او مؤسسه او دولة ، حيث يكون لهذا القرار تاثير على قرارات اخرى او اعمال اخرى .من هنا يجب توخي الحذر و بذل اقصى جهد ممكن قبل اتخاذ مثل هذه القرارات . و براينا ان القرار الصحيح هو ذلك القرار المبني على دراسة و تحليل الواقع بشكل علمي و منطقي صحيحين. ان هذا التحليل للواقع يحتاج الى معلومات و التي بدورها تستنتج من البيانات . فالبيانات هي المادة الخام التي تستنتج منها المعلومات.

نحصل على المعلومات لاتخاذ قرار ما او لاجراء بحث ما من معارفنا المتراكمة حيث ان المعرفة هي تراكم للمعلومات ،ولكن نادرا ما يلم شخص ما بكافة المعلومات و عن كل شيء فهنا هو بحاجة الى جمع البيانات عن الموضوع او المشكلة قيد البحث ليحصل على معلومات . يتضح ان الغاية الاساسية اذا هو ليس جمع البيانات ولكن الحصول على معلومات لاتخاذ القرار و هذا يتم عن طريق تحليل البيانات التي تم جمعها و ترتيبها بشكل يمكن من تحليلها ، من هنا نرى ان تلك االعملية تمر بعدد من الخطوات و التي هي بحد ذاتها خطوات البحث العلمي.

1 _ تحديد المشكلة

2 _ جمع البيانات الاحصائية عن المشكلة

3 _ عرض البيانات الاحصائية

4 _ تحليل البيانات واتخاذ القرار

كما يمكن اعتمادا على تلك الخطوات تعريف الاحصاء بانه العلم الذي يقدم البيانات لاتخاذ القرار و ذلك من خلال جمع البيانات العددية و عرضها و تحليلها .وبالتالي يمكن ان نعرف الاحصاء ايضا بانه تعلم لغة البيانات من اجل فهم ما تقوله البيانات او من اجل ترجمتها لمن لا يعرف هذه اللغة .

اكتسب علم الاحصاء هذه التعاريف الموسعة نتيجة ادخال استخدام الرياضيات و نظرية الاحتمالات والرسوم البيانية و الحاسب الالكتروني في الاحصاء .

ليس من الضروي اننا نستخدم الاحصاء بالبحث العلمي وحسب ،فهو يستخدم من قبلنا وبشكل مستمر دون ان نسميه احصاء .من التصرفات التي يقوم بها كل انسان يوميا هو قيامه بالتسوق ، فهنا عليه القيام بدراسة السوق بجمع البيانات عن السلعة واستعراض هذه البيانات واجراء مقارنات عدة للوصول للقرار بالشراء .هذا هو الاحصاء .

من الضرورة التمييز بين الاحصاء الوصفيDescriptive statistics والاستدلالي Inferential statistics.يتناول الاول كيفة جمع وعرض ووصف البيانات ،في حين يستخدم الثاني لتعميم نتائج العينة على المجتمع ككل .

سوف نتناول في هذا الكتاب وبشيء من التفصيل عملية جمع البيانات العددية و عرضها وكيفة حساب الاحتمالات ، اما التحليل فستتناول جزء منه في هذا المقرر ،ويتضمن وصف البيانات العددية بيانيا وجدوليا ورقميا، كما يتضمن المقرر دراسة الارتباط والانحدار والارقام الفياسية وجاء ذلك في عشرة فصول، قام د.منذر العواد بكتابة الفصول الاول ،الثاني ،الثالث ،الرابع التاسع ، العاشر.اما الفصول الخامس ، السادس ،السابع، و الثامن كتبها د. حسام كمرجي .

الفصل الاول

تعاريف و مفاهيم احصائية هامة

Essential statistical definitions and concepts

الاهداف والمخرجات التعليمية :

بعد الانتهاء من دراسة هذا الفصل سيكون الطالب قادرا على:

1- إدراك الفرق بين :احصاء وصفي واستدلالي، مجتمع وعينة ،بيانات اولية وثانو

2- التمييز بين انواع المتغيرات.

3- قادرا على سحب العينات .

4- جمع البيانات العددية .

ملخص الفصل:

يتناول هذا الفصل التعريف بعلم الإحصاء والتعريف بالمفاهيم والمصطلحات الاحصائية الاساسية ، كما يتناول هذا الفصل الخطوات العملية الاولى الواجب القيام بها للبدء ببحث علمي : تحديد المشكلة والمتغيرات وطرق الجمع ومصدر البيانات واسلوب الجمع ،وجمع البيانات.

جمع البيانات العددية Collection Of Data

تعتبر عملية جمع البيانات العددية من اكثر المراحل كلفة ووقتا حيث ان المراحل التالية يمكن ان تنجز بوقت بسيط جدا باستخدام الحاسب الالكتروني لذلك يجب في هذه المرحله بذل اقصى درجات الجهد و العنايه لتكون البيانات التي سنقوم بجمعها صحيحة وبالتالي تكون مراحل البحث التالية صحيحة و من ثم القرار المتخذ صحيحا.

و اننا نرى للحصول على بيانات صحيحه ضرورة الاعتماد على بعض الاسس و التي هي عبارة عن عدد من التعاريف والاجراءات والتي سنتعرف عليها بالتفصيل فيما يلي .

1.1 _المتغيرات Variables :

يعرف المتغير بانه الصفة او الخاصية و التي يمكن ان تتغير من مفردة الى اخرى سواء كان بالقيم او بالشكل...

فلكل شيء من حولنا (سنطلق عليه اسم مفردة) عدد لا نهائي من الصفات و بنفس الوقت يمكن ان تختلف أي من هذه الصفات من مفردة لاخرى، فلمجموعة من الناس نجد المتغيرات (الصفات) التالية: انهم بشر – يختلفون من حيث الطول و الوزن و الجنس و مستوى المعيشة ومكان الاقامة و متوسط الانفاق و متوسط الدخل ولون الشعر ولون البشرة ..... .

يتضح من هذا ضرورة التعرف على المتغيرات والتفريق بينها من اجل جمع البيانات . فبعد التعرف على المتغير يقوم الباحث بجمع البيانات عن المفردات التي تتمتع بهذه الصفة و يستبعد المفردات الاخرى و بالتالي يكون قد اختصر البيانات الى البيانات الضرورية .

تقسم المتغيرات الى قسمين :

1.1.1 متغيرات نوعية او وصفية Qualitative Variables:

هي تلك المتغيرات التي يعبر عنها بكلمة او جملة او رمز مثل ذكر و انثى والالوان – طاولة _ مدرسة _ جامعة_ .....الخ، او هي تلك المتغيرات التي تتمتع بها المفردة او لا و لا يمكن لها ان تتمتع بها بدرجات مختلفة .

2.1.1 متغيرات كمية Quantitative Variables:

المتغيرات الكمية هي تلك الصفات التي يمكن قياسها بوحدات القياس الوزن _ الطول _ الحجم _ الزمن _ القيم النقدية ..... ويمكن القول هي تلك المتغيرات التي تتمتع بها المفردات بدرجات مختلفة. و تقسم الى قسمين :

_ متغيرات مستمرة **Continuous Variables**:

وذلك عندما ياخذ المتغير أي قيمة ضمن مجال معين فطول انسان بالغ يمكن ان ياخذ أي قيمة ضمن المجال 150 سم و 200 سم ،وذلك بسبب وجود اجزاء ،و كذلك طول شخص ووزنه وعمره ودخله وانفاقه كلها متغيرات كمية مستمرة .

_ متغيرات منقطعة **Discrete Variables** :

و ذلك عندما ياخذ المتغير عدد محدود من القيم فعدد افراد الاسرة هو 2 او 3 و عدد الاطفال في الاسرة 0، 1، 2....... هو متغير منقطع ،لاحظ ان هذا المتغير قيم صحيحة .
و اخيرا نقول يمكن ان تكون المتغيرات بسيطة عندما يعبر عنها بصفة واحدة و هنا يمكن ان نستخدم الوحدات التالية :
1- وحدات العد   2 - وحدات القياس   3 - الوحدات النقدية
و يمكن ان تكون المتغيرات مركبة و ذلك عندما نعبر عنها باكثر من صفة مثل سرعة السيارة في وحدة الزمن و استهلاك السيارة في وحدة المسافة ..الخ.

2.1 - المجتمع الاحصائي Population والعينة Sample والوحدة الاحصائية :Unit

المجتمع بالاحصاء هو جميع المفردات التي تتمتع بصفة ما او خاصية (متغير) ما و هذه المفردات قد تكون بشرا او اشياءا او ظواهر و هذا مايميز المجتمع الاحصائي من المجتمع بالمعنى الاجتماعي و الذي هو مجموعة من الناس تربط بينهم علاقة ما . قبل البدء بجمع البيانات العددية عن المجتمع لا بد من تحديد المجتمع الاحصائي حتى نحصل على البيانات الضرورية فقط ، و يتحدد المجتمع الاحصائي بتحديد المكان و الزمان معا و عدا ذلك لا يتحدد المجتمع الاحصائي و بهذا نرى ان مفهوم المجتمع الاحصائي مفهوما مرنا حيث يمكن ان يوسع المجتمع او العكس، بتوسيع المكان او الزمان او كليهما و كامثلة على المجتمع الاحصائي : اسر مدينة دمشق عام 2015 و طلاب التعليم الجامعي عام 2019
و طلاب جامعه دمشق عام 2018. نرى من جميع هذه الامثلة ان هناك تحديدا للمجتمع من حيث الزمان و المكان .
- قد يكون المجتمع الاحصائي محدودا Finite و ذلك عندما نستطيع حصر عدد مفرداته و بالتالي نستطيع جمع بيانات عنها مثل طلاب جامعة دمشق _ سكان مدينة دمشق .
_ وقد يكون المجتمع الاحصائي غير محدود Infinit و ذلك عندما لانستطيع حصر عدد مفرداته : عدد اسماك البحر الابيض المتوسط او الاحتياطي من النفط او انتاج الة المفترض انتاجه ، وهذه المجتمعات يطلق عليها عادة تسمية مجتمعات نظرية Hypothetical populations.
اما العينة فهي جزئ من المجتمع الاحصائي تتم دراسته للتعرف على المجتمع الكبير كما سنرى لاحقا ،في حين ان الوحدة الاحصائية او المفردة هي اصغر جزئ يتمتع بالصفات او المتغيرات موضوع البحث .

3.1 تحديد المشكلة :

الغاية من جمع البيانات هو الوصول لقرار وعادة هو حل مشكلة بحثية او ادارية او....فعند تحديد المشكلة بشكل صحيح ودقيق يقتصر جمع البيانات على البيانات الضرورية والمتعلقة بالمشكلة ،مثلا لمعرفة نسبة المدخنين في الجامعة ،هنا المشكلة واضحة ومحددة بدقة وبالتالي تقتصر مهمتنا على جمع البيانات المتعلقة بصفة واحدة هي التدخين .

4.1 _ الدراسة التحضرية:

يقصد بالدراسة التحضيرية الاطلاع المسبق على كل الجوانب المتعلقة بالمشكلة موضوع الدراسة للتعرف على مصطلحاتها الفنية و نشأتها التاريخية و تعتبر هذه العملية هامة حيث تساعد في اختصار البيانات
واقتصارها على الضرورية و الهامة منها و ذلك على النحو التالي :
قد يجد الباحث ان هناك من قام بهذه الدراسة سابقا و توصل الى نتائج و ما عليه سوى تطوير هذه النتائج ، قد يجد الباحث ان هناك من اكد استحالة هذا البحث و بالتالي يمكن الاستفادة من هذه النتيجة.
بالاضافة لذلك تساعد الدراسة التحضرية بالتعرف على المشكلة بشكل اكثر و التعرف على المصطلحات الفنية الخاصة بها .

5.1 – مصادر البيانات Data Sources :

يمكن تقسيم البيانات حسب مصادرها الى قسمين : بيانات ثانوية و بيانات الاولية .

1.5.1- البيانات الاولية Primary Data :

هي البيانات التي يجمعها الباحث بنفسه من الميدان ، وتمتاز هذه البيانات بملائمتها لمشكلة البحث ،ولكنها تحتاج الى وقت وجهد وتكاليف .

2.5.1- البيانات الثانوية Secondary Data :

هي كل البيانات المتوفرة حول المشكلة موضوع البحث بشكل جاهز منشورة او غير منشورة ورقية او الكترونية و يمكن الحصول عليها من مصادر غير مباشرة ،أي هي عبارة عن جميع البيانات المتوفرة والمعدة مسبقا من قبل الغير كالإحصائيات التي تعدها الوزارات المختلفة في كل دولة والإحصائيات التي تعدها الدول مثل المجموعة الاحصائية السنوية الصادرة عن المكتب المركزي للاحصاء في الجمهورية العربية السورية .وتمتاز هذه البيانات بقلة الكلفة والوقت اللازم للحصول عليها ،أما مساوئها فتتلخص بان هذه البيانات عادة ما تجمع لغرض مختلف عن موضوع الدراسة وبالتالي لابد من اجراء بعض التعديلات عليها اذا امكن ذلك ،كما ان الثقة بهذه البيانات تكون حسب الجهة المعدة لها.

6.1–اسلوب جمع البيانات Methods Data Collection :

بعد ان يحدد الباحث مصدر البيانات لبحثه يجب ان يحدد الاسلوب الواجب استخدامه لجمع البيانات . وتجمع البيانات باحد اسلوبين هما :اسلوب الحصر الشامل واسلوب العينة ،واعتمادا على مزايا ومساوئ كل اسلوب و ظروف كل بحث يختار الباحث الاسلوب المناسب . لذلك سنتناول ادناه وبشئ من التفصيل كلا الاسلوبين .

1.6.1–اسلوب الحصر الشامل Survey:

حسب هذا الاسلوب نقوم بجمع البيانات العددية عن كافة مفردات المجتمع الاحصائي ويستخدم هذا الاسلوب عادة في التعدادات العامة مثل تعداد السكان Issue.وان مزايا هذا الاسلوب هي دقة النتائج حيث نحصل على بيانات عن المجتمع الاحصائي كما هو فعلا باستثناء بعض الاخطاء الحسابية ،بالمقابل فان لهذا الاسلوب بعض المساوئ هي ان الجمع حسب هذا الاسلوب يحتاج لوقت طويل وتكاليف عالية عندما يكون حجم المجتمع كبيرا.

2.6.1–اسلوب العينة Sample :

العينة هي جزء يحوي عدة مفرادات من المجتمع و يحدد عدد هذه المفردات حسب قواعد علمية سنمر على ذكرها في مقررات اخرى ،و نقوم بدراسة العينة للحصول منها على معلومات عن المجتمع الاحصائي الذي سحبت منه العينة ، ومن الضروري حتى تكون هذه المعلومات صحيحة عن المجتمع ،ان تكون العينة ممثلة لمجتمعها تمثيلا صحيحا أي تحوي على نفس الصفات الموجودة بالمجتمع الاحصائي و بنفس النسب ، و يمكن الحصول على هذه العينة باستخدام الاساليب العلمية عند سحبها و هذا ما سنعرضه ادناه .

3.6.1 – اسلوب الحصر الشامل ام اسلوب العينة .

هناك عدة اعتبارات يجب اخذها بعين الاعتبار لاختيار اسلوب الجمع الواجب اتباعه :

– لكل بحث وقت لانجازه و تكاليف مخصصة له فاعتمادا على الوقت المخصص و التكاليف يمكن للباحث ان يقرر او يختار الاسلوب المناسب مع الاخذ بعين الاعتبار ان اسلوب الحصر الشامل يعطي نتائج اكثر دقة .

– عندما يكون المجتمع الاحصائي غير محدود او افتراضي فمن الصعوبة الوصول الى كل مفرداته لدراستها مثل تحليل ماء البحر فهنا يجب استخدام اسلوب العينة في الدراسة .

– اذا كانت الدراسة ستؤدي الى تلف الوحدات موضوع الدراسة يجب استخدام اسلوب العينة ، مثلا تحليل دم الانسان او فحص شحنة مستوردة من السمك المعلب فان الدراسة باسلوب الحصر الشامل معناه سحب دم الشخص بشكل كامل و فتح كل علب الشحنة المستوردة من المعلبات و بالتالي نكون في كلا الحالتين قد اتلفنا و حدات المجتمع .

7.1 – تقنيات المعاينة Sampling Techniques:

عرف استخدام العينات منذ القدم ، ادى و يؤدي هذا الاسلوب الغرض في كثير من الاحيان و بنجاح دون العودة لاسلوب الحصر الشامل ،من المشاهد اليومية لاستخدام اسلوب العينة، فحص الزمرة الدمويه و تناول الادوية و كل الفحوصات المخبرية و النماذج التسويقية......الخ . تقسم العينات الى نوعيين :

- **عينات عشوائية** Random Samples.
- **عينات غير عشوائية** Non-Probability Samples .

ان الغاية من سحب العينات هو دراسة المجتمع بشكل سهل ،ولتحقيق هذه الغاية يجب تكون العينة ممثلة لمجتمعها وذلك عندما تحتوي على كافة الصفات الموجودة في المجتمع و بنفس النسب، يمكن ان يتحقق هذا اذا كانت العينة عشوائية و المقصود بالعشوائية هنا ان تعطى لكل مفردة من مفردات المجتمع الاحصائي حظاً متساويا للظهور بالعينه . يتم هذا باستخدام اساليب علمية عند السحب تدعى باساليب السحب العشوائي او الاختيار العشوائي Random Selection.

قبل ان نستعرض كيفية السحب بهذه الاساليب لنلاحظ الفرق بين الامثلة التالية : عند اخذ عينة صغيرة من دم الانسان تعتبر هذه العينة ممثله لمجتمعها و نعطي حكما عن المجتمع وهو دم هذا الانسان من خلال هذه العينة ،مثال اخر :هناك بعض الادوية كتب عليها يرجى رج الزجاجة قبل تناول الدواء ،والغاية من هذه العملية هي تحقيق التجانس بالمحلول ككل وعندما ناخذ جزء منه يكون ممثلا لمجتمعه .

لو اردنا فحص نسبة المعدن في فلز ما فيتوجب علينا طحن الفلز وخلطه جيدا حتى نحصل على عينة ممثلة لمجتمعها . لكننا نصادف في حياتنا العملية حالات لايمكن اجراء عمليات الطحن والرج والخلط عليها لتحقيق التجانس في المجتمع ، مثل سحب عينة من الاسر او الاشخاص او السيارات ،فهنا لابد من استخدام اساليب

تقنية وعلمية لتحقيق هذا التجانس اولا وتحقيق عشوائية العينة ثانيا وهي ماندعوه باساليب السحب العشوائي وهي :

1- اسلوب القرعة :

تتم عملية اجراء القرعة كما يلي: نقوم بترقيم مفردات المجتمع N..........1-2-3 **حيث N هي حجم المجتمع وبعدها نقوم بكتابة هذه الارقام او اسماء المفردات اذا امكن على قصاصات من الورق وتوضع في كبسولات متماثلة ، كل ورقة في كبسولة وقبل البدء في السحب تخلط الكبسولات جيدا في وعاء ونبدأ بسحب الكبسولات واحدة تلو الاخرى مع تكرار الخلط قبل كل عملية سحب وهكذا حتى نحصل على عددا من الكبسولات مساو لحجم العينة المطلوب n.**

يمكن استخدام اسلوب القرعة وبنجاح كبير اذا كان حجم المجتمع الاحصائي صغيرا ،اما عندما يكون حجم المجتمع كبيرا جدا فانه من الصعوبة تطبيق هذا الاسلوب .

2- دواليب الحظ :

لايختلف هذا الاسلوب بجوهره عن الاسلوب السابق ولكنه اسهل بكثير . نقوم بترقيم مفردات المجتمع 1-2-3........N **حيث** N **هي حجم المجتمع ونبدا بالسحب بتدوير الدواليب وعندما تتوقف الدواليب نقرأ الرقم فتكون المفردة في المجتمع الاحصائي التي تحمل ذلك الرقم هي المفردة التي ستدخل في العينة وهكذا نكرر عملية تدوير الدواليب** n **مرة بحجم العينة المطلوب سحبها .**

لقد اصبح هذا الاسلوب اكثر تطبيقا في وقتنا الحاضر بسبب انتشار الحاسبات الالكترونية حيث اغلب البرامج تحوي على هذه الدواليب الالكترونية او تعمل بشكل مماثل للدواليب .

3- جداول الارقام العشوائية :

من الطرق القديمة للسحب العشوائي ،وهي عبارة عن جداول في كل صفحة عدد من الاسطر والاعمدة للارقام من 0الى 9 ،وهذه الجداول معدة بشكل عشوائي .يكفي لاستخدام الجدول ان نبدا باي سطر او عمود ونبدا قراءة الارقام وكاننا ندور دواليب الحظ.

مثال **:لسحب عينة غشوائية بحجم 19 عمال ، من عمال احد الشركات والبالغ عددهم 200 عامل (حجم المجتمع).باستخدام الجدول المرفق ادناه (1.1) ،مبتدئين بالعمودين الاول والثاني من اليسار يكون افراد العينة هم اصحاب الترتيب التالي :95 49 75 58 39 .**

جدول الأرقام العشوائية

9588	0198	3820	0894	6705	2742	4569
4900	7466	7496	2909	4978	3722	3079
7521	2725	3902	4063	0949	8802	7023
5843	8325	0528	9800	2068	9029	9665
3949	5233	7119	8639	8131	2475	7942
8715	3299	6692	9803	4528	3506	4945
8401	7484	8384	0224	1242	3068	3077
9207	0149	8058	6870	8529	7588	6798
7736	7574	8588	1319	1872	8386	2689
6520	7841	7373	6333	0333	7685	6023
4165	7946	3533	1031	3257	7900	7313
4838	1276	1246	7801	6225	0457	4350
7947	9142	2410	4292	9920	7496	3669
0253	3400	6426	1920	4342	1146	5190
8348	7762	5936	2109	1036	0439	5598
6720	7203	2655	2512	6737	4896	0476
0628	4188	5390	4364	9341	5278	5372
3661	2808	6778	9171	3486	0070	8395
9784	8036	3352	4960	3833	1634	2902
6119	7492	6927	0885	5729	5241	3936
7482	6518	9499	2966	7333	3457	5776
5938	9499	8154	6110	2970	2117	8560
4725	7083	3201	7851	0122	0435	5401
4298	3395	5982	9762	3821	6014	4061
8745	3018	6243	7117	6795	7009	4831
2084	5778	1475	9227	6156	7372	7152
2578	7162	3083	1147	2882	2377	9183
0045	7556	2820	1923	6061	6569	1805
8138	7994	0808	5852	3319	8838	7975
4187	5360	9691	0736	2104	5497	3998
9239	6149	2790	8068	0064	5110	2736
6640	0233	9391	4707	6740	8089	4086
5314	1778	0038	4353	3054	3854	7375
1982	4704	4771	5948	5567	8559	6118

جدول (1.1)

1.7.1 – انواع العينات العشوائية

1 – العينة العشوائية البسيطة **Simple Random Sample:**

تعتبر عملية سحب هذه العينة الابسط بين سحب العينات اذ نستطيع القول انها تلك العينة التي يتم سحبها باحدى طرق السحب العشوائي السابقة دون أي اجراءات اخرى ،ولكن الى جانب هذه البساطة و التي هي اهم مزاياها هناك بعض المساوئ و التي هي :

– لايمكن سحب عينة بهذا الاسلوب الا اذا كان حجم المجتمع محدود و معلومة كل مفرداته .

– في بعض الاحيان قد لا نحصل على نفس نسب الصفات الموجودة في المجتمع مثلا سحب عينة من طلاب الجامعة الموزعيين كما يلي 40 % ذكور و 60 % اناث فان العينة البسيطة قد تعطي نسبا تختلف عن الموجودة في المجتمع كأن نحصل على عينة تحوي 20% اناث و ذلك بسبب العشوائية .

2 _ العينة المنتظمة **Systematic Sample:**

نقوم بتقسيم مفرادات المجتمع المرتبة و المرقمة N -.......1-2-3 الى مجموعات نطلق على كل مجموعة اسم فترة السحب و عدد مفراداتها k،ونحصل على طول فترة السحب بتقسيم حجم المجتمع Nعلى حجم العينة n اي $k = \frac{N}{n}$.

ثم نختار مفردة بشكل عشوائي من مفردات اول فترة سحب و لتكن المفردة s , فتكون المفردة s اول مفردة في العينة ، و للحصول على باقي مفردات العينة نضيف اليها على التتالي فترة السحب k فتكون بهذا الشكل مفردات العينة هي المفردات ذات الارقام :

$$s\ ,\ s+k,\ s+2k,\ s+3k,\ .\ .\ .\ s+\ (n-1)k$$

مثال :مجتمع حجم مفرداته 100 مفردة نريد سحب عينة منتظمة بحجم 5مفردة :

$$k=\frac{N}{n}=\frac{100}{5}=20$$

نختار من ضمن فترة السحب الاولى و البالغ عددها 20 مفردة ، مفردة بشكل عشوائي باحدى الطرق التي تعرفنا عليها سابقا و لتكن هذه المفردة ذات الرقم 12 و عليه تكون مفردات العينة هي :

12 32 52 72 92

مزايا هذه العينة هي قلة التكاليف وسهولة سحبها كما وانها تعطي نفس النسب الموجودة في المجتمع لصفة ما.ولكن الى جانب هذه المزايا هناك بعض المساوئ هي انه لايمكن سحبها الا اذا كان المجتمع محدود ومفرداته مرتبة ومرقمة ،كما انه في بعض الاحيان قد تعطي عينة متحيزة اذا كانت البيانات في المجتمع الاحصائي مرتبة بشكل دوري يتوافق الدور مع فترة السحب ،مثلا عند اختيار عينة من البيوت في منطقة مخططة بطريقة الشطرنج ،فاذا كانت وحدة البدء بيت يقع على الشارع العام وكان طول فترة السحب يساوي عدد البيوت في رتل البيت الذي يقع على الشارع العام فستكون العينة كلها من البيوت الواقعة على الشارع وبذلك نكون اغفلنا تمثيل البيوت الغير الواقعة على الشارع العام .

3- العينة الطبقية **Stratified Sample:**

عندما يكون المجتمع غير متجانس نقوم بتقسيمه الى مجموعات مؤلفة من مفردات متجانسة تدعى كل مجموعة طبقة (Strata) ،ثم ناخذ من كل طبقة عينة عشوائية حيث مجموع هذه العينات الماخوذة من الطبقات يشكل العينة الطبقية .اما نسبة تمثيل كل طبقة في العينة فيمكن ان يكون حسب نسبة الطبقة في المجتمع :

اذا كان N **حجم المجتمع وهو مؤلف من 3 طبقات مثلا ، عدد مفرداتها** $N_1,N_2,N_3,$ **حيث ان :** $N_1+N_2+N_3=N$ **و** n **حجم العينة الطبقية الواجب سحبها .في هذه الحالة يتوجب سحب 3 عينات ،عينة من كل طبقة احجامها هي** n_1,n_2,n_3 **بحيث يكون** $n_1+n_2+n_3=n$**. وللحصول على قيم** n_1,n_2,n_3 **نطبق نسب الطبقات في المجتمع كما يلي :**.........$\frac{n}{N}=\frac{n_1}{N_1}=\frac{n_2}{N_2}=\frac{n_3}{N_3}$ او $n_i=\frac{N_i*n}{N}$

مثال : نود سحب عينة من طلاب الجامعة لمعرفة رايهم بدراسة و تدريس اللغة الاجنبية، و من المعروف بشكل مسبق ان هناك اختلاف بصعوبة و ممارسة هذه اللغات. فاذا علمنا ان عدد الطلاب 1000 طالب ،600 طالب لغتهم الانكليزية و300 طالب لغتهم الفرنسية و الباقي لغات مختلفة و نود سحب عينة بحجم 100 طالب فما هو عدد الطلاب من كل طبقة .

$$\frac{n}{N}=\frac{n_1}{N_1}=\frac{n_2}{N_2}=\frac{n_3}{N_3}$$

$$\frac{100}{1000}=\frac{n_1}{600}=\frac{n_2}{300}=\frac{n_3}{100}$$

ومنه : $n_1=60, n_2=30, n_3=10$

تمتاز العينة الطبقية بانه عنداستخدامها تمثل كل الطبقات و بنفس النسب، كما انها تمكن من دراسة الطبقات بشكل مستقل اذا كان ذلك مفيدا بالاضافة الى دراسة المجتمع، ولكن من مساوءها انه لايمكن سحبها الا اذا كان معلوما مفردات المجتمع ككل و عددها و كذلك مفردات كل طبقة و عددها .

4 - العينة العنقودية **Cluster Sample:**

رأينا ان العينات الثلاث اعلاه لايمكن سحبها الا اذا توفرت قوائم بمفردات المجتمع ككل ،ولكن هناك حالات عندما لايتوفر لدينا تلك القوائم بمفردات المجتمع فيتوجب علينا اعداد هذه القوائم و هو غير ممكن في بعض الحالات، لذلك لا بد من استخدام اساليب اخرى لسحب العينات منها اسلوب العينة العنقودية . حسب هذا الاسلوب نقوم بتقسيم المجتمع الى مجموعات او اجزاء حسب خاصة ما او صفة ما و تعتبر هذه الاجزاء هي

وحدات السحب و التي يدعى كل منها عنقودا ،ثم نختار بشل عشوائي عددا من هذه العناقيد لنجمع البيانات منها .
مثلا سحب عينة من اسر مدينة ما لدراسة رايها بشكل سلعة فاننا نقسم المدينة الى احياء (عناقيد) ونسحب بشكل عشوائي عنقود او عدة عناقيد ،حيث الاسر بهذه العناقيد هي العينة المطلوبة .

5 – العينة المتعددة المراحل **Multi_stage Sample :**

تستخدم هذه العينة في تلك الحالات التي لا يتوفر قائمة لدينا عن مفردات المجتمع –كما في العينة العنقودية السابقة. و كذلك عندما يكون الوقت المتاح قصيرا و الامكانيات المادية المتاحة قليلة .
حسب هذه الطريقة يتم سحب العينة على عدة مراحل ،كمثال عليها لدراسة متوسط انفاق الاسرة في الجمهورية العربية السورية نختار عدة محافظات بشكل عشوائي (كمرحلة اولى) بعد ذلك من ضمن المحافظات المختارة في المرحلة الاولى نختار عدة مناطق (كمرحلة ثانية) ثم نختار عدة قرى و احياء بشك عشوائي من المناطق المختارة بالمرحلة الثانية (كمرحلة ثالثة)،و اخيرا كمرحلة رابعة نختار عددا من الاسر من القرى و الاحياء المختارة في المرحلة الثالثة و تجمع البيانات من هذه الاسر المختارة بالمرحلة الرابعة فقط .

2.7.1 – العينات الشخصية (غير الاحتمالية):Non-Probability Samples

يتم اختيار بعض العينات بشكل غير عشوائي أي عمدي ،فهذا النوع من العينات هو غير عشوائي و بالتالي لا يمكن حساب احتمال الخطأ باستخدام نظرية الاحتمالات، من هذه العينات :

1 – عينة الحصص:

يحدد الباحث عددا من المفردات لكل صفة و هي بهذا الشكل تتشابه مع العينة الطبقية و تختلف معها بانها ليست عشوائية .

2 – العينة المنتقاة :

يستخدم هذا النوع بشكل اوسع من العينة السابقة و ذلك لما حققه هذا النوع من نجاح .ويتوقف نجاح هذه العينة على مهارة الباحث في اختيار المفردات التي ستدخل في العينة و من الامثلة على هذا النوع ،عينة السلع الداخلة في حساب الرقم القياسي للاسعار.

3 – العينات كبيرة الحجم:

و هي عينات تحوي عددا كبيرا من المفردات مختارة بشك غير عشوائي و يدعي مؤيدي هذا النوع من العينات انه كلما كبر حجم العينة كانت اكثر تمثيلا للمجتمع ، ولكن هذا غير صحيح الا اذا كانت العينة عشوائية .اما في حال العينات غير العشوائية فان التحيز سوف يبقى و يمكن ان يزداد مادام سحبها يتم بشكل غير عشوائي .

8.1 – اخطاء العينة :Error Of Sample

نوهنا اعلاه باننا نقوم بسحب عينة للتعرف منها على صفات المجتمع الاحصائي ، ولكن في اغلب الاحيان نحصل على نتائج تختلف عن النتائج الممكن الحصول عليها فيما لو قمنا بالحصر الشامل و ذلك لان العينة عرضة لنوعين من الخطأ هما :

- **خطأ الحظ و الصدف** Randon Error او الخطأ العشوائي.
- **خطأ التحيز** Bias Error.

1– الخطأ العشوائي :

ينتج هذا الخطأ بسبب العشوائية ذاتها حيث تختلف نتائج العينة عن القيم الموجودة في المجتمع كما تختلف من عينة لاخرى، و لتبيان ذلك نأخذ المثال التالي :

لدينا اربعة اشخاص A,B,C,D **اعمارهم على التوالي** 22 , 21 , 19 , 18 **عام فلو اعتبرنا ان هؤلاء الاشخاص يؤلفون مجتمعا متوسط الوزن فيه :**

$$\mu = \frac{18+19+21+22}{4} = 20$$

لو سحبا كل العينات الممكنة من حجم 3 – انظر الجدول رقم (2.1).

من تدقيق متوسط الاوزان التي ستعطتها العينات نلاحظ ان هناك فروقا بين متوسط العينات و متوسط المجتمع لكل العينات الممكنة و هذا الفرق ناتج عن عملية الاختيار و نطلق عليه خطأ الحظ و الصدف، لكن لاحظ ان هذا الخطأ الناتج هو لبعض العينات موجب و لبعضها سالب أي يحدث في الاتجاهين و ان مجموع الخطأ يساوي الصفر . من هنا يمكن القول ان القوى التي ادت الى ظهور هذا الخطأ متساوية و انها تلغي بعضها بعضا، من جهة اخرى عندما يزداد حجم العينة فان قيمة هذا الخطأ تقل ، ففي مثالنا السابق اكبر خطأ كان 0.66 سنة اما لو سحبنا عينة بحجم اصغر و ليكن مفردتين هما A , B **سيكون متوسطهما 18.5 سنة و الخطأ سيكون ((**18.5– 20 = -1.5 **)) أي ان الخطأ اكبر عندما**

العينة	مفردات العينة	مجموع اوزان العينة	متوسط الوزن	الخطأ = (متوسط المجنمع – منوسط العينة)
1	A B C	18+19+21=58	19.33	**19.33-20= -0.66**
2	A B D	18+19+22=59	19.66	**19.66-20= -0.33**
3	A C D	18+21+22=61	20.33	**20.33-20= 0.33**
4	B C D	19+21+22=62	20.66	**20.66-20= 0.66**

الجدول (2.1 **)**

يكون حجم العينة اصغر، و هذا يؤدي بنا للقول انه كلما كبر حجم العينة قل مقدار خطأ الحظ و الصدف . و اخيرا نقول ان هذا الخطأ لابد و ان يقع لانه ناتج عن العشوائية نستطيع تقليل قيمته بزيادة حجم العينة كما نستطيع حساب احتمال الوقوع به كما سيمر معنا لاحقا .

2 – خطأ التحيز:

يمكن وصف التحيز بانه الحصول على عينة كل مفراداتها او اغلب مفراداتها تحمل نفس الصفات مثلا سحب عينة لقياس متوسط الطول و كل مفراداتها اشخاص طوال القامة، و التحيز يحدث دوما باتجاه واحد سالب او موجب و يعتبر التحيز من اخطر الاخطاء التي يمكن ان تؤثر على البحث ، و تكمن خطورته بانه لا يمكن التنبؤ به و قياسه .

و نجمل اسباب التحيز بما يلي :

– عدم السحب العشوائي او استخدام اطار معيب عند سحب العينة ، أي السحب ليس من كامل المجتمع .

– عدم تنفيذ العينة بشكل كامل – اهمال بعض وحدات العينة .

–الاستعاضة بوحدات مكان وحدات عند سحب العينة .

و يمكن تصحيح خطأ التحيز بادخال بعض التعديلات على مفردات العينة عندما يكون معلوما اتجاه التحيز و حجمه و الا وجب اعادة سحب العينة .

9.1– اداة البحث – الاستمارة الاحصائية Statistical Questionnaire:

لكل بحث يجب اعداد استمارة خاصة به يطلق عليها الاستبيان او الاستمارة الاحصائية Questionnaire،وهي عبارة عن مجموعة من الاسئلة حول متغيرات مشكلة البحث . تعتبر عملية صياغة الاستمارة عملية هامة جدا ومعقدة حيث يتوقف النجاح في الحصول على البيانات الهامة على صحة صياغة هذه الاستمارة ، وهنا يمكن ذكر بعض القواعد الواجب مراعاتها عند اعداد الاستمارة لكي تكون صحيحة قدر الامكان وبالتالي نضمن الحصول على بيانات صحيحة ،ويمكن تقسيم هذه القواعد الى جزاين :

1.9.1– قواعد متعلقة بشكل الاستمارة :

1– يجب كتابة الاستمارة على ورق جيد بحيث تلفت انتباه المستجوب وتثير اهتمامه .
2– يجب ان يكون ترتيب الاستمارة متناسبا مع شكل الورقة وتكتب بخط واضح .
4– ذكر المشرف او الهيئة المشرفة على البحث في بداية الاستمارة وعنوان الجهة المشرفة.
5– الاشارة الى سرية البيانات وهي غير اسمية ، وعدم استخدامها لاغراض اخرى.
6– وجود محفز معنوي او مادي في الاستمارة لتحفيز المستجوب على الاجابة .
6– يجب اعطاء الاسئلة ارقاما متسلسلة حتى يسهل التعرف عليها عند تفريغ الاجابات و يمكن ان يعطى لكل سؤال رمز في حال كان بالامكان استخدام الحاسبات في تفريغ النتائج .

7 – ترتب الاسئلة بشك منطقي و يجب ان تبدا بالاسئلة السهلة مثل الاسئلة الخاصة :الاسم ، الجنس السن ،.... . و يمكن تقسيم الاسئلة الى مجموعات تكون الاسئلة ضمن كل مجموعة مشتركة بعامل ما .
8 – عدم كتابة اكثر من سؤال في السطر الواحد .
9 – يمكن تضمين الاستمارة جداول لطرح الاسئلة بشكل اسهل .

2.9.1 – قواعد متعلقة بالاسئلة :

1 – يجب طرح الاسئله الهامة ذات العلاقة بموضوع البحث حتى لا نحصل على بيانات زائدة تعرقل الخطوات التالية .
2– يجب صياغة الاسئلة بشكل سهل وواضح بحيث لاتقبل اللبس او اساءة الفهم او التأويل مثلا السؤال التالي (الفرع ؟) فهل المقصود بفرع الدراسة ادبي او علمي ام المقصود فرع المنظمة الشعبية ام فرع هيئة سياسية . . .
3 –يجب صياغة الاسئلة بحيث تكون الاجابة عليها قاطعة مانعة و يفضل ان تكون الاجابة بعدد او بكلمة صح او خطأ او الاشارة اليها (✓, X).
4 – في حال كان هناك خيارات للاجابة اكثر من خيارين صح او خطأ يجب ذكر هذه الاجابة و يترك للقارئ الاختيار بينها .
مثال : الحالة الاجتماعية ؟
- □ **عازب**
- □ **متزوج**
- □ **مطلق**
- □ **ارمل**
- □ **خاطب (مخطوبة)**

5 – اذا كان هناك خيارات كثيره و متعددة للاجابة و كان اهتمامنا ينصب على بعضها فقط نذكر بعض الاجابات الهامة و نفرد للباقي اجابة: عدا ذلك.
مثال : ما هي الهواية المفضلة ؟
- □ **الرياضة**
- □ **السباحة**
- □ **السفر**
- □ **عدا ذلك**

6 – يجب ان لا تتطلب الاسئلة عمليات حسابية مطولة فلا نسأل كم كان عمرك في 2017/03/20 مثلا يكفي هنا ان نسأل عن تاريخ التولد.
7 – يجب ان لا نسأل اكثر من سؤال في السؤال الواحد .فلا نسأل (متزوج وعدد الاولاد الذكور و الاناث ؟)يجب هنا توجيه ثلاث او اربعة اسئلة:
1 – متزوج او الحالة الاجتماعية
2 – عدد الاولاد ؟
3 – الذكور ؟
4 – الاناث ؟

8 - يجب قدر الامكان تجنب الاسئلة التي تقلق المستجوب .
مثل : ماهو دخلك ؟ رصيدك في البنوك ؟
بهذه الاسئلة يتحسب المستجوب او يبدأ بالتفكير بالضرائب او بأمور اخرى .
9 - ان لا نسأل اسئلة تثير تحيز المبحوث او المستجوب كأن نسأله :
_ هل تتأخر عن المدرسة احيانا بسبب المواصلات ؟
_ هل الدخل قليل ؟
_ هل تقرأ الصحف يوميا ؟
_ هل رسبت بسبب صعوبة الاسئلة ؟
10 - يجب ذكر الوحدات المستخدمة و توضيحها فعند السؤال عن الدخل يجب توضيح المقصود بالدخل هل هو الراتب الشهري لرب الاسرة او لافراد الاسرة ام هو مجموع الدخول لهذه الاسره من الراتب و العقارات و الاملاك . . الخ ومن ثم توضيح الوحدة المستخدمة في قياس الدخل و هل هو الدخل الشهري او السنوي . . الخ.
11 _ و اخيرا يحبذ اضافة اسئلة راجعة و هي توجيه أسئلة طرحت بطريقة اخرى للكشف عن صحة الاجابة من عدمها:
مثلا السؤال هل تقرأ الجرائد يوميا ؟

السؤال الراجع . كم تصرف من الوقت لقراءة الجرائد ؟او ماهي الجرائد التي تقرأها؟
3.9.1- ملء الاستمارة الاحصائية :

يتم ملء الاستمارة الاحصائيةQuestionnaire باحدى الطرق التالية :
1-المقابلة الشخصية .
2-المراسلة .
3-الهاتف .
4-الانترنيت .
ولكل من هذه الطرق مزاياها ومساوئها بحيث تجعلها ممكنة في بعض الحالات وغير ممكنة في حالات احرى ،لذلك سنطلع على كل طريقة بشئ من التفصيل .
1- المقابلة الشخصية :
حسب هذه الطريقة يقوم الباحث بنفسه او اشخاص مدربين على البحث بشكل جيد بمقابلة المبحوث وتوجيه الاسئلة اليه بشكل مباشر سؤالا فسؤلا حسب ورودها بالاستمارة او ملاحظة المبحوث بشكل مباشر وتسجيل الاجابات .وتعتبر هذه الطريقة من اكثر الطرق شيوعا وفائدة في الحالات التالية :
- **في حال كثر عدد الاميين بين افراد البحث .**
- **في حالة الاسئلة التي تتطلب الاجابة عليها ملاحظة مباشرة مثل ملاحظة نتائج تجربة ما او ملاحظة حركات المستجوب ، او معرفة نظافة المنزل ......الخ**
- **بالاضافة لما ورد اعلاه فان لهذا الاسلوب مزايا اخرى هي ان الباحث سيساعد المبحوث من خلال توجيه الاسئلة التوضيحية وبالتالي نضمن صحة الاجوبة.**
الى جانب المزايا المذكورة اعلاه لهذا الاسلوب بعض المساوئ نذكر منها :

– قد لايصلح هذا الاسلوب في الحصول على اجوبة صحيحة اذا كانت الاسئلة محرجة كأن نسأل عن الامراض الجلدية او العلاقات الزوجية .
– من الممكن ان نحصل على اجوبة متحيزة بسبب ثاثير الباحث على المستجوب عن طريق الايحاء بجواب معين.
– تحتاج عملية الجمع بحسب هذا الاسلوب الى وقت طويل وكلفة عالية وخاصة اذا كان البحث يتطلب عددا كبيرا من العاملين .

2– المراسلة :

ترسل الاستمارة بالبريد او توزع باليد على المبحوثين او يمكن ان تنشر بالصحف ،ويقوم المبحوث بملىء الاستمارة بنفسه بالاجابة على الاسئلة الواردة فيها بالترتيب .ولهذا الاسلوب مزاياه ومساوئه والتي نجملها بما يلي :

1–سهولة الوصول الى المبحوثين وقلة التكاليف .
2–يصلح في تلك الحالات التي تحوي فيها الاستمارة على اسئلة محرجة.
3–التخلص من تاثير الباحث على المبحوث عن طريق الايحاء وبالتالي التخلص من تحيز الباحث.
4–يعطي هذا الاسلوب وقتا للمبحوث للتفكير بالاجوبة .

اما مساوئ هذا الاسلوب فهي :

1–لايصلح في تلك الحالة التي يكثر فيها عدد الاميين بين افراد البحث .
2–قد يؤدي هذا الاسلوب الى الحصول على اجوبة متحيزة وذلك لان الاجوبة ستعود من بعض فئات من الناس وهم الفئة الذكية من الناس والذين يريدون توجيه البحث باتجاه معين، كما سيكون هناك غياب لبعض الاجابات من الناس المهملين . لذلك ننصح عند استخدام هذا الاسلوب باضافة بعض الاجراءات الى صياغة الاستمارة والتي تكفل تحريض المهملين على الاجابة منها:
–ارسال طابع بريدي ومغلف مكتوب عليه عنوان الجهة التي تقوم بالبحث او اجراء اتفاق مع مؤسسة البريد بالدفع عند وصول الرسائل وذكر هذا على الاستمارة.
–ذكر بعض العبارات في بداية الاستمارة والتي يمكن ان تحرض المبحوث على الاجابة :(اخي المواطن نحن نثق بوعيك وحبك لوطنك وبالتالي حبك للتطوير والتقدم للامام .لذلك نرجو منك المساعدة بالاجابة على الاسئلة الواردة بهذه الاستمارة .....).
–من الممكن استخدام اساليب محرضة اخرى كأن تعطى للاستمارة ارقام ونخصص بعض الجوائز لعدد من الاستمارات يتم اختيارها بعد عودتها بطريقة السحب العشوائي الوارد ذكرها اعلاه.

3–استخدام الهاتف لملء الاستمارة :

حسب هذا الاسلوب يقوم الباحث بتوجيه الاسئلة الواردة بالاستمارة للمبحوث عن طريق الهاتف . مزايا هذا الاسلوب هي:

- **السرعة في الحصول على البيانات.**
- **يسمح هذا الاسلوب بتوضيح الاسئلة للمبحوث.**

اما مساوئه فهي:

- **لايصلح الا للاتصال مع الاشخاص الذين يملكون هواتف .**
- **يحتاج الى تكاليف عالية احيانا .**
- **لايصلح في تلك الحالات التي تتطلب ملاحظة مباشرة للمبحوث.**

4–استخدام الانترنيت **Internet** لملء الاستمارة:

يتم حسب هذا الاسلوب نشر الاستمارة البريدية على شبكة الانترنيت ويطلب من المبحوثين ملء هذه الاستمارة .يستخدم هذا الاسلوب بشكل واسع من قبل المحطات التلفزيونية والصحف والمجلات لاستطلاع الراي . مزايا هذا الاسلوب السرعة الفائقة في الحصول على البيانات وفي معالجتها ايضا كذلك في قلة التكاليف ،اما المساوئ فتتمثل بعدم امتلاك كل الناس لاجهزة الحاسب ولخطوط الانترنيت بالاضافة لامية بعض الافراد باستخدام الحاسب ، كما ان استخدام هذا الاسلوب يتطلب الاعلان عن وجود هذه الاستمارة على شبكة الانترنيت بواسطة الصحف او الاذاعة او التلفزيون .

عرضنا في هذا الفصل بعض الارشادات و الوسائل الفنية التي يمكن استخدامها في عملية جمع البيانات و التي سوف تساعد في الحصول على البيانات الصحيحة و الضرورية للبحث مما يسهل عمل الباحث في المراحل اللاحقة حيث ان الهدف من عملية جمع البيانات هو ليس غاية بحد ذاته و انما وسيلة و خطوة للانتقال للمراحل الاخرى : عرض البيانات و تحليلها و اتخاذ القرار .
تنتهي عملية جمع البيانات بالقيام بمراجعة الاستمارات (Revison) التي تم ملؤها باحدى الطرق السابقة ، واحدة واحدة. ان الغاية من هذه العملية هو اكتشاف الاستمارات التي تحوي على نقص فى الاجابات او تناقض و هنا يجب استكمال الاجوبة في الميدان او الغاءها ، كما يمكن خلال عملية المراجعة اجراء بعض الحسابات اللازمة التي وفرنا على المبحوث القيام بها.
نتيجة مرحلة جمع البيانات هي الحصول على عددا كبيرا من البيانات ندعوها بشكلها هذا دون أي معالجة لها بيانات خام Raw data، فاذا ما عرضت بهذا الشكل ضمن بحث ما او على الصحف او على الشاشة لن تقدم أية فائدة للباحث او للقارئ و لذا كان من الضروري عرض هذه البيانات بشكل علمي و فني و سهل و هذا ما سنناقشه في الفصل التالي .

اسئلة وتمارين غير محلولة

1	أي من المتغيرات التالية متغير نوعي (وصفي) **C)** عمر الانسان **B)** أطوال الموظفين **A)** عدد افراد الاسرة **D)** مكان اقامة الموظفين
2	**أي من المتغيرات التالية متغير كمي منقطع** C) **قلم الكتابة** B) **أطوال الموظفين** **عدد افراد الاس** D) **طول فترة استخدام قلم الكتابة**
3	**أي من البيانات التالية متغير كمي مستمر :** C) **التي يتقنها** B) **عدد أفراد الأسرة** **اطوال عينة من الموظف** D) **لون قلم الكتابة** **الطالب**
4	**في اي من الحالات التالية يجب استخدام اسلوب العينة بجمع البيانات :** B) **دراسة متوسط عدد أفراد الأسرة** A) **دراسة نفقات الاسرة** D) **دراسة مدى رضى العاملين في مديرية** C) **نسبة المادة الفعالة في معجون الاسنان لماركة ما**
5	**سحب عينة منتظمة بحجم 50 من العاملين في شركة والبالغ عددهم 1000 .اذا علمت ان العامل الاول ف (وحدة البدء) اختير عشوائيا و هو رقم 15، فأن العامل الثاني في العينة هو رقم :** A) **35** B) **55** C) **16** D) **115**
6	**سحب عينة طبقية بحجم 50 من العاملين في شركة والبالغ عددهم 1000، 600 ذكور و400 اناث ،فار** **الذكور الواجب اختيارهم عشوائيا هو :** A) **20** B) **50** C) **30** D) **40**
7	**ما الفرق بين الخطأ العشوائي وخطأ التحيز .**
8	**ما هي مزايا ومساوئ البيانات الاولية والثانوية .**
9	**بين ما هي الحالات التي يفضل بها اسلوب المقابلة الشخصية لجمع البيانات .**
10	**بين ما هي الحالات التي يفضل بها اسلوب المراسلة لجمع البيانات.**
11	**ما هي الحالات التي يجب عندها استخدام اسلوب العينة لجمع البيانات .**
12	**ما هي الغاية من القيام بدراسة تحضيرية .**
	حالات التي يجب استخدام العينات العنقودية .

مراجع الفصل الأول :

1- حميدان، عدنان؛ الجاعوني، فريد؛ ناصر آغا، عمار؛ العواد، منذر، 2003، مبادئ الإحصاء، منشورات جامعة دمشق.

2- Groebner David F., Shannon Patrick W. and Fry Phillip C. (2017)," Business Statistics :A Decision-Making approach", 10 th Edition. person.

3- Weiss, N.A.(1999), Introductory Statistics. Addison Wesley.

الفصل الثاني

عرض البيانات الاحصائية

Presentation of Statistical Data

-1

ملخص الفصل:

يتناول هذا الفصل طرق تلخيص البيانات واختصارها جدوليا او بيانيا .سنتعرف على كيفية تكوين الجدول الاحصائي وكيفية انشاء الرسوم البيانية او المنحنيات لتمثيل البيانات العددية .

راينا اعلاه انه بعد الانتهاء من مرحلة جمع البيانات نحصل على كم هائل من البيانات احيانا والتي دعوناها بيانات خام .حتى تؤدي هذه البيانات الغاية المرجوة منها لابد من تلخيصها واختصارها، وهذا يتم من خلال عرض هذه البيانات. تستخدم اساليب عرض البيانات الاحصائية بشكل واسع في كل المجالات والعلوم ، وتبلغ اهمية هذه الاساليب اهمية علم الاحصاء لهذه العلوم بل هي احيانا ذات اهمية اكبر وذلك بسبب شعبية هذه الاساليب ،فلا يخلو بحث او كتاب او صحيفة او نشرة اخبار تلفزيونيه او اعلان تسويقي من احد هذه الاساليب .و اكتسبت اساليب العرض اهمية اكبر في و قتنا الحاضر و ذلك لقدرة الحاسبات الالكتروني على القيام بها بشكل سريع وجميل ملفت للانتباه . هذا جعل معرفة اساليب العرض من المعارف الضرورية لكل انسان معاصر ،لذلك تم اعطاء هذا الفصل اهمية خاصة وبالتالي التوسع في عرض وسائل و اساليب عرض البيانات الاحصائية .

تعرض البيانات الاحصائية عرضا جدوليا و بيانيا و لكل منهما عدة اساليب . سوف نتعرف على هذه الاساليب من خلال تقسيمها حسب طبيعة البيانات، و تقسيمنا هذا سيكون لاسباب درسية ليتمكن الطالب بسهولة من معرفة اماكن استخدام كل اسلوب. سوف نقسم هذه الاساليب كالتالي :

- **عرض البيانات النوعية.**
- **عرض البيانات الكمية.**
- **عرض بيانات السلاسل الزمنية.**

1.2 – عرض البيانات النوعية (الوصفية):

تعرفنا في الفصل السابق على المتغيرات النوعية ورأينا انها بيانات تعود لمفردات تتصف او لا تتصف بصفة ما، او هي التي يمكن التعبير عنها بكلمة او جملة : شجرة ليمون ،طالب ، ذكر انثى ،جيد ، ضعيف ..... الخ . سنرى في هذا الفصل كيفية عرض البيانات النوعية حيث يمكن عرضها بطريقتين :

1- العرض الجدولي للنوعيات .

2- العرض البياني للنوعيات .

1.1.2- العرض الجدولي للمتغيرات النوعية :

يقصد بالعرض الجدولي اختصار البيانات وتلخيصها من خلال تنظيمها بشكل جداول ، وتدعى هذه العملية بالتبويب Tabulation .

- **التبويب** Tabulation :

تمر عملية التبويب بثلاث مراحل:

1- تقسيم البيانات الى مجموعات حسب صفاتها المشتركة ، وبما ان هذه البيانات هي نوعية فهذا التقسيم متوفر حسب الصفة او المتغير .

2- عد مفردات كل صفة .

3- انشاء الجدول الاحصائي ونقل البيانات اليه بحيث يظهر فيه عدد مفردات او تكرارات frequency **كل صفة .**

مثال :

من عينة بحجم 20 طالب حصلنا على البيانات الخام التالية عن لغة الطالب :

اللغة	E	G	F	G	E	F	E	E	G	G	E	F	E	F	E	E	E	F	F	E

E :الانكليزية F :الفرنسية G:الالمانية

المطلوب : تبويب البيانات في جدول احصائي.

باتباع الخطوات الثلاث السابقة لانشاء الجدول نحصل على الاعمدة الثلاثة من يسار الجدول (2−1)

توزيع مفردات العينة حسب اللغة

اللغة	تخدام عصا العد y	التكرار **frequency** f_i	التكرار النسبي **Relative Frequency**	التكرار المئوي **Percentage %**
الانكليزية	𝍸 𝍸	**10**	0.5	50
الفرنسية	𝍸 𝍷	**6**	0.3	30
الالمانية	𝍷𝍷𝍷𝍷	**4**	0.2	20
المجموع		**20**	**1**	100

الجدول (2−1) **المصدر :فرضي**

من ملاحظة البيانات في الجدول يتبين لنا فوائد عملية التبييب ،وهي اختصار البيانات ومعرفة تمركز البيانات من قيم التكرار f_i ،كما يمكن قراءة البيانات بشكل نسبي Relative Frequency **و بشكل مئوي .**

يحسب التكرار النسبي RF **من العلاقة :** $RF = \dfrac{f_i}{\sum f_i}$ ، **مثلا:** $RF_1 = \dfrac{f_1}{\sum f_i} = \dfrac{10}{20} = 0.5$

حيث f_1 **هو التكرار الاول و** $\sum f_i = f_1 + f_2 + f_3 + ...$

اما التكرار المئوي Percentage **يحسب كما يلي :**

$$percentage = \frac{f_i}{\sum f_i} * 100$$

لاحظ ان مجموع التكرار النسبي هو 1 ،في حين مجموع التكرار المئوي هو 100.

ان نتيجة عملية التبويب هي الجدول الاحصائي والذي يمكن تعريفه بانه ترتيب البيانات في اسطر واعمدة حسب صفاتها المشتركة بشكل نستطيع قراءة هذه البيانات بشكل عمودي او افقي او كليهما ويفيد هذا العرض في اختصار البيانات وملاحظة صفانها العامة .وحتى يؤدي الجدول الاحصائي الاغراض المرجوة منه لابد من مراعاة بعض القواعد عند اعداده واهمها :

1 − يجب ان يكون للجدول عنوان واضح يعبر عن محتوى الجدول وزمن الحصول على البيانات ومكان الحصول عليها وبشكل اخر يجب ان يجيب عنوان الجدول على الاسئلة التالية :ماذا يحتوي الجدول ؟ ومن اين تم الحصول على محتوياته؟وزمن الحصول عليها؟

2 − يجب ذكر وحدات القياس المستخدمة بالجدول .

3− يجب وضع عنوان لكل عمود ولكل سطر ،وفي حال وضع الجدول باكثر من صفحة يجب اعادة العناوين.

4− يمكن ترقيم الاسطر والاعمدة لتسهيل العودة اليها اذا كان لابد من العودة اليها عند شرح محتوياتها .

5− تجب الاشارة الى المصدر الذي اخذت منه بيانات الجدول .

6− يجب وضع رقم للجدول اذا كان سيوضع ضمن كتاب وذلك للعودة اليه عند شرح البيانات الموجودة ضمنه .

7− يجب ان يكون شكل الجدول متناسبا مع شكل الصفحة التي سيعرض عليها وذلك حتى لايضطر القارئ لتغيير وضع الكتاب اثناء القراءة .

8- يجب عرض البيانات في الجدول الاحصائي حسب ترتيب ما ،حيث يمكن ان تعرض حسب الحجم تصاعديا او تنازليا او حسب التاريخ او يمكن ان تعرض جغرافيا ،و الغاية من استخدام ترتيب ما هو سهولة الحصول على البيانات عند البحث في متن الجدول ،و قد ترتب البيانات ضمن الجدول على اكثر من اساس واحد فقد يكون الترتيب زمنيا و جغرافيا معا .

الجدول التالي (2-2) هو جدول احصائي تتوفر فيه بعض واهم القواعد المذكورة اعلاه .

توزيع عمال احد المعامل حسب وقت العمل و الجنس للعام 2018 (الوحدة عامل)

وقت ال	ذكر	انثى	المجموع
نهاري	1000	600	1600
مسائي	300	100	400
المجموع	1300	700	2000

جدول (2-2) المصدر **:فرضي**

1.1.1.2- انواع الجداول الاحصائية:

يمكن تقسيم الجداول الاحصائية حسب الغاية من اعدادها وكذلك حسب عدد المتغيرات (الصفات) المعروضة في الجدول .

1- حسب الغاية من اعداد الجدول الاحصائي ، تقسم الجداول الى جداول عامة وجداول خاصة :

- الجداول العامة **:هي تلك الجداول التي تعد دون أي هدف سوى العد والاختصار وعادة ما توضع هذه الجداول من قبل الوزارات و الدول لتوفير البيانات الهامة عن النشاطات المختلفة في البلد وكمثال عليها الجداول الواردة في المجموعة الاحصائية السنوي التي تصدر عن المكتب المركزي للاحصاء في الجمهورية العربية السورية .ويمكن استخدام البيانات الواردة في هذه الجداول في الابحاث العلمية كبيانات ثانوية .**
- الجداول الخاصة **:هي تلك الجداول التي يضعها الباحث بنفسه لدراسة او بحث مشكلة معينة ويمكن ان تستمد بيانات هذه الجداول من الجداول العامة .**

2- اما حسب عدد المتغيرات في الجدول الاحصائي فتقسم الجداول الى جداول بسيطة بمتغير وحيد كما في الجدول (2-1) **ومزدوجة بمتغيرين كما في الجدول** (2-2) **،ويمكن ان تكون مركبة اذا كانت باكثر من متغيرين ،حيث تضاف المتغيرات التي تزيد عن 2 من خلال تقسيم الاعمدة او الاسطر انظر الجدول جدول** (2-3) **، حيث تمت اضافة متغير ثالث هو الحالة العائلية.**

توزيع عمال احد المعامل حسب وقت العمل و الجنس والحالة العائلية للعام 2018 (الوحدة عامل)

الجنس / وقت العمل	ذكر		انثى		المجموع
	عازب	**متاهل**	**عازبة**	**متاهلة**	
نهاري	800	200	400	200	1600
مسائي	250	50	80	20	400
المجموع	1050	250	480	120	2000

جدول (2-3) المصدر **:فرضي**

2.1.2– العرض البياني للمتغيرات النوعية:

يقصد بالعرض البياني استخدام الرسوم والمنحنيات والاشكال الهندسية للتعبير عن البيانات العددية ،وتفيد هذا الطريقة في اظهار البيانات و ملاحظة التغيرات والفروق فيها بشكل يجذب الانتباه و يساعد على تذكرها ، كما تبين هذه الطريقة في بعض الحالات العلاقة بين الظواهر و المتغيرات التي ندرسها . تعرض المتغيرات النوعية بيانياً باستخدام :

– الاعمدة البيانية البسيطة و المجزأة.

– الرسوم الدائرية.

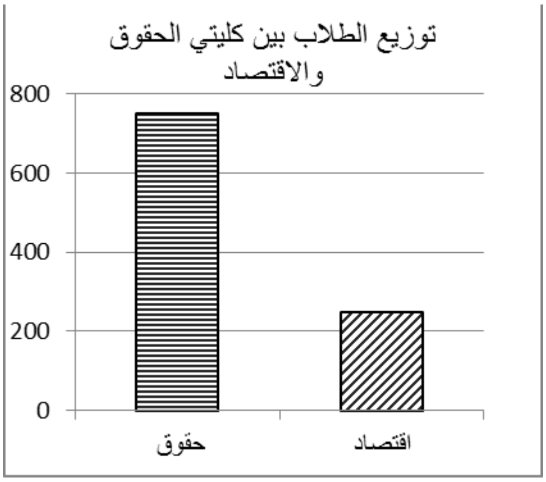
1.2.1.2 – الاعمدة البيانية **Bar Charts** :

تستخدم الاعمدة البيانية للتعبير عن البيانات العددية بمستطيلات على المحاور الاحداثية ، قواعد المستطيلات على محور السينات و ارتفاعاتها تتناسب مع قيمة (قيم مطلقة او نسبية او تكرارات) الظاهرة التي ندرسها .و تستخدم الاعمدة البيانية لعرض ظاهرتين او اكثر لنفس الفترة الزمنية مثلا مقارنة عدد طلاب كلية الاقتصاد و عدد طلاب كلية الحقوق لعام 2018 ،او مقارنة عدد الطلاب في اقسام كلية الادارة ،كما في المثال:

مثال: في الجدول التالي توزيع طلاب كليتي الحقوق والاقتصاد في احدى الجامعات 2018:

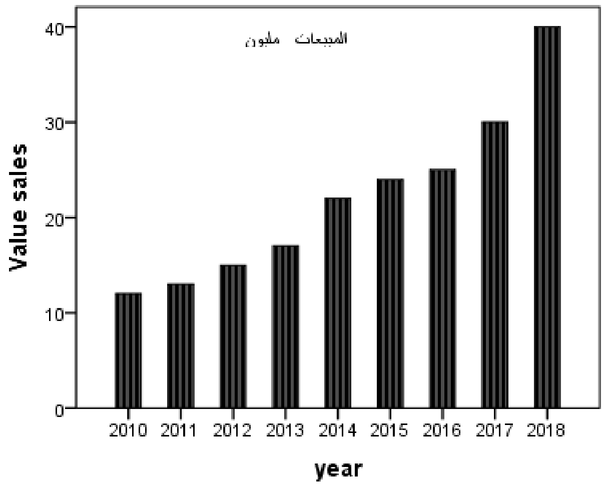
	لاب	مئوية %

فيمكن عرض البيانات بشكل اعمدة بيانية كما يلي ،الشكل(2-1):



الشكل(2-1):

كما تستخدم الاعمدة البيانية لعرض بيانات لظاهرة واحدة لفترات زمنية مختلفة من اجل ملاحظة تطور الظاهرة خلال تعاقب الزمن كما في الشكل (2-2**)حيث عرضنا المبيعات السنوية لاحد المتاجر للاعوام 2010 حتى 2018 م .**



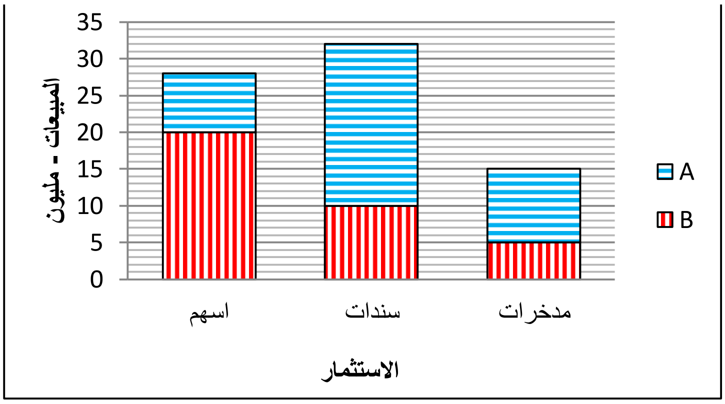
الشكل (2-2)

يمكن استخدام الاعمدة البيانية عند عرض عدة متغيرات وذلك باستخدام الاعمدة البيانية المتلاصقة Clustered Bar Charts **او المقسمة** Stacked**،كما في المثال التالي :**

مثال: فيمايلي الحقيبة الاستثمارية لمستثمرين :

الاستثمار (مليون)	المستثمر		المجموع
	A	B	
اسهم	**8**	**20**	28
سندات	**22**	**10**	32
مدخرات	**10**	**5**	15
المجموع	**40**	**35**	75

يتم عرض هذه البيانات بالاعمدة المتلاصقة كما في الشكل (3-2) **،او بالاعمدة المقسمة كما في الشكل** (4-2).



(3-2)

(4-2)

اهم القواعد و الملاحظات الواجب مراعاتها عند رسم الاعمدة البيانية :

1_ يجب ان تكون قاعدة الاعمدة متساوية و تفصل بين الاعمدة مسافة لاتقل عن قاعدة العمود .

2 _ يفضل تميز الاعمدة عن بعضها بألوان مختلفة او بالتظليل في بعض الحالات .

3 _ يستحسن اظهار تقسيمات محور العينات على المستوى الاحداثي و ذلك لبيان قيمة الظاهرة بشكل اكثر دقة .

4- يجب ان تبدأ تقسيمات محور العينات من الصفر و الا سوف نحصل على مقارنة خاطئة ، وعند الضرورة يجب كسر المحور او العمود .

5 _ يمكن استبدال الاعمدة بشكل مستمد من الظاهرة نفسها و ضمن الشروط و القواعد نفسها التي تستخدم بها الاعمدة و ذلك كأن نعرض عدد الاشجار بشكل شجرة بدلا من العمود و عدد السكان بشكل ادمي و حجم المبيعات برزمة من القطع النقدية . .الخ ،وهذا الشكل الاكثر استخداما في وقتنا الحاضر سواء في اعمال الدعاية و الاعلان او في المقالات او البرامج التلفزيونية .

2.2.1.2 – الرسوم الدائرية **Pie Charts :**

تعتبر الرسوم الدائرية شكل من اشكال العرض البياني للمتغيرات النوعية، و فيه نقسم الدائرة الى قطاعات ويعبر كل قطاع عن جزء من البيانات.حيث تمثل البيانات بدائرة و تقسم الى قطاعات تتناسب مساحتها مع حجم الاجزاء المكونة لها، تلون القطاعات بالوان مختلفة لتميزها عن بعضها مع الانتباه الى وضع وسيلة ايضاح تبين ما يمثله كل لون .

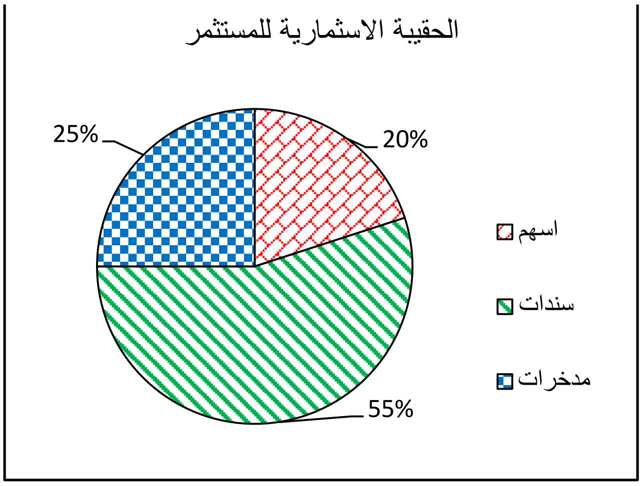
يتم تحديد مساحة كل جزء عن طريق تقسيم الزوايا المركزية للدائرة و البالغة 360 درجة الى أجزاء ، حيث مجموع أجزاء البيانات (الظاهرة ككل) يقابل 360 درجة وبالتناسب نستطيع حساب زاوية كل قطاع .

مثال:فيما يلي الحقيبة الاستثمارية لمستثمر :

وية القطاع x°	المئوي% النسبي	ستثمار	ر (مليون)
			٤

لعرض البيانات بقطاعات دائرية يجب اولا حساب قياس الزاوية لكل قطاع من العلاقة:

$x^{\circ} = \frac{f_i}{\sum f_i} * 360$ **ومنه قياس زاوية القطاع الاول** $x_1^{\circ} = \frac{8}{40} * 360 = 72$



(5-2)

تعتبر الرسوم الدائرية أكثر فائدة عندما ينصب اهتمامنا على البنية النوعية للظاهرة و تطورها حيث تمثل الاجزاء بشكل مئوي (%) كما في الشكل (2-5).

2.2 – عرض البيانات الكمية:

البيانات الكمية هي بيانات تعود لمتغيرات كمية و كنا قد عرفنا المتغيرات الكمية اعلاه و قسمناها الى قسمين متغيرات كمية مستمرة و منقطعة و سنبين في فيما يلي كيف يتم عرض المتغيرات الكمية عرضا جدوليا و كذلك بيانيا .

1.2.2– العرض الجدولي للمتغيرات الكمية:

تتميز المتغيرات الكمية بان عدد القيم التي يأخذها المتغير يكون كبيرا و احيانا غير محدود كما في حالة المتغيرات المستمرة ، و تختلف التكرارات او عدد المفردات التي تأخذ قيمة ما و بالتالي هي تشكل توزيعا نطلق عليه توزيعا تكراريا (Frequency Distribution) . تبوب البيانات الكمية في جداول احصائية تدعى بالجداول التكرارية (Frequency Table) و تمر عملية التبويب بعدة مراحل كما في حال المتغيرات النوعية ،و نميز هنا حالتين حالة المتغيرات المنقطعة و المتغيرات المستمرة .

1.1.2.2 _ تبويب المتغيرات الكمية المنقطعة :

تتشابه عملية تبويب المتغيرات الكمية المنقطعة مع المتغيرات النوعية التي مرت اعلاه حيث عدد القيم التي يأخذها المتغير محدود مثلا عدد افراد الاسرة هو 2 _3 _ . . . . .،او عدد التلاميذ في الصف . . . الخ .

وعلى هذا يمكن القول ان تبويب المتغيرات الكمية المنقطعة يمر بالمراحل التالية :

1– تقسيم البيانات الى مجموعات (كل مجموعة هي قيمة واحدة فقط) .

2 – عد مفرادات كل قيمة .

3 – انشاء الجدول التكراري .

مثال :لدينا البيانات التالية عن عدد الاطفال في الاسرة لعينة مؤلفة من 20 اسرة ،والمطلوب تبويبها في جدول تكراري :

0 3 3 2 1 1 2 3 2 1 1 2 0 0 2 0 1 1 3 0

من مراجعة البيانات اعلاه نرى ان عدد الاطفال هو بين 0 و 3 وبالتالي يمكن تقسيم البيانات الى 4 مجموعات .ننشا جدول ونضع في العمود الاول قيم المتغير x_i (عدد افراد الاسرة) .

عدد الاطفال x_i	باستخدام عصا ا **Tally**	التكرار **frequency** f_i	التكرار النسبي **Relative Frequency**	التكرار المئوي **Percentage %**
0	卌	**5**	0.25	25
1	卌 |	**6**	0.3	30
2	卌	**5**	0.25	25
3	||||	**4**	0.2	20
المجموع		**20**	**1**	100

الجدول (2-4)

بعد وضع قيم المتغير في العمود الاول نبدأ بتفريغ البيانات الخام وذلك بقراءة المفردات واحدة تلو الاخرى ووضع اشارة في العمود الثاني مقابل القيمة المساوية لقيمة المفردة، وعندما يبلغ عدد الاشارات 5 نشكل منها حزمة بوضع الاشارة الخامسة بشكل مائل او نضع 5 **اشارات كحزمة ونترك مسافة بين الحزم ،انظر الجدول (2−4) .اخيرا نقوم بعد الاشارات المقابلة لكل قيمة ونكتبها بشكل عشري في العمود الثالث عمود التكرارات** FREQUENCY **او** f_i **. وبذلك نكون حصلنا على الجدول التكراري والذي سوف يعرض فيما بعد بعمودين فقط الاول والثالث ويحذف عمود التفريغ (العد) .**

بعد عرض البيانات في الجدول يمكن ملاحظة فوائد هذا العرض من اختصار للبيانات ومعرفة تمركزها ومداها ،كما يمكن معرفة تكراراتها النسبية العشرية والمئوية كما في الجدول.

2.1.2.2 _ تبويب المتغيرات الكمية المستمرة :

يعتبر تبويب المتغيرات الكمية المستمرة اكثر تعقيدا من المنقطعة حيث يكون عدد القيم التي ياخذها المتحول المستمر غير محدود وبالتالي يجب تقسيم البيانات الى مجموعات تدعى كل مجموعة فئة Class .

تمر عملية انشاء الجدول التكراري للمتغيرات الكمية المستمرة بالخطوات التالية:

1− ترتيب البيانات الخام بشكل تصاعدي او تنازلي و يفيد ذلك في تسهيل التعرف على اكبر مفردة و اصغر مفردة و بالتالي المدى العددي للبيانات حيث ان المدى (Range) **يحسب كما يلي:-** $Range = MAX - MIN$

حيث MAX **:هي اكبر القيم و** MIN **:اصغرها.**

كما يساعد الترتيب في معرفة مدى اختلاف البيانات او تجانسها و بالتالي يساعد على تقسيمها الى مجموعات وذلك عند تفريغ البيانات في الجدول . هذا بالاضافة الى ان الترتيب يعتبر شكل مفيدا من اشكال العرض يكون ذو فائدة كبيرة اذا كان عدد المفردات قليلا ،اما في حال كان عدد المفردات كبيرا فهو عديم الفائدة لانه لايختصر البيانات .

2 − حساب عدد الفئات و طول كل فئة :

بما ان عدد القيم التي يأخذها المتغير غير محدود او كبير نقوم بتقسيم المدى الذي اوجدناه اعلاه الى فئات Classes او فترات Intervals،و ليس هناك من قاعدة ثابتة لتحديد طول الفئة و بالتالي عدد الفئات .

ولكن نقول يجب ان لايكون عدد الفئات كبيرا و بالتالي تضيع الفائدة المرجوة من الجدول و هي اختصار البيانات .كما انه اذا كان عدد الفئات صغير تضيع معالم التوزيع .

يقترح بعض المؤلفين ان يكون عدد الفئات بين 5 و 20 فئة و بالتالي نحسب طول الفئة بتقسيم المدى على عدد الفئات .

او باستخدام طريقة اخرى لتحديد طول الفئة و هي علاقة سترجس (Sturges)حيث تعتبر اكثر ملائمة لانها تاخذ المدى و عدد المفردات بعين الاعتبار ، وطول الفئة حسب هذه العلاقة:

$$c = \frac{Range}{1 + 3.32(Logn)}$$

حيث n عدد المفردات الخام الواجب تبويبها و C طول الفئة و Range المدى و الباقي هي ثوابت . وبعد حساب او تحديد عدد الفئات ،فان طول الفئة يحسب بتقسيم المدى على عدد الفئات ، و يجب مراعاة ان يكون عدد الفئات عددا صحيحا، و في حال استخدام علاقة سترجس يمكن تقريب طول الفئة ،ولا ننسى هنا ان التقريب بهذه الحالة دوما للاعلى Round off.

3 - اختيار حدود الفئات و طريقة كتابتها :

يؤثر اختيار حدود الفئات كثيرا على نتائج تحليل البيانات و الجدول الصحيح او الجدول السوي هو ذلك الجدول الذي يعطي نتائج تماما او قريبة جدا من النتائج التي يمكن حسابها من البيانات قبل التبويب ، ويمكن الحصول على جدول سوي باختيار حدود الفئات بشكل صحيح و هذه العملية معقدة و تتطلب عددا كبيرا من الحسابات و المقارنات و ذلك لان المفردات في كل فئة ستعامل لاحقا بان قيمتها تساوي وسط الفئة ،لذلك لابد من اختيار حدود الفئات بحيث يكون الوسط الحسابي لمفردات كل فئة يساوي قيمة منتصفها .او يمكن بطريقة اخرى مراجعة الترتيب و اختيار حدود الفئات بحيث يكون منتصفها يساوي القيم الاكثر تكرارا .

بعد اختيار حدود الفئات نختار احدى الطرق التالية لكتابة حدود الفئات جدول (2-5).

من ملاحظة الجدول (2-5) اعلاه نرى ان الفرق الاساسي في الطرق السابقة لكيفية كتابة حدود الفئات هو في كتابة الحد الاعلى للفئة وهنا يمكن ذكر التوضيحات التالية :في الطريقة الاولى ذكرت حدود الفئات وذكر توضيح للقيم المنتمية الى كل فئة ،فينتمي الى الفئة كل المفردات التي قيمها اكبر او تساوي الحد الادنى واقل من الحد الاعلى أي ان مجال الفئة هو من الشكل $]10 - 5]$مفتوح من الاعلى . اما في الطريقة الثانية فالعبارة التوضيحية للقيم المنتمية للفئة تبين ان المفردات التي قيمها مساوية للحد الاعلى للفئة تنتمي الى الفئة أي ان مجال الفئة هو من الشكل [9-5] وتستخدم هذه الطريقة في حال المتغيرات المنقطعة .ولتسهيل كتابة حدود الفئات نفضل عدم ذكر العبارات التوضيحية التي تدل على القيم المنتمية لكل فئة ونكتفي بكتابة حدود الفئات بينها اشارة معترضة كما في الطرق الرابعة والخامسة (الجدول 2−5)،حيث جرى العرف على قراءة حدود الفئات كما يلي :اذا تكررت كتابة الحد الاعلى للفئة كحد ادنى للفئة التالية فهذا يدل على مجال الفئة مفتوح من الاعلى أي ان المفردات التي قيمها تساوي الحد الاعلى للفئة لاتدخل في الفئة ،اما اذا لم تتكرر كتابة الحد الاعلى للفئة كحد ادنى للفئة التالية فمجال الفئة مغلق من الطرفين والمفردات التي قيمها مساوية للحد الاعلى للفئة تنتمي لهذه الفئة ، ويلجئ بعض المؤلفين الى حل وسط بين الطريقتين الرابعة والخامسة وذلك بعدم ذكر الحد الاعلى للفئة كما في الطريقة الثالثة في الجدول وهنا يقرا الحد الادنى للفئة كحد اعلى للفئة السابقة .

كيفية كتابة حدود الفئات

1	2	3	4	5
5 الى اقل من 10 **10 الى اقل من 15** **او** **5 الى 10 ولكن لاتشمل 10** **10 الى 15 ولكن لاتشمل 15**	**5 الى 9** **10 الى 14** **15 الى 19**	5 - 10 – 15 – 20 –	5 –10 10 –15 15 –20	**5 –9** **10 –14** **15 –19**

جدول (2−5)

4− التفريغ :

بعد تحديد عدد الفئات واختيار طريقة كتابتها نرسم جدول من ثلاثة اعمدة نكتب في العمود الاول حدود الفئات وفي العمود الثاني نفرغ البيانات بوضع اشارة (عصا العد Tally **) لتدل على ان مفردة ما تنتمي الى هذه الفئة وذلك كما مر معنا اعلاه .اما في العمود الثالث نترجم الاشارات الى ارقام عشرية ويكون عنوانه التكرار** Frequency **او بالرمز** f_i .

ولبيان كيف تتم هذه العملية نورد المثال التالي :

مثال : **لدينا البيانات التالية عن درجات 30 طالب . والمطلوب عرضها في جدول تكراري.**

1	0	4	5	3	4.5	6	8	7	8
6	**7**	**7**	**5**	**5**	**5**	**3**	**4**	**6**	5
9	**9**	**8**	7	**6**	**8**	**6.6**	**2**	**3**	0

لاعداد الجدول التكراري نتبع الخطوات التالية :

1 − نرتب البيانات ترتيب تصاعدي لمعرفة مدى البيانات و لكي يساعدنا الترتيب في عملية التفريغ

0	0	1	2	3	3	3	4	4	4.5
5	**5**	**5**	**5**	**5**	**6**	**6**	**6**	**6**	6.6
7	**7**	**7**	**7**	**8**	**8**	**8**	**8**	**9**	9

نحسب المدى : RANGE=MAX-MIN =9-0=9

2 −نحسب طول الفئة من علاقة سترجس :

$$c = \frac{Range}{1+3.32(Log n)} =$$

$$\frac{9}{1+3.32(Log 30)} = \frac{9}{1+3.32*1.698} = 1.52$$

نقرب الطول الى 2 فيكون هو طول الفئة .

3 −نكتب حدود الفئات حيث سنختار اصغر مفردة كحد ادنى للفئة الاولى و نضيف اليها طول الفئة 2 على التوالي فنحصل على حدود الفئات و سوف نختار كتابة الحدود بالشكل :

0−2

2−4

|

|

ويتضح هنا ان شكل المجال هو [2 – 0[**كما بينا اعلاه .**

4 − نقوم بتفريغ البيانات في الجدول كما هو في الجدول(2−6).

فئات الدرجات	العد **Tally**	التكرار f_i
0−2	ǀǀǀ	3
2−4	ǀǀǀǀ	4
4−6	𝍸 ǀǀǀ	**8**
6−8	𝍸 ǀǀǀǀ	9
8−10	𝍸 ǀ	**6**
المجموع		30

جدول (2− 6)

3.1.2.2 – صفات وخصائص الجدول التكراري :

– اوساط الفئات :

عند اعداد الجدول التكراري و نقل المفردات اليه تضيع المعالم الاصلية للبيانات ، وكل ما يخبرنا به الجدول هو ان المفردة تنتمي الى هذه الفئة او تلك الفئة ، لذلك سوف نفترض ان قيمة كل المفردات في فئة ما متساوية، و تساوي وسط تلك الفئة(مركز الفئة) x_i و يتم حساب وسط الفئة كما يلي :

$$x_i = (\text{الحد الاعلى} + \text{الحد الادنى})/2$$

– الجدول المنتظم و غير المنتظم :

عند حساب طول الفئة حسب علاقة سترجس او بالطريقة الاخرى ، و هي تحديد عدد الفئات بين 10 – 20 ،سوف نحصل على فئات متساوية الطول للجدول كما في الجدول (6.2) ويدعى الجدول التكراري في هذه الحالة جدولا منتظما . و لكن في بعض الحالات نرى ان هناك فئات فارغة من المفردات او قليلة المفردات فندمجها مع بعضها فنحصل على جدول بأطوال فئات غير متساوية الطول و بهذه الحالة يدعى الجدول غير منتظم ، ويفضل ان يكون الجدول بأطوال فئات متساوية لتسهيل الحسابات و تسهيل التمثيل البياني للتوزيع كما سنرى لاحقا .

– الجدول المغلق و الجدول المفتوح :

يدعى الجدول التكراري مغلقا اذا علمت حدود كل فئاته ، اما اذا وجد في الجدول التكراري ان الفئة الاولى بدون حد ادنى دعي الجدول مفتوحا من الادنى ، و يدعى الجدول مفتوحا من الاعلى اذا كانت آخر الفئات بدون حد أعلى . ان السبب في ترك أحد الحدود غير معلوما احيانا ،هو بسبب تباعد البيانات في الفئة الاولى او الاخيرة :مثلا توزيع الدخل ، فمعلوما ان أكثر الدخول في حدود معينة و هناك بعض الدخول تكون كبيرة بحيث يجب ان نخصص لها عددا من الفئات ، فيفضل في هذه الحالة ترك الحد الاعلى للفئة الاخيرة مفتوحا ، وقد يكون سبب ترك احد الحدود مفتوحا هو سرية البيانات .
من عيوب الجداول التكرارية المفتوحة هو اننا لا نستطيع ايجاد بعض المقاييس الاحصائية كما اننا لا نستطيع ايجاد التمثيل البياني للتوزيع كما سنرى لاحقا .

– التكرار التجميعي **Cumulative Frequency:**

يبين الجدول التكراري عدد المفردات في كل فئة ، ولكن قد نرغب بمعرفة عدد المفردات التي تقل عن حد معين او عدد المفردات التي تزيد عن حد معين وهذا ما نحصل عليه بايجاد التكرار التجميعي ، والتكرار التجميعي اما ان يكون صاعدا او هابطا .

نعرف التكرار التجميعي الصاعد $CF\uparrow$ بانه عدد المفردات في التوزيع والتي تقل قيمتها عن الحد الاعلى لكل فئة ونحصل عليه بجمع تكرار الفئات الى بعضها على التتالي .اما التكرار التجميعي الهابط $CF\downarrow$ فيعرف بانه عدد المفردات التي قيمها اكبر او تساوي الحد الادنى لكل فئة ،ونحصل عليه بطرح تكرار كل فئة من المجموع مبتدئين من اول فئة .

يمكن ان ناخذ التكرار التجميعي بشكل نسبي ومئوي ، نحصل عليه بتقسيم التكرار التجميعي على مجموع التكرارات . فالتكرار التجميعي الصاعد المئوي $CF\uparrow\%$ يعطي نسبة المفردات التي تقل قيمتها عن الحد الاعلى لكل فئة اما التكرار التجميعي الهابط المئوي $CF\downarrow\%$ فيعطي نسبة المفردات التي قيمتها تساوي او اكبر من الحد الادنى لكل فئة . انظر الجدول (2-7) .

كيفية حساب بعض المقاييس الاحصائية من الجدول التكراري

ات الدرجات	f_i كرار	وسط الفئة x_i	RF	Percentage%	$CF\uparrow$	$CF\downarrow$	$CF\uparrow\%$	$CF\downarrow\%$
0−2	3	**1**	**0.1**	**10**	**3**	**30**	**10**	100
2−4	**4**	**3**	**0.133**	**13.333**	**7**	**27**	**23.33**	90
4−6	8	**5**	**0.266**	**26.666**	**15**	**23**	**50**	76.66
6−8	9	**7**	**0.3**	**30**	**24**	**15**	**80**	50
8−10	6	**9**	**0.2**	**20**	**30**	**6**	**100**	20
المجموع	**30**		**1**	**100**				

الجدول (2−7)

–التكرار النسبي **Relative Frequency :**

نحصل على التكرار النسبي RF بتقسيم تكرار كل فئة على مجموع التكرارات ،والتكرار النسبي عبارة عن نسبة المفردات التي قيمها تنتمي لكل فئة ويمكن التعبير عنه بشكل عشري RF او بشكل مئوي (%) Percentage وذلك بضرب التكرار النسبي بمائة .انظر الجدول (2−7).

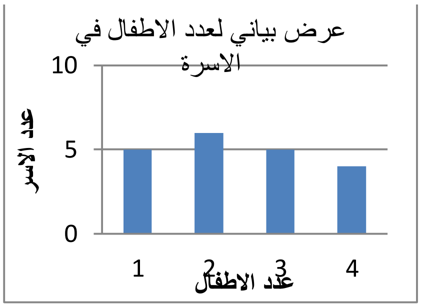
– مزايا و مساوئ الجدول التكراري:

يساعد الجدول التكراري على اختصار البيانات بشكل كبير حيث نستطيع حصر عدد كبير من المفردات في جدول واحد و بصفحة واحدة، كما يساعدنا الجدول على ملاحظة تمركز القيم في الجدول و ملاحظة انتشارها . ولكن بالاضافة الى هذه المزايا هناك سيئة خطيرة للجدول التكراري يجب الانتباه اليها حين اعداد الجدول و كذلك عند تحليل البيانات ، هي انه تضيع معالم المفردات في الجدول التكراري و تعطى قيما جديدة هي وسط الفئه التي تنتمي اليها المفردة و نتيجة ذلك تختلف المقاييس المحسوبة من البيانات قبل التبويب عن المقاييس المحسوبة بعد التبويب ، للتخلص من هذه السيئة يتم اعداد جدول تكراري صحيح او سوي باختيار حدود الفئات بحيث يكون وسط الفئة مساويا لوسط القيم المنتمية اليها وبذلك نحصل على جدول تكراري يعطي فرقا بسيطا بين المقاييس الاحصائية المحسوبة قبل التبويب و بعده او حتى يمكن ان يكون هذا الفرق معدوما .

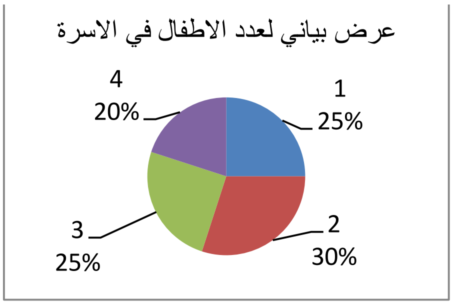
2.2.2– العرض البياني للمتغيرات الكمية :

1.2.2.2– العرض البياني للمتغيرات الكمية المنقطعة :

تعرض المتغيرات الكمية المنقطعة بيانيا كما تعرض المتغيرات النوعية التي تعرفنا عليها اعلاه، ويتم ذلك باستخدام الاعمدة البيانية والرسوم الدائرية.انظر الشكلين الشكل (2−6)و الشكل (2−7) حيث تم عرض بيانات الجدول (2−4).



الشكل (2−6)



الشكل (2−7)

2.2.2.2− العرض البياني للمتغيرات الكمية المستمرة :

تعرض المتغيرات الكمية المستمرة بيانيا باحد الاساليب التالية:

−المدرج التكراري

−المضلع التكراري

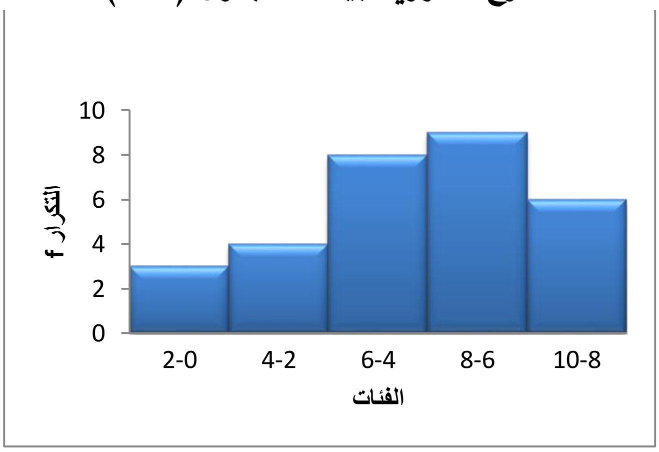
−المنحني التكراري

و سوف نرى كيف يمكن انشاء كل نوع من انواع العرض هذه .

−المدرج التكراري **Histogram** :

المدرج التكراري هو عرض بيانات الجدول التكراري على المحاور الاحداثية ، حيث تمثل الفئات على محور السينات و التكرارات F على محور العينات ، و نعبر عن تكرار كل فئة بمستطيل عمودي قاعدته مساوية لطول الفئة و ارتفاعه مساوٍ لتكرار الفئة .

المدرج التكراري لبيانات الجدول (2−6)



الشكل (2−8)

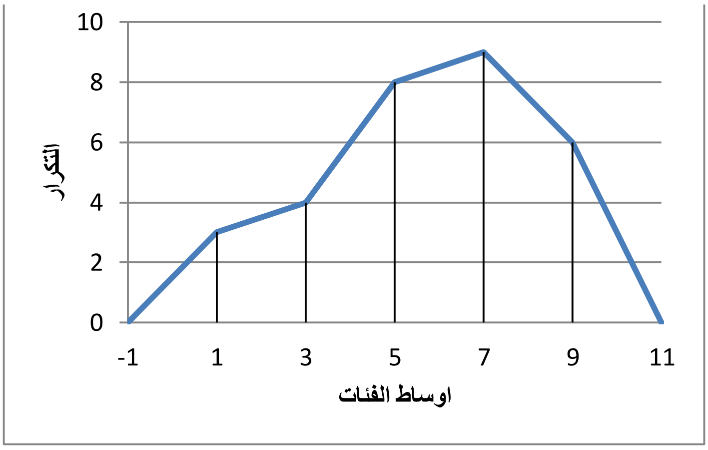
ان التمثيل البياني للتكرار F او التكرار النسبي Fr او التكرار المعدل بطول الفئة (حاصل تقسيم التكرار على طول الفئة) يعطي نفس الشكل عندما تكون أطوال الفئات متساوية و الشكل الناتج هو الشكل الحقيقي للتوزيع التكراري .

لاحظ ان الفرق بين المدرج التكراري والاعمدة البيانية ،هو انه في الاول الاعمدة متلاصقة بسبب وجود حدود مشتركة بين الفئات، في حين ان الاعمدة متباعدة .

- المضلع التكراري **Frequency Polygon :**

المضلع التكراري هو عرض للبيانات بشكل هندسي مغلق متعدد الاضلاع ويتم الحصول عليه كما يلي. نمثل النقاط التي هي اوساط الفئات وتكرار فئات الجدول التكراري على المحاور الاحداثية حيث نضع اوساط الفئات على محور السينات و التكرارات على محور العينات كما في الشكل 5 حيث قمنا بتمثيل بيانات الجدول (2−6) .

المضلع التكراري لبيانات الجدول (2−6)



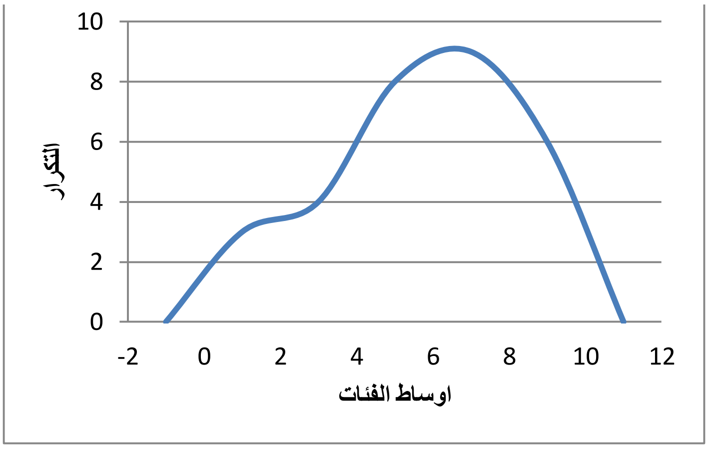
الشكل (2−8)

لاحظ انه من اجل اغلاق الشكل، قمنا باضافة فئتين و هميتين تكرار كل منها يساوي الصفر احداهما قبل اول فئة و الاخرى بعد آخر فئة.

- المنحنى التكراري **Frequency Curve:**

المنحنى التكراري هو مضلع تكراري بعد بتمهيد الخط المنكسر ليصبح منحني ، انظر الشكل 2.16 .ان التبرير العلمي لهذا التمهيد هو ، انه كلما كان عدد الفئات كبيرا و طولها قصيرا يقترب المضلع التكراري من المنحني التكراري .او بشكل اخر انه عندما يتناهى عدد الفئات الى اللانهاية يتناهى طول الفئة للصفر وبالتالي يتناهى طول الضلع للصفر .

المنحنى التكراري لبيانات الجدول (2−6)



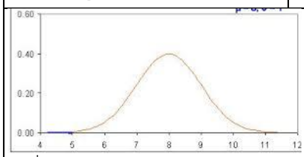
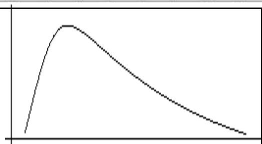
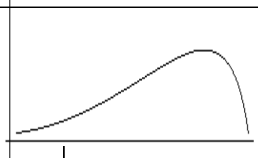
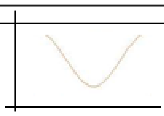
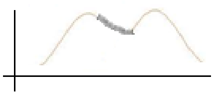
الشكل (2-9)

مايجب الانتباه اليه عند النظر الى المنحني التكراري هو ان ارتفاعه يشير الى التكرار دوما فابتعاد المنحني عن محور السينات يدل على تكرارات اكثر و اقترابه على تكرارات اقل .

-اشكال المنحنيات التكرارية :

تختلف اشكال المنحنيات حسب البيانات لكن يمكن ان نعطي تسميات للمنحنيات الاكثر شيوعا مع امثلة كمافي الجدول (2-8).

اشكال المنحنيات التكرارية مع الامثلة

	اسم المنحنى	بيانات لها شكل هذا المنحنى	شكل المنحنى
1	طبيعي Normal	1- نتائج امتحان عادل 2- اطوال عينة من الذكور	
2	ملتوي نحو اليمين Skewed to rig	1- نتائج امتحان صعب . 2- توزيع الدخل في مجتمع فقير.	
3	ملتوي نحو اليسار Skewed to le	1- نتائج امتحان سهل . 2- توزيع الدخل في مجتمع غني.	
4	منحنى نوني U shape	1- توزع الوفيات حسب العمر. 2- توزيع الدخل في الازمات.	
5	متعدد المنوال Multimodal	لما تكون البيانات غير متجانسة مأخوذة من مجتمعين او اكثر فاننا نحصل على منحني ذو قمتين (ثنائي المنوال Bimodal) او متعدد المنوال (Multimodal)، مثل الطول لعينة من الذكور والاناث.	

الجدول (2-8)

-خصائص و صفات التوزيعات التكرارية :

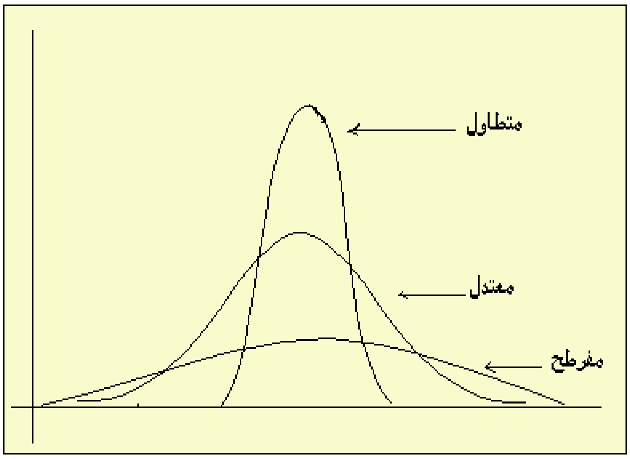
عندما تكون البيانات الاحصائية متجانسة و مأخوذة من مجتمع احصائي واحد فان التوزيع يتمتع بالخصائص التاليه:

1 –هناك قيمة في التوزيع تتجمع عندها القيم و هذا مايدعى بالنزعة المركزية (Central Tendency)و تقاس النزعة المركزية بمقاييس النزعة المركزيه و التي سندرسها بالتفصيل في الفصل القادم .

2 –بالاضافة الى تجمع القيم حول قيمة معينة يبدي التوزيع اختلاف في القيم و يبدو ذلك بابتعاد طرفي المنحني عن الوسط و هذا ما يدعى بالتشتت Variation و يقاس بمقاييس التشتت التي سندرسها لاحقا ايضا .

3 – بعض التوزيعات لها شكل التوزيع الطبيعي و بعضها ملتوي الرسوم A، C، D شكل (2.17).

4 – تختلف التوزيعات من حيث شدة التمركز فبعض التوزيعات شديدة التمركز و قليلة التشتت يدعى توزيعا متطاولا (Leptokurtic) و بعضها على العكس قليل التمركز و كثير التشتت يدعى توزيعا مفرطحا (Platykurtic) الشكل(2–10).



الشكل(2–10)

مسائل غير محلولة

1 –قم بانشاء جدول تكراري بخمس فئات لاوزان عينة بحجم 40 طفلا الواردة ادناه بالجدول (كيلو غرام):

30	24	30	34	20	34	23	26
28	40	28	28	20	23	34	28
25	30	25	30	25	32	22	31
32	29	32	34	26	40	25	33
34	31	34	44	32	41	32	34

2 –قم بانشاء جدول تكراري للبيانات التالية عن عدد الاطفال في الاسرة والماخوذة من عينة عشوائية بحجم 50 اسرة .

3– لدينا البيانات التالية والممثلة لدرجات الطلاب بمقرر الرياضيات لعينة حجمها 200:

الدرجات	عدد الطلاب
0-20	20
20-40	25
40-60	70
60-80	60
80-100	25
المجموع	200

المطلوب :

1. اوجد التكرار التجميعي الصاعد والهابط لبيانات الجدول السابق .
2. احسب اوساط الفئات
3. اوجد التكرار النسبي .
4. اوجد التكرار التجميعي الصاعد النسبي وكذلك الهابط النسبي.
5. اوجد عدد الطلاب الذين درجاتهم اقل من 80 درجة.
6. اوجد عدد الطلاب الذين درجاتهم اكثر او تساوي 40 درجة .
7. احسب نسبة الطلاب الذين درجاتهم اقل من 60 درجة .
8. مثل بيانيا البيانات السابقة مستخدما كل الاساليب الممكنة .

4– حدد طبيعة التوزيع للبيانات التالية من خلال رسم المنحنى التكراري :

الدخل الشهري (الف وحدة)	عدد الاسر
2 – 6 **6 – 10**	10 30 40

10 – 14 14 – 18 18 – 22	60 50
المجموع	180

5- فيما يلي حجم التداول في احد الاسواق المالية حسب نوع السهم في احدى الجلسات:

نوع السهم	حجم التداول (مليون)
مصارف	100
اتصالات	40
تامين	120
صناعية	90
	30

المطلوب:1- عرض البيانات بيانيا باستخدام الاعمدة البيانية.

2- عرض البيانات بيانيا باستخدام القطاعات الدائرية مع النسب المئوية .

يعرض الدول التالي الحقيبة لاستثمارية لثلاث مستثمرين (مليون) :

الاستثمار	المستثمر			المجموع
	A	B	C	
اسهم	**20**	**30**	**10**	60
سندات	**40**	**40**	**30**	110
مدخرات	**25**	**15**	**50**	90
اخرى	**15**	**5**	**20**	40
المجموع	**100**	**90**	**110**	300

المطلوب:

عرض بيانات المستثمر A بيانيا باستخدام الاعمدة البيانية.

عرض بيانات المستثمر A بيانيا باستخدام القطاعات الدائرية.

عرض البيانات للمستمرين الثلاثة بيانيا في شكل واحد باستخدام الاعمدة البيانية.

مراجع الفصل الثاني :

4- حميدان، عدنان؛ الجاعوني، فريد؛ ناصر آغا، عمار؛ العواد، منذر، 2003، مبادئ الإحصاء، منشورات جامعة دمشق.

5- Groebner David F., Shannon Patrick W. and Fry Phillip C. (2017)," Business Statistics :A Decision-Making approach", 10 th Edition. person.

6- Weiss, N.A.(1999), Introductory Statistics. Addison Wesley.

الفصل الثالث

مقاييس النزعة المركزية ومقاييس الموضع

Measures of central tendency and location measures

ملخص الفصل:

في هذا الفصل سنتعرف على كيفية وصف تمثيل البيانات بمقاييس رقمية ، من خلال حساب مقاييس النزعة المركزية :الوسط الحسابي ،الوسيط، المنوال، الوسط الهندسي والتوافقي ،ومقاييس الموضع الربيعيات والمئينيات ،ومقاييس التشتت :المدى التباين الانحراف المتوسط والربيعي والمعياري .وسنرى كيف نحدد شكل التوزيع باستخدام هذه المقاييس.

1.3- مقدمة :

لقد رأينا في الفصل السابق كيف يتم تلخيص البيانات الاحصائية و عرضها بشكل بياني كالمنحنى التكراري ، ومنه نصل لشكل التوزيع .ولكن ألا نستطيع وصف هذه البيانات أو المنحنى التكراري بشكل أدق بحيث نعطي مقاييس لوصف لهذه البيانات ؟نعم نستطيع ذلك من خلال المقاييس التي سوف ندرسها في هذا الفصل و الفصل القادم و هي مقاييس النزعة المركزية و مقاييس التشتت .

و تعتبر مقاييس النزعة المركزية من أدوات التحليل الاحصائي الهامة فنستطيع بواسطتها ايجاد قيمة ممثلة للبيانات ، يمكن استخدامها في مقارنة البيانات لعينات مختلفة أو مجتمعات مختلفة فاذا اردنا مقارنة مستوى الطلاب في صف مع مثيلهم في صف اخر. فكيف تتم المقارنة هل نقارن درجات جميع الطلاب مع درجات جميع الطلاب في الصف الاخر و هذا مستحيل ،فهنا يجب استخدام قيمة واحدة تعتبر ممثلة لمستوى الطلاب في كل من الصفين و هذه القيمة هي احدى مقاييس النزعة المركزية التي سوف ندرسها في هذا الفصل : الوسط الحسابي ،الوسيط ،المنوال ، الوسط الهندسي ، الوسط التوافقي ، و لكل من هذه المقاييس مزاياه و مساوءه والتي سنتعرف عليها لمعرفة المقياس المناسب لكل حالة .

عند حساب اي من المقاييس الاحصائية يجب التمييز بين حالتين هما:

- **بيانات غير مبوبة** Ungrouped Data
- **بيانات مبوبة** Grouped Data

البيانات غير المبوبة هي تلك البيانات التي تكون فيها على شكل مفردات ولا يتم تجميعها .اما المبوبة فعدد تكرارات كل مفردة ياخذ قيما مختلفة ،وهي التكرارات Frequency .

سنتناول فيما يلي بالتعرف وبالتفصيل على مقاييس النزعة المركزية الوسط الحسابي الوسيط المنوال الوسط الهندسي والوسط التوافقي بالإضافة الى مقاييس الموضع الربيعيات والمئينيات.

2.3Mean - الوسط الحسابي :

الوسط الحسابي هو اسهل مقاييس النزعة المركزية و أكثرها فهما ووضوحا و استخداما و شعبية فكل الناس تعرف استخدامه دون علم منه بالاحصاء و بالنزعة المركزية، فالمتسوق العادي يقارن اسعار اليوم بالامس باستخدام الوسط الحسابي ،والفلاح يقول بعت الكيلو من البطاطا اليوم ب 100 ليرة لاعوض خسارة البارحة حيث بعته ب50 ليرة أي وسطيا 75 ليرة . . .

و يعرف الوسط الحسابي بأنه مجموع القيم مقسوما على عددها ، و سنرى أدناه أن كل القوانين التي يتم تطبيقها في كل الحالات لحساب الوسط الحسابي، هي معبرة عن هذا التعريف و تطبيقا له .

- حساب الوسط الحسابي لبيانات غير مبوبة Ungrouped Data:

يتم حساب الوسط الحسابي لبيانات غير مبوبة كتطبيق مباشر للتعريف :

لتكن لدينا القيم : $x_1, x_2, x_3, \ldots x_n$

حيث n : عدد المفردات وهو حجم العينة .

فان الوسط الحسابي للعينة يساوي حسب التعريف :

$$\bar{X} = \frac{x_1 + x_2 + x_3 + \ldots\ldots + x_n}{n}$$

او يمكن كتابتها بالشكل :

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

اما الوسط الحسابي للمجتمع يساوي حسب التعريف :

$$\mu = \frac{x_1 + x_2 + x_3 + \ldots\ldots + x_N}{N}$$

حيث N : عدد المفردات وهو حجم المجتمع .

او

$$\mu = \frac{\sum x_i}{N}$$

مثال :لتكن لدينا البيانات التالية و التي هي عبارة عن الدخل الشهري لخمسة اسر ،30، 35، 35، 40، 60 الف ليرة فان متوسط دخل الاسرة هو :

$$\bar{X} = \frac{\sum X_i}{n} = \frac{30+35+35+40+60}{5} = 40 \quad \text{الف}$$

- حساب الوسط الحسابي المرجح **Weighted Mean**:

عند حساب الوسط الحسابي بالطريقة المباشرة كنا نعتبر أن كل المفردات لها نفس الاهمية (التكرار) أو هي متساوية بالاهمية ، لكن في بعض الحالات تكون المفردات مختلفة بالاهمية (التكرار) و لابد من أخذ ذلك بعين الاعتبار عند حساب الوسط الحسابي، ويتم ذلك عن طريق ترجيح القيم باهميتها النسبية أي بتكراراتها و هذا ما ندعوه بالوسط المرجح .

ويحسب المتوسط المرجح كما يلي:

لنفرض انه لدينا القيم: x_1 اهميتها او تكرارها w_1

x_2 اهميتها او تكرارها w_2

| |

x_n اهميتها او تكرارها w_n

فيكون الوسط الحسابي:

$$\bar{x} = \frac{x_1 w_1 + x_2 w_2 + ... + x_n w_n}{w_1 + w_2 + ... + .w_n}$$

او يمكن كتابتها :

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

مثال : فيما يلي الاجر الساعي لعينة من 10 عمال (وحدة نقدية)

ان الوسط الحسابي لهذه البيانات كبيانات غير مبوبة هو :

$$\bar{X} = \frac{\sum x_i}{n} = \frac{4+4+4+9+9+8+8+8+8+8=}{10} = \frac{70}{10} = 7$$

ولكن لو دققنا بهذه البيانات لراينا انه يمكن التعامل معها بطريقة اخرى ،حسب تكراراتها او اهميتها .القيمة 4 تتكرر 3 مرات ،والقيمة 9 مرتان اما القيمة 8 تتكرر 5 مرات وعليه يكون الوسط المرجح:

$$\bar{X} = \frac{\sum x_i w_i}{\sum w_i} = \frac{4*3+9*2+8*5}{3+2+5} = \frac{70}{10} = 7$$

- وسط الاوساط

اذا كان لدينا عدة عينات اوساطها $\bar{X}_1 - \bar{X}_2 - ....\bar{X}_n$ **وحجومها هي :** n_1 $n_2...n_n$ **على الترتيب، هنا يمكن اعتبار جم العينة هو قيم الترجيحات W ، وعليه فان الوسط الحسابي العام لهذه الاوساط يعطى بالعلاقة:**

$$\bar{X} = \frac{\bar{X}_1 n_1 + \bar{X}_2 n_2 + ... + .\bar{X}_n n_n}{n_1 + n_1 + ... + .n_n}$$

$$\bar{X} = \frac{\sum \bar{X}_i n_i}{\sum n_i}$$

مثال : **تبين من عينة عشوائية بحجم 100 عامل من مدينة دمشق ان متوسط الدخل هو 60 الف ليرة و من عينة عشوائية أخرى بحجم 150 عامل من حلب ان متوسط الدخل هو 70 ألف ليرة ،احسب متوسط الدخل في المدينتين.**

$$\bar{X} = \frac{\bar{X}_1 n_1 + \bar{X}_2 n_2}{n_1 + n_1} =$$

$$\bar{X} = \frac{60*100+70*150}{100+150} = \frac{16500}{250} = 66$$ **الف ليرة**

- الوسط الحسابي للبيانات المبوبة :

البيانات المبوبة هي بيانات على الاغلب معروضة في جدول تكراري ، وبالتالي يختلف تكرار القيم للمفردات ، أي يمكن بهذه الحالة استخدام الوسط الحسابي المرجح واعتبار التكرار هو قيم الترجيح .و عليه يحسب الوسط الحسابي للبيانات المبوبة من العلاقة :

$$\bar{X} = \frac{\sum f_i x_i}{\sum f_i}$$

حيث: x_i قيمة المفردة .

f_i تكرار المفردة .

$\sum f_i$مجموع التكرارات ويساوي حجم العينة n .

$\sum f_i\, x_i$: مجموع الجداء (نجري جداء f بـ x اولا ثم نجمع)

سنميز في الامثلة بين حالتين ،متغير كمي منقطع ومتغير كمي مستمر .

مثال:**لدينا البيانات التالية عن عدد افراد الاسرة لعينة عشوائية من الاسر. جدول(3−1):**

عدد افراد الاسرة x_i	f_i	$f_i x_i$
1	2	**2**
2	15	**30**
3	18	**54**
4	28	**112**
5	20	**100**
6	17	**102**
المجموع	100	**400**

جدول(3−1)

المطلوب :حساب الوسط الحسابي .

في هذا المثال لدينا متغير كمي منقطع هو عدد افراد الاسرة x_i وكل قيمة تتكرر عدد من المرات ، في عدد من الاسر f_i .اعتمادا على ذلك فان الوسط الحسابي يعطى بالعلاقة :

$$\bar{X} = \frac{\sum f_i x_i}{\sum f_i} = \frac{400}{100} = 4$$

اي ان متوسط عدد افراد الاسرة هو 4 افراد.

مثال:**لدينا البيانات التالية عن الدخل الشهري لعينة عشوائية من الاسر ، جدول(3−2):**

فئات الاجر (الف وحدة نقدية)	x_i	f_i	$f_i x_i$
4 - 8	6	**1**	6
8 - 12	10	**11**	110
12 -16	14	**22**	308
16 -20	18	40	720
20 -24	22	30	660
24 -28	26	16	416
المجموع		**120**	**2220**

جدول(3-2)

المطلوب :حساب الوسط الحسابي .

في هذا المثال لدينا متغير كمي مستمر هو الدخل الشهري للاسرة مقسم الى فئات ،لذلك سناخذ وسط الفئة x_i كقيمة للمفردات في الفئة والتي عددها ، اي عدد الاسر هو f_i .

اعتمادا على ذلك لايجاد الوسط الحسابي نوجد اوساط الفئات اولا ، ثم نضرب اوساط الفئات لكل فئة بتكرار الفئة $f_i x_i$ ثم نطبق العلاقة اعلاه كما يلي:

$$\bar{X} = \frac{\sum f_i x_i}{\sum f_i} = \frac{2220}{120} = 18.5 \text{ الف}$$

- مزايا ومساوئ المتوسط الحسابي :

1-مجموع الانحرافات عن الوسط الحسابي يساوي الصفر $\sum (X_i - \bar{X}) = 0$.

2- مجموع مربعات الانحرافات عن الوسط الحسابي هي أصغر ما يمكن $\sum (X_i - \bar{X})^2 \to \min$.

3- ان اضافة عدد ثابت c الى القيم Xi تؤدي الى زيادة الوسط الحسابي بنفس المقدار .

4- اذا ضربنا كل من القيم Xi بالعدد C فان الوسط الحسابي سيكون السابق مضروبا بنفس العدد.

5- عند حساب الوسط تؤخذ كافة القيم بعين الاعتبار .

6-الوسط الحسابي محدد جبريا بدقة ويتمتع بخواص جبرية لا يتمتع بها غيره من المقاييس مثل ضرب الوسط الحسابي بعدد القيم يساوي مجموع القيم ..... الخ.

7-هو مفهوم واضح وسهل الفهم اذ يمكن فهمه بسهولة على انه تلك القيمة التي يمكن اعطاؤها لكل مفردة ليبقى المجموع نفسه.

8-هو معيارٍ تقارن به القيم فنقول اقل من الوسط أو أكثر من الوسط ....

بالاضافة للمزايا السابقة للوسط بعض المساوئ التي تجعل منه مقياساً غير مرضياً منها :

1- لايمكن حسابه بدقة اذا كان الجدول التكراري مفتوحا، لانه لايمكن ايجاد وسط الفئة المفتوحة اذ يجب اغلاقها لنحصل على وسط فئة تقديري .

2- لا يمكن حساب الوسط الحسابي للبيانات النوعية .

3- الوسط الحسابي شديد التأثر بالقيم المتطرفة ، فهو في هذه الحالة لا يعتبر ممثلا للبيانات بشكل أفضل-في هذه الحالة هنا مقاييس افضل كما سنرى لاحقا .

المنوال **-3.3** Mode :

يعرف المنوال بأنه القيمة أو الصفة او المفردة الأكثر تكرارا او شيوعا بين البيانات و بالتالي نرى أنه من السهولة معرفة المنوال لبيانات غير مبوبة .

لنلاحظ الامثلة التالية لبيانات غير مبوبة ونستنتج بعض الخصائص للمنوال .

مثال 1 **:كانت درجات عينة من التلاميذ كما يلي :**

6 ، 5 ، 6 ، 7 ، 4 ، 6 ، 8 ، 9 ، 9 ، 9 ، 9 ، 10

و بالتالي يكون منوال الدرجات هو 9 درجة ، القيمة الاكثر تكرارا.

مثال 2**: كانت درجات عينة ثانية من التلاميذ كما يلي :**

6 ، 6 ، 6 ، 7 ، 4 ، 6 ، 8 ، 9 ، 9 ، 9 ، 9 ، 10

هنا يوجد قيمتان للمنوال هما 6 و 9 درجة ، القيم الاكثر تكرارا.

مثال 3**: كانت درجات عينة ثالثة من التلاميذ كما يلي :**

0 ، 3 ، 6 ، 7 ، 5 ، 1 ، 8 ، 9 ، 10

لا يوجد منوال .

مثال 4 **: كانت درجات عينة من الطلاب الجامعيين كما يلي:**

$A \quad A^+ \quad B^+ \quad B^+ \quad B^+ \quad B \quad B \quad C \quad C \quad C^-$

هنا المنوال الصفة (الرمز) الاكثر تكرارا هو B^+.

من هذه الامثلة نستنتج ما يلي :

1- في بعض الحالات يكون للبيانات منوال وحيد، او اكثر من قيمة منوالية ،ويمكن ان لايكون للبيانات اي منوال، في حين للبيانات وسط حسابي وحيد.

2- يمكن ايجاد المنوال للبيانات النوعية ،ولا يمكن ايجاد الوسط الحسابي .

- المنوال لبيانات مبوبه:

* المنوال للبيانات المبوبة النوعية والكمية المنقطعة:

يمكن ايجاد المنوال للبيانات المبوبة النوعية والكمية المنقطعة كما في البيانات غير المبوبة ،القيمة او الصفة الاكثر تكرارا.كما في الامثلة التالية.

مثال **1: في احد الصفوف تم تبويب الطلاب حسب اللغات الاجنبية كما يلي :**

فان المنوال هو اللغة الانكليزية،الاكثر تكرارا .

مثال2: **بالعودة الى بيانات الجدول(3−1) والمتضمن عدد افراد الاسرة لعينة عشوائية من الاسر، نرى ان المنوال هو الاسر التي عدد افرادها 4 افراد ،وهو العددالاكثر تكرارا.**

- المنوال للبيانات المبوبة الكمية المستمرة:

لايجاد المنوال من الجدول التكراري للبيانات الكمية المستمرة ، فان الخطوة الاولى تكون بتحديد الفئة المنوالية و هي الفئة المقابلة للتكرار الاكبر ،ضمن هذه الفئة يقع المنوال .سنعطي لهذه الفئة الرمز m وللفئة السابقة m-1 والتالية m+1 .

لتحديد قيمة المنوال يمكن استخدام عددا من الطرق أهمها طريقة الفروق حيث يحسب المنوال كما يلي:

$$Mod = L_m + \left[\frac{\Delta_1}{\Delta_1 + \Delta_2}\right] * C_m$$

حيث Mod : **المنوال**

L_m: **الحد الأدنى للفئة المنوالية .**

Δ_1 : **الفرق بين تكرار الفئة المنوالية وتكرار الفئة السابقة** $\Delta_1 = f_m - f_{m-1}$.

Δ_2 : **الفرق بين تكرار الفئة المنوالية و تكرار الفئة التالية** $\Delta_2 = f_m - f_{m+1}$.

C_m: **طول الفئة المنوالية .**

مثال : **في الجدول التالي(3−2) بيانات عن الدخل الشهري لعينة من الأسر .**

الفئة	فئات الدخل (الف وحدة)	عدد الاسر Fi
1	5 – 10	**15**
2	10 – 15	**25**
3	15 - 20	**30**
4	20 – 25	**40**
5	25 – 30	**10**
	المجموع	**120**

الفئة المنوالية

جدول(3−2)

لحساب المنوال فانه يجب اولا تحديد الفئة المنوالية m، وهي الفئة المقابلة للتكرار الاكبر وهي الفئة الرابعة في مثالنا ومن ثم نطبق العلاقة كما يلي :

$$\Delta_1 = f_m - f_{m-1} = 40 - 30 = 10$$
$$\Delta_2 = f_m - f_{m+1} = 40 - 10 = 30$$

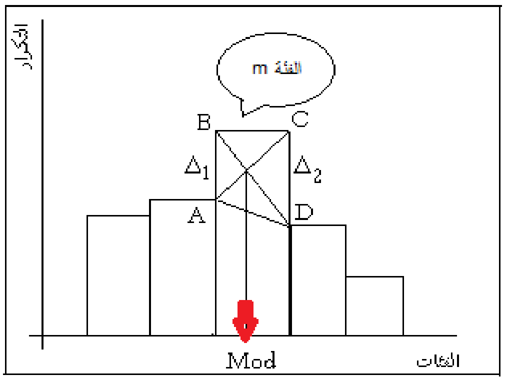
$$Mod = L_m + \left[\frac{\Delta_1}{\Delta_1 + \Delta_2}\right] * C_m$$

$$Mod = 20 + \left[\frac{10}{10+30}\right] * 5 = 21.25$$

و يفسر هذا الرقم بأن الأسر التي دخلها 21.25 الف وحدة هي الأكثر تكرارا .

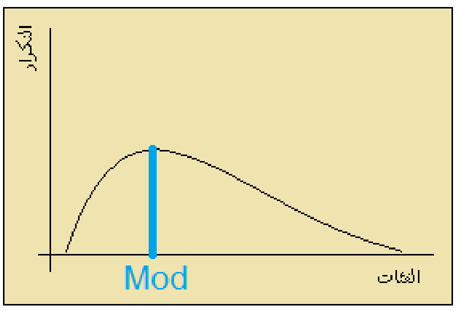
– تحديد المنوال بيانيا :

يمكن تحديد موقع المنوال بيانيا و ذلك بالفئة المقابلة لأكبر تكرار في المدرج التكراري ،الفئة المنوالية m . لحساب قيمته بدقة نصل بين أعلى تكرار الفئة السابقة و اعلى تكرار الفئة التالية فنحصل على الشكل الرباعي A B C D (انظر الشكل 3−1) نوجد مركز هذا الرباعي برسم أقطاره ثم نسقط المركز على محور السينات فيلاقيه عند نقطة ، هي قيمة المنوال.



الشكل (3−1)

كما يمكن من خلال المنحنى التكراري تحديد قيمة المنوال حيث تكون قيمته هي القيمة على محور السينات المقابلة لأعلى تكرار كما في الشكل (3−2) .



الشكل (3−2)

– مزايا و مساوئ المنوال :

1 – يمكن حسابه اذا كانت احدى فئات الجدول مفتوحة باستثناء الفئة المنوالية .

2 – لايتأثر بالقيم المتطرفة اذ أن القيم المتطرفة لا تؤثر على حسابه .

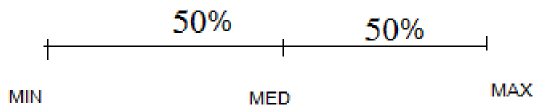
3 – يستخدم في حال البيانات النوعية .

4 – غير محدد جبريا بدقة اذ ان هناك عدة طرق لحسابه من الجداول التكرارية و كل طريقة تعطي قيمة مختلفة ، كما أنه لا يتمتع بمزايا جبرية يستفاد منها في حسابات أخرى.

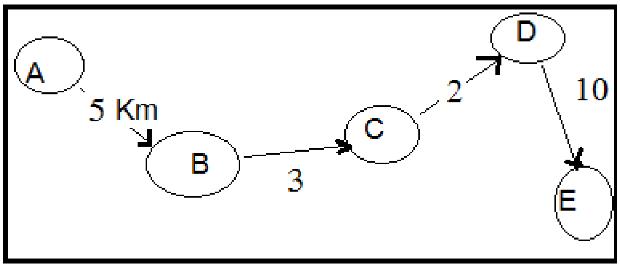
5 – تتغير قيمة المنوال بسرعة متأثرة بتغير أطوال الفئات في الجدول التكراري .

- الوسيط **MEDIAN**

الوسيط: يعرف الوسيط بأنه القيمة التي تقسم <u>الترتيب</u> **التصاعدي للبيانات الى قسمين متساويين بحيث عدد المفردات التي تسبق الوسيط يساوي عدد التي تليه .ويمكن التعبير عن ذلك بالشكل التالي :**



مثال : **لدينا البيانات التالية عن الطريق الواصل بين 5 مدن وطول الطريق بين كل مدينتين بالكيلو متر :**



ان وسيط المسافة هي النقطة التي تقسم الطريق الى قسمين متساويين ،وهذا يتحقق عند المدينة D قبلها 10 كيلو متر وبعدها 10 .اما وسيط عدد المدن فهو المدينة C ،عدد المدن قبلها يساوي عدد المدن بعدها .

- حساب الوسيط لبيانات غير مبوبة .

يتضح من التعريف بأن قيمة الوسيط تتحدد بموقعه (بموضعه) فهو من مقاييس الموضع و لحسابه يجب تحديد موقعه أولا ، القيمة بهذا الموقع هي الوسيط . لتحديد موقع الوسيط نطبق العلاقة التالية:

اذا كان عدد المفردات فردي : $Loc(Med) = \frac{n+1}{2}$

أما اذا كان عدد المفردات زوجي فلدينا موقعين وسيطين الموقع الأول $Loc(Med_1) = \frac{n}{2}$ **و الموقع الثاني** $Loc(Med_2) = \frac{n}{2} + 1$ **و قيمتين وسيطيتين وسطهما الحسابي هو الوسيط .**

مثال1 : **فيما يلي ترتيب 7 طلاب حسب درجاتهم باحد المقررات:**

65 67 70 72 80 85 93

موقع الوسيط لبيانات هذا المثال هو: $Loc(Med) = \frac{n+1}{2} = \frac{7+1}{2} = 4$

وبالتالي المفردة رقم 4 ، وهي 72 درجة قيمة الوسيط.

مثال 2 :**احسب الوسيط للبيانات التالية الممثلة لأوزان 6 اطفال ،كغ :**

20، 38 ، 28 ، 42 ، 32، 34

الحل :يجب اولا ترتيب البيانات بشكل تصاعدي :42 38 34 32 28 20

نلاحظ ان عدد البيانات زوجي فالموقع الوسيطي الأول $Loc(Med_1)=\frac{n}{2}$ وهو الموقع الثالث والقيمة الوسيطية الاولى هي 32 .

و الموقع الوسيطي الثاني $Loc(Med_2)=\frac{n}{2}+1$ وهو الموقع الرابع وبالتالي تكون القيمة الوسيطية الثانية هي 34 .اما الوسيط فهو الوسط الحسابي للقيمتين الوسيطيتين: $Med=\frac{32+34}{2}=33$

- حساب الوسيط للبيانات المبوبة :

تكون البيانات في الجدول التكراري مرتبة ترتيباً تصاعدياً ضمن الفئات ،بالتالي فان موقع الوسيط هو المفردة التي ترتيبها $\frac{\sum f}{2}$، أي ان هذه المفردة هي ضمن مفردات احدى الفئات، تدعى هذه الفئة بالفئة الوسيطية وسنعطيها الرمز m. بعد تحديد الفئة الوسيطية نطبق العلاقة التالية لحساب قيمة الوسيط:

$$Med=L_m+\left[\frac{\sum f/2-CF_{m-1}}{f_m}\right]*C_m$$

حيث : L_m :الحد الادنى للفئة الوسيطية

cf_{m-1}:التكرار التجميعي الصاعد للفئة قبل الوسيطية .

f_m :تكرار الفئة الوسيطية .

C_m :طول الفئة الوسيطية .

و سنبين من خلال المثال التالي كيف يتم حساب قيمة الوسيط .

مثال : **في الجدول التالي (3−3) بيانات عن الاجر الشهري ل 100 عامل من عمال احدى الشركات المطلوب حساب الوسيط:.**

التجميعي الصا cf↑	التجميعي الها cf↓	عدد العمال fi	فئات الاجر (1000 ليرة)	الفئة
10	**100**	**10**	0 - 6	
25	**90**	**15**	6 - 12	
55	**75**	**30**	12 – 18	
90	**45**	**35**	18 – 24	
100	**10**	**10**	24 – 30	
		100		

الفئة الوسيطية m وهي اول فئة بالتكرار الصاعد يكون في التكرار

جدول(3−3)

لمعرفة الفئة التي تحوي الوسيط، أي المفردة ذات الترتيب $50 = \frac{100}{2} = \frac{\sum f}{2}$ ندقق التكرار التجميعي الصاعد، و من تدقيقه نرى أن عدد العمال الذين يحصلون على أقل من 6 ألف هو 10 عامل و عدد العمال الذين يحصلون على أقل من 12 ألف هو 25 عامل، عدد العمال الذين يحصلون على أقل من 18 ألف هو 55 ، فالعامل ذو الترتيب$50 = \frac{100}{2} = \frac{\sum f}{2}$ هو ضمن عمال الفئة الثالثة ، بالتالي أصبح لدينا فكرة مبدئية بأن قيمة الوسيط هي أكبر أو تساوي 12 ألف (الحد الأدنى للفئة الوسيطية) ولن تزيد عن 18 ألف (الحد الأعلى للفئة الوسيطية) . فالوسيط هو الحد الادنى للفئة الوسيطية مضافا اليها جزء من طول الفئة الوسيطية يتناسب مع بعد الوسيط عن الحد الادنى للفئة الوسيطية و يحسب كما يلي :

$$Med = L_m + \left[\frac{\sum f/_{2} - CF_{m-1}}{f_m} \right] * C_m$$

تذكر: **أن كل المتغيرات في هذا القانون تتعلق بتحديد الفئة الوسيطية** m

لنحسب الوسيط لبيانات المثال اعلاه الجدول (3−3):

$$Med = 12 + \left[\frac{100/_{2} - 25}{30} \right] * 6 = 17$$ **الف ليرة**

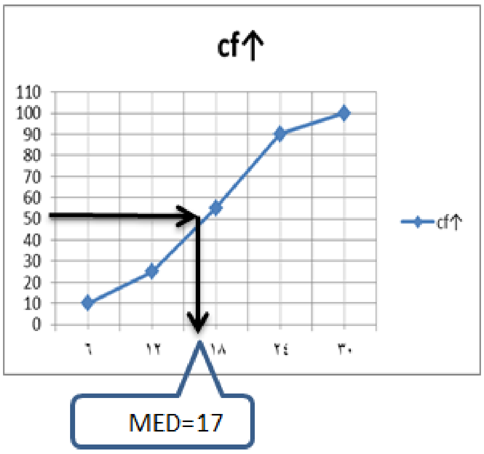
و اعتمادا على تعريف الوسيط نستطيع تفسير هذه القيمة بأن نصف العمال أو 50% منهم يتقاضون ما قيمته أقل أو يساوي 17 ألف ليرة و نصف العمال أو 50% منهم يتقاضون ما قيمته أكثر أو يساوي 17 ألف ليرة .

- تحديد الوسيط بيانيا :

يمكن تحديد الوسيط بيانيا باسلوبين هما :

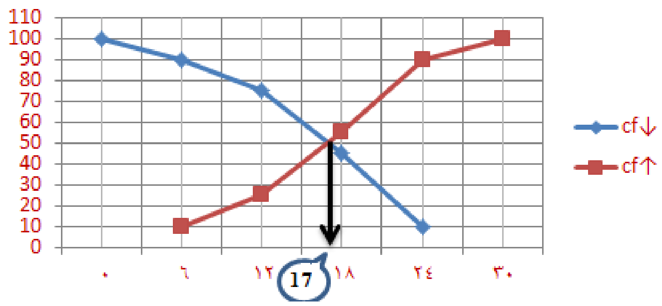
1 _ بتمتيل التكرار التجميعي الصاعد على المحاور الاحداثية و من ثم ننشئ مستقيم موازي لمحور السينات من النقطة $\sum f/_{2}$ على محور العينات باتجاه منحنى التكرار التجميعي الصاعد ، ومن نقطة تقاطع هذا المستقيم مع المنحنى ننزل عمود على محور السينات فيتقاطع معه عند قيمة الوسيط ،انظر الشكل (3−3) حيث تم تمثيل بيانات الجدول (3−3).

تحديد الوسيط بيانيا



(الشكل 3-3)

2- بتمثيل التكرار التجميعي الصاعد و الهابط على المحاور الاحداثية فيتقاطعان عند الوسيط ننزل عمود من نقطة التقاطع على محور السينات فيلاقيه عند الوسيط (الشكل 3-4) - بيانات الجدول (3-3).



(الشكل 3-4)

- مزايا و مساوئ الوسيط :

1 - تتحدد قيمة الوسيط بموقعه و ليس بقيمته او القيم الاخرى لذلك فهو لايتأثر بالقيم المتطرفة .

2 - يمكن حسابه اذا كانت احدى الفئات مفتوحة .

3- يجب النظر اليه بحذر في حالة العينات الصغيرة اذ أن قيمته يمكن أن تتغير بسرعة بمجرد اضافة أو طرح قيمة واحدة .

5- لايتمتع بأي خواص جبرية بحيث يمكن الاستفادة منه في حسابات لاحقة .

- مقاييس الموضع الاخرى :

رأينا أن الوسيط يقسم الترتيب الى قسمين متساويين 50 % من البيانات قبله و 50 % بعده ، بالمثل يمكن ايجاد قيم تقسم الترتيب بنسب مئوية مختلفة من البيانات منها :

1 _ الربيع الادنى (الاول)**First Quartile** و الاعلى (الثالث) **Third Quartile:**

يعرف الربيع الادنى بأنه المفردة في الترتيب التي يسبقها 25% من البيانات و يليها 75 %منها وواضح بأن الربيع يحدد بموقعه مثل الوسيط و لايجاده يجب تحديد موقعه اولا. بينما الربيع الثالث هو المفردة او القيمة في الترتيب التي يسبقها 75% من البيانات و يليها 25 %.

ملاحظة :الربيع الثاني هو الوسيط ،لذلك نادرا ما يدعى بهذه التسمية .

- الربيعات لبيانات غير مبوبة :

موقع الربيع الادنى او الاول : $Loc(Q1) = \frac{n+1}{4}$

موقع الربيع الاعلى أو الثالث : $Loc(Q3) = \frac{3(n+1)}{4}$

مثال : **لتكن لدينا البيانات التالية عن اوزان 11 طفل (كغ) :**

12 ، 13 ، 15 ، 16 ، 16 ، 17 ، 17.7 ، 18 ، 18.5 ، 19، 20

الربيع الاول :

موقع الربيع الأدنى ، هو الثالث $Loc(Q1) = \frac{n+1}{4} = \frac{11+1}{4} = 3$ **و قيمته 15 كغ .**

الربيع الثالث:

موقع الربيع الاعلى ، هو التاسع $Loc(Q3) = \frac{3(n+1)}{4} = \frac{3(11+1)}{4} = 9$ **و قيمته 18.5 كغ .**

عندما يكون موقع الربيع عددا ليس صحيحا ،يكون الربيع هو الوسط الحسابي للمفردتين في الموقعين الصحيحين ،السابق واللاحق .

_الربيعات لبيانات مبوبة :**كما فعلنا لايجاد الوسيط لبيانات مبوبة يمكن ان نفعله هنا لحساب الربيعات حيث يجب ايجاد فئة الربيع الادنى و فئة الربيع الاعلى و من ثم نطبق الدساتير التالية المشابهة لدستور الوسيط باستثناء أن المتغيرات فيها تتعلق بفئة الربيع :**

$$Q_1 = L_{Q1} + \left[\frac{\sum f / 4 - CF_{Q_1 - 1}}{F_{Q1}} \right] * C_{Q1}$$

حيث : Q_1 : الربيع الاول أو الادنى

L_{Q1} : الحد الأدنى لفئة الربيع الأدنى .

CF_{Q1-1} : التكرار التجميعي الصاعد للفئة السابقة لفئة الربيع الأدنى .

C_{Q1} : طول فئة الربيع الأدنى .

f_{Q1} :تكرار فئة الربيع الأدنى .

و يتم ايجاد فئة الربيع الأدنى بتدقيق التكرار التجميعي الصاعد ،و تكون فئة الربيع الادنى الفئة المقابلة لأول قيمة تصادفنا تساوي أو أكبر من $\sum f \,/\, 4$.

اما الربيع الاعلى او الثالث فيتم حسابه من العلاقة :

$$Q_3 = L_{Q3} + \left[\frac{3\sum f \big/ 4 - CF_{Q3-1}}{F_{Q3}} \right] * C_{Q3}$$

حيث : Q_3 :الربيع الاعلى او الثالث .

L_{Q3} :الحد الأدنى لفئة الربيع الاعلى .

CF_{Q3-1} :التكرار التجميعي الصاعد للفئة السابقة لفئة الربيع الاعلى .

C_{Q3} :طول فئة الربيع الاعلى.

f_{Q3} :تكرار فئة الربيع الاعلى.

و يتم ايجاد فئة الربيع الاعلى بتدقيق التكرار التجميعي الصاعد، حيث تكون فئة الربيع الاعلى هي الفئة المقابلة لأول قيمة نصادفنا تساوي أو أكبر من $3\sum f \,/\, 4$.

مثال : **فيما يلي درجات الطلاب في مقرر الاحصاء لعينة من 120 طالب.**

الفئة	الفئات	التكرار f_i	التكرار الصاعد $Cf_{\uparrow}$
1	0-20	**15**	15
2	20-40	**20**	35
3	40-60	**30**	65
4	60-80	**40**	105
5	80-100	**15**	200
	المجموع	**120**	

جدول(3−4)

المطلوب :**حساب الربيع الادنى و الأعلى لدرجات الطلاب أعلاه.**

نوجد التكرار التجميعي الصاعد في الجدول ، ثم نوجد موقع الربيع الاول ($\sum f/4 = 30$) . أي ان المفردة رقم 30 هي الربيع الادنى ، وتقع في الفئة الثانية . ومنه يمكن تطبيق العلاقة:

$$Q_1 = L_{Q1} + \left[\frac{\sum f/4 - CF_{Q_1-1}}{F_{Q1}} \right] * C_{Q1}$$

$$Q_1 = 20 + \left[\frac{30-15}{20} \right] * 20 = 35$$

وتفسر هذه القيمة بان 25%من البيانات هي اصغر اوتساوي 35 درجة و75% من البيانات اكبر او تساوي 35 درجة .

لحساب الربيع الثالث نوجد موقعه اولا :($3\sum f/4 = 90$)

اما فئة الربيع الثالث فهي الرابعة حيث تكرارها الصاعد 105، اول تكرار اكبر من 90 وعليه:

$$Q_3 = L_{Q3} + \left[\frac{3\sum f/4 - CF_{Q3-1}}{f_{Q3}} \right] * C_{Q3}$$

$$Q_3 = 60 + \left[\frac{90-65}{40} \right] * 20 = 72.5$$

وتفسر هذه القيمة بان 75%من البيانات هي اصغر اوتساوي 72.5 درجة، و25% من البيانات اكبر او تساوي 72.5 درجة .

2- المئينيات :

المئينيات هي القيم التي تقسم الترتيب الى 100 قسم بحيث كل قسم يحوي 1% من البيانات ، و بالتالي نستطيع القول أنه يسبق المئيني الأول 1% من البيانات و يليه 99% و يسبق الثاني 2 % و يليه 98% و يسبق الخامس 5% و يليه 95% و هكذا . . .

و يحدد موقع المئيني i لبيانات غير مبوبة كما يلي:: $Loc(P_i) = \frac{i(n+1)}{100}$

أما لبيانات مبوبة فيحدد موقعه بالعلاقة : $Loc(P_i) = \frac{i * \sum f}{100}$

ويحسب المئيني P_i من العلاقة:

$$P_i = L_{Pi} + \left[\frac{i*\sum f/100 - CF_{pi-1}}{f_{Pi}} \right] * C_{Pi}$$

حيث : i :رقم المئيني.

L_{pi} : الحد الادنى لفئة المئيني.

f_{Pi}: تكرار فئة المئيني.

C_{pi}:طول فئة المئيني.

و تحدد فئة المئيني بنفس طريقة تحديد فئة الوسيط و الربيع .

و يمكن أن نستنتج أن المئيني الخمسين يساوي الوسيط و المئيني الخامس والعشرين يساوي الربيع الاول ....

مثال **:لدينا البيانات التالية عن درجات عينة عشوائية حجمها 200 طالب في مقرر اللغة الاجنبية ماخوذة من احد معاهد اللغات، الجدول(3−4) والمطلوب :**

1− **احسب الوسيط و المئيني الخمسين للدرجات.**

2− **احسب المئيني الثلاثين للدرجات وفسر هذه القيمة .**

3− **احسب الدرجة التي يليها 10 % من الدرجات الاعلى.**

4− **قررت ادارة المعهد تنظيم دورة لـ 20 %من الطلاب اصحاب الدرجات الاقل ،اوجد مجال درجات الطلاب الواجب تنسيبهم للدورة .**

$Cf_{\uparrow}$	العدد f_i	فئات الدرجات	رقم الفئة
20	20		
50	30		
100	50		
170	70		
200	30		
	200	الم	

الجدول (3−5)

الحل:**نحسب التكرار التجميعي الصاعد** CF↑ **كما في الجدول وذلك من اجل تحديد فئة الوسيط والمئيني .**

− لحساب الوسيط نوجد موقع الوسيط **اولا** : $\text{Loc. Med } \frac{\sum f}{2} = \frac{200}{2} = 100$ **اي ان الفئة الثالثة هي فئة الوسيط،**

نحسب الوسيط من العلاقة: $$Med = L_m + \left[\frac{\sum f/2 - CF_{m-1}}{f_m} \right] * C_m$$

$$Med = 4 + \left[\frac{100-50}{50} \right] * 2 = 6$$ **درجة**

− **لحساب المئينئ الخمسين نوجد** موقع المئيني 50، هو $Loc.P_i = \frac{i * \sum f}{100} = \frac{50*200}{100} = 100$ اي **ان الفئة الثالثة هي فئة المئيني الخمسين .**

نحسب المئيني الخمسين من العلاقة:

$$P_i = L_{Pi} + \left[\frac{i\sum f / 100 - CF_{pi-1}}{f_{Pi}}\right] * C_{Pi}$$

$$P_{50} = 4 + \left[\frac{100-50}{50}\right] * 2 = 6$$

- لحساب المئيني **30** ، نوجد فئته اولا $Loc.P_i = \frac{i*\sum f}{100} = \frac{30*200}{100} = 60$ ، أي هي الفئة الثالثة ومنه يحسب المئيني الثلاثون :

$$P_{30} = L_{P_{30}} + \left[\frac{30*200 / 100 - CF_{p30-1}}{f_{P30}}\right] * C_{P30}$$

$$P_{30} = 4 + \left[\frac{60-50}{50}\right] * 2 = 4.4$$

- لحساب الدرجة التي يليها **10 %** من الدرجات الاعلى، يجب حساب المئيني تسعين .

موقع المئيني **90** هو: $\text{Loc.}\, P_i = \frac{i*\sum f}{100} = \frac{90*200}{100} = 180$

$$P_{90} = L_{P_{90}} + \left[\frac{90*200 / 100 - CF_{p90-1}}{f_{P90}}\right] * C_{P90}$$

$$P_{90} = 8 + \left[\frac{90*200 / 100 - 170}{30}\right] * 2 = 8.666$$

اي ان **10%** من الطلاب درجاتهم اكبر او تساوي 8.666 و**90%** اصغر منها او تساويها.

- لحساب مجال درجات الطلاب الواجب تنسيبهم للدورة والذي يضم **20 %**من الطلاب اصحاب الدرجات الاقل يجب ايجاد المئيني عشرين .

موقع المئيني **20** هو: $\text{Loc.}\, P_i = \frac{i*\sum f}{100} = \frac{20*200}{100} = 40$اي هو في الفئة الثانية ،ومنه:

$$P_{20} = L_{P_{20}} + \left[\frac{20*200 / 100 - CF_{p20-1}}{f_{P20}}\right] * C_{P20}$$

$$P_{20} = 2 + \left[\frac{40-20}{30}\right] * 2 = 3.333$$

اي ان الطلاب الذين يجب تنسيبهم للدورة هم من حصل على **3.333** واقل.

- الوسط الهندسي **(Geometric Mean):**

يعرف الوسط الهندسي لمجموعة من المفردات عددها n بأنه الجذر النوني لجداء هذه القيم .

فلو رمزنا لهذه القيم ب $x_1, x_2, x_3, ...x_n$ **و للوسط الهندسي ب G فان الوسط الهندسي يحسب من العلاقة:**

$$G = \sqrt[n]{x_1 * x_2 * .... * x_n}$$

– الوسط الهندسي للبيانات غير المبوبة :

يحسب الوسط الهندسي للبيانات غير المبوبة من العلاقة أعلاه بشكل مباشر .

مثال : **احسب الوسط الهندسي للقيم 1000 ، 40 ، 25 :**

$$G = \sqrt[3]{25*40*1000} = \sqrt[3]{1000000} = 100$$

كما يمكن حسابه باستخدام اللوغاريتم و ذلك بتحويل العلاقة الى لوغاريتمية عن طريق رفع الطرفين في العلاقة السابقة الى القوة n ، ثم أخذ لوغاريتم الطرفين :

$$G^n = x_1 * x_2 * .... * x_n$$
$$nLogG = Logx_1 + Logx_2 + ... + Logx_n$$
$$LogG = \frac{\sum Logx_i}{n}$$

ويجب الانتباه الى أن هذه العلاقة تعطينا لوغاريتم الوسط الهندسي و للحصول على الوسط الهندسي يجب اجراء العملية المعاكسة للوغاريتم $G = 10^{\log G}$.

مثال : **سنحل المثال السابق بالطريقة اللوغاريتمية:**

$$LogG = \frac{Log25 + Log40 + Log1000}{3} = \frac{1.398 + 1.602 + 3}{3} = \frac{6}{3} = 2$$

اي لوغاريتم الوسط الهندسي يساوي2 ومنه فان الوسط الهندسي يساوي $G = 10^{\log G} = 10^2 = 100$.

الآن لنفترض ان هذه القيمة 25 ، 40 ، 1000 هي أجور ساعية لثلاث عمال مختلفين عن بعضهم بالمؤهلات ونحسب الوسط الحسابي لهذه القيم و نقارن النتائج .

$\bar{X} = \frac{1000+25+40}{3} = 355$ **وحدة نقدية في حين ان الوسط الهندسي G=100 وحدة نقدية.الآن نستطيع تفسير الوسط الحسابي بأنه لو تساوت الأجور الساعية لهؤلاء العمال الثلاث لكان أجر كل منهم 355 و لكان مجموع مايتقاضاه العمال الثلاث هو 1065 وحدة نقدية .أما الوسط الهندسي فيمكن تفسيره رياضيا بأنه**

لو تساوت أجور هؤلاء العمال الثلاث و ساوى كل منها 100 وحدات لكان جداء الاجور يساوي الجداء السابق 1000000، و هنا نلاحظ ان قيمة الوسط الهندسي بهذه الحالة هي قيمة مجردة بلا معنى تطبيقي .اذا متى يكون للوسط الهندسي معنى أو قيمة تطبيقية ؟.

لنأخذ المثال التالي[1] : حوض ماء ابعاده 2 ، 4 ، 1 متر فتكون سعته أو حجمه 8 متر مكعب ، نريد تغير ابعاد هذا الحوض بحيث نحافظ على سعته اي ليصبح مكعب (متساوي الأضلاع)، ماهي القيمة التي سنستخدمها كطول للضلع ، هل هي الوسط الحسابي أم الوسط الهندسي .

اذا استخدمنا الوسط الحسابي لطول الضلع 2.666 **فان حجم الحوض الناتج سيكون** $\{(2.666)^3=18.95\}$ **متر مكعب و بالتالي نرى أننا لم نحصل على نفس الحجم .**

أما اذا استخدمنا الوسط الهندسي لطوال الضلع والمساوي الى مترين فنرى ان حجم الحوض الجديد سيكون($2*2*2 = 8$) متر مكعب و هو الحجم السابق. اذا يجب استخدام الوسط الهندسي لايجاد ابعاد الحوض الجديد و ان استخدام الوسط الحسابي لايجاد قيمة تمثيلية للقيم هو استخدام خاطئ . هذا المثال يبين لنا أن هناك حالات يجب استخدام الوسط الهندسي لايجاد القيمة التمثيلية و استخدام اي مقياس آخر هو استخدام خاطئ ، فما هي هذه الحالات ؟

ببساطة نقول : بكل الحالات التي يكون فيها لناتج الجداء معنى او قيمة تطبيقية يكون للوسط الهندسي معنى او قيمة تطبيقية، فجداء الأجور بلا معنى و لكن جداء أضلاع المكعب أو متوازي المستطيلات له معنى. لذلك سندرس بعض الحالات التي يكون فيها لناتج الجداء معنى و بالتالي يجب عندها استخدام الوسط الهندسي.

1 – متوسط معدل النمو :

معدل

سنرى من خلال بيانات المثال التالي جدول(3–6) و الممثلة لإنتاج معمل في الاعوام 2012 حتى 2016 كيف نحسب معدل النمو السنوي و متوسط معدل النمو خلال فترة من الزمن .

انتاج المعمل 2012–2016

معدل النمو بالنسبة للعام السابق %	الانتاج بالنسبة للعام السابق			الانتاج (طن)	العام
	(1+r)	(1+r)			

[1] المؤلف واخرون. (2004). *مبادئ الاحصاء – كلية الاقتصاد– جامعة دمشق*. دمشق: مديرية الكتب الجامعية.

2012	60	-	-	-	
2013	120	2	$(1+r_1)$	1+1	100
2014	90	0.75	$(1+r_2)$	1-0.25	-0.25
2015	90	1	$(1+r_3)$	1+0	0
2016	135	1.5	$(1+r_4)$	1+0.5	50
المجموع	495				125

جدول (3−6) **المصدر: فرضي**

نرى من بيانات الجدول ان معدلات النمو لكل سنة بالنسبة للعام السابق تختلف من عام لاخر ففي عام 2013 كان المعدل 100 %، وفي عام 2014 كان سالبا 25% ،وفي عام 2015 كان معدوما وارتفع الى 50% عام 2016.الان اذا اردنا ايجاد قيمة تمثيلية (معدل نمو) لهذه المعدلات المختلفة بحيث اذا طبقنا هذه القيمة على بيانات الجدول نحصل على نفس الناتج ، فاي من المقاييس نستخدم ؟.

للجواب على السؤال السابق يجب ان نتذكر ان النمو يمكن ان يكون حسب قانون الفائدة المركبة او حسب الفائدة البسيطة . لو فرضنا ان النمو يتم حسب الفائدة المركبة فيتم حساب ذلك كما يلي .

لاحظ مايلي من الجدول (3−6):اذا رمزنا لانتاج عام 2012 **بالرمز** y_{12} **ولانتاج عام 2013 بالرمز** y_{13} **وهكذا .....فان :**

$$y_{13} = y_{12} * (1+r_1)$$

$$y_{14} = y_{13} * (1+r_2)$$

$$y_{14} = y_{12} * (1+r_1) * (1+r_2)$$

$$y_{15} = y_{14} * (1+r_3)$$

$$y_{15} = y_{12} * (1+r_1) * (1+r_2) * (1+r_3)$$

$$y_{16} = y_{12} * (1+r_1) * (1+r_2) * (1+r_3) * (1+r_4)$$

وبالتالي هناك جداء اربعة اقواس في السطر الاخير فالمتوسط المناسب هو الهندسي :

$$(1+r) = \sqrt[4]{(1+1)*(1-0.25)*(1+0)*(1+0.5)} = 1.2247$$

$$(1+r) = 1.2247 \Rightarrow r = 1.2247 - 1 = 0.2247$$

الان نستطيع القول ان متوسط معدل النمو السنوي خلال الفترة المعطاة 2012 **حتى** 2016 **يبلغ** 22.47 **%**

. وللتاكد من ذلك نطبق هذا المعدل الناتج باستخدام قانون الفائدة المركبة :

$$y_{16} = y_{12} * (1+r)^4$$

$$y_{16} = 60*(1+0.2247)^4 = 135$$

نلاحظ هنا بانه كان من الممكن ايجاد متوسط معدل النمو كما يلي من قانون الفائدة المركبة :

$$y_n = y_0(1+r)^n$$

$$(1+r) = \sqrt[n]{\frac{y_n}{y_0}} \Rightarrow r = \sqrt[n]{\frac{y_n}{y_0}} - 1$$

حيث ان y_n **قيمة الظاهرة في الزمن** n .

و y_0 **قيمة الظاهرة في الزمن** 0 .

$$r = \sqrt[4]{\frac{135}{60}} - 1 = 0.2247$$

2- متوسط معدل الفائدة :

من المعروف أن التوظيفات المالية توظف في اغلب الأحيان بمعدل فائدة ثابت طيلة فترة التوظيف و يطبق عليها قانون الفائدة المركبة $K_n = K_0(1+i)^n$ **و لكن في بعض الحالات تطبق معدلات فائدة مختلفة من سنة لاخرى و تحسب جملة المبلغ :**

$$K_n = K_0(1+i_1)*(1+i_2)*.......*(1+i_n)$$

واضح هنا أن هناك علاقة جداء بين الأقواس فلايجاد متوسط معدل الفائدة يجب استخدام الوسط الهندسي كما في الحالة السابقة.

3 _ الرقم القياسي للأسعار[2] :

يقيس الرقم القياسي للأسعار التغير النسبي لأسعار السلع و هو يأخذ بعين الاعتبار حوالي 160 سلعة ويحسب بطرق مختلفة (كما سنرى لاحقا) . هنا للسهولة سنحسب الرقم القياسي لسلعتين فقط لنرى مدى ملائمة استخدام الوسط الهندسي ، و سنناقش ذلك من خلال المثال التالي: نفرض أنه لدينا السلعتين A , B و اسعارهما في عام 2000 و عام 2012 م كما يلي:

السلعة	عر سنة الاساس 2000		سعر سنة الدراسة 2012	
	السعر (وحدة)	سبة لسنة الاس	السعر (وحدة)	سبة لسنة الاس
A	5	100	10	200
B	10	100	5	50
الوسط الهندسي		100		100
الوسط الحسابي		100		125

[2] راجع فصل الارقام القياسية من هذا المقرر .

نلاحظ أن الوسط الحسابي يبين أن الاسعار قد ارتفعت عام 2012 بنسبة 25 % عنها في عام 2000 و هذا غير صحيح حيث ما كان يدفع مقابل السلعتين في عام 2000 هو نفسه ما يدفع عام 2012 و هو 15 وحدة . أما اذا أخذنا الوسط الهندسي فإننا نلاحظ ثبات الاسعار حيث الوسط الهندسي لأسعار عام 2012 يشكل 100 %من أسعار 2000 أي أن الاسعار ثابتة و هذا صحيح ، حيث الزيادة في سعر سلعة يعوضها نقص في سلعة أخرى .

– الوسط الهندسي للبيانات المبوبه

يعطى الوسط الهندسي في الجداول التكرارية بالعلاقة التالية :

$$G = \sqrt[\Sigma f_i]{x_1^{f_1} * x_2^{f_2} * ...... * x_n^{f_n}}$$

أو بالشكل اللغاريتمي :

$$LogG = \frac{\Sigma f_i Logx_i}{\Sigma f_i}$$

حيث x_i **أوساط الفئات في الجدول التكراري للمتغيرات الكمية المستمرة او قيمة المتغير للمتغيرات الكمية المنقطعة ، و** F_i **تكرار القيم و يجب أن نتذكر أن هذه العلاقة تعطينا لغاريتم الوسط الهندسي ويجب ايجاد الوسط الهندسي كما مر سابقا.**

مثال:**احسب الوسط الهندسي لبيانات الجدول التالي (3−7):**

الفئات	f_i	x_i	$Logx_i$	f_iLogx_i
0 -2	10	1	0	0
2 -4	20	3	0.47712	9.542
4 -6	30	5	0.699	20.969
6 -8	35	7	0.845	29.578
8 -10	5	9	0.954	**4.771**
المجموع	100			64.86

جدول (3−7)

الحل:**لايجاد الوسط الهندسي لبيانات الجدول التكراري نوجد اولا اوساط الفئات و كذلك لوغاريتم اوساط الفئات ، ثم نضرب لوغاريتم وسط كل فئة بتكرار الفئة** $F_i \, Log \, x_i$ **وبعد ذلك نطبق على العلاقة التالية التي تعطينا لوغاريتم الوسط الهندسي :**

$$LogG = \frac{\Sigma f_i Logx_i}{\Sigma f_i}$$

$$LogG = \frac{64.86}{100} = 0.6486$$

ومنه الوسط الهندسي هو $G = 10^{LogG} = 10^{0.6486} = 4.452$

- مزايا و مساوئ الوسط الهندسي :

1 _ محدد جبريا بدقة .

2 _ هو مفهوم وسطي مثله مثل الوسط الحسابي .

3 _ قليل التأثر بالقيم المتطرفة نحو اليمين بعكس الوسط الحسابي لذلك اذا كان التوزيع ملتوي نحو اليمين فيكون الوسط الهندسي افضل المقاييس لتمثيل البيانات.

4 _ هو المقياس الوحيد الواجب استخدامه في بعض الحالات كما رأينا .

5 _لايمكن حسابه للجداول المفتوحة .

6_لايمكن حسابه اذا كانت احدى القيم سالبة أو معدومة .

7_ هو قيمة مجردة و غير واضحة في بعض الحالات .

- – الوسط التوافقي **Harmonic Mean**

1– الوسط التوافقي لبيانات غير مبوبة : يعرف الوسط التوافقي بأنه مقلوب الوسط الحسابي لمقلوبات القيم ، حيث مقلوب القيمة x هو $\frac{1}{x}$.

لنفرض أن لدينا القيم $x_1, x_2, x_3, x_4.........x_n$ **عددها n قيمة فان وسطها التوافقي حسب التعريف ، هو :**

$$H = \frac{1}{\frac{\sum 1/x_i}{n}} = \frac{n}{\sum \frac{1}{x_i}}$$

و يعتبر الوسط التوافقي من مقاييس النزعة المركزية قليلة الاستخدام و لكن في بعض الاحيان يوفر الحسابات اللازمة لحل مسألة ما بطريقة مختصرة مثل ايجاد متوسط السرعة.

مثال : **تم جمع البيانات التالية عن حافلة ركاب خلال رحلتها بين 3 مدن** ABC :

الرحلة	المسافة كيلو متر	الزمن	سرعة:كيلومتر/بالساعة
$A \to B$	**300**	**2**	150
$B \to C$	**300**	**3**	100
	600	**5**	250

المطلوب متوسط السرعة (كيلومتر/بالساعة)التي قطعت بها الحافلة المسافة بين A وC .

الحل : باستخدام الوسط الحسابي :

$$\bar{x} = \frac{\sum x}{n} = \frac{250}{2} = 125$$ **كم /ساعة**

الحل : باستخدام الوسط التوافقي: $H = \frac{n}{\sum \frac{1}{X_i}} = \frac{2}{\frac{1}{150} + \frac{1}{100}} = \frac{2}{0.0266} = 120$ **كم /ساعة**

لنتساءل الان ايهما الاصح في هذه الحالة ،الوسط الحسابي ام التوافقي .الجواب هو التوافقي لانه يمكن ان نحسب متوسط السرعة كما يلي (المسافة /الزمن)=(120=5/600)وهي قيمة الوسط التوافقي وليس الحسابي .

2 _ الوسط التوافقي للبيانات المبوبة :

تصبح علاقة الوسط التوافقي للبيانات المبوبة بعد اخذ التكرار بعين الاعتبار كما يلي :

$$H = \frac{\sum F_i}{\sum \frac{F_i}{X_i}}$$

حيث : x_i **اوساط الفئات او القيم ،**f_i **تكرارتها ، فعند الحل يجب ايجاد القيم** $\frac{f_i}{x_i}$ **أولا ثم نطبق العلاقة اعلاه لحساب الوسط التوافقي .**

مثال:**احسب الوسط التوافقي لبيانات المثال السابق والتي اعدنا كتابتها بالجدول التالي (3−7):**

فئات الاجر(الف)	f_i	x_i	$\frac{f_i}{x_i}$
4 -6	10	5	2
6 -8	35	7	5
8 -10	45	9	5
10 -12	22	11	2
12 -14	26	13	2
المجموع	138		16

جدول (3−8)

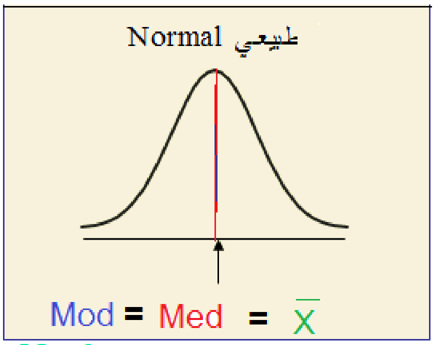
الحل:بعد ان اوجدنا اوساط الفئات وحسبنا لكل فئة القيم $\frac{f_i}{x_i}$ **نطبق العلاقة كما يلي:**

$$H = \frac{\sum f_i}{\sum \frac{f_i}{x_i}} = \frac{138}{16} = 8.625$$ **الف**

4.3- العلاقة بين الوسط الحسابي و الوسيط و المنوال :

- اذا كان التوزيع التكراري طبيعياً تتحقق العلاقة التالية : $\overline{X}=Med=Mod$

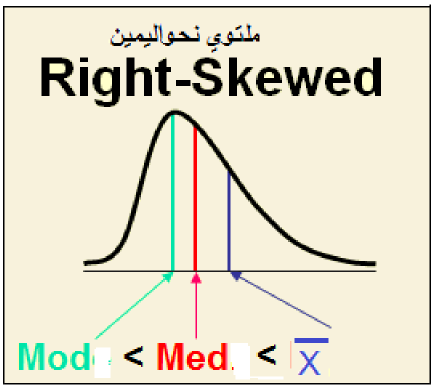
ويكون المنحنى التكراري كما في الشكل (3-5).



الشكل (3-5)

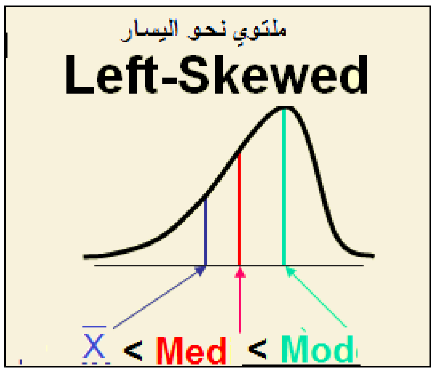
- اذا كان التوزيع ملتوياً تتحقق علاقة بيرسون التالية : $\overline{X}-Mod=3(\overline{X}-Med)$

ويكون توضع هذه المقاييس كما في الشكل (3-6) اذا كان التوزيع ملتوياً نحو اليمين .



الشكل (3-6)

ويكون توضع هذه المقاييس كما في الشكل (3-7) اذا كان التوزع ملتوياً نحو اليسار .



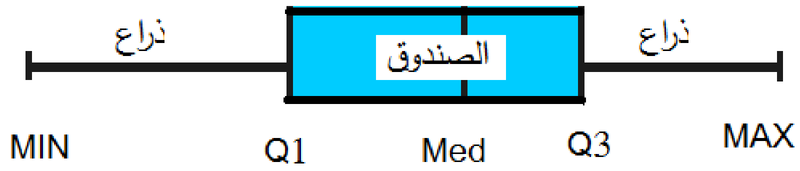
الشكل (3-7)

شكل الصندوق والاذرع 5.3- Box and Whisker Plot:

يمكن تمثيل البيانات باستخدام 5 قيم لمعرفة شكل التوزيع ،هذه القيم هي :

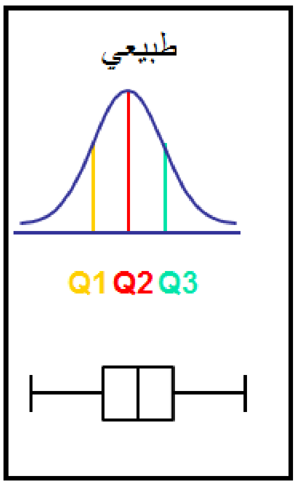
Minimum ----- Q1 ------- Median ------ Q3 ------ Maximum

وذلك كما يلي :



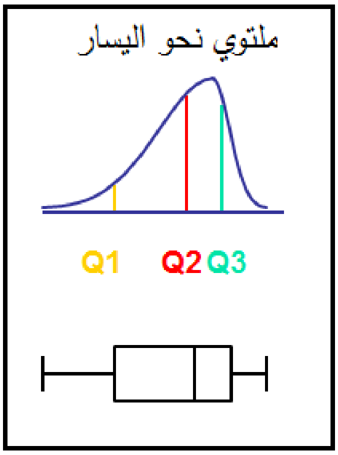
يختلف هذا الشكل من توزيع لاخر معبراً عن شكل التوزيع التكراري ،وذلك كما يلي :

- **اذا كان التوزيع طبيعياً ، يكون الصندوق في المنتصف بين اصغر واكبر القيم وكذلك خط الوسيط في منتصف الصندوق ،الشكل (3-8).**



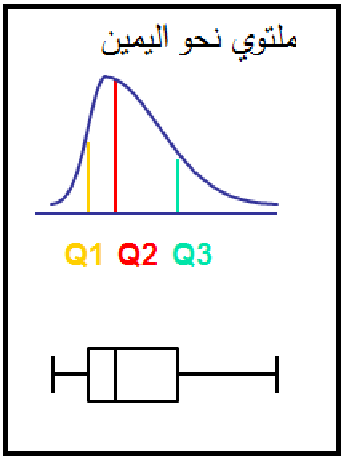
الشكل (3-8)

- **اذا كان التوزيع ملتويا نحو اليسار يكون الصندوق اقرب الى MAX في يمين الشكل بين اصغر واكبر القيم وخط الوسيط اقرب الى Q_3،الشكل (3-9).**



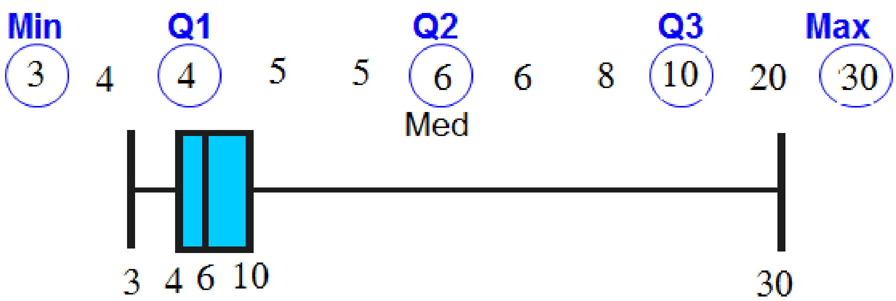
الشكل (3−9)

- **اذا كان التوزيع ملتويا نحو اليمين يكون الصندوق اقرب الى MIN في يسار الشكل وخط الوسيط اقرب الى Q_1،الشكل (3−10).**



الشكل (3−10)

مثال:فيما يلي اوزان عينة من 11 طفل :



من الشكل نرى ان الصندوق على يسار الشكل وكذلك خط الوسيط على يسار الصندوق ،او بشكل اخر، نرى ان الذراع الطويل يقع الى اليمين ،وبالتالي ان شكل التوزيع ملتوي نو اليمين .

مسالة محلولة :

1− **بغية دراسة الوضع المعاشي للاسرة في احدى المدن تم سحب عينة عشوائية بحجم 100اسرة وتم جمع وتبويب البيانات التالية :**

فئات الدخل (الف)	عدد الاسر f_i	وسط الفئة x_i	$f_i x_i$	$cf^{\uparrow}$
2-6	10	4	40	**10**
6-10	20	8	160	**30**
10-14	30	12	360	**60**
14-18	35	16	560	**95**
18-22	5	20	100	**100**
المجموع	**100**		1220	

جدول (3−8)

المطلوب :

1. **احسب متوسط الدخل للاسرة .**
2. **احسب الوسيط للدخل وفسر هذه القيمة.**
3. **احسب المنوال للدخل وفسر هذه القيمة.**
4. **اوجد حدود المجال الذي يضم 25 %من الاسر الاقل دخلا.**
5. **اوجد حدود المجال الذي يضم 20 %من الاسر الاكثر دخلا.**
6. **حدد طبيعة التوزيع التكراري للدخل باستخدام مقاييس النزعة المركزية التي اوجدتها اعلاه .**

الحل:1 − **متوسط الدخل:** $\bar{X} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{1220}{100} = 12.2$ **الف وحدة**

2 −**احسب الوسيط للدخل وفسر هذه القيمة.نوجد التكرار التراكمي في الجدول لنحدد فئة الوسيط ، ثم نحدد موقع الوسيط** $\text{loc of Med} = \frac{\Sigma f}{2} = \frac{100}{2}$ **، أي بالفئة الثالثة وهي ما دعوناها بالفئة الوسيطية** m **.بعد ذلك نحسب الوسيط من العلاقة:**

$$Med = L_m + \left[\frac{\sum f / 2 - CF_{m-1}}{f_m} \right] * C_m$$

$$Med = 10 + \left[\frac{50-30}{30} \right] * 4 = 12.666$$

أي وسيط الدخل 12.666 الف وحدة نقدية ، هذه القيمة تقسم البيانات الى قسمين ،نصف البيانات او نصف الاسر دخلها اصغر او يساوي 12.666 الف والنصف الاخر اكبر او يساوي هذه القيمة .

3− احسب المنوال للدخل وفسر هذه القيمة.

لحساب المنوال فانه يجب اولا تحديد الفئة المنوالية m **،وهي الفئة المقابلة للتكرار الاكبر وهي الفئة الرابعة في مثالنا .ومن ثم نطبق العلاقة كما يلي :**

$$\Delta_1 = f_m - f_{m-1} = 35 - 30 = 5$$
$$\Delta_2 = f_m - f_{m+1} = 35 - 5 = 30$$

$$Mod = L_m + \left[\frac{\Delta_1}{\Delta_1 + \Delta_2} \right] * C_m$$

$$Mod = 14 + \left[\frac{5}{5+30} \right] * 4 = 14.571$$

و يفسر هذا الرقم بأن الأسر التي دخلها 14.571 الف وحدة هي الأكثر تكرارا .

4 –اوجد حدود المجال الذي يضم 25 %من الاسر الاقل دخلا.

بما ان هذا المجال هو 25 % فيجب حساب الربيع الادنى : نوجد موقع الربيع الاول ($\sum f/4 = 25$) .

أي ان المفردة ذات الترتيب 25 هي الربيع الادنى ، وتقع في الفئة الثانية . ومنه يمكن تطبيق العلاقة:

$$Q_1 = L_{Q1} + \left[\frac{\sum f / 4 - CF_{Q_1 - 1}}{F_{Q1}} \right] * C_{Q1}$$

$$Q_1 = 6 + \left[\frac{25 - 10}{20} \right] * 4 = 9$$

حدود المجال الذي يضم 25 % من الاسر الاقل دخلا هو 9 الف وحدة و اقل.

5–حدود المجال الذي يضم 20 %من الاسر الاكثر دخلا ،نحسب المئيني الثمانين الذي يليه 20%من المفردات ، موقع المئيني 80 هو : $\text{Loc.}\, P_i = \frac{i * \sum f}{100} = \frac{80*100}{100} = 80$**، اي ان فئته الرابعة وبالتالي نطبق العلاقة :**

$$P_{80} = L_{P_{80}} + \left[\frac{80*100/100 - CF_{p80-1}}{f_{P80}} \right] * C_{P80}$$

$$P_{80} = 14 + \left[\frac{80*100/100 - 60}{35} \right] * 4 = 16.285$$

اي ان المجال الذي يضم 20%من الاسر الاكثر دخلا هو 16.285 الف وحدة فما فوق .

6–لتحديد طبيعة التوزيع نقارن قيم الوسط الحسابي و المنوال والوسيط ،نرى ان قيمة المنوال تساوي 14.571 الف وحدة وهي اكبر من الوسط الحسابي 12.2 واكبر من الوسيط وبالتالي فان شكل التوزيع ملتوي نحو اليسار.

تمارين غير محلولة

لدينا البيانات التالية عن الاجور الساعية لعينة من 7 عمال :

العامل	1	2	3	4	5	6	7	المجموع
الاجر الساعي	8	7	8	9	8	10	13	63

المطلوب:

ان الوسط الحسابي هو:

E) غير ذلك	D) 7.5	C) 9	B) 8	A) 10

ان الوسيط هو:

E) غير ذلك	D) 7.5	C) 9	B) 8	A) 10

ان المنوال هو:

E) غير ذلك	D) 7.5	C) 9	B) 10	A) 8

ان الوسط الهندسي يساوي:

E) غير ذلك	D) 7.3335	C) 9.333	B) 10.22	A) 8.832

ان الوسط التوافقي يساوي:

E) غير ذلك	D) 8.686	C) 8.333	B) 7.22	A) 9.832

فيما يلي درجات عينة عشوائية بحجم 120طالب في مقرر الاحصاء:

عدد الطلاب f_i	فئات الدرجات
2	0-20
8	20-40
20	40-60
70	60-80
20	80-100
120	المجموع

المطلوب :1 –احسب متوسط الدرجات .

2 –احسب الوسيط للدرجات وفسر هذه القيمة.

3- احسب المنوال للدرجات وفسر هذه القيمة.

4 –اوجد حدود المجال الذي يضم 75 %من الطلاب الاقل الدرجات.

5 –اوجد حدود المجال الذي يضم 25 %من الطلاب الاكثر الدرجات.

6- اوجد حدود المجال الذي يضم 25 %من الطلاب الاقل الدرجات.

–حدد شكل التوزيع التكراري للدرجات .

في الجدول التكراري التالي عدد الاطفال في عينة عشوائية من اسر مدينة دمشق.

عدد الاطفال	0	1	2	3	4	المجموع
عدد الاسر	20	45	70	45	20	200

المطلوب :1 –احسب متوسط عدد الاطفال في الاسرة.

2 –احسب وسيط عدد الاطفال وفسر هذه القيمة.

3- احسب المنوال لعدد الاطفال في الاسرة وفسر هذه القيمة.

4 –اوجد الربيع الاول لعدد الاطفال في الاسرة وفسر هذه القيمة.

5 – اوجد الربيع الثالث لعدد الاطفال في الاسرة وفسر هذه القيمة.

6- اوجد المئيني 95 لعدد الاطفال في الاسرة وفسر هذه القيمة.

فيما يلي بيانات عن رحلة حافلة ركاب للنقل العام بين دمشق وحمص :

السرعة كم/سا	الرحلة
80	دمشق- حمص

100	حمص- دمشق

المطلوب :احسب متوسط السرعة في الرحلتين مستخدما المقياس المناسب .

الجدول التالي يعرض بيانات عن حجم الانتاج لاحد المعامل في الاعوام 2010-2016 (طن).

العام	2010	2011	2012	2013	2014	2015	المجموع
حجم الانتاج Y	200	210	226.8	249.28	311.6	342.76	1540.44

المطلوب:

احسب في كل سنة معدل النمو بالنسبة للعام السابق.

احسب في كل سنة مقدار التغير في حجم الانتاج بالنسبة للعام السابق.

احسب متوسط التغير السنوي .

احسب متوسط معدل النمو معتبرا النمو حسب قانون الفائدة المركبة.

احسب متوسط معدل النمو معتبرا النمو حسب قانون الفائدة البسيطة.

مساعدة :ان نسبة متوسط التغير السنوي بالنسبة للعام الاول ،هي متوسط معدل النمو باعتبار ان النمو يتم حسب قانون الفائدة البسيطة .ويمكن حساب هذا المعدل ايضا r من العلاقة $y_t = y_0(1 + nr)$

مراجع الفصل الثالث:

7- حميدان، عدنان؛ الجاعوني، فريد؛ ناصر آغا، عمار؛ العواد، منذر، 2003، مبادئ الإحصاء، منشورات جامعة دمشق.

8- **Groebner David F., Shannon Patrick W. and Fry Phillip C. (2017)," Business Statistics :A Decision-Making approach", 10 th Edition. person**.

9- **Weiss, N.A.(1999), Introductory Statistics. Addison Wesley**.

10- **Deborah Rumsey,(2010)" Statistics Essentials For Dummies" Wiley**.

الفصل الرابع

مقاييس التشتت

Measures Of Variation

الاهداف والمخرجات التعليمية :

بعد الانتهاء من هذا الفصل سيصبح الطالب قادرا على:

1- حساب مقاييس التشتت :المدى ،الانحراف الربيعي ،التباين ، الانحراف المعياري. كيف ومتى تستخدم.
2- مقارنة البيانات من حيث اختلافها او تجانسها .
3- وصف وتمثيل البيانات بشكل اكثر شمولا باستخدام مقاييس رقمية اكثر دقة .

ملخص الفصل

في هذا الفصل سنتعرف على صفة جديدة من صفات البيانات وهي اختلاف البيانات او تماثلها وهو ما ندعوه التشتت .سنعرض مقاييس التشتت

راينا في الفصل الثاني ان التوزيعات التكرارية للبيانات تختلف عن بعضها بالنزعة المركزية ، والتي تم مناقشتها والتعرف عليها في الفصل الثالث .و كذلك تختلف عن بعضها باختلاف البيانات او بتشتت البيانات . ما هو التشتت وكيف يتم حسابه وما هي استخداماته ،هذا ما سنتعرف عليه في هذا الفصل.

1.4 -: تعريف التشتت واستخداماته

يعرف تشتت البيانات بأنه اختلاف قيم مفردات هذه البيانات أو تباعدها أو انتشارها ، و يعبر عن ذلك بطريقتين :

- التشتت هو الاختلاف بين قيمة أكبر مفردة و أصغر مفردة و هذا يدعى المدى Range، أو الاختلاف بين الربيع الأدنى Q_1 و الأعلى Q_3 ويدعى المدى الربيعي Interquartile Range.
- يعبر عن التشتت في الطريقة الثانية ، بانه اختلاف جميع القيم عن أحد مقاييس النزعة المركزية الوسط الحسابي او الوسيط ،كما سترى .

لنرى الآن ماهي استخدامات التشتت و كيف يوظف التشتت في التحليل الاحصائي.

الغاية من مقياس النزعة المركزية هي الوصول الى قيمة تمثيلية للبيانات او معيار يساعد على مقارنة بيانات العينات أو المجتمعات و لكن هل يكفي مقياس النزعة المركزية لهذه الغاية ؟

سنجيب على هذا السؤال من خلال المثال التالي :

لنفرض أنه لدينا البيانات التالية عن اعمار لاعبي فريقين لكرة اليد :

اعمار الفريق الاول :30 ، 32 ، 28 ، 18 ، 17 سنة .

اعمار الفريق الثاني : 25 ، 25 ، 25 ، 25 ، 25 سنة .

ان الوسط الحسابي لاعمار الفريق الاول هو 25 سنة وكذلك لافراد الفريق الثانية، فهل نستطيع الحكم على اعمار الفريقين بانها متساوية او مختلفة من خلال الوسط الحسابي فقط ،الجواب لا ولكن لو علمنا ان الوسط الحسابي لاعمار الفريق الثاني 25 سنة واختلافها (تشتتها) يساوي الصفر، والوسط الحسابي لاعمار الفريق الاول يبلغ 25 سنة وتشتتها يبلغ قيمة معينة غير الصفر (سنرى كيفية حسابها لاحقا) فنستطيع عندها اعطاء وصفا افضل للبيانات وبالتالي مقارنة افضل . يتبين من هذا ان احد استخدامات التشتت هو استخدامه بجانب مقياس النزعة المركزية لوصف البيانات بشكل افضل . اذا كان التشتت معدوما كما في بيانات الفريق الثاني بالمثال اعلاه فهذا يعني ان البيانات متساوية ومقياس النزعة المركزية يمثلها افضل تمثيل وعلى العكس كلما زاد التشتت دل على اختلاف البيانات و على انخفاض القدرة التمثيلية لمقياس النزعة المركزية ، وهذا هو الاستخدام الثاني للتشتت :الحكم على القدرة التمثيلية لمقياس النزعة المركزية .

يدل اختلاف البيانات لمتغير ما في بعض الحالات على عدم استقرار هذا المتغير، مثل بيانات الاسعار أو بيانات عن مواصفات سلعة منتجة . . . فقيمة التشتت هنا تدل على الاستقرار أو عدم الاستقرار لهذا المتغير .

بالاضافة الى هذه الاستخدامات يستخدم التشتت بشكل واسع في اساليب التحليل الأخرى مثل الارتباط والانحدار و الاستدلال كما سنرى بعضها في فصول تالية من هذا المقرر.
تستخدام مقاييس التشتت بشكل مطلق و هي بهذه الحالة تقاس بنفس الوحدات المستخدمة في البيانات ، و تستخدم بهذا الشكل المطلق عند مقارنة بيانات لها نفس الوحدات و الفارق بين الأوساط الحسابية لكلا المجموعتين ليس كبيرا انظر المثال في الفقرة 2.4 ، أما في ما عدا ذلك فيجب استخدام مقاييس التشتت بشكل نسبي و ذلك باخذ التشتت المطلق كنسبة مئوية من مقاييس النزعة المركزية .
سندرس أدناه مقاييس التشتت التالية :

1. **المدى .**
2. **المدى الربيعي و الانحراف الربيعي .**
3. **الانحراف المتوسط .**
4. **التباين و الانحراف المعياري .**

المدى − 2.4 Range:

يعرف المدى بأنه الفرق بين أكبر قيمة في البيانات Max و أصغر قيمة Min .

$$R = Max - Min$$

يتضح من تعريف المدى سهولة فهم هذا المقياس و سهولة حسابه ، يستخدم المدى بشكل واسع في خرائط مراقبة جودة الانتاج ، ولكن ما يؤخذ على هذا المقياس هو تركيزه على القيم الأكثر تطرفا و اهماله تشتت القيم بينهما ، كما أنه لايمكن حسابه للجداول التكرارية المفتوحة ، حيث يحسب بأنه الفرق بين الحد الاعلى الفئة الأخيرة و الحد الأدنى للفئة الاولى .
يستخدم المدى بشكل نسبي اذا كانت وحدات البيانات مختلفة ، أو كان الفرق بين أوساطها كبيرا .
و يحسب المدى النسبي كما يلي :

$$PR = \frac{R}{\bar{X}} * 100$$

مثال:**فيمايلي اطوال واوزان عينة عشوائية مؤلفة من 5 اطفال .احسب المدى و المدى النسبي وقارن النتائج:**

الطفل	1	2	3	4	5	المجموع
الوزن K.g	3	7	9	11	15	45
الطول C.m	50	52	60	65	88	315

نوجد اولا المدى والوسط الحسابي والمدى النسبي للوزن:

$$R = Max - Min = 15 - 3 = 12K.g$$
$$\bar{X} = \frac{\sum x}{n} = \frac{45}{5} = 9K.g$$
$$PR = \frac{12}{9} * 100 = 133.33\%$$

نوجد المدى والوسط الحسابي والمدى النسبي للطول:

$$R = Max - Min = 88 - 50 = 38C.m$$
$$\bar{X} = \frac{\sum x}{n} = \frac{315}{5} = 65C.m$$
$$PR = \frac{38}{65} * 100 = 58.46\%$$

نلاحظ أن المدى للوزن هو 12 وهو اصغر من المدى للطول و البالغ 38 .أما اذا أخذنا المدى النسبي لأن البيانات من وحدات مختلفة اولا، كما ان الفرق كبير بين الأوساط الحسابية ثانيا ، نرى أن تشتت الوزن اكبر من تشتت الطول .

لاحظ لو تمت المقارنة على اساس المدى لتوصلنا الى نتيجة مختلفة ومضللة

3.4Quartile Range المدى الربيعي – الانحراف الربيعي Quartile Deviation:

يعرف المدى الربيعي بانه الفرق بين الربيع الاول و الربيع الثالث ، يرمز له بالرمز QR ويحسب من العلاقة $QR = Q_3 - Q_1$:

وباعتبار ان المدى الربيعي كبير يمكن اخذ نصف المدى الربيعي وندعوه الانحراف الربيعي ، نرمز له اختصارا QD ويحسب من العلاقة :

$$QD = \frac{QR}{2} = \frac{Q_3 - Q_1}{2}$$

و لأغراض المقارنة يمكن حساب الانحراف الربيعي النسبي و ذلك بنسب الانحراف الربيعي الى الوسيط :

$$PQD = \frac{QD}{Med} * 100$$
$$PQD = \frac{Q_3 - Q_1}{2Med} * 100$$

و اذا ما كان الوسيط مجهولا يمكن تقديره بالوسط الحسابي للربيع الأدنى و الأعلى مع العلم أن هذه العلاقة محققة اذا كان التوزيع طبيعياً $\{Med = \frac{Q_3 - Q_1}{2}\}$:

$$PQD = \frac{Q_3 - Q_1}{Q_3 + Q_1} * 100$$

يجب استخدام احدى العلاقتين لحساب الانحراف الربيعي النسبي لكلا البيانات الواجب مقارنتها .

مثال 1 **:احسب الانحراف الربيعي والانحراف الربيعي النسبي لعدد القطع التي انجزها العمال في اليوم لعينة من العمال حجمها 8 حيث كان انتاجها اليومي كما يلي :**

العامل	1	2	3	4	5	6	7	8
دد القطع	**25**	**28**	**30**	**32**	**34**	**35**	**41**	42

الحل:1- **نحسب الوسيط بايجاد موقعه اولا** $Loc(Med) = \frac{n+1}{2} = \frac{9}{2} = 4.5$ **وبالتالي نكون قيمة الوسيط هي الوسط الحسابي للمفردتين الرابعة والخامسة 33 .**

2-**نوجد الربيع الادنى والاعلى بتحديد موقعهما اولا ومن ثم ايجاد قيمتيهما :**

$Loc(Q1) = \frac{n+1}{4} = \frac{9}{4} = 2.25$ **وبالتالي تكون قيمة الربيع الاول عبارة عن الوسط الحسابي للمفردتين بالموقع الثاني والثالث أي:كيلو** $\frac{28+30}{2} = 29$.

وبالمثل تكون قيمة الربيع الاعلى هي الوسط الحسابي للمفردتين بالموقع السادس والسابع:

$$Loc(Q3) = \frac{3(n+1)}{4} = \frac{3*9}{4} = 6.75$$

$\frac{35+41}{2} = 38$ **قطعة** .

3-نحسب الانحراف الربيعي :قطعة $QD = \frac{Q3-Q1}{2} = \frac{38-29}{2} = 4.5$.

4- نحسب الانحراف الربيعي النسبي: $QD = \frac{Q3-Q1}{2*Med} * 100 = \frac{38-29}{2*33} * 100 = 13.63\%$

مثال 2 **:احسب الانحراف الربيعي والانحراف الربيعي النسبي لبيانات الجدول التالي(1.4) والذي يحوي بيانات عن المدة الزمنية التي استغرقها 100 عامل لانجاز عمل محدد :**

الزمن المستغرق (ساعة)	عدد العمال fi	التكرار التجميعي الصاعد f↑
1 -2	**10**	**10**
2 -3	25	35
3 -4	**35**	70
4-5	20	90
5 فما فوق	**10**	100
	100	

الجدول (1.4)

الحل : 1-لحساب الانحراف الربيعي يجب حساب الربيع الاول والثاني اولا :

-حساب الربيع الاول :نحدد فئة الربيع الادنى وذلك بتدقيق عمود التكرار التجميعي الصاعد فنجد ان اول قيمة اكبر او تساوي $\sum f/_4 = 100/_4$ هي مقابل الفئة الثانية وبالتالي الفئة الثانية هي فئة الربيع الاول، ومن ثم نطبق العلاقة:

$$Q_1 = L_{Q1} + \left[\frac{\sum f/_4 - CF_{q1-1} \uparrow}{f_{Q1}} \right] * C_{Q1}$$

$$Q_1 = 2 + \left[\frac{25-10}{25} \right] * 1 = 2.6$$

من ثم نحدد فئة الربيع الاعلى بنفس الطريقة ولكن يجب ان نقارن التكرار الصاعد بالقيمة $3\sum f/_4 =$ 75ونرى ان هذا يتحقق عند الفئة الرابعة، ومنه:

$$Q_3 = L_{Q3} + \left[\frac{3\sum f/_4 - CF^{\uparrow}_{q3-1}}{F_{Q3}} \right] * C_{Q3}$$

$$Q_3 = 4 + \left[\frac{300/_4 - 70}{20} \right] * 1 = 4.25$$

2- نحسب الانحراف الربيعي من العلاقة :

$$QD = \frac{Q_3 - Q_1}{2} = \frac{4.25 - 2.6}{2} = 0.825$$

3-من اجل حساب الانحراف الربيعي النسبي يجب حساب الوسيط ، موقع الوسيط في الفئة الثالثة لان التكرار التجميعي الصاعد المقابل لها اكبر من 50 = $\sum f/_2$:

$$Med = L_M + \left[\frac{\sum f/_2 - CF^{\uparrow}_{m-1}}{f_M} \right] * C_M$$

$$Med = 3 + \left[\frac{100/_2 - 35}{35} \right] * 1 = 3.43$$

$$PQD = \frac{Q_3 - Q_1}{2Med} * 100 = \frac{4.25 - 2.6}{2 * 3.43} * 100 = 24.05\%$$

مزايا ومساوئ الانحراف الربيعي:

1. **يمكن حسابه اذا كان الجدول التكراري مفتوحا .**
2. **تخلص من أحد عيوب المدى الذي يعتمد على القيم الأكثر تطرفا .**

3. لايزال هذا المقياس يهمل تشتت هذه القيم التي تم اسقاطها على طرفي التوزيع كما يهمل تشتت القيم الواقعة بين الربيع الأدنى و الأعلى.

4. لايتمتع بمزايا جبرية.

4.4Mean Deviation _الانحراف المتوسط :

يعرف الانحراف المتوسط بأنه الوسط الحسابي لانحراف قيم المفردات عن أحد مقاييس النزعة الوسط الحسابي أو الوسيط مع اهمال الاشارات الجبرية . و غالبا ماتؤخذ هذه الانحرافات عن الوسيط Med ،ويتم اهمال الاشارات الجبرية كي لاتلغي هذه الانحرافات بعضها بعضا .

- **الانحراف المتوسط للبيانات غير المبوبة :**

يحسب الانحراف المتوسط MD للبيانات غير المبوبة بالعلاقة التالية :

$$MD = \frac{\sum |di|}{n}$$

حيث: $di = |X_i - Med|$

ولأغراض المقارنة يتم استخدام الانحراف المتوسط النسبي :

مثال 1 : **احسب الانحراف المتوسط و الانحراف المتوسط النسبي للنفقات الشهرية للاسر التالية:**

الاسرة	1	2	3	4	5	6	7
نفقات (الف)	**40**	**38**	**30**	**50**	**70**	**60**	**40**

الحل : نرتب البيانات تصاعديا من أجل ايجاد قيمة الوسيط و من ثم نحسب الانحرافات و الانحراف المتوسط :

X_i	$PMD = \frac{MD}{Med} * 100$
30	10
38	2
40	0
40	0
50	10
60	20
70	30

المجموع	72

الوسيط هو المفردة الرابعة : Med = **40**

$$MD = \frac{\sum |di|}{n} = \frac{72}{7} = 10.28$$

أي أن النفقات الشهرية تختلف وسطيا بمقدار 10.28 ألف عن الوسيط .

$$PMD = \frac{MD}{Med} * 100 = \frac{10.28}{40} * 100 = 25.71\%$$

و يفسر على أن الانحراف المتوسط يشكل ما قيمته 25.71 % من قيمة الوسيط .ولكن لاغراض المقارنة بين تشتت البيانات مع بيانات اخرى يجب مقارنته مع الانحراف المتوسط النسبي للبيانات الاخرى .

مثال 2 :**احسب الانحراف المتوسط لبيانات الجدول التالي(2.4) و الذي يحوي بيانات عن الدخل الشهري ل 100 اسرة .**

فئات الدخل (الف)	f_i	x_i	$cf \uparrow$	$PMD = \frac{MD}{Med} * 100$	$\lvert d_i * f_i \rvert$
10-12	**5**	**11**	**5**	**5**	**25**
12-14	**20**	**13**	**25**	**3**	**60**
14-16	**25**	**15**	**50**	**1**	**25**
16-18	**40**	**17**	**90**	**1**	**40**
18-20	**10**	**19**	**100**	**3**	**30**
المجموع	100				**180**

جدول (2.4)

الحل : 1 _ نحسب الوسيط:

$$Med = L_M + \left[\frac{\sum f/_2 - CF^{\uparrow}_{m-1}}{f_M} \right] * C_M$$

$$Med = 14 + \left[\frac{100/_2 - 25}{25} \right] * 2 = 16$$ **الف**

2 _ نحسب أوساط الفئات في الجدول .

3_ نحسب انحراف أوساط الفئات عن الوسيط $di = \lvert X_i - Med \rvert$ **و من ثم نضرب بتكرار الفئة** d_iF_i .

4 _نحسب الانحراف المتوسط:الف $MD = \frac{\sum \lvert diF_i \rvert}{\sum F_i} = \frac{180}{100} = 1.8$

5 _ نحسب الانحراف المتوسط النسبي :

$$PMD = \frac{MD}{Med} * 100 = \frac{1.8}{16} * 100 = 11.25\%$$

– مزايا و مساوئ الانحراف المتوسط :

1 – سهل، معناه واضح اذ هو متوسط انحراف القيم عن مقياس النزعة المركزية .

2 _ هو مفهوم وسطي تؤخذ قيم كافة المفردات بعين الاعتبار عند حسابه .

3 _ لايمكن حسابه للجداول التكرارية المفتوحة .

4 _ ان اهمال الاشارات الجبرية عملية جبرية غير مرضية .

التباين **–5.4** Variation:

يعرف التباين بأنه متوسط مربعات انحرافات القيم عن الوسط الحسابي حصرا . ويتم تربيع الانحرافات للتخلص من القيم السالبة وبالتالي حتى لا يكون مجموع هذه الانحرافات معدوما $\sum(x_i - \bar{x}) = 0$ ، التباين بهذه العملية يتشابه مع الانحراف المتوسط ، حيث هنا نقوم بتربيع الانحرافات بدلا من اهمال الاشارات الجبرية . اذا نستطيع تعريف التباين بانه الوسط الحسابي لمربعات انحرافات القيم عن وسطها الحسابي ،و نرمز له بالرمز σ^2 للمجتمع الاحصائي و s^2 للعينة .و اعتمادا على هذا التعريف يمكن حساب التباين كما يلي.

_ التباين للبيانات غير المبوبة :

حسب التعريف يمكن حساب التباين للعينة و المجتمع من العلاقات التالية :

- **تباين بيانات المجتمع الاحصائي :**

$$\sigma^2 = \frac{\sum(x_i - \mu)^2}{N}$$

حيث :σ^2 تباين بيانات المجتمع

µ الوسط الحسابي للمجتمع

N حجم المجتمع

- **عند حساب تباين بيانات العينة نميز بين حالتين :**

اذا كان جم العينة كبيرا $n \geq 30$ فان التباين يعطى بالعلاقة التالية:

$$S^2 = \frac{\sum(x_i - \bar{x})^2}{n}$$

حيث :S^2 تباين بيانات العينة

$\bar{x}$ الوسط الحسابي للعينة

n حجم العينة

اما اذا كان جم العينة صغيرا $n < 30$ فان التباين يعطى بالعلاقة التالية:

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

حيث: S^2 تباين بيانات العينة

$\bar{x}$ الوسط الحسابي للعينة

n حجم العينة

مثال: **احسب التباين لبيانات عينة حجمها 5:**

Xi	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$
3	-5	25
7	-1	1
8	0	0
10	2	4
12	4	16
40	0	46

$\sum (x_i - \bar{x}) = 0$

الحل : 1 – نحسب الوسط الحسابي: $\bar{X} = \frac{\sum X_i}{n} = \frac{40}{5} = 8$

2 – نحسب انحراف القيم عن الوسط الحسابي : $(X_i - \bar{X})$

3 – نربع الانحرافات: $(X_i - \bar{X})^2$

4 – نحسب التباين من العلاقة:

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{40}{5-1} = 10$$

لاحظ اننا استخدمنا علاقة التباين للعينات الصغيرة n=5 .

–التباين للبيانات المبوبة :

يتم حساب التباين للبيانات المبوبة كما يلي :

- **تباين بيانات المجتمع الاحصائي :**

حيث: σ^2 **تباين بيانات المجتمع**

µ الوسط الحسابي للمجتمع

$\sum f$ حجم المجتمع

- عند حساب تباين بيانات العينة نميز بين حالتين :

اذا كان جم العينة كبيرا $n \geq 30$ فان التباين يعطى بالعلاقة التالية:

$$S^2 = \frac{\sum f_i (x_i - \bar{x})^2}{\sum f_i}$$

حيث: S^2 تباين بيانات العينة

$\bar{x}$ الوسط الحسابي للعينة

$\sum f$حجم العينة

اما اذا كان جم العينة صغيرا $n < 30$ فان التباين يعطى بالعلاقة التالية:

$$S^2 = \frac{\sum f_i (x_i - \bar{x})^2}{\sum f_i - 1}$$

حيث: S^2 تباين بيانات العينة

$\bar{x}$ الوسط الحسابي للعينة

$\sum f$ حجم العينة

مثال :أوجد التباين لبيانات الجدول(3.4) وهي بيانات عن النفقات اليومية لافراد عينة عشوائية مؤلفة من **100** سائح في أحد المنتجعات .

1 – نحسب متوسط النفقات اليومية للسائح: $\bar{X} = \frac{\sum F_i X_i}{\sum F_i} = \frac{1200}{100} = 12$.

2 – نحسب الانحرافات المربعة لأوساط الفئات عن الوسط الحسابي $(X_i - \bar{X})^2$.

3 _ نضرب مربع الانحرافات المربعة لكل فئة بالتكرار F_i.

ت اليومية (الف	د السواح Fi	Xi	FiXi	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$	$F_i(X_i - \bar{X})^2$
6-8	12	7	84	-5	25	300
8-10	17	9	153	-3	9	153
10-12	20	11	220	-1	1	20
12-14	29	13	377	1	1	29
14-16	10	15	150	3	9	90
16-18	6	17	102	5	25	150
18-20	6	19	114	7	49	294
المجموع	100		1200			**1036**

جدول (3.4)

4 _ نحسب التباين من علاقة التباين للعينات الكبيرة :

$$S^2 = \frac{\sum F_i (X_i - \bar{X})^2}{\sum F_i} = \frac{1036}{100} = 10.36$$

الانحراف المعياري-6.4Standard Deviation:

عند ايجاد التباين قمنا بتربيع الانحرافات حتى لانحصل على مجموع معدوم للانحرافات ، بالنتيجة نحصل على التباين وهو عبارة عن متوسط الانحرافات المربعة . بالعودة الى الجذر من المرتبة الاولى نكون قد تخلصنا من التربيع والناتج يدعى الانحراف المعياري ، يمكن تعريف الانحراف المعياري بشكل مبسط ، هو الجذرالتربيعي للتباين ويحسب كمايلي :

- **الانحراف المعياري للمجتمع بيانات غير مبوبة:**

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

- **الانحراف المعياري للمجتمع بيانات مبوبة:**

$$\sigma = \sqrt{\frac{\sum f_i (x_i - \mu)^2}{\sum f_i}}$$

- **الانحراف المعياري للعينة بيانات غير مبوبة:**

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$ **او للعينات الصغيرة** $$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

- **الانحراف المعياري للعينة بيانات مبوبة:**

$$S = \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{\sum f_i}}$$ **او للعينات الصغيرة** $$S = \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{\sum f_i - 1}}$$

مثال **:أوجد الانحراف المعياري لدرجات عينة من الطلاب حجم 6 ،كانت درجاتهم كما يلي :**

الطالب	1	2	3	4	5	6	المجموع
الدرجة **x**	**60**	**65**	**70**	**87**	**90**	**90**	**462**
	-17	-12	-7	10	13	13	**0**
$(x_i-\bar{x})^2$	**289**	**144**	**49**	**100**	**169**	**169**	**920**

1 – نحسب متوسط الدرجات : $\bar{X} = \frac{\sum X_i}{n} = \frac{462}{6} = 77$.

2 – نحسب الانحرافات المربعة لأوساط الفئات عن الوسط الحسابي $(X_i - \bar{X})^2$.

3- نحسب الانحراف المعياري من العلاقة للعينات الصغيرة :

$$S=\sqrt{\frac{\sum(x_i-\bar{x})^2}{n-1}}$$

$$S=\sqrt{\frac{920}{6-1}}=\sqrt{184}=13.56$$

مثال :**أوجد الانحراف المعياري لبيانات الجدول(4.4) و التي هي بيانات عن الاجور اليومية لافراد عينة عشوائية مؤلفة من 80 عامل في أحدى المدن.**

1 – نحسب متوسط الاجور اليومية للعامل: $\bar{x}=\frac{\sum f_i x_i}{\sum f_i}=\frac{1200}{80}=15$.

2 – نحسب الانحرافات المربعة لأوساط الفئات عن الوسط الحسابي $(x_i-\bar{x})^2$.

3 _ نضرب مربع الانحرافات المربعة لكل فئة بالتكرار F_i.

ر اليومية (الف	دد العمال Fi	Xi	FiXi	$(x_i-\bar{x})$	$(x_i-\bar{x})^2$	$f_i(x_i-\bar{x})^2$
6-10	13	8	104	-7	49	637
10-14	18	12	216	-3	9	162
14-18	28	16	448	1	1	28
18-22	18	20	360	4	16	288
22-26	3	24	72	8	64	192
المجموع	80		1200			**1307**

جدول (3.4)

4_ نحسب الانحراف المعياري من العلاقة للعينات الكبيرة :

$$S=\sqrt{\frac{\sum f_i(x_i-\bar{x})^2}{\sum f_i}}$$

$$S=\sqrt{\frac{1307}{80}}=\sqrt{16.3375}=4.042$$

7.4Coefficient Of Variation: _ معامل الاختلاف (الانحراف المعياري النسبي)

عند مقارنة تشتت البيانات لعينات مختلفة أو لمجتمعات مختلفة رأينا أن هناك حالات يجب استخدام مقياس التشتت النسبي . و بالمثل فان الانحراف المعياري يعطى بشكل نسبي بمقياس يدعى معامل الاختلاف ،ويحسب كما يلي :

حيث :

C.V **الانحراف المعياري النسبي او معامل الاختلاف.**

S **الانحراف المعياري.**

$\bar{x}$ **الوسط الحسابي.**

مثال:

فيما يلي بيانات عن اطوال و اوزان عينة من 6 اعضاء من احدى فرق الرقص الشعبي:

الراقص	1	2	3	4	5	6	المجموع
الطول C.m	**60**	**68**	**65**	**70**	**58**	**57**	**378**
الوزن K.g	**160**	**165**	**170**	**174**	**178**	**179**	**1026**
للطول $(X_i - \bar{X})^2$	**121**	**36**	**1**	**9**	**49**	**64**	**280**
للوزن $(X_i - \bar{X})^2$	**9**	**25**	**4**	**49**	**25**	**36**	**148**

المطلوب:

ايهما اكثر تشتتا بيانات الطول ام الوزن .او بطريقة اخرى باي الصفات هذه الفرقة اكثر تجانسا (تقاربا).

الحل :للاجابة على السؤال،باي الصيغتين اعلاه المطلوب مقارنة تشتت بيانات الطول مع تشتت بيانات الوزن ،هنا يجب استخدام مقياس تشتت نسبي وهو معامل الاختلاف لان البيانات من وحدات مختلفة.

- **نحسب الاوساط الحسابية اولا للطول وللوزن :**

$$\bar{X} = \frac{\sum X_i}{n} = \frac{1026}{6} = 171C.m$$

$$\bar{X} = \frac{\sum X_i}{n} = \frac{378}{6} = 63K.g$$

- **نحسب الانحرافات المربعة للقيم عن الوسط الحسابي (في جدول البيانات) .**
- **نحسب الانحراف المعياري للطول وللوزن من العلاقة للعينات الصغيرة :**

$$S=\sqrt{\frac{\sum\left(x_i-\bar{x}\right)^2}{n-1}}=\sqrt{\frac{280}{6-1}}=\sqrt{56}=7.48C.m$$

$$S=\sqrt{\frac{\sum\left(x_i-\bar{x}\right)^2}{n-1}}=\sqrt{\frac{148}{6-1}}=\sqrt{29.6}=5.44K.g$$

- **نحسب معامل الاختلاف للطول وللوزن :**

معامل الاختلاف للطول

$$C.V=\frac{S}{\bar{X}}*100=\frac{7.48}{171}*100=4.37\%$$

معامل الاختلاف للوزن

$$C.V=\frac{S}{\bar{X}}*100=\frac{5.44}{63}*100=8.63\%$$

بمقارنة معامل الاختلاف للطول والبالغ 4.37%،مع معامل الاختلاف للوزن والبالغ 8.63% . نرى ان معامل الاختلاف للوزن هو الاكبر و بالتالي بيانات الوزن اكثر تشتتا وهي اقل تجانسا (تقاربا).

- مزايا و مساوئ الانحراف المعياري :

1- هو مقياس محدد جبريا بدقة و يتمتع ببعض المزايا الجبرية تساعد على استخدامه في مقاييس أخرى .

2- اذا أضفنا أو طرحنا عدد ثابت c الى كافة القيم Xi فان قيمة الانحراف المعياري تبقى نفسها.

3- اذا ضربنا كافة القيم بعدد ثابت cفان التباين الناتج هو السابق مضروبا بـ C^2 و الانحراف المعياري هو السابق مضروبا ب c .

هذه المزايا للانحراف المعياري تجعله افضل مقاييس التشتت واكثرها استخداما بين مقاييس التشتت .

اما مساوءه:

1- لايمكن حسابه للجداول التكرارية المفتوحة .

2- هو مفهوم مجرد صعب الفهم .

3- لايستخدم الا للمتغيرات الكمية .

مسائل و تمارين الفصل الرابع

1 – تبين من عينة عشوائية حجمها 10 سواح تم سحبها بشكل عشوائي من أحد فنادق مدينة اللاذقية البيانات التالية عن عدد الايام التي قضاها كل منهم في المدينة :

2	4	5	5	7	8	8	9	10	10

المطلوب:

احسب المدى.

احسب المدى النسبي.

احسب المدى الربيعي.

احسب الانحراف الربيعي .

احسب الانحراف الربيعي النسبي.

احسب التباين.

احسب الانحراف المعياري.

احسب معامل الاختلاف .

2– بغية التنبؤ بنتيجة مباراة مقبلة بكرة السلة بين الفريقين **A ,B** تم اخذ اطوال اللاعبين في كلا الفريقين فكانت:

اطوال الفريق **A**	**186**	**190**	195	189	170	178	198
اطوال الفريق **B**	195	196	200	205	195	198	197

المطلوب :

احسب متوسط الطول لكلا الفريقين وفسر النتائج.

لاعبي أي من الفريقين اكثر تجانسا(تقاربا) من حيث الطول برر ذلك بالحسابات اللازمة.

3- في دراسة عن الوضع المعاشي للاسرة في احدى الدول تم جمع البيانات التالية عن النفقات الشهرية للاسرة وعدد افراد الاسرة :

عدد افراد الاسرة	عدد الاسر	النفقات الشهرية (وحدة نقدية)
2	10	**2-4**
3	25	**4-6**
4	30	**6-8**
5	25	**8-10**
6	10	**10-12**
	100	المجموع

مساعدة (لاحظ في الجدول متغيرين كميين احدهما مستمر والاخر منقطع)

المطلوب :

احسب متوسط النفقات الشهرية للاسرة .

احسب متوسط عدد افراد الاسرة .

احسب الانحراف المعياري للنفقات الشهرية للاسرة .

احسب الانحراف المعياري لعدد افراد الاسرة .

أي البيانات اكثر تستتاً ، بيانات الانفاق الشهري ام عدد افراد الاسرة .

4-لدينا البيانات التالية عن اطوال واوزان 6 مشتركات في احدى الفرق الشعبية :

المشتركة	1	2	3	4	5	6
الطول **C.m**	**170**	**175**	169	180	185	183
الوزن**K.g**	70	69	71	72	68	68

المطلوب:

احسب الانحراف المعياري للطول.

احسب الانحراف المعياري للوزن.

بين باي الصفات عناصر هذه الفرقة اكثر تجانسا.

5- من عينة عشوائية بحجم 120 شجرة زيتون متقاربة في مواصفاتها تم جمع البيانات التالية عن غلة الشجرة :

الغلة (كغ)	عدد الاشجار
30-40	**15**
40-50	**20**
50-60	**25**
60-70	**30**
70-80	**20**
80فما فوق	**10**
المجموع	120

المطلوب 1- احسب وسيط الغلة و فسره .

2 - ماهو مقياس التشتت المناسب لبيانات هذه العينة ثم احسبه.

4- تبين من عينة أخرى مأخوذة من محافظة أخرى أن وسيط الغلة للاشجار المماثلة هو 64 كغ بانحراف ربيعي نسبي 20% . فبيانات أي العينتين أكثر تجانسا.

مراجع الفصل الرابع:

11- حميدان، عدنان؛ الجاعوني، فريد؛ ناصر آغا، عمار؛ العواد، منذر، 2003، مبادئ الإحصاء، منشورات جامعة دمشق.

12- Groebner David F., Shannon Patrick W. and Fry Phillip C. (2017)," Business Statistics :A Decision-Making approach", 10 th Edition. person.

13- Weiss, N.A.(1999), Introductory Statistics. Addison Wesley.

14- Deborah Rumsey,(2010)" Statistics Essentials For Dummies" Wiley.

الفصل الخامس

التجربة والحادث

المخرجات والأهداف التعليمية:

بعد الانتهاء من هذا الفصل ،سيكون الطالب قادرا على:

1- التمييز بين التجربة ونتائج التجربة .
2- التمييز بين انواع الحوادث .
3- يستطيع الوصول الى اجتماع وتقاطع وفرق الحوادث.

يعتبر هذا الفصل مدخل لعلم الاحتمالات حيث تم العرض فيه لبعض المفاهيم الهامة في الاحتمالات كالتجربة الاحصائية وفضاء العيتة والحوادث. كما تم توضيح العلاقات بين الحوادث كالحوادث المستقلة والشرطية وكذلك الحوادث المتنافية وغير المتنافية . وتم عرض العمليات الأساسية على الحوادث كالتقاطع والاجتماع والفرق، وأخيرا تم توضيح مفهوم التجزئة لحادث ما.

1.5- فضاء العينة والحوادث :

1.1.5- التجربة الإحصائية:

يستخدم في الإحصاء كلمة تجربة عشوائية للدلالة على أية عملية نقوم بها لدراسة ظاهرة معينة، تقودنا إلى مجموعة من النتائج أو البيانات. فعند إلقاء قطعة نقدية في الهواء نجد أننا أمام نتيجتين منتظرتين «ظهور الصورة» أو «ظهور الكتابة» وعند فحص مصباح كهربائي من مجموعة مصابيح مخزنة نجد أن النتيجتين المنتظرتين «المصباح جيد» أو «المصباح تالف».

2.1.5- فضاء العينة:

يطلق على مجموعة كل النتائج الممكنة لتجربة عشوائية أسم فضاء العينة، ويُرمز لها بالرمزΩ. فمثلاً في عملية رمي قطعة نقود في الهواء نجد أن ناتج هذه التجربة هو إمّا ظهور الصورة (الشعار) H أو ظهور الكتابة T ومنه يكون فضاء العينة أو الفضاء الاحتمالي بالشكل:

$$\Omega = \{H,T\}$$

وعند إلقاء حجر نرد (زهر) وبفرض أننا نهتم بالرقم الظاهر على الوجه العلوي للحجر نجد أن الفضاء الاحتمالي هو:

$$\Omega =\{1,2,3,4,5,6\}$$

أمّا إذا كان اهتمامنا منصباً على صفة الرقم الظاهر «زوجي» أو «فردي» فإن الفضاء الاحتمالي عندئذٍ سيكون بالشكل:

Ω={زوجي ،فردي}

عند رمي قطعتي نقود دفعة واحدة (أو رمي قطعة نقود مرتين متتاليتين) نجد أن فضاء العينة يتكون من كافة الثنائيات التالية:

$$\Omega = \{HH,HT,TH,TT\}$$

أخيراً عند رمي حجري نرد متمايزين في آن واحد فإننا نجد أن فضاء العينة يتكون من كافة الثنائيات التالية:

$$\Omega = \{(1,1),(1,2),\ldots,(6,6)\}$$

والتي يكون عددها /36/ ثنائية.

3.1.5- الحوادث وأنواعها:

يُعرّف الحادث على أنه مجموعة جزئية من فضاء العينة Ω ، ويرمز للأحداث بأحرف لاتينية كبيرة مثل ...,A,B . كما يعرف الحادث البسيط (أو الابتدائي) على أنه مجموعة جزئية من فضاء العينة ذات عنصر وحيد.

نستنتج مما سبق أن فضاء العينة Ω هو حادث لأن: $\Omega \subseteq \Omega$

كذلك نجد أن المجموعة الخالية ϕ هي أيضاً حادث لأن: $\phi \subset \Omega$

كما ينتج أن الحادث البسيط هو الحادث الذي لا يمكن تجزئته الى حوادث ابسط منه والحادث المركب هو الحادث المكون من عدة حوادث بسيطة مثل الحادث A "ظهور عدد زوجي عند إلقاء حجر النرد" حيث يكون: $A = \{2,4,6\}$ فهو حادث مركب من الحوادث البسيطة: $\{2\},\{4\},\{6\}$.

عند إجراء تجربة عشوائية ما مرةً واحدة فقط سنحصل على نتيجة واحدة فإذا فرضنا أن هذه النتيجة هي x فإننا نقول عن الحادث E بأنه قد وقع إذا وفقط إذا انتمت x إلى E. كما نقول عن أي حادث آخر تنتمي إليه x أنه قد وقع أيضاً.

فمثلاً عند رمي حجر النرد، إذا كان العدد الظاهر على الوجه العلوي هو 4 فإننا نقول بأن الحادث: A = {2,4,6} قد وقع . كذلك الأمر بالنسبة للحادث{3,4} = B أو الحادث {4} = C أو الحادث {1,2,3,4,5,6} = $\boldsymbol{\Omega}$.

مما سبق نجد بأنه عند إجراء تجربة ما فإن الحادث $\boldsymbol{\Omega}$ سيقع بكل الاحوال لأنه أيا كانت النتيجة x فإنها ستنتمي له ويسمى بالحادث الأكيد .كذلك نجد بأنه أيا كانت النتيجة x فلن تنتمي للحادث $\boldsymbol{\Phi}$ الذي لايمكن أن يقع ، لذلك يسمى بالحادث المستحيل .

4.1.5- الحوادث المتنافية والحوادث المتتامة:

يُسمى الحادثان اللذان يستحيل وقوعهما معاً (في أن واحد) بالحادثين المتنافيين ، أي يكون الحادثان A و B متنافيين إذا كان: $A \cap B = \phi$

أمّا الحادثان المتتامان فهما اللذان يقع أحدهما إذا وفقط إذا لم يقع الآخر. أي يكون الحادثان A و B متتامان إذا كان: $A \cap B = \Phi \,, A \cup B = \Omega$

أي: $B = \Omega - A = \bar{A}$ حيث $\bar{A}$ الحادث المتمم للحادث الأصلي A .

فمثلاً في تجربة رمي حجر النرد يكون فضاء العينة $\boldsymbol{\Omega}$ بالشكل التالي:

$$\boldsymbol{\Omega} = \{1,2,3,4,5,6\}$$

فإذا كانت A و B و C الأحداث الإحصائية التالية:

A = {1,3,5}

B = {2,4,6}

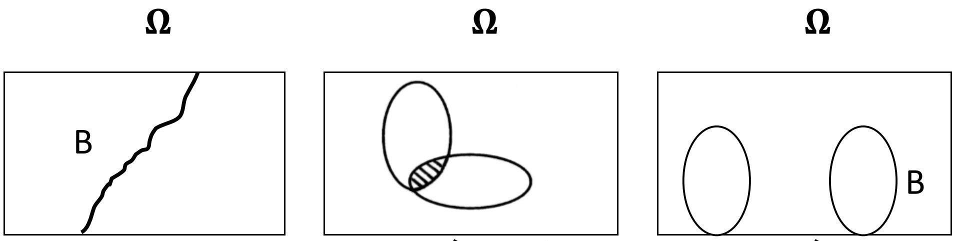
C = {2,4}

فعندئذٍ نلاحظ أن الحادثين A و C متنافيان لأن: $A \cap C = \Phi$ كما نلاحظ أن الحادثين A,B متتامان لأن:

$$A \cup B = \Omega \,, \mathrm{A} \cap \mathrm{B} = \Phi$$

ممّا سبق نستنتج أن الحادثين المتتامين هما حادثان متنافيان والعكس غير صحيح بالضرورة.

ملاحظة(1): يمكن توضيح مفهوم فضاء العينة والحوادث بمخططات فن (Venn) حيث يتم تمثيل فضاء العينة $\boldsymbol{\Omega}$ بمستطيل وتمثيل الحوادث بدوائر أو أشكال بيضوية أو مناطق جزئية واقعة داخل هذا المستطيل. الشكل التالي يبين بعض الحالات الممثلة لنوع الارتباط بين حادثين مفروضين A و B:



(متتامين) (غير متنافيين) (متنافيين)

الشكل (1-5): توضيح العلاقة بين حدثين احصائيين مفروضين

5.1.5- الحوادث المتكافئة والحوادث غير المتكافئة:

الحوادث المتكافئة هي الحوادث التي لها نفس الفرصة في الوقوع مثل الحادث "ظهور الصورة" والحادث "ظهور الكتابة" عند رمي قطعة نقود متوازنة. والحادثين "ظهور كرة حمراء" و "ظهور كرة سوداء" عند سحب كرة من صندوق يحوي كرات حمراء وسوداء بشكل متساوي حيث يكون الحادثان متكافئان أيضا . وبالعكس تكون الحوادث غير متكافئة اذا كانت فرصة وقوعها غير متساوية فعند سحب كرة بشكل عشوائي من صندوق يحوي 20 كرة حمراء و 5 كرات سوداء نجد أن الحادثين "ظهور كرة حمراء" و"ظهور كرة سوداء" هما حادثين غير متكافئين.

6.1.5- الحوادث المستقلة والحوادث الشرطية (غير المستقلة):

الحوادث المستقلة هي الحوادث التي يكون وقوع كل منها لا يؤثر ولا يتأثر بوقوع بقية الحوادث. فالحادث الدال على ظهور الصورة في الرمية الأولى لقطعة نقود مستقل عن الحادث الدال على ظهور الصورة في الرمية الثانية لنفس القطعة. و الحادث الدال على ظهور كرة حمراء عند سحب كرة من صندوق فيه كرات حمراء وبيضاء مع إعادة الكرة المسحوبة لايؤثر على الحوادث " ظهور كرة حمراء" أو " ظهور كرة بيضاء " عند السحب مرة ثانية.

أما الحوادث الشرطية (غير المستقلة) فهي الحوادث التي وقوعها أو عدم وقوعها يؤثر و يتأثر بوقوع أو عدم وقوع الحوادث الأخرى . فمثلا في تجربة سحب كرتين بشكل متتالي من صندوق يحوي كرات بيضاء وحمراء وبدون إعادة للكرة المسحوبة للصندوق، نجد أن الحادث " الكرة المسحوبة ثانيا حمراء " مرتبط (أو غير مستقل) بالحادث " الكرة المسحوبة أولا حمراء". وتسمى مثل هذه الحوادث بالحوادث الشرطية.

فمثلاً نجد أن هطول المطر مرتبط بالغيوم وحركة الغيوم ترتبط بالرياح لذا فهذه الأحداث الجوية الثلاثة غير مستقلة. وفي تجربة رمي قطعة نقود متزنة نجد أن ظهور الصورة ينفي ظهور الكتابة وبالتالي نلاحظ أن ظهور الكتابة متعلق بعدم ظهور الصورة لذا فالحدثان (ظهور الصورة وظهور الكتابة) غير مستقلين على الرغم من كونهما متنافيين.

أمّا ظهور الصورة H على قطعة نقود متزنة لا ينفي ولا يتعلق بظهور صورة أخرى H على قطعة نقود أخرى وعلى هذا الأساس نجد أن الحدثين ظهور H على قطعة نقود ما، وظهور H على قطعة أخرى هما حدثان

مستقلان. وبالمثل ظهور العدد 5 عند رمي حجر نرد وظهور الصورة H عند رمي قطعة نقود، هما حدثان مستقلّان.

وكذلك النجاح في مادة الإحصاء والنجاح في مادة الجغرافيا هما حدثان مستقلان. واصطدام سيارة في شارع ما من مدينة دمشق وأخرى في شارع ما من مدينة حمص، هما حدثان مستقلان. وهكذا...

2.5– العمليات على الحوادث :

1.2.5–اجتماع وتقاطع الحوادث – فرق حادثين:

إن اجتماع الحادثين A و B هو حادث يتحقق إذا وقع أحدهما على الأقل أي إذا وقع A أو B أو كليهما معا ، ونرمز له بالرمز $A \cup B$.

أما تقاطع الحادثتين A , B فهو الحادث الذي يتحقق عند وقوع كل من الحادثين A و B في أن واحد ونرمز له بالرمز $A \cap B$.

وبشكل عام فإن اجتماع عدة حوادث هو حادث يتحقق إذا وقع أحد هذه الحوادث على الأقل . و تقاطع عدة حوادث هو حادث يتحقق إذا وقعت كل الحوادث معا .

نتيجة:

تقاطع الحوادث المتنافية يساوي الحدث المستحيل Φ .

مثال (1):

اذا كان A هو حادث إصابة هدف بالطلقة الأولى و B هو حادث الإصابة للهدف نفسه بالطلقة الثانية فإن $A \cup B$ يكون الحادث الذي يقع إذا تمت الإصابة للهدف بأحد الطلقتين على الأقل، كما يكون $A \cap B$ الحادث الذي يقع إذا تمت إصابة الهدف بكلا الطلقتين.

فرق حادثين A , B هو حادث يتحقق إذا وقع أحدهما ولم يقع الآخر. فالحادث

A-B يقع إذا وقع A و لم يقع B. والحادث B-A فهو يقع إذا وقع B و لم يقع A .

مثال (2):

في المثال (1) السابق نجد أن:

A-B : هو الحادث الذي يتحقق إذا تمت إصابة الهدف بالطلقة الأولى ولم يصاب بالطلقة الثانية، أما:

B-A: فهو الحادث الذي يتحقق إذا أصيب الهدف بالطلقة الثانية ولم يصاب بالطلقة الأولى.

2.2.5- تجزيء حادث:

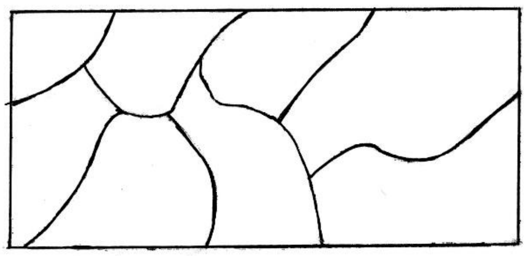
نقول عن الحوادث.... , $B_1 , B_2 , B_3 , . , . , . , . , B_n$ إنها تشكل تجزيئا للحادث Ω إذا كانت تحقق ما يلي :

1 - $B_i \cap B_j = \Phi \ , \ i \neq j$ **أي الحوادث متنافية مثنًا مثناً .**

2 - $\bigcup_{i\geq 1} B_i = \Omega$ **أي أن اجتماع الحوادث يشكل الحادث الأكيد.**

الشكل التالي يوضح تجزيئا لفضاء العينة Ω إلى ثمانية حوادث.

Ω



الشكل (5.1)

مثال(3):

أوجه حجر النرد تشكل تجزيئا لأن الحوادث متنافية مثنى من جهة واجتماعها الحادث الأكيد من جهة أخرى.

كما أن اختيار صندوق من بين أربع صناديق مختلفة تحوي كل منها منتج ما يشكل تجزيئا .

تمارين الفصل الخامس

1- ألقينا قطعة نقود، فإذا كان الناتج «صورة» أعدنا إلقاءها ثانية، أمّا إذا كان «كتابة» فإننا نلقي حجر النرد. أكتب فضاء العينة.

2- أربعة طلاب انتخبوا عشوائياً من صف ما، فإذا رمزنا بالرمز M للطالب و F للطالبة، أكتب فضاء العينة Ω_1 . ثم أكتب فضاء عينة آخر Ω_2 عناصره تمثل عدد الإناث المنتخبة.

3- إذا كان الحدثان A,B معرفان على فضاء عينة ما. فعبّر عن الرموز التالية بعبارات واضحة.

$$A', B',\ A \cup B,\ A \cap B,\ A' \cap B',\ A' \cup B',\ (A \cap B)',\ (A \cup B)'$$

4- إذا كانت الحوادث A,B,C معرَّفة على فضاء عينة ما. فعبّر عن الحوادث التالية رياضياً:

أ- وقوع الحادث A مع عدم وقوع كلاً من الحدثين C , B.

ب- وقوع الحادث A مع عدم وقوع أحد الحدثين C , B.

جـ - وقوع الحوادث الثلاثة معاً.

د - عدم وقوع أي من الحوادث الثلاثة.

هـ - وقوع أحد الحوادث الثلاثة على الأقل.

5- استخدم مخططات فن لإثبات العلاقات التالية بين الحوادث المذكورة:

أ- $A = A \cup (A \cap B)$

ب- $A = (A \cap B) \cup (A \cap B')$

جـ - $(A \cap B)' = A' \cup B'$

د - $(A \cup B)' = A' \cap B'$

هـ - $(A \cup B) = (A \cap B) \cup (A \cap B') \cup (A' \cap B)$

و - $A\cap(B\cup C)=(A\cap B)\cup(A\cap C)$

6 - **ألقي حجري نرد، الأول أخضر والثاني أبيض، وسجلنا النتائج والمطلوب:**

أ - أكتب فضاء العينة.

ب- أكتب عناصر الحادث A الذي يقع إذا كان المجموع للرقمين الظاهرين على الحجرين أقل من 6 .

ج- أكتب عناص الحادث B الذي يقع إذا كان مجموع الرقمين الظاهرين أكبر من 7.

د- أكتب عناص الحادث C الذي يقع إذا كان مجموع الرقمين الظاهرين يقبل القسمة على 3.

هـ- أوجد ما يلي: $\bar{A}$, $C-A$, $A-C$, $A\cup C$, $A\cap C$, $A\cap B$

مراجع الفصل الخامس :

1- حميدان، عدنان؛ الجاعوني، فريد؛ ناصر آغا، عمار؛ العواد، منذر، 2003، مبادئ الإحصاء، منشورات جامعة دمشق.

1) حسام ،كمرجي،2011 - الاحتمال والاحصاء - جامعة دمشق – كلية الهندسة الميكانيكية والكهربائية .

2) Groebner David F., Shannon Patrick W. and Fry Phillip C. (2017)," Business Statistics :A Decision-Making approach", 10 th Edition. person.

الفصل السادس

الاحتمالات و العمليات عليها

المخرجات والأهداف التعليمية:

بعد الانتهاء من هذا الفصل ،سيكون الطالب قادرا على:

1- التمييز بين انواع الاحتمالات .
2- حساب الاحتمالات للحوادث البسيطة والمركبة .
3- على جمع وضرب الاحتمالات.
4- الوصول الى الاحتمال الكلي .
5- حساب الاحتمال الشرطي.

يتضمن هذا الفصل توضيح لمفهوم الاحتمال بكافة حالاته حيث تم العرض لمفهوم الاحتمال التقليدي والنسبي والهنسي والرياضي كما تم توضيح احتمال الاجتماع للحادثين المتنافيين وغير المتنافيين وكذلك احتمال التقاطع للحادثين المستقلين والغير مستقلين وذلك بعد تعريف الاحتمال الشرطي. وأخيرا تم العرض لمفهومي الاحتمال الكلي واحتمال السبب.

1.6 – تعريف الاحتمال:

رأينا أن الاحتمال مرتبط بوقوع حادث ما و قد يتحقق هذا الحادث و قد لا يتحقق و قد يكون الحادث ذو فرصة كبيرة للوقوع و قد تكون فرصته قليلة أو ضعيفة فكيف نقيس هذه الفرصة التي يملكها حادث ما للوقوع ؟. يمكن معرفة ذلك من خلال حساب احتمال الحادث، فالاحتمال هو قياس حظ أو فرصة حادث ما للوقوع و سوف نرمز له بالرمز P ، فاحتمال وقوع الحادث A هو P(A).

1.1.6 – الاحتمال النظري (التقليدي):

إن تسمية الاحتمال النظري أو أحيانا الاحتمال القبلي آتية من أننا نستطيع حساب احتمال تحقق حادث ما بشكل نظري قبل القيام بالتجربة، وحسب هذا التعريف فإن احتمال وقوع حادث ما A يعطى بالعلاقة :

$$P(A)=\frac{\text{عدد المرات المواتية لوقوع } A}{\text{عدد المرات الممكنة}}$$

مثال(1):

إن احتمال ظهور الصورة عند رمي قطعة النقود المتوازنة هو $\frac{1}{2}$ لأن لقطعة النقود وجهان فقط الصورة والكتابة ومنه تكون عدد الحالات المواتية لظهور الصورة هي واحد وعدد المرات الممكنة (الكلية) هي أثنان ومنه فإن احتمال ظهور الصورة يكون:

$$P(H)=\frac{1}{2}$$

مثال(**2**):

صندوق يحوي 12 كرة منها 8 زرقاء اللون و **4** بيضاء اللون. تم سحب كرة عشوائيا من الصندوق، ما هو احتمال كون الكرة المسحوبة زرقاء اللون.

الحل:

نستطيع حساب الاحتمال قبل سحب الكرة أو حتى دون وجود الصندوق حيث نلاحظ أن عدد النتائج الممكنة هو12 (مجموع الكرات بالصندوق) وعدد المرات المواتية لوقوع الحادث (ظهور كرة زرقاء) هو 8 ، فإذا رمزنا لحادث ظهور كرة زرقاء بالرمز A فيكون:

$$P(A) = \frac{8}{12} = \frac{2}{3} = 0.667$$

مثال(3):

أراد شخص استخدام عدد من أربع خانات لتشكيل عددٍ سريٍّ للدخول إلى حاسبه الشخصي ما احتمال أن يحتوي هذا العدد على ثلاثة أرقام زوجية ورقم فردي علماً بأن للأرقام نفس الفرصة في الاختيار.

الحل:

إذا رمزنا بالرمز E للرقم الزوجي والرمز O للرقم الفردي فإن فضاء العينة يكون:

$\mathbf{\Omega}$ ={EEEE, EEEO, EEOE, EEOO, EOEE, EOEO, EOOE, EOOO,

OEEE, OEEO, OEOE, OEOO, OOEE, OOEO, OOOE, OOOO}

حيث يكون الحدث المطلوب هو:

A = {EEEO, EEOE, EOEE, OEEE}

وبما أن للأرقام نفس الفرصة في الاختيار فإن:

$$P(A) = \frac{4}{16} = \frac{1}{4}$$

مما سبق يمكننا ملاحظة الخصائص التالية للاحتمال النظري:

1 – لا يمكن تطبيق هذا التعريف لحساب الاحتمال إلا إذا كان معلوما عدد المرات المواتية لوقوع الحادث وعدد النتائج الممكنة للتجربة حيث أنه هناك حالات يكون فيها فضاء العينة غير محدود وبالتالي يتعذر حساب عدد المرات الممكنة.

2– حتى يطبق هذا التعريف لحساب الاحتمال يجب أن تكون الحوادث البسيطة للتجربة حوادث متكافئة في الوقوع و إلا تطبيقه سيعطي نتائج خاطئة. فمثلا لايمكن تطبيق تعريف الاحتمال النظري في تجربة رمي قطعة نقد مغشوشة (مثقلة من أحد وجهيها).

3– قيمة هذا الاحتمال لا تقل عن الصفر ولا تزيد عن الواحد: $0 \leq P(A) \leq 1$. وذلك لأن عدد المرات المواتية لا يمكن أن يقل عن الصفر (عندما يكون الحادث مستحيل P(Φ) = 0) ولا يمكن أن يزيد هذا العدد عن النتائج الممكنة للفضاء Ω (حيث: P(Ω)=1).

2.1.6– التردد النسبي (الاحتمال التجريبي):

دعي هذا الاحتمال بهذا الاسم لأنه يتم حسابه بعد القيام بالتجربة ، ويتم حسابه بعد التجربة حسب عدد مرات وقوع الحادث اذا تكررت التجربة N مرة وذلك وفق العلاقة التالية:

$$P(A)= \frac{\text{عدد مرات وقوع الحادث } A}{\text{عدد مرات التجربة } N}$$

مثال(**4**):

ما هو احتمال ظهور الكتابة عند رمي قطعة نقود مغشوشة اذا علمت ان الكتابة ظهرت 20 مرة عند رمي هذه القطعة 100 مرة.

الحل:

بما ان هذه القطعة مغشوشة وبالتالي الحوادث البسيطة لهذه التجربة غير متكافئة فيجب استخدام الاحتمال التجريبي ، فإذا رمزنا لظهور الكتابة بالرمز A فيكون الاحتمال المطلوب هو:

$$P(A) = \frac{20}{100} = 0.2$$

مثال(5):

تبين من عينة عشوائية حجمها 800 شخص ممن تناولوا دواء" معينا أن هناك 16 شخصا أصيبوا بأعراض جانبية من تناول هذا الدواء. والمطلوب ما هو احتمال إصابة شخص ما بأعراض جانبية من جراء تناول هذا الدواء.

الحل:

$$P(A) = \frac{16}{800} = 0.02$$

مما سبق يمكننا ملاحظة الخصائص التالية للاحتمال التجريبي:

1 - يستخدم الاحتمال التجريبي إذا كان عدد المرات المواتية لوقوع الحادث و عدد النتائج الممكنة مقادير مجهولة .

2- يتم استخدامه إذا كانت الحوادث البسيطة للتجربة غير متكافئة .

3- قيمته تقع بين الصفر و الواحد، أي أن: $0 \leq P(A) \leq 1$.

4- ينتهي الاحتمال التجريبي إلى الاحتمال النظري عندما يكون عدد مرات التجربة كبيرا .

3.1.6 – الاحتمال الهندسي:

يستخدم هذا الاحتمال بطريقة مشابهة لاستخدام الاحتمال النظري ولكن باستخدام وحدات قياس المساحة.

مثال(6):

ما هو احتمال وقوع سهم في نصف الدائرة العلوي إذا كان لابد وأن يصيب الدائرة.

الحل:

حسب تعريف الاحتمال النظري فهو يساوي مساحة النصف العلوي مقسوما على مساحة الدائرة ويساوي $\frac{1}{2}$.

مثال(7):

ما هو احتمال سقوط الشهب على اليابسة.

الحل:

هذا الاحتمال يساوي مساحة اليابسة (المساحة المواتية) مقسوما على مساحة الكرة الأرضية (المساحة الممكنة) ويساوي 0.29.

4.1.6- التعريف الرياضي للاحتمال:

الفضاء الاحتمالي ثلاثية $(\Omega, \rho(\Omega), P)$ فيها Ω مجموعة غير خالية و $\rho(\Omega)$ مجموعة أجزاء Ω (مجموعة كل المجموعات الجزئية من Ω) و P دالة لـ $\rho(\Omega)$ في R (مجموعة الأعداد الحقيقية) تحقق الخصائص التالية:

1- $0 \leq P(A) \leq 1$ **وذلك مهما يكن A من $\rho(\Omega)$.**

2- $P(\Omega) = 1$

3- **إذا كانت $A_1, A_2, ..., A_n, ...$ متتالية منتهية أو غير منتهية من عناصر $\rho(\Omega)$ المتنافية مثنى مثنى كان:**

$$P\left(\bigcup_{i\geq1} A_i\right) = \sum_{i\geq1} P(A_i)$$

وعندئذٍ نقول عن الدالة P أنها دالة احتمالية وعن P(A) ، قيمة الدالة P عند A ، أنها احتمال وقوع الحادث A.

نتيجة(1):

1- اعتماداً على التعريف السابق للدالة الاحتمالية نستنتج أنه إذا كانت $A_1, A_2, ..., A_n$ مجموعة منتهية من الحوادث المتنافية مثنى مثنى فإن:

$$P(A_1 \cup A_2 \cup ... \cup A_n) = P(A_1) + P(A_2) + ... + P(A_n)$$

2- إذا كان فضاء التجربة S يحتوي على n نتيجة $x_1, x_2, ..., x_n$ وبما أن الحوادث الابتدائية

$\{x_1\},\{x_2\}...\{x_n\}$ تشكل جماعة من الحوادث المتنافية مثنى فإن:

$$P(S)=P(\bigcup_{i=1}^{n}\{x_i\}=\sum_{i=1}^{n}P(\{x_i\})$$

فإذا رمزنا للاحتمال $P(\{x_i\})$ بالرمز $P(x_i)$ اختصاراً، من أجل الاحتمال لحدث ابتدائي وبملاحظة أن P(S)=1 نجد أن: $1=\sum_{i=1}^{n}P(x_i)$

حيث نستنتج في الحالة التي تكون فيها جميع الحوادث الابتدائية $\{x_i\}$ لها نفس الفرصة في الظهور (متساوية الاحتمال)، أن:

$$1=nP(x_i)\Rightarrow P(x_i)=\frac{1}{n}$$

وبالتالي فإن احتمال أي حدث A عدد عناصره k في هذه الحالة هو $P(A)=\frac{k}{n}$

وهذا ما وجدناه من خلال التعريف التقليدي للاحتمال مع التذكير بأن ما ذكر أعلاه يبقى مشروطاً بالحالة التي تكون فيها جميع نتائج فضاء التجربة متساوية الفرصة في الظهور.

مثال(8):

عند إلقاء قطعة نقود متوازنة مرتين متتاليتين. أوجد احتمال ظهور صورة واحدة على الأقل.

الحل:

بما أن قطعة النقود متوازنة فإن «ظهور الكتابة» و «ظهور الصورة» متساوياً الفرص في الظهور. إن فضاء العينة هو:

$$\Omega = \{HH, HT, TH, TT\}$$

حيث تكون جميع نتائج Ωمتساوية الاحتمال. إن الحادث المطلوب في هذا المثال هو:

$$A = \{HT, TH, HH\}$$

وبالتالي فإن :

$$P(A)=\frac{3}{4}$$

مثال(9):

ألقي حجر نرد بحيث أن فرصة ظهور العدد الزوجي تساوي ضعف فرصة ظهور العدد الفردي. ما احتمال الحادث A الذي يقع إذا وفقط إذا كان الناتج أكبر من 3. وما هو احتمال الحادث B الذي يقع إذا وفقط إذا كان الناتج يقبل القسمة على 3. ثم ما هو احتمال الحادث C الذي يقع إذا وفقط إذا كان الناتج عدداً زوجياً؟

الحل:

إن فضاء العينة Ω هو:

$$\Omega = \{1,2,3,4,5,6\}$$

فإذا فرضنا أن احتمال ظهور العدد الفردي هو x فإن احتمال ظهور العدد الزوجي يكون 2x وبالتالي فإن:

$$P(\Omega) = P(1) + P(2) + P(3) + P(4) + P(5) + P(6)$$

$$1 = x + 2x + x + 2x + x + 2x$$

$$1 = 9x \rightarrow x = \frac{1}{9}$$

وبالعودة لنص المثال نلاحظ أنه لدينا الحوادث التالية:

$$A = \{4,5,6\}$$

$$B = \{3,6\}$$

$$C = \{2,4,6\}$$

حيث نجد أن:

$$P(A) = P(4) + P(5) + P(6) = \frac{2}{9}+\frac{1}{9}+\frac{2}{9}=\frac{5}{9}$$

$$P(B) = P(3) + P(6) = \frac{1}{9}+\frac{2}{9}=\frac{3}{9}=\frac{1}{3}$$

$$P(C) = P(2) + P(4) + P(6) = \frac{2}{9}+\frac{2}{9}+\frac{2}{9}=\frac{6}{9}=\frac{2}{3}$$

2.6 – العمليات على الاحتمالات:

إن حساب الاحتمال بالعلاقات التي تعرفنا عليها أعلاه يتم للحوادث البسيطة بأبسط أشكالها مثل نتيجة رمي قطعة نقد ، أو الوجه الذي يظهر عند رمي حجر النرد ، أو لون الكرة الذي سيظهر عند سحب كرة واحدة من صندوق ... ولكن إذا كانت الحوادث مركبة فلحساب احتمال حادث مركب لابد من القيام بعمليات جبرية لمعرفة احتمال وقوع هذا الحادث المركب و هذا ما سنتعرف عليه هنا.

1.2.6– جمع الاحتمالات أو احتمال الاجتماع :

1.1.2.6– في حالة الحوادث المتنافية :

إذا كان عدد المرات المواتية لوقوع الحادث A هو r مرة و عدد المرات المواتية لوقوع الحادث B هو m مرة و عدد المرات الممكنة لكل منهما هو n فإن عدد المرات المواتية لوقوع الحادث A أو B هو (r+m) مرة و بالتالي فإن :

$$P(A\cup B)=\frac{r+m}{n}=\frac{r}{n}+\frac{m}{n}=P(A)+P(B)$$

ومنه نحصل على العلاقة: $P(A\cup B)=P(A)+P(B)$

مثال **(10):**

صندوق يحوي 6 كرات حمراء و 4 كرات بيضاء و 8 كرات زرقاء. سحبنا كرة بشكل عشوائي، ما هو احتمال أن تكون الكرة المسحوبة حمراء أو بيضاء.

الحل:

اذا رمزنا لحادث ظهور كرة حمراء بالرمز A و كرة بيضاء بالرمز B فإن:

$$P(A\cup B)=P(A)+P(B)$$

لأن الحدثان متنافيان ومنه فإن:

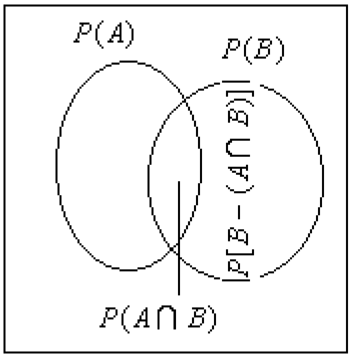
$$P(A\cup B)=\frac{6}{18}+\frac{4}{18}=\frac{10}{18}=\frac{5}{9}=0.556$$

2.1.2.6– في حالة الحوادث غير المتنافية:

بهذه الحالة يمكن للحوادث أن تقع مع بعضها و بالتالي فان عدد المرات المواتية لوقوع الحادث ستزداد بشكل مزدوج عدد مرات وقوع الاول و عدد مرات وقوع الثاني بالإضافة إلى أنه إذا وقعا معا فإن ذلك يعتبر تحقق واحد للحادث و ليس تحققين و بالتالي يجب أن نخفض المجموع بمقدار عدد مرات التي يقع بها الحادثين معا ومنه:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

و يمكن توضيح ماسبق من خلال الشكل (5.1) التالي:



الشكل (6.1)

حيث نجد أن:

$$P(A \cup B) = P(A) + P[B - (A \cap B)]$$

ومنه:$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

مثال **(11):**

إذا علمت أن عدد طلاب المستوى الرابع 1000 طالب بينهم 300 طالبة وأن عدد طلاب شعبة المحاسبة 800 طالب بينهم 200 طالبة ، أما عدد طلاب شعبة الإدارة فهو 200 . قمنا باختيار طالب بشكل عشوائي من المستوى الرابع ، والمطلوب: ما هو احتمال أن يكون من شعبة المحاسبة (طالب او طالبة) أو يكون من الطالبات (شعبة محاسبة او إدارة).

الحل:

إذا رمزنا لاختيار طالب محاسبة (طالب أو طالبة) بالرمز A و لاختيار طالبة (محاسبة أو إدارة) بالرمز B فإن:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B) = \frac{800}{1000} + \frac{300}{1000} - \frac{200}{1000} = \frac{900}{1000} = 0.9$$

مثال **(12):**

صندوق يحوي 8 كرات خضراء مرقمة من 1 الى 8 و 12 كرات بيضاء مرقمة من 1 الى 12. سحبنا عشوائيا كرة من الصندوق، ما هو احتمال أنها بيضاء اللون أو تحمل رقم زوجي.

الحل:

نرمز لظهور كرة تحمل رقم زوجي بالرمز A و لظهور كرة بيضاء بالرمز B ومنه فإن:

$$P(A) = \frac{10}{20} \quad , \quad P(B) = \frac{12}{20} \quad , \quad P(A \cap B) = \frac{6}{20}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B) = \frac{10}{20} + \frac{12}{20} - \frac{6}{20} = \frac{16}{20} = \frac{4}{5}$$

نتيجة (2):

اجتماع الحادث و متممه يعطي Ω ومنه فإن:

$$P(\,\Omega\,) = P(A \cup \bar{A}) = P(A) + P(\bar{A})$$

لأن الحادث ومتممه متنافيان وبالتالي فإن:

$$P(A) + P(\bar{A}) = 1$$

أو: $P(\bar{A}) = 1 - \ P(A)$

نتيجة (3):

احتمال اجتماع الحوادث المشكلة تجزيئا للحدث الأكيد Ω هو الواحد (لأنها متنافية مثنى) أي أن:

$$\bigcup_{i=1}^{n} A_i = \Omega$$

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) = 1$$

نتيجة (4):

يمكن تعميم علاقة احتمال الاجتماع لأكثر من حادثين كما يلي :

- اذا كانت الحوادث متنافية مثنى فإن:

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) = P(A_1) + P(A_2) + ...$$

- اذا كانت الحوادث غير متنافية مثنى ولنأخذ حالة ثلاث حوادث، ومنه فإن:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) \ - P(A \cap C) \ - P(B \cap C) + P(A \cap B \cap C)$$

مثال(13):

سحبنا ورقة من ورق اللعب المؤلف من 52 ورقة موزعة على أربع أنواع (كبة – بستوني – ديناري – سباتي) ما هو احتمال أن تكون الورقة المسحوبة من نوع غير الديناري.

الحل:

هنا الحوادث متنافية والحادث المطلوب يقع إذا كانت الورقة المسحوبة بستوني B أو كوبا K أو سباتي S ومنه فإن:

$$P(B \cup K \cup S) = P(B) + P(K) + P(S)$$

$$P(B \cup K \cup S) = \frac{13}{52} + \frac{13}{52} + \frac{13}{52} = \frac{3}{4}$$

ويمكن حساب الاحتمال السابق أيضا كما يلي:

$$1 - P(D) = 1 - \frac{13}{52} = \frac{3}{4}$$

حيث D هو حادث اختيار ورقة الديناري .

مثال **(14):**

من اجل معرفة نسبة الذين يتكلمون لغات أجنبية في إحدى المدن العربية تم سحب عينة بحجم 350 شخص فأعطت النتائج التالية: (سنرمز للإنكليزية E والفرنسية F والألمانية G):

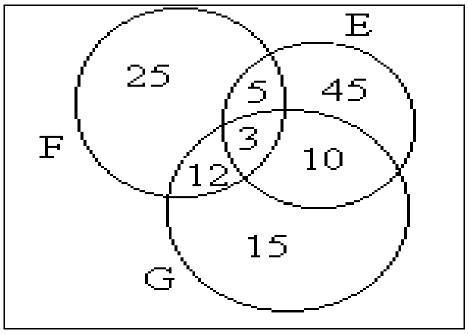
E فقط	F فقط	G فقط	E+G	E+F	F+G	E+F+G	فقط عربية	المجموع
63	45	40	13	8	15	3	202	350

اخترنا شخص من هذه العينة بشكل عشوائي و المطلوب :

1- احسب احتمال أنه يتكلم الإنكليزية، 2-احسب احتمال أنه يتكلم الإنكليزية أو الفرنسية، 3- احسب احتمال أنه يتكلم لغة أجنبية واحدة على الأقل.

الحل:

1- لتسهيل الحل سنستعين بالشكل التالي(6.2) لتوضيح البيانات:



الشكل (6.2)

ومنه فإن عدد الذين يتكلمون الإنكليزية هو: 45+10+5+3=63 ، وبالتالي فإن:

$$P(E) = \frac{63}{350}$$

1− عدد الذين يتكلمون الفرنسية هو: 25+5+12+3=45 ومنه:

$$P(F) = \frac{45}{350}$$

وبالتالي فإن احتمال أن يتكلم الشخص المختار الإنكليزية أو الفرنسية هو:

$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

$$P(E \cup F) = \frac{63}{350} + \frac{45}{350} - \frac{8}{350} = \frac{100}{350} = \frac{2}{7}$$

3−

$$P(E \cup F \cup G) = P(E) + P(F) + P(G) - P(E \cap F) - P(E \cap G) - P(F \cap G) + P(E \cap F \cap G)$$

$$P(E \cup F \cup G) = \frac{63}{350} + \frac{45}{350} + \frac{40}{350} - \frac{8}{350} - \frac{13}{350} - \frac{15}{350} + \frac{3}{350} = \frac{115}{350}$$

2.2.6 – ضرب الاحتمالات أو احتمال التقاطع :

قبل أن نستعرض قوانين ضرب الاحتمالات نعرف الاحتمال الشرطي، بأنه الاحتمال الذي يتم حسابه لحادث ما بعد وقوع حادث آخر قبله و نرمز له بالرمز $P(A/B)$ و يقرأ احتمال الحادث A علماً بأن الحادث B قد وقع .

مثال (15):

صندوق يحوي 3 كرات حمراء اللون و6 كرات بيضاء اللون، فإذا تم سحب كرتين على التتالي بدون إعادة فإن احتمال أن تكون الكرة المسحوبة أولا" حمراء اللون هو:

$$P(A) = \frac{3}{9} = \frac{1}{3}$$

و احتمال أن تكون الكرة المسحوبة ثانيا" حمراء اللون علما بأن الاولى كانت حمراء هو:

$$P(B/A) = \frac{2}{8} = \frac{1}{4}$$

و احتمال أن تكون الكرة المسحوبة في المرة الثانية حمراء اللون علما بأن الكرة المسحوبة في المرة الاولى بيضاء اللون هو:

$$P(B/\bar{A}) = \frac{3}{8}$$

نتيجة**(5)**:

من التعريف السابق للاحتمال الشرطي نستنتج أن:

1– $P(A/B) \neq P(B/A)$

2– إذا كان الحادثان A,B مستقلين فإن:

$$P(B/A) = P(B), \& P(A/B) = P(A)$$

1.2.2.6− احتمال التقاطع للحوادث الشرطية:

ليكن لدينا الحادثان B , A وبفرض أن عدد المرات المواتية لوقوع الحادث A هو m مرة و عدد المرات المواتية لوقوع الحادث B هو l مرة و عدد المرات المواتية

لوقوعهما معا هو k مرة وعدد المرات الممكنة لكل منهما هو n. ومنه فان : $P(A) = \frac{m}{n}$, $P(B) = \frac{l}{n}$, $P(A \cap B) = \frac{k}{n}$

إذا علمنا أن A قد وقع فان عدد المرات الممكنة لوقوع B حتى يقعا معاهوm مرة ومنه: $P(B \cap A) = \frac{k}{m}$

وبالتالي فإن:

$$P(A \cap B) = P(A)P(B/A)$$

$$P(A \cap B) = \frac{m}{n} . \frac{k}{m} = \frac{k}{n}$$

ومنه فإن احتمال التقاطع لحوادث شرطية يساوي احتمال الأول مضروبا بالاحتمال الشرطي للثاني.

مثال **(16):**

صندوقان متماثلان يحوي الأول 8 كرات مرقمة من 1 الى 8 و الثاني 11 كرة مرقمة من 1 حتى 11. اخترنا صندوق بشكل عشوائي و سحبنا منه كرة عشوائيا. ما هو احتمال أن تكون الكرة المسحوبة من الصندوق الأول و تحمل رقم فردي.

الحل:

بفرض A الحادث الدال عل السحب من الصندوق الأول و B الحادث الدال على كون الكرة المسحوبة تحمل رقم فردي، ومنه فإن الحادث المطلوب هو $A \cap B$

حيث نجد أن:

$$P(A) = \frac{1}{2} \quad , P(B/A) = \frac{4}{8} = \frac{1}{2}$$

وبالتالي فإن احتمال سحب كرة تحمل رقم فردي ومن الصندوق الأول يكون:

$$P(A \cap B) = P(A)P(B/A)$$

$$P(A \cap B) = P(A)P(B/A) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = 0.25$$

نتيجة(6):

$$P(A \cap B) = P(A)P(B/A) = P(B)P(A/B)$$

2.2.2.6− احتمال التقاطع للحوادث المستقلة:

إذا كانت الحوادث مستقلة فإن:

$$P(A/B) = P(A)$$
$$P(B/A) = P(B)$$

ويعطى احتمال التقاطع في هذه الحالة بالعلاقة:

$$P(A \cap B) = P(A).P(B)$$

مثال **(17):**

احسب احتمال الحصول على صورتين عند رمي قطعة نقود متوازنة مرتين متتاليتين.

الحل:

بما أن الحادثان، ظهور صورة في الرمية الأولى وظهور صورة في الرمية الثانية مستقلين فإن:

$$P(A \cap B) = P(A).P(B)$$

$$P(A \cap B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

مثال(18):

إذا كان احتمال إصابة محمد لهدف معين هو $\frac{3}{4}$ واحتمال إصابة فايز لنفس الهدف هو $\frac{1}{3}$، فاحسب كلاً من الاحتمالات التالية:

أن لا يصيب محمد الهدف، أن يصيبا الهدف معاً، أن يصيب أحدهما على الأقل، أن لا يصيبا الهدف في آنٍ واحد (أي أن يصيب أحدهما فقط أو لا يصيبا معاً)، أن يصيب محمد الهدف ولا يصيبه فايز

الحل:

نفرض أنّ A الحادث الذي يقع إذا أصاب محمد الهدف و B الحادث الذي يقع إذا أصاب فايز الهدف عندئذٍ نجد أنّ:

$$P(A)=\frac{3}{4} \quad \text{و} \quad P(B)=\frac{1}{3}$$

وبالتالي فإنّ:

احتمال أن لا يصيب محمد الهدف هو: $p(\bar{A}) = 1 - P(A) = 1 - \frac{3}{4} = \frac{1}{4}$

واحتمال أن يصيبا الهدف معاً هو: $P(A \cap B) = P(A)P(B) = \frac{3}{4}.\frac{1}{3} = \frac{1}{4}$

وذلك بملاحظة أن B و A حدثين مستقلين وغير متنافيين.

أمّا احتمال أن يصيب الهدف أحدهما على الأقل فهو:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{3}{4} + \frac{1}{3} - \frac{1}{4} = \frac{5}{6}$$

ولحساب احتمال أن لا يصيبا الهدف في آنٍ واحد يجب أن نحسب الاحتمال للحدث $(A \cap B)'$ **حيث نجد أنّ:**

$$P(A \cap B)' = 1 - P(A \cap B) = 1 - \frac{1}{4} = \frac{3}{4}$$

وأخيراً نجد أن احتمال أن يصيب محمد الهدف ولا يصيبه فايز هو:

$$P(A \cap B') = P(A) - P(A \cap B) = \frac{3}{4} - \frac{1}{4} = \frac{1}{2}$$

مثال (19):

صندوق يحوي 6 كرات حمراء و 9 كرات بيضاء. سحبنا كرتين و ذلك بالأشكال التالية:

1 – سحبنا الكرة الثانية بعد إعادة الاولى .

2 – سحبنا الكرة الثانية دون إعادة الاولى .

و المطلوب: احسب احتمال الحصول على كرتين حمراء اللون في كلا الحالتين.

الحل:

بفرض A الحادث الدال على سحب كرة حمراء في المرة الأولى و B الحادث الدال على سحب كرة حمراء في المرة الثانية فيكون الحادث المطلوب هو $A \cap B$ حيث نجد ما يلي:

1 – في حالة الإعادة يكون الحادثان مستقلين ومنه فإن:

$$P(A \cap B) = P(A).P(B)$$

$$P(A \cap B) = \frac{6}{15} \cdot \frac{6}{15} = \frac{4}{25}$$

2 – في حالة السحب بدون إعادة يكون الحادثان غيرمستقلين (شرطيين) ومنه فإن:

$$P(A \cap B) = P(A)P(B/A) = \frac{6}{15} \cdot \frac{5}{14} = \frac{1}{7}$$

مثال (20):

صندوق فيه 25 مصباحاً كهربائياً 5 منها تالف. سحبنا عشوائياً ثلاثة مصابيح. ما احتمال أن تكون المصابيح المسحوبة جميعها تالفة في الحالتين التاليتين:

1– إذّا تمت عملية السحب على التتالي (دون إعادة)

2– إذّا تمت عملية السحب بحيث يُعاد المصباح المسحوب للصندوق.

الحل:

ليكن A الحادث الذي يقع إذّا كان المصباح المسحوب أولاً تالف، و B الحادث الذي يقع إذا كان المصباح المسحوب ثانياً تالف و C الحادث الذي يقع إذّا كان المصباح المسحوب ثالثاً تالف. عندئذٍ يكون الحادث المطلوب هو $A\cap B\cap C$ حيث نميّز الحالتين التاليتين:

1- إذّا تمت عملية السحب بدون إعادة فإنّ:

$$P(A\cap B\cap C)=P(A)P(B|A)P(C|(A\cap B))$$

حيث نلاحظ أنّ:

$$P(A)=\frac{5}{25}=\frac{1}{5}$$

$$P(B|A)=\frac{4}{24}=\frac{1}{6}$$

$$P(C|(A\cap B))=\frac{3}{23}$$

ومنه:

$$P(A\cap B\cap C)=(\frac{1}{5})(\frac{1}{6})(\frac{3}{23})=\frac{1}{230}\cong 0.004$$

2- إذّا تمت عملية السحب بحيث يُعاد المصباح المسحوب للصندوق فإنّ الأحداث A, B, C في هذه الحالة تكون مستقلة ومتساوية فرص الوقوع حيث يكون:

$$P(A)=\frac{5}{25}=\frac{1}{5}\quad,\quad P(B)=\frac{5}{25}=\frac{1}{5}\quad,\quad P(C)=\frac{5}{25}=\frac{1}{5}$$

وبالتالي فإنّ: $P(A\cap B\cap C)=P(A)P(B)P(C)=\frac{1}{5}.\frac{1}{5}.\frac{1}{5}=\frac{1}{125}=0.008$

مثال(21):

لدى شخص جهاز كهربائي حسّاس يحتاج إلى منظم كهربائي. فإذّا علمت أنّ احتمال شراء الشخص للمنظم هذا العام هو 0.70، واحتمال أن يتعطل الجهاز إذا وصل معه منظم هو 0.20، وأنّ احتمال تعطل الجهاز في حال عدم وصله بمنظم هو 0.85. ما احتمال أن يتعطل الجهاز هذا العام.

الحل:

ليكن A الحادث الذي يقع إذا وفقط إذا تعطل الجهاز هذا العام و B الحادث الذي يقع إذا وفقط إذا اشترى الشخص منظم هذا العام أيضاً. عندئذٍ يكون الحادث المطلوب حساب احتماله هو تعطل الجهاز هذا العام سواءً اشترى الشخص منظماً أم لم يشتر منظماً أي الحادث $A\cap(B\cup B')$ حيث نجد أنّ:

$$P[A\cap(B\cup B')]=P[(A\cap B)\cup(A\cap B')]=P(A\cap B)+P(A\cap B')$$

وذلك بملاحظة أنّ الحدثين $A\cap B\ ,\ A\cap B'$ متنافيان. ومنه نجد أنّ:

$$P[A\cap(B\cup B')]=P(B)P(A|B)+P(B')P(A|B')$$

$$=(0.70)(0.20)+(0.30)(0.85)=0.395$$

3.2.6– الاحتمال الكلي:

ليكن لدينا مجموعة الحوادث B_i التي تشكل تجزيئا للحادث B وليكن A حادث آخر مرتبط بهذه الحوادث عندئذ نجد أن احتمال وقوع الحادث A و الذي ندعوه بالاحتمال الكلي هو:

$$P(A)=P(B_1)*P(A/B_1)+P(B_2)*P(A/B_2)+............$$

وذلك بملاحظة أن الحوادث B_i متنافية مثنى (تشكل تجزيئا للحادث B) وأن الحوادث B_i و A هي حوادث شرطية. و منه يمكن كتابة العلاقة السابقة بالشكل :

$$\boxed{P(A)=\sum_{i=1}^{n}P(B_i)*P(A/B_i)}$$

مثال **(22):**

لدينا صندوقان يحوي الأول 20 كرة منها 10 بيضاء و الباقي سوداء و يحوي الثاني 10 كرات منها 8 بيضاء و الباقي سوداء . سحبنا بشكل عشوائي كرة من أحد الصناديق فما هو احتمال أن تكون بيضاء اللون.

الحل:

بفرض A الحادث الذي يقع إذا كانت الكرة المسحوبة بيضاء، B_1 و B_2 الحدثان الدالان على اختيار الصندوق الأول أو الثاني على الترتيب. عندئذ نجد أن:

$$P(A) = P(B_1) * P(A/B_1) + P(B_2) * P(A/B_2)$$
$$P(A) = \frac{1}{2} * \frac{10}{20} + \frac{1}{2} * \frac{8}{10} = \frac{10}{40} + \frac{8}{20} = \frac{26}{40}$$

ومنه: $P(A) = \frac{13}{20}$

3.6– احتمال السبب:

يعتبر احتمال السبب متابعة للاحتمال الكلي فقد رأينا أن الأخير يعطينا احتمال حدوث الحادث A الذي يقع بعد وقوع أحد الحوادث B_i التي تشكل تجزيئا للحادث B. أما احتمال السبب فيعطي احتمال وقوع الحادث B_i بعد ان علمنا عن وقوع الحادث A، وبمعنى آخر يبين السبب الذي أدى الى وقوع A ومن هنا جاءت التسمية احتمال السبب.

من قانون احتمال التقاطع رأينا ان:

$$P(A \cap B) = P(A)P(B/A) = P(B)P(A/B)$$

ومنه فإن:

$$P(B/A) = \frac{P(B)P(A/B)}{P(A)}$$

ولكن P(A) هو الاحتمال الكلي، اي أن:

$$P(A) = \sum_{i=1}^{N} P(B_i) * P(A/B_i)$$

وبما أننا نريد حساب احتمال وقوع B_i قبل A فإن:

$$\boxed{P(B_i/A) = \frac{P(B_i) * P(A/B_i)}{\sum_{i=1}^{N} P(B_i) * P(A/B_i)}}$$

مثال (23):

لدى فريق كرة قدم 4 فرص لأن يلعب بإحدى المجموعات A ، B، C، D . إذا علمت أن احتمال وصوله إلى النهائيات إذا لعب بكل من هذه المجموعات هو 0.7، 0.5 ، 0.4 ، 0.2 على الترتيب .أحسب ما يلي إذا كان اختيار المجموعة يتم بشكل عشوائي:

1 – احتمال وصوله إلى النهائيات.

2 – إذا علمت أنه وصل إلى النهائيات ما هو احتمال أنه لعب في المجموعة A.

الحل:

1 –

$$P(A) = \Sigma P(B_i) * P(A/B_i)$$

$$P(A) = \frac{1}{4} * \frac{7}{10} + \frac{1}{4} * \frac{5}{10} + \frac{1}{4} * \frac{4}{10} + \frac{1}{4} * \frac{2}{10} = \frac{18}{40}$$

2 –

$$P(B_i / A) = \frac{P(B_i) * P(A/B_i)}{P(A)}$$

$$P(B_i / A) = \frac{\frac{1}{4} * \frac{7}{10}}{\frac{18}{40}} = \frac{7}{40} * \frac{40}{18} = \frac{7}{18}$$

مثال (24)

ثلاث آلات M_1, M_2, M_3 تنتج 40% ، 35% ، 25% من انتاج مصنع وذلك على الترتيب. وجد أنّ 2٪ و 4٪ و 5٪ من إنتاج هذه الآلات رديء وذلك على الترتيب أيضاً. اخترنا سلعة من إنتاج هذا المصنع فكانت رديئة، فما هو احتمال كون هذه السلعة من انتاج الآلة M_1؟

الحل:

لنرمز بالرمز A_i للحادث الذي يقع إذا كانت السلعة من إنتاج الآلة M_i، $i=1,2,3$ وبالرمز B للحادث "السلعة المسحوبة رديئة". من الواضح أن الأحداث A_i تشكل تجزيئا للحدث الشامل وأنّ المطلوب حسابه هو: $P(A_1|B)$. حيث لدينا:

$$P(A_1) = 0.25\ ,\ P(A_2) = 0.35\ ,\ P(A_3) = 0.40$$

$$P(B|A_1) = 0.05\ ,\ P(B|A_2) = 0.04\ ,\ P(B|A_3) = 0.02$$

$$P(A_1|B)\ \frac{P(A_1)P(B|A_1)}{\sum_{j=1}^{3} P(A_j)P(B|A_j)} = \frac{(0.25)(0.05)}{(0.25)(0.05)+(0.35)(0.04)+(0.40)(0.02)}$$

$$= \frac{0.0125}{0.0345} = \frac{125}{345} \cong 0.362$$

تمارين الفصل السادس

1 – ألقي حجري نرد، الأول أخضر والثاني أبيض، وسجلنا النتائج والمطلوب:

أ – أكتب الفضاء الاحتمالي.

ب – أكتب عناصر الحادث A الذي يقع إذا كان المجموع للرقمين الظاهرين على الحجرين أقل من 6 ثم احسب احتمال A.

جـ – احسب احتمال الحادث B الذي يقع إذا ظهر الرقم 2 على الحجر الأخضر.

د – احسب احتمال الحادث C الذي يقع إذا ظهر الرقم 4 على أحد حجري النرد.

هـ – احسب احتمال الحادث $B \cap C$

و – احسب احتمال الحادث $A \cup B \cup C$

2 – جندي احتمال إصابته للهدف 0.7 ، أطلق على هدف ثابت 3 طلقات و المطلوب :

1) – اكتب الفضاء الاحتمالي للتجربة.

2)– احسب احتمال إصابة الهدف بطلقة واحدة فقط .

3)– احسب احتمال إصابة الهدف.

4)– احسب احتمال عدم إصابة الهدف .

3 – سحبنا ورقة بشكل عشوائي من ورق اللعب المكون من 52 ورقة. ما هو احتمال أن تكون صورة أو كبة.

4 – رمينا قطعة نقد مرتين على التتالي و المطلوب :

1– احسب احتمال أن تكون نتيجة القطعة الأولى صورة و الثانية كتابة.

2– احسب احتمال أن نحصل على صورتين.

3– احسب احتمال أن نحصل على صورة واحدة.

5 – يطير طائر فوق ثلاث صيادين A , B , C لدى كل منهم طلقة واحدة و احتمال إسقاط كل منهم للطائر هو 0.5 ، فإذا أطلقوا النار على الطائر بالتتالي اعتبارا من الصياد A .

المطلوب :

1– اكتب الفضاء الاحتمالي للتجربة .

2– احسب احتمال إسقاط الطائر.

3– احسب احتمال عدم إسقاط الطائر.

6– اطلق رامي على هدف 200 طلقة فأصابه بـ 120 طلقة. والمطلوب :

1– احسب احتمال إصابة هذا الرامي للهدف.

2 – احسب عدد الطلقات الواجب رميها ليكون احتمال إصابته للهدف 0.9 على الاقل.

7– اذا علمت أن 60%من طلاب أحد المستويات يتكلمون الإنكليزية و 30% يتكلمون الفرنسية و 20% يتكلمون الإنكليزية و الفرنسية معا و باقي الطلاب لا يتكلمون أي لغة أجنبية .

المطلوب:

– احسب احتمال أن يكون طالب لا يتكلم الإنكليزية .

–احسب احتمال أن يكون طالب يتكلم الإنكليزية فقط .

–احسب احتمال أن يكون طالب يتكلم لغة أجنبية على الاقل .

– احسب احتمال أن يكون طالب لا يتكلم أي لغة أجنبية .

8 – أتفق لاعبان A, B على رمي قطعة نقد سوية حتى اللانهاية و يربح من يحصل على الشعار أولا و المطلوب:

1 –اكتب الفضاء الاحتمالي لهذه التجربة اذا بدأ A بالرمي .

2–احسب احتمال فوز كل منهما .

9 – نرمي حجري نرد معا و المطلوب :

– احسب احتمال الحصول على مجموع أكبر من 10.

– احسب احتمال الحصول على مجموع أكبر من 10 إذا علمت أن أحد الحجرين استقر على الوجه 5.

– احسب احتمال الحصول على مجموع أكبر من 10 اذا علمت أن الحجر الأول فقط استقر على الوجه 5.

10 – صنعت قطعة نقود بحيث أن احتمال ظهور الصورة هو $P(H)=\frac{1}{3}$ **واحتمال ظهور الكتابة** $P(T)=\frac{2}{3}$ **فإذا ألقيت هذه اقطعة مرة واحدة وظهرت الصورة فإننا نختار عدداً عشوائياً من 1 إلى 11، أمّا إذا ظهرت الكتابة فإننا نختار عدداً عشوائياً من 1 إلى 7. فما هو احتمال أن يكون العدد المختار فردياً؟**

11 – معمل يحوي خطي إنتاج تنتج علبا متماثلة تحوي كل علبة 24 قطعة. ينتج الخط الأول 60% من الإنتاج اليومي و الباقي ينتجها الخط الثاني ،كما ينتج الخط الأول قطعتين رديئتين في كل علبة و ينتج الخط الثاني قطعة رديئة في كل علبة . و المطلوب:

1– سحبنا علبة من إنتاج أحد الأيام و أخذنا منها قطعة بشكل عشوائي فما هو احتمال أن تكون رديئة الصنع.

2– سحبنا علبة من إنتاج أحد الأيام و أخذنا منها قطعة فكانت رديئة فما هو احتمال انها من إنتاج الخط الثاني.

12 – لدينا صندوقان متماثلان يحوي الأول 20 كرة حمراء و 10 بيضاء و يحوي الثاني 15كرة حمراء و 5 بيضاء و المطلوب:

1 – اخترنا صندوق عشوائيا و سحبنا منه كرة بشكل عشوائي أيضا فما احتمال أن تكون الكرة المسحوبة حمراء اللون.

2 – سحبنا كرة بشكل عشوائي من أحد الصناديق فكانت حمراء اللون فما هو احتمال كونها من الصندوق الأول.

3– سحبنا كرة من الصندوق الأول و وضعناها بالثاني ثم سحبنا كرة من الصندوق الثاني فما هو احتمال أن تكون حمراء اللون.

13 – تنوي أسرة قضاء إجازة نهاية أسبوع في إحدى الأماكن السياحة A ، B ،أو

C فإذا كان احتمال سقوط المطر في A هو 0.6 وفي B هو 0.7 وفي C هو 0.5

وإذا اختارت الأسرة مكان الاجازة عشوائياً فأحسب:

أ – احتمال أن تكون الإجازة ممطرة.

ب – إذا كانت الإجازة ممطرة فما هو احتمال أن هذه الإجازة كانت في المكان B.

14– ينتج أحد المصانع 300 بدلة رجالية و 900 بدلة نسائية كما ينتج مصنع آخر 500 بدلة رجالية و 500 بدلة نسائية فإذا أختيرت بدلة واحدة فقط من كل مصنع لقياس جودة الإنتاج، فما هو احتمال أن تكون إحدى البدلتين رجالية على الأقل

15 – إذا كان معدل إنتاج ثلاثة مصانع مختلفة للبطاريات هي 500، 700، 900

بالترتيب وكان عدد البطاريات التالفة في هذا الانتاج هو 20 ، 30 ، 60 على

الترتيب حيث يُباع هذا الإنتاج في محل لبيع المفرق فما هو احتمال أن تكون البطارية

المشتراة تالفة. ثم ما هو احتمال أن تكون البطارية المشتراة التالفة هي من إنتاج

المصنع الأول؟

مراجع الفصل السادس :

1- حميدان، عدنان؛ الجاعوني، فريد؛ ناصر آغا، عمار؛ العواد، منذر، 2003، مبادئ الإحصاء، منشورات جامعة دمشق.

2- حسام ،كمرجي،2011 - الاحتمال والاحصاء - جامعة دمشق – كلية الهندسة الميكانيكية والكهربائية .

3- Groebner David F., Shannon Patrick W. and Fry Phillip C. (2017)," Business Statistics :A Decision-Making approach", 10 th Edition. person.

الفصل السابع

المتغيرات العشوائية والتوزيعات الاحتمالية

المخرجات والأهداف التعليمية:

بعد الانتهاء من هذا الفصل سيكون الطالب:

1- قادرا على التمييز بين انواع المتغيرات العشوائية .
2- قد الم بمفهوم التوزيعات الاحتمالية .
3- قادرا على حساب الاحتمالات الممكنة للمتغير العشوائي .
4- تعلم كيف يستخدم التوزيعات الاحتمالية في التطبيقات الادارية والاقتصادية .

يعتبر الفصل السابع استكمالا للفصلين السابقين حيث يتم من خلاله التعرف على مفهوم المتغير العشوائي المنقطع والمستمر والذي يعتبر اكثر ديناميكية من استخدام مفهوم الحوادث. كما يتم استعراض بعض المفاهيم الهامة المرتبطة بالمتغيرات العشوائية كالتوقع الرياضي والتباين ودالة التوزيع المجمعة. وأخيرا تم التطرق لبعض التوزيعات الرئيسية كالتوزيع الثنائي وتوزيع بواسون والتوزيع الطبيعي والتي يوجد لها تطبيقات متعددة.

1.7- المتغيرات العشوائية و قانون التوزيع الاحتمالي :

1.1.7 - تمهيد:

إذا كان لدينا تجربة عشوائية ما فإنّ هذه التجربة تقودنا إلى مجموعة من النتائج أو البيانات. إنّ هذه النتائج قد تكون كمية أو كيفية. فمثلاً عند رمي قطعتي نقود دفعة واحدة (أو رمي قطعة واحدة مرتين متتاليتين) فإنّ فضاء العينة يكون: $\Omega_1 = \{HH,HT,TH,TT\}$

وذلك إذا كان اهتمامنا هو النتيجة الظاهرة على القطعتين من حيث كونها صورة أو كتابة وفي حالة اهتمامنا بعدد الصور الظاهرة يصبح فضاء العينة بالشكل:

$$\Omega_2 = \{0,1,2\}$$

حيث نلاحظ أنّ البيانات الناتجة عن التجربة العشوائية في Ω_1 هي بيانات كيفية وفي Ω_2 هي بيانات كمية. ولكن نحتاج أحياناً لرد البيانات الكيفية إلى رقمية، حيث يُساعد ذلك الإحصائي على تسهيل الدراسة الإحصائية واستقراء النتائج. فيمكن أن نقرن كل عنصر من فضاء العينة بعدد يعبر عن صفة معينة في العنصر. فمثلاً في مثالنا السابق يمكن أن نقرن كل عنصر من عناصر Ω_1 بعدد مرات ظهور الكتابة.

2.1.7 – المتغير العشوائي:

المتغير العشوائي هو دالة من فضاء عينة ما إلى مجموعة الأعداد الحقيقية R. حيث يقابل كل عنصر من فضاء العينة عدداً حقيقياً هو قيمة المتغير العشوائي عند هذا العنصر.

يُرمز للمتغيرات العشوائية عادةً بأحرف لاتينية كبيرة مثل X, Y, Z,... ويُرمز لقيم هذه المتغيرات العشوائية بأحرف لاتينية صغيرة مثل x, y, z... .

فمثلاً في تجربة رمي قطعتي نقود دفعة واحدة، إذا كان X المتغير العشوائي الدال على عدد مرات ظهور الكتابة فإنّ X يعتبر دالة من فضاء العينة Ω_1 المبين في الفقرة السابقة إلى مجموعة الأعداد الحقيقية R حيث يقرن المتغير X كل عنصر من Ω_1 بعدد مرات ظهور الكتابة في هذا العنصر. وبالتالي نجد أنّ:

$$X(HH) = 0 \,,\, X(HT) = X(TH) = 1 \,,\, X(TT) = 2$$

وللمتغير العشوائي نوعان:

1 – متغير عشوائي منقطع : وهو المتغير الذي يأخذ عددا محدودا من القيم في مجال تغيره، مثل عدد الصور الناتجة عند رمي قطعة نقود خمس مرات أو عدد الناجحين بمقر ما من مجموعة طلاب عددها 100 طلاب.

2– متغير عشوائي مستمر : إذا أمكن للمتغير أن يأخذ عددا غير منته من القيم في مجال تغيره يسمى متغيرا مستمرا مثل الزمن اللازم لإنجاز عمل معين ، طول شخص، زمن احتراق مصباح كهربائي...

3.1.7– قانون التوزيع الاحتمالي :

يعرف قانون التوزيع الاحتمالي بأنه العلاقة التي تربط بين القيم الممكنة للمتغير العشوائي و احتمالاتها و تعطى هذه العلاقة كما يلي :

في حالة المتغيرات المنقطعة تعطى العلاقة الاحتمالية بجدول كما يلي:

x_n	...	...	...	x_3	x_2	x_1	x_i

P_i	P_1	P_2	P_3	...	...	...	P_n

حيث: $P_i = P(X = x_i)$ **هو الاحتمال عند القيمة** x_i **للمتغير العشوائي** X.

ويجب أن يحقق قانون التوزيع الاحتمالي الشرطين التاليين:

1 - $0 \leq P_i \leq 1$

2 - $\sum_{i=1}^{n} P_i = 1$

وفي حالة المتغير العشوائي المستمر فتعطى الاحتمالات لقيمه من خلال علاقة رياضية (دالة رياضية) كما يلي:

$$P_i = \begin{cases} f(x) & , \quad x \in [a,b] \\ 0 & , \quad x \notin [a,b] \end{cases}$$

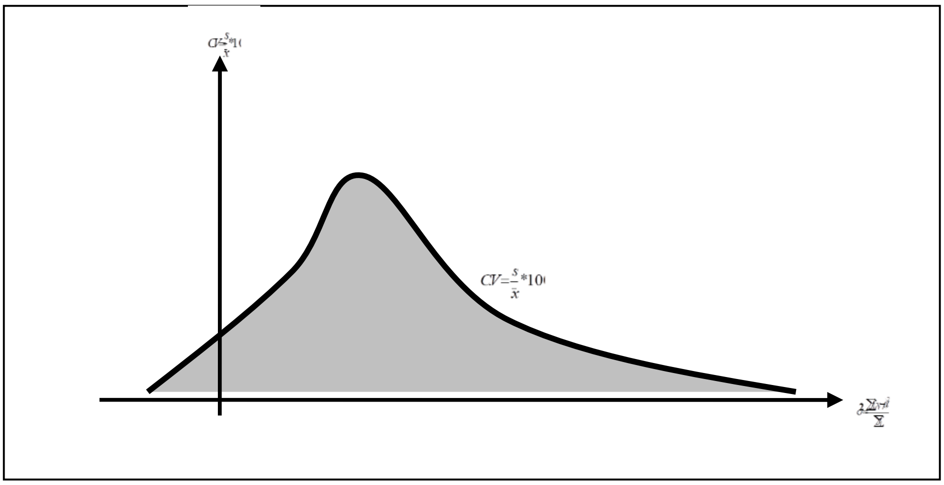
حيث: $P_i = P(X = x_i) = f(x_i)$

وتحقق هذه الدالة الشرطين التاليين:

1 - $0 \leq f(x) \leq 1$

2 - $\int_a^b f(x)dx = 1$

أي أن المساحة الاحتمالية بين منحني الدالة $f(x)$ **والمحور ox تساوي الواحد كما هو موضح بالشكل (6.1) . وتسمى الدالة** $f(x)$ **في هذه الحالة بدالة الكثافة الاحتمالية.**



الشكل(6.1)

مثال(1) :

ليكن X المتغير العشوائي الدال على القيم الناتجة عند رمي حجر نرد متوازن. ومنه نجد أن قيم X هي 1,2,3,4,5,6 وجدول التوزيع الاحتمالي للمتغير Xهو:

x_i	1	2	3	4	5	6
P_i	1/6	1/6	1/6	1/6	1/6	1/6

حيث نلاحظ أن: $0 \leq P_i \leq 1$

$$\sum_{i=1}^{n} P_i = 1$$

مثال(2):

في تجربة رمي قطعتي نقود متوازنتين دفعة واحدة، ليكن X المتغير العشوائي الدال على عدد الصور الظاهرة. أوجد جدول التوزيع الاحتمالي لـ X.

الحل:

إنّ الفضاء الاحتمالي في هذا المثال هو: Ω={HH,HT,TH,TT}

حيث يدل الرمز H على ظهور الصورة والرمز T على ظهور الكتابة. ومنه فإنّ المتغير العشوائي X يأخذ القيم صفر أو 1 أو 2. وبالتالي نجد أنّ جدول التوزيع الاحتمالي لهذا المتحول هو:

x_i	0	1	2
P_i	1/4	2/4	1/4

ملاحظة(1):

عند تشكيل جدول التوزيع الاحتمالي للمتغير المنقطع في المثالين السابقين تم حساب احتمالات قيم المتغير بأخذ عدد الحالات المواتية على عدد الحالات الكلية بإعتبار أن فرص الظهور لعناصر فضاء العينة في هذين المثالين متساوية . وفي الحالة التي تكون فيها فرص الظهور غير متساوية يتم حساب الاحتمالات باستخدام التوافيق كما في المثال التالي.

مثال(**3**):

صندوق فيه عشرة مصابيح كهربائية، اثنان منها تالف. اشترى شخص مصباحين أختيرا عشوائياً. بفرض المتغير العشوائي X يدل على عدد المصابيح التالفة التي أشتراها الشخص. أوجد جدول التوزيع الاحتمالي لـ X.

الحل:

إنّ فضاء العينة في هذا المثال هو: $\mathbf{\Omega} = \{GG, GB, BG, BB\}$

حيث يدل الرمز B على المصباح التالف والرمز G على المصباح غير التالف. ومنه فإنّ المتغير العشوائي X يأخذ القيم صفر أو 1 أو 2. والتي تحسب احتمالاتها كما يلي:

$$f(0) = P(X=0) = \frac{C_8^2 C_2^0}{C_{10}^2} = \frac{28}{45}$$

$$f(1) = P(X=1) = \frac{C_8^1 C_2^1}{C_{10}^2} = \frac{16}{45}$$

$$f(2) = P(X=2) = \frac{C_8^0 C_2^2}{C_{10}^2} = \frac{1}{45}$$

وبالتالي فإنّ جدول التوزيع الاحتمالي المنقطع للمتغير العشوائي X هو:

x	0	1	2
$f(x)=P(X=x)$	$28/45$	$16/45$	$1/45$

مثال(4):

ليكن X المتغير العشوائي المستمر والمعرف بدالة الكثافة الاحتمالية التالية:

$$f(x)=\begin{cases}\frac{1}{2}x & , \quad x\in[0,2]\\ 0 & , \quad x\notin[0,2]\end{cases}$$

عندئذ نلاحظ بسهولة بأن:

$$0\le f(x)\le 1$$

$$\int_0^2 f(x)dx=\int_0^2\frac{1}{2}x\ dx=\frac{1}{2}\left[\frac{x^2}{2}\right]_0^2=1$$

2.7–القيم المميزة للمتغير العشوائي:

1.2.7– التوقع الرياضي:

يعرف التوقع الرياضي لمتغير عشوائي منقطع، بأنه مجموع جداءات القيم باحتمالاتها و يرمز له بالرمز E(X) ، ومنه فإن:

$$E(X)=\sum_{i=1}^{n}x_iP_i$$

وواضح هنا أن التوقع هو عبارة عن المتوسط الحسابي لقيم X المرجح بالاحتمال.

مثال(5):

احسب التوقع الرياضي للمتغير العشوائي X المعطى بالمثال (1) والدال على نتيجة رمي حجر النرد .

الحل:

$$E(X) = \sum_{i=1}^{6} x_i P_i = \frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6} = 3.5$$

مثال(6):

احسب التوقع الرياضي للمتغير العشوائي X المعطى بالمثال (2) والدال على عدد الصور الظاهرة عند رمي قطعتي نقود متوازنتين.

الحل:

$$E(X) = \sum x_i P_i = 0(\frac{1}{4}) + 1(\frac{2}{4}) + 2(\frac{1}{4}) = 1$$

أي أننا نتوقع على المدى الطويل بأنه عند رمي قطعتي نقود متوازنتين فإن عدد الصور الظاهرة ستكون واحد.

مثال(7):

احسب التوقع الرياضي للمتغير العشوائي X المعطى بالمثال (3) والدال على عدد المصابيح التالفة المشتراة من صندوق فيه عشرة مصابيح أثنان منها تالف.

الحل:

$$E(X) = \sum x_i P_i = 0(\frac{28}{45}) + 1(\frac{16}{45}) + 2(\frac{1}{45}) = \frac{18}{45} = 0.4$$

أي أننا نتوقع على المدى الطويل بأنه عند سحب مصباحين على التتالي من صندوق فيه عشرة مصابيح أثنان منها تالفان فإن 0.4 سيكون من المصابيح التالفة.

أما التوقع الرياضي للمتغير العشوائي المستمر فيعطى بالعلاقة:

$$E(X) = \int_a^b xf(x)dx$$

مثال(8):

احسب التوقع الرياضي للمتغير العشوائي X المعطى بالمثال (4) والمعرف بالدالة:

$$f(x) = \begin{cases} \frac{1}{2}x & , \quad x \in [0,2] \\ 0 & , \quad x \notin [0,2] \end{cases}$$

الحل:

$$E(X) = \int_a^b xf(x)dx = \int_0^2 x\left(\frac{1}{2}x\right)dx = \frac{1}{2}\left[\frac{x^3}{3}\right]_0^2 = \frac{8}{6} = 1.33$$

2.2.7.التباين و الانحراف المعياري:

يعطى التباين للمتغير العشوائي المنقطع X بإحدى العلاقتين التاليتين:

$$\sigma_X^2 = \sum_{i=1}^{n} [x_i - E(X)]^2 P_i$$

$$\sigma_X^2 = E(X^2) - [E(X)]^2$$

وفي حالة المتغير العشوائي المستمر X فيعطى التباين بإحدى العلاقتين التاليتين:

$$\sigma_X^2 = \int_a^b [x - E(X)]^2 f(x)dx$$

$$\sigma_X^2 = E(X^2) - [E(X)]^2$$

أما الانحراف المعياري للمتغير X فهو الجذر التربيعي الموجب للتباين:

$$\sigma_X = \sqrt{\sigma_X^2}$$

مثال(9):

احسب التباين للمتغير العشوائي المنقطع X المعطى بجدول التوزيع الاحتمالي التالي باستخدام العلاقتين السابقتين ثم أوجد الانحراف المعياري.

	2	5	7
P_i			

الحل:

نحسب أولا التوقع الرياضي:

$$E(X) = \sum x_i P_i = \frac{2}{4} + \frac{10}{3} + \frac{7}{12} = \frac{53}{12} = 4.4166$$

ولنحسب الأن التباين باستخدام العلاقة الاولى:

$$\sigma_X^2 = \sum_{i=1}^{n} [x_i - E(X)]^2 P_i$$

$$\sigma_X^2 = (2-4.4166)^2 * \frac{1}{4} + (5-4.4166)^2 * \frac{2}{3} + (7-4.4166)^2 * \frac{1}{12} = 2.24$$

كذلك نجد أن التباين باستخدام العلاقة الثانية هو:

$$\sigma_X^2 = E(X^2) - [E(X)]^2$$

$$\sigma_X^2 = \frac{4}{4} + \frac{50}{3} + \frac{49}{12} - \left(\frac{53}{12}\right)^2 = 2.24$$

ومنه نجد أن الانحراف المعياري هو: $\sigma_X = \sqrt{\sigma_X^2} = \sqrt{2.24} = 1.497$

دالة التوزيع الاحتمالي المجمعة:-3.7

تعرف دالة التوزيع الاحتمالي المجمعة للمتغير العشوائي X والتي نرمز لها بالرمز $F(x)$ بالعلاقة التالية:

$$F(x) = P(X \le x)$$

فإذا كان x متغيراً عشوائياً منقطعاً فإنّ:

$$F(x_s) = P(X \le x_s) = \sum_{i=1}^{s} P_i$$

ومنه يكون:

$$P(X > x_s) = 1 - F(x_s)$$

أما في حالة المتغير العشوائي المستمر، فإن دالة التوريع الاحتمالي المجمعة تعطى بالعلاقة التالية:

$$F(x_s) = P(X \le x_s) = \int_a^{x_s} f(x)dx$$

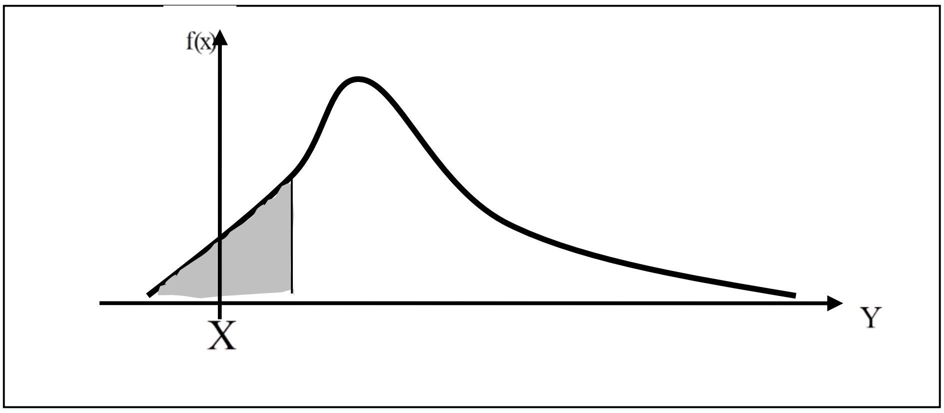
ومنه يكون:

$$P(X > x_s) = \int_{x_s}^{b} f(x)dx = 1 - F(x_s)$$

وإذا كانت x_s قيمة متغيرة عندئذ نرمز لها بالرمز x وتصبح علاقة دالة التوزيع المجمعة هي:

$$F(x) = P(X \le x) = \int_a^{x} f(x)dx$$

حيث تمثل المساحة المظللة المبينة بالشكل (7.2)التالي:



الشكل (7.2)– دالة التوزيع المجمعة

مثال(10):

بفرض أن X المتغير العشوائي الدال على الرقم الظاهر عند رمي حجر نرد، احسب احتمال ان تكون النتيجة أصغر أو تساوي 4، وإحتمال كونها أكبر من 4.

الحل:

احتمال كون النتيجة أصغر أو تساوي 4 هي:

$$F(4) = P(X \leq 4) = \sum_{i=1}^{4} P_i = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{4}{6} = \frac{2}{3}$$

أما احتمال أن تكون النتيجة أكبر من 4 فهو:

$$P(X > 4) = 1 - F(4) = 1 - \frac{4}{6} = \frac{2}{6}$$

مثال(11):

أوجد احتمال أن تكون قيمة المتغير X أصغر أو تساوي الواحد ، واحتمال كونها أكبر من الواحد، وذلك للمتغير العشوائي المستمر المعطى بقانون التوزيع الاحتمالي التالي:

$$f(x) = \begin{cases} \frac{1}{2}x & , \quad x \in [0,2] \\ 0 & , \quad x \notin [0,2] \end{cases}$$

الحل:

احتمال أن تكون قيمة المتغير X أصغر او تساوي 1 باستخدام مفهوم دالة التوزيع المجمعة هو:

$$F(1) = P(X \leq 1) = \int_0^1 \frac{1}{2} x \, dx = \left[\frac{x^2}{4}\right]_0^1 = 0.25$$

واحتمال أن تكون قيمة المتغيرX أكبر من 1 فهي:

$$P(X > 1) = \int_1^2 \frac{1}{2} x \, dx = \left[\frac{x^2}{4}\right]_1^2 = 0.75$$

أو:

$$P(X > 1) = 1 - P(X \leq 1) = 1 - 0.25 = 0.75$$

4.7– العمليات على المتغيرات العشوائية:

1 – ضرب متغير عشوائي بثابت **C**:

إن ضرب متغير عشوائي X بثابت C يعطي متغير عشوائي جديد، قيمة هي عبارة عن ضرب القيم x_iللمتغير X بالعدد C واحتمالاتها هي نفس الاحتمالات السابقة.

مثال(12):

ليكن X المتغير العشوائي المنقطع والمعطى بقانون التوزيع الاحتمالي التالي:

x_i	2	3	5
P_i	1/4	1/2	1/4

إن ضرب المتغير X بالعدد C=2 يعطي متغيرا جديدا 2X له قانون التوزيع الاحتمالي التالي:

$2x_i$	4	6	10
P_i	1/4	1/2	1/4

2– جمع متغيرين عشوائيين: **جمع متغيرين عشوائين منقطعين X,Y ، كل منهما يخضع لقانون توزيع احتمالي يعطي متغير عشوائي منقطع جديد قيمه عبارة عن كل المجاميع الممكنة لقيم كل من المتغيرين السابقين و احتمالاتها هي عبارة عن احتمال تحقق هذه المجاميع وذلك كما يلي:**

ليكن لدينا المتغيرين العشوائيين X,Y المعطيين بقانوني التوزيع الاحتمالي التاليين:

x_i	x_1	x_2	…	…	…	x_n
P_i	P_1	P_2	…	…	…	P_n

y_j	y_1	y_2	…	…	…	y_m

P_j	P_1	P_2	...	...	...	P_m

عندئذ نجد أن قانون التوزيع الاحتمالي للمتغير العشوائي الناتج عن مجموعهما X+Y هو:

$x_i + y_j$	$x_1 + y_1$	$x_1 + y_2$	...	$x_1 + y_m$	$x_2 + y_1$	...	$x_n + y_m$
P_{ij}	P_1*P_1	P_1*P_2	...	P_1*P_m	P_2*P_1	...	P_n*P_m

مثال(13):

ليكن لدينا المتغيرين العشوائيين الخاضعين لقانوني التوزيع الاحتماليين التاليين:

y_j	5	7
P_j	1/3	2/3

x_i	2	3	5
P_i	1/4	1/2	1/4

عندئذ نجد أن مجموع المتغيرين X+Y هو متغير عشوائي يخضع للقانون الاحتمالي التالي:

$x_i + y_j$	5+2	5+3	5+5	7+2	7+3	7+5
P_{ij}	$\frac{1}{3}*\frac{1}{4}$	$\frac{1}{3}*\frac{1}{2}$	$\frac{1}{3}*\frac{1}{4}$	$\frac{2}{3}*\frac{1}{4}$	$\frac{2}{3}*\frac{1}{2}$	$\frac{2}{3}*\frac{1}{4}$

وباصلاح القانون يكون:

$x_i + y_j$	7	8	9	10	12
P_{ij}	1/12	2/12	2/12	5/12	2/12

لاحظ ان :

$$0 \leq P_{i,j} \leq 1$$

$$\sum P_{i,j} = 1$$

3- ضرب متغيرين عشوائيين: **ليكن لدينا المتغير العشوائي X الخاضع لقانون التوزيع الاحتمالي التالي:**

x_i	x_1	x_2	…	…	…	x_n
P_i	P_1	P_2	…	…	…	P_n

والمتغير العشوائي Y الخاضع لقانون التوزيع الاحتمالي التالي:

y_j	y_1	y_2	…	…	…	y_m
P_j	P_1	P_2	…	…	…	P_m

عندئذ نجد أن حاصل ضرب المتغيرين X وY هو المتغير العشوائي XY الخاضع لقانون التوزيع الاحتمالي التالي:

$x_i * y_j$	$x_1 * y_1$	$x_1 * y_2$	…	$x_1 * y_m$	$x_2 * y_1$	…	$x_n * y_m$
P_{ij}	P_1*P_1	P_1*P_2	…	P_1*P_m	P_2*P_1	…	P_n*P_m

مثال(14):

ليكن لدينا المتغيرين العشوائيين XوY الخاضعين لقانوني التوزيع الاحتمالي التاليين:

y_j	5	7
P_j	1/3	2/3

x_i	2	3	5
P_i	1/4	1/2	1/4

عندئذ نجد أن قانون التوزيع الاحتمالي للمتغير XY هو:

$x_i * y_j$	2*5	2*7	3*5	3*7	5*5	5*7
P_{ij}	$\frac{1}{3}*\frac{1}{4}$	$\frac{1}{4}*\frac{2}{3}$	$\frac{1}{2}*\frac{1}{3}$	$\frac{2}{3}*\frac{1}{2}$	$\frac{1}{3}*\frac{1}{4}$	$\frac{2}{3}*\frac{1}{4}$

أو:

$x_i * y_j$	10	14	15	21	25	35
P_{ij}	$\frac{1}{12}$	$\frac{2}{12}$	$\frac{2}{12}$	$\frac{4}{12}$	$\frac{1}{12}$	$\frac{2}{12}$

5.7– قوانين التوزيع الاحتمالي الرئيسية:

تعرفنا فيما سبق على المتغيرات العشوائية المنقطعة والمستمرة وعلى التوزيع الاحتمالي لها حيث يصبح حساب الاحتمالات للمتغيرات المنقطعة أصعب عندما تكون فرص الظهور غيرمتساوية وعدد قيم X كبيرة، كذلك لاحظنا بأن الاحتمالات في حالة المتغير العشوائي المستمر تعتمد على عمليات تكاملية لدالة الكثافة الاحتمالية حيث تزداد صعوبة حساب الاحتمالات بازدياد صعوبة حساب التكامل للكثير من دوال الكثافة .

لذلك يعتمد الإحصائيون على أخذ النماذج (المتغيرات) الأكثر ظهورا في التجارب الإحصائية ووضع قوانين ثابتة لحساب احتمالات قيمها وهو مايسمى بالتوزيعات الاحتمالية الرئيسية لبعض المتغيرات المنقطعة والمستمرة الشهيرة. وسندرس فيما يلي أهم هذه التوزيعات.

1.5.7 – قوانين التوزيع الاحتمالية للمتغيرات المنقطعة:

1.1.5.7 – قانون التوزيع المنتظم:

اذا كان المتغير العشوائي المنقطع X يأخذ قيمه الممكنة باحتمالات متساوية، فنقول أنه يخضع لقانون التوزيع الاحتمالي المنتظم كما في الجدول التالي:

x_i	x_1	x_2	…	…	…	x_n
P_i	$\frac{1}{n}$	$\frac{1}{n}$	…	…	…	$\frac{1}{n}$

حيث:

$$0 \leq \frac{1}{n} \leq 1$$

$$\sum_{i=1}^{n} \frac{1}{n} = 1$$

وتكون القيم المميزة للمتغير المنقطع المنتظم X هي:

التوقع الرياضي: $E(X) = \frac{\sum x_i}{n}$

التباين: $\sigma_X^2 = E(X^2) - [E(X)]^2$

ودالة التوزيع الاحتمالي المجمعة:

$$F(x_s) = P(X \leq x_s) = \sum_{i=1}^{s} P_i = \sum_{i=1}^{s} \frac{1}{n} = \frac{s}{n}$$

مثال(15):

المتغير العشوائي X الدال على نتيجة رمي حجر النرد هو متغير عشوائي يخضع لقانون التوزيع الاحتمالي المنتظم التالي:

x_i	1	2	3	4	5	6
P_i	1/6	1/6	1/6	1/6	1/6	1/6

حيث نجد أن:

التوقع الرياضي: $E(X) = \frac{\sum x}{n} = \frac{21}{6} = 3.5$

التباين : $\sigma_X^2 = E(X^2) - [E(X)]^2 = 15.16 - (3.5)^2 = 2.619$

كما نجد أن احتمال أن تكون قيمة X أصغر أو تساوي 4 هو:

$$F(x_s) = P(X \leq x_s) = \frac{s}{n}$$

$$F(4) = P(X \leq 4) = \frac{4}{6} = \frac{2}{3}$$

2.1.5.7– قانون التوزيع الثنائي:

هناك بعض الحالات التي تتصف فيها التجربة الإحصائية بما يلي :

1 – لكل تجربة نتيجتان فقط ، نجاح (تحقق الصفة) أو فشل (عدم تحقق الصفة).

2– إذا كان احتمال النجاح p واحتمال الفشل q فإن: p+q = 1 .

3– عدد المحاولات منته ويساوي n ونتائج المحاولات مستقلة.

4– نهتم بعدد مرات النجاح خلال n تجربة .

تسمى كل تجربة تتصف بالصفات السابقة بالتجربة الثنائية. بفرض X المتغير العشوائي الدال على عدد مرات النجاح خلال n محاولة في تجربة ثنائية، فإن X سيكون متغير عشوائي منقطع قيمه هي 0,1,2,...,n ويمكن حساب احتمالات هذه القيم كما يلي:

بفرض أنه قمنا بالتجربة 3 مرات وأن المتغير العشوائي X يمثل عدد مرات النجاح عندئذ ينتج الجدول التالي لقيم X واحتمالاتها:

و منه نرى أن الاحتمالات P_x ما هي إلا عبارة عن منشور ثنائي حدي نيوتن:

$$(p+q)^3 = \sum_{X=0}^{3} C_3^X p^X q^{3-X}$$

وبالتالي نستنتج أن قانون التوزيع الاحتمالي لحساب احتمالات قيم X هو:

$$\boxed{P_x = P(X = x) = C_n^x p^x q^{n-x}}$$

حيث يسمى X بالمتغير العشوائي الثنائي، وتسمى العلاقة السابقة بقانون التوزيع الثنائي.

من تعريف التوزيع الثنائي نجد بأن القيم المميزة للمتغير العشوائي الثنائي X هي:

التوقع الرياضي: E(X)=np:

التباين: $\sigma_X^2 = npq$

الانحراف المعياري: $\sigma_X = \sqrt{npq}$

مثال**(16):**

إذا علمت أن احتمال إصابة جندي للهدف هو 0.6 ، وأنه قام بإطلاق3 طلقات، والمطلوب:

1- احسب عدم اصابة الهدف خلال جميع المحاولات.

2- احسب احتمال إصابة الهدف مرتين فقط.

3- احسب احتمال إصابة الهدف بطلقة على الأقل .

4-احسب التوقع الرياضي والتباين والانحراف المعياري لعدد مرات إصابة الهدف .

الحل:

إذا كان p احتمال اصابة الهدف فإن: p=0.6 واحتمال عدم الإصابة q=0.4. ومنه لدينا توزيع ثنائي حيث n=3 وX هو المتغير العشوائي الدال على عدد الإصابات، وبالتالي فإن:

-1 $$P_0 = C_3^0\,(0.6)^0(0.4)^3 = 1(1)(0.064) = 0.064$$

-2 $$P_2 = C_3^2(0.6)^2(0.4) = 3(0.36)(0.4) = 0.432$$

3- احتمال اصابة الهدف بطلقة على الأقل:

$$P_1 + P_2 + P_3 = 1 - P_0 = 1 - 0.064 = 0.936$$

- 4 E(X)=np=3*0.6=1.8

$$\sigma_X^2 = npq = 3(0.6)(0.4) = 0.72$$

$$\sigma_X = \sqrt{0.72} = 0.849$$

مثال(17):

وجد في إنتاج أحد المصانع أنّ من بين كل 1000 وحدة منتجة توجد 150 وحدة تالفة، فإذا أخذت عينة مكونة من 5 وحدات فاحسب كلاً من الاحتمالات التالية:

أ –الوحدات المختارة كلها سليمة.

ب – وحدة واحدة على الأكثر تالفة.

ج – وحدتين تالفتين على الأقل.

الحل:

نلاحظ بأنّ لدينا تجربة ثنائية حيث يكون الإنتاج تالفاً أو غير تالف. فإذا افترضنا أنّ X هو المتغير العشوائي الممثل لعدد الوحدات التالفة ضمن العينة العشوائية المختارة. فإننا نجد أنه يأخذ القيم التالية: $x = 0,1,2,3,4,5$

كما نجد: $p = \frac{150}{1000} = 0.15 \quad , \quad q = 1 - p = 0.85 \quad , \quad n = 5$

وباستخدام علاقة التوزيع الثنائي نجد أنّ:

أ – احتمال أن تكون جميع الوحدات المختارة سليمة هو:

$$P(X = 0) = f(0, 5, 0.15) = c_5^0\,(0.15)^0(0.85)^5 = 0.4437$$

ب – احتمال وجود وحدة واحدة تالفة على الأكثر هو:

$$P(X \le 1) = P(X = 0) + P(X = 1) = 0.4437 + c_5^1\,(0.15)^1(0.85)^4 = 0.8352$$

ج – احتمال وجود وحدتين تالفتين على الأقل: $P(X \ge 2) = 1 - P(X \le 1) = 1 - 0.8352 = 0.1648$

مثال(18):

أوجد التوقع الرياضي والتباين والانحراف المعياري للمتغير العشوائي الثنائي المعرّف في المثال(17).

الحل:

بـــالعودة غلـــى قـــوانين التوزيـــع الثنـــائي نجـــد أنّ التوقـــع الرياضـــي للمتغيـــر الثنـــائي X هـــو:

$$E(X) = np = (5)(0.15) = 0.75$$

كما نجد أنّ التباين لهذا المتغير هو : $\sigma_x^2 = npq = (5)(0.15)(0.85) = 0.6375$

وبالتالي فإنّ الانحراف المعياري يكون: $\sigma_x = 0.798$

أي أننا نتوقع وجود وحدة واحدة تالفة تقريباً بين الوحدات الخمس المختارة وبانحراف معياري قدره 0.798

لأنّ $0.75 - 0.798 = -0.048 \le \mu \le 0.75 + 0.798 = 1.548$

3.1.5.7– قانون توزيع بواسون (الظواهر النادرة):

هناك بعض الحالات تتمتع بنفس شروط التجربة الثنائية التي يخضع لها قانون التوزيع الثنائي و لكن عدد المحاولات يكون فيها غير محدود ∞ ←n و احتمال تحقق صفة ما p يكون صغير جدا بحيث إذا قمنا بالتجربة عدد كبير من المرات فإن $np=\lambda$ هوالمعدل الوسطي للنجاح حيث يتناهى التوزيع الثنائي الى توزيع بواسون عندما ∞ ←n اي أن:

$$\lim_{n\to\infty} C_n^x p^x q^{n-x} = e^{-\lambda}\frac{\lambda^x}{x!}$$

وبالتالي نستنتج أن قانون التوزيع الاحتمالي لحساب احتمالات المتغير العشوائي البواسوني X هو:

$$\boxed{P_x = P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}}$$

إن التوقع الرياضي والتباين للمتغير البواسوني هو:

$$E(X) = \sigma_X^2 = \lambda$$

وبالتالي فإن الانحراف المعياري يكون: $\sigma_X = \sqrt{\lambda}$

مثال(15):

إذا علمت أن احتمال أن يولد طفل مصاب (بالمنغوليا) لأم تبلغ من العمر أكثر من 40 عاما هو 0.006 ، وإذا كان عدد الولادات في أحد المجتمعات لأمهات يبلغن من العمر أكثر من 40 عاما هو 1000 ولادة فالمطلوب :

1 – احسب احتمال أن يكون عدد الأطفال المصابين 10 فقط .

2– احسب احتمال أن يكون عدد الأطفال المصابين 3 على الأكثر.

3– احسب التوقع الرياضي لعدد المصابين.

الحل:

لدينا توزيع بواسون حيث يمثل X عدد الاطفال المصابين و n=1000 , p = 0.006 ومنه فإن المعل الوسطي للاصابات: λ= np = 6 وبالتالي فإن:

1– احتمال أن يكون عدد الأطفال المصابين 10 هو:

$$P_{10} = e^{-6}\frac{6^{10}}{10!} = 0.0413$$

2– احتمال أن يكون عدد الأطفال المصابين 3 على الأكثر هو:

$$P(X \leq 3) = P_0 + P_1 + P_2 + P_3$$

حيث:

$$P_0 = e^{-6}\frac{6^0}{0!} = 0.00248$$

$$P_1 = e^{-6}\frac{6^1}{1!} = 0.01487$$

$$P_2 = e^{-6}\frac{6^2}{2!} = 0.04462$$

$$P_3 = e^{-6}\frac{6^3}{3!} = 0.08923$$

ومنه فإن:

$P(X \leq 3) = P_0 + P_1 + P_2 + P_3 = 0.00248 + 0.01487 + 0.04462 + 0.08923 = 0.1512$

3-التوقع الرياضي لعدد المصابين: E(X)=np=1000*0.006=6

2.5.7- قوانين التوزيع الاحتمالي للمتغيرات المستمرة:

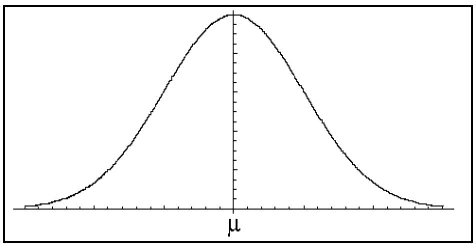
هناك الكثير من قوانين التوزيع الاحتمالي للمتغيرات المستمرة و كلها هامة و لها استخدامات مختلفة و لكن أهم هذه التوزيعات و أكثرها استخداما هو التوزيع الطبيعي لأن أكثر المتغيرات تخضع لهذا التوزيع من جهة، ومن جهة أخرى فأن أكثر التوزيعات الاحتمالية للمتغيرات المستمرة و المنقطعة تنتهي إليه إذا توفرت الظروف المناسبة، لذلك سنكتفي هنا بدراسة التوزيع الطبيعي.

1.2.5.7- قانون التوزيع الطبيعي العام:

لقد تعرفنا على قانون التوزيع الثنائي و الحالات التي يستخدم فيها و لكن لنتساءل:

لو قمنا بالتجربة عدد كبير من المرات و ليكن مليون مرة فما احتمال أن نحصل على ألف مرة نجاح، أي المطلوب حساب القيم التالية: توافيق المليون المأخوذة ألفا فألفا. و P مرفوعة للقوة ألف و q مرفوعة للقوة مليون ناقص ألف و هذا مستحيل فما العمل اذا ؟.

لو قمنا بزيادة عدد مرات التجربة بالتدريج n=3 ومن ثم 4 ... حتى حد معقول سنرى أن شكل المنحنى للتوزيع في كل مرة يتوسع حتى يقترب من الشكل (7.1) التالي:



الشكل (7.3)- منحني دالة الكثافة الطبيعية

الممثل للتوزيع الطبيعي. و لهذا التوزيع صفات سنراها أدناه، نستطيع من خلالها حساب الاحتمالات المطلوبة.

ليكن X متغير عشوائي طبيعي، عندئذ نجد أن X متغير مستمر يتمتع بالصفات التالية:

1 – الصيغة الرياضية لدالة الكثافة الاحتمالية للمتغير X هي:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

حيث :

μ: **الوسط الحسابي (التوقع الرياضي) لقيم المتغير الطبيعي x.**

σ: **الانحراف المعياري للمتغير X.**

e: **العدد الطبيعي(النيبري).**

2– يأخذ المتغير X قيمه في المجال: $]-\infty, \infty[$

3– **يبلغ المنحني** $f(x)$ **نهايته العظمى عند** $x = \mu$

4– محور السينات هو خط مقارب للمنحني$f(x)$.

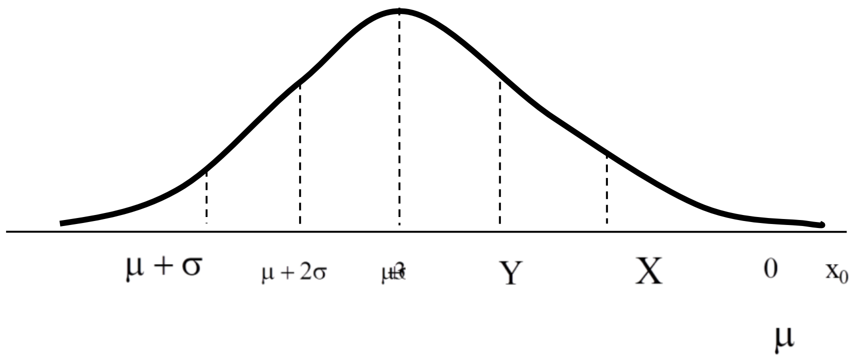
5– للمنحني نقطتا انعطاف عند: $x = \mu+\sigma$ **و** $x = \mu-\sigma$

6– المنحني متناظر بالنسبة للمستقيم $x=\mu$.

7– المساحة الكلية تحت المنحني تساوي الواحد، أي أن:

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

8– اضافة وطرح ثلاث انحرافات للوسط الحسابي μ يغطي تقريبا كامل التوزيع الطبيعي كما هو موضح بالشكل (7.4) التالي:



الشكل(7.4)

9- دالة التوزيع المجمعة للمتغير الطبيعي X هي:

$$F(x_s) = P(X \leq x_s) = \int_{-\infty}^{x_s} f(x)dx$$

إذا كان X متغير طبيعي وسطائه μ , σ^2 فإننا نكتب: $X \sim N(\mu, \sigma^2)$

مما سبق نستنتج بأنه يمكن حساب الاحتمال الطبيعي من خلال العلاقة:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

ولكن حساب هذا الاحتمال مرتبط بقيم وسطاء التوزيع μ وσ^2 في علاقة دالة الكثافة الاحتمالية الطبيعية السابقة. لذلك لحساب مثل هذه الاحتمالات نستعين بالتوزيع الطبيعي المعياري و الذي نعتبر فيه أن $\mu = 0$ و $\sigma^2 = 1$ ونعبر عنه بالرمز: $X \sim N(0,1)$

2.2.5.7- قانون التوزيع الطبيعي المعياري:

المقصود بالقيمة المعيارية هو تحويل الفرق بين أي قيمة x_i والوسط الحسابي إلى عدد من الانحرافات المعيارية، ونطلق على هذه القيمة الجديدة أسم الدرجة المعيارية ونرمز لها بالرمز z حيث يتم حسابها كما يلي:

$$z = \frac{x_i - \mu}{\sigma}$$

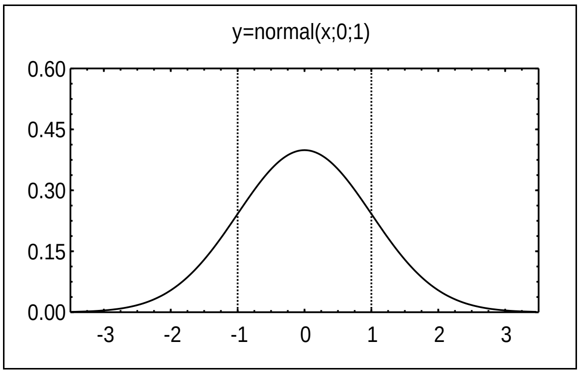
بهذه العملية نكون قد تخلصنا من وحدات القياس للمتغير العشوائي فلا نقول أن القيمة x مثلا تبعد 20 سم عن الوسط الحسابي بل نقول انها تبعد 1 أو 2 .... انحراف معياري. فإذا قمنا بتحويل كل قيم المتغير الطبيعي X الى قيم معيارية يتحول عندئذ التوزيع الطبيعي العام إلى توزيع طبيعي معياري.

صفات قانون التوزيع الطبيعي المعياري:

1-هو توزيع طبيعي وسطائه $\mu = 0$ و $\sigma^2 = 1$ ومنه فإن دالة كثافته الاحتمالية هي:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

ومنحني هذه الدالة يعطى بالشكل التالي:



الشكل (7.5)- منحني دالة الكثافة الطبيعية المعيارية

2- يبلغ منحنيه نهايته العظمى عندما $x = 0$

3- المساحة تحت المنحني تساوي الواحد أي أن:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1$$

4- محور السينات هو خط مقارب لمنحنيه.

5 - دالة التوزيع المجمعة للمتغير المعياري تعطى بالشكل:

$$F(x_s) = P(X \le x_s) = \int_{-\infty}^{x_s} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

وباستخدام الرمز z بدلا من x للدلالة على التوزيع المعياري، تصبح العلاقة الأخيرة بالشكل التالي:

$$F(z) = P(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

و قد قام الاحصائيون بإعداد جداول إحصائية تعطينا قيمة هذا الاحتمال حسب قيمة z. فيكفي أن نعرف بعد القيمة x_s عن الوسط الحسابي مقدرة بالانحرافات المعيارية وذلك بحساب z حتى نستطيع استخراج المساحة تحت المنحني من بدايته حتى هذه النقطة وذلك من جدول المساحات. وقد الحقنا في نهاية هذا الفصل جدول يعطينا المساحات تحت المنحني الطبيعي المعياري حسب قيمة z.

6 – تنحصر قيم z في التوزيع الطبيعي المعياري بين z = -3.6 و z = 3.6 تقريبا.

مثال(16):

تبين في إحدى الجامعات أن متوسط الطول للطلاب هو 175 سم بانحراف معياري قدره 5 سم. والمطلوب احسب احتمال أن يكون طول طالب:

1 – أكثر من 175سم.

2 – أقل من 175 سم.

3 – أقل من 180 سم.

4 – أكثر من 180سم.

5 – بين 170 و185 سم.

الحل:

بفرض أن X هو المتغير الدال على طول الطالب فإن X يكون متغيرا طبيعيا وسطائه $\mu = 175$, $\sigma^2 =$ 25 . وحتى نستطيع حساب كل من الاحتمالات السابقة يجب أولا تحويل التوزيع الطبيعي المعطى إلى توزيع معياري Z وذلك من خلال العلاقة:

$$Z = \frac{X - \mu}{\sigma}$$

حيث ينتج توزيع طبيعي معياري وسطائه: $\mu = 0\ ,\ \sigma^2 = 1$

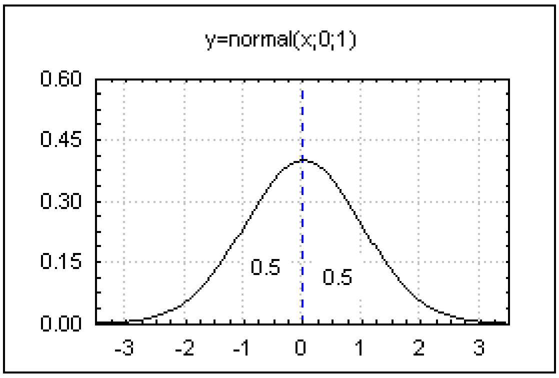
ومنه فإن:

1– لحساب الاحتمال $P(X>175)$ **نحسب أولا z حيث:**

$$z = \frac{x-\mu}{\sigma} = \frac{175-175}{5} = 0$$

$$P(X > 175) = P(Z > 0) = 0.5$$

كما هو موضح بالشكل (7.6) التالي:



الشكل (7.6)

–2

$$P(X < 175) = P(Z < 0) = 0.5$$

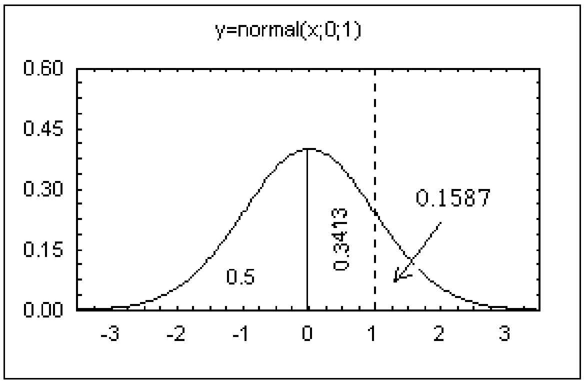
–3

$$z = \frac{x - \mu}{\sigma} = \frac{180 - 175}{5} = 1$$

ومنه فإن:

$$P(X < 180) = P(Z < 1) = 0.8413$$

حيث تم التوصل لهذه القيمة بحساب المساحة تحت المنحني المعياري من بدايته حتى النقطة z=1 وذلك من خلال الجدول المعياري المرفق والذي يعطي الاحتمال المجمع قبل القيمة. كما هو موضح بالشكل (7.7) التالي:



الشكل (7.7)

4 –

$$P(X > 180) = P(Z > 1) = 1 - P(Z \leq 1) = 1 - 0.8413 = 0.1587$$

5 – لنحسب الأن الاحتمال: P(170<X<185) كما يلي:

$$\mathrm{P}(170 < \mathrm{X} < 185) = P(X \leq 185) - P(X \leq 170)$$

وبالتحويل للتوزيع المعياري نجد أن:

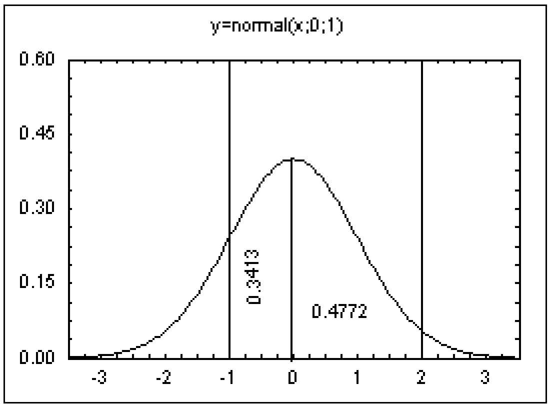
$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{170 - 175}{5} = -1$$

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{185 - 175}{5} = 2$$

ومنه فإن:

$$P(170 < X < 185) = P(-1 < Z < 2) = P(Z \leq 2) - P(Z \leq -1) = 0.9772 - 0.1587 = 0.8185$$

كما هو موضح بالشكل (7.8) التالي:



الشكل (7.8)

مثال(17):

ينتج معمل للبطاريات، بطاريات صغيرة متوسط أعمارها 30 ساعة بانحراف معياري قدره 5 ساعات. بفرض أنّ أعمار البطاريات تتوزع وفق التوزيع الطبيعي، أوجد احتمال أن تعيش بطارية ما أقل من 23 ساعة؟

الحل:

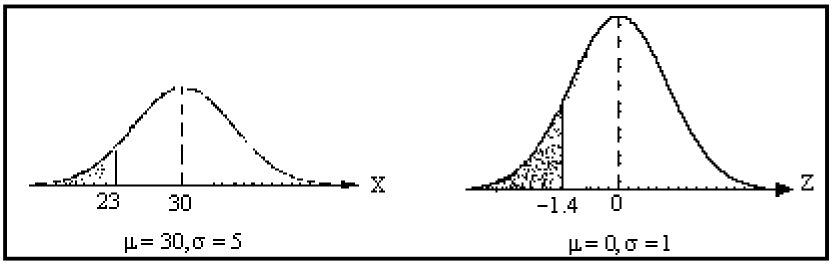
لدينا توزيع طبيعي حيث: $x = 23$ **و** $\mu = 30$ **و** $\sigma = 5$. **وبالتحويل إلى التوزيع الطبيعي المعياري نجد أنّ:**

$$z = \frac{x - \mu}{\sigma} = \frac{23 - 30}{5} = -1.4$$

وبالتالي نجد أنّ الاحتمال المطلوب هو: $P(X < 23) = P(Z < -1.4)$

وباستخدام جداول التوزيع الطبيعي المعياري نجد أنّ: $P(X < 23) = 0.0808$

كما هو موضح في الشكل التالي.



الشكل(7.9)

تمارين الفصل السابع

1 – **في كل حالة من الحالات التالية وضّح ما إذا كان المتغير العشوائي منقطعاً أم مستمرا":**

أ – المتغير العشوائي يدّل على عدد السيارات التي تم صيانتها في ورشة ما في يوم ما.

ب – المتغير العشوائي يدّلُ على كمية هطول الأمطار في دمشق في يوم ما من أيام فصل الشتاء.

جـ – المتغير العشوائي يدّلُ على عدد الإجابات الصحيحة التي يحصل عليها طالب ما في امتحان مؤلف من 20 سؤالاً.

د– الزمن الذي يأخذه ميكانيكي سيارات لتركيب جهاز العادم لسيارة ما.

2 – إذا كان X متغيراً عشوائياً يدّلُ على مجموع العددين الظاهرين عند رمي حجر نرد مرتين متتاليتين، أكتب جميع القيّم التي يمكن أن يأخذها المتغير X.

3 – إذا رميت قطعة نقود ثلاث مرّات متتالية، وإذا كان المتغير X يرمز إلى الفرق بين عدد الصور وعدد الكتابات في الثلاث رميات المذكورة. أكتب القيم الممكنة للمتغير العشوائي X.

4 – في كل حالة من الحالات التالية، وضّح ما إذا كانت الدالة المعطاة يمكن اعتبارها دالة كثافة احتمالية لمتغير عشوائي ما أم لا مع ذكر الأسباب:

(a)

$$f(x) = \begin{cases} \quad , & x=3 \quad \text{عندما} \\ \quad , & x=4 \\ \quad , & x=5 \end{cases}$$

(b)

$$f(x) = \begin{cases} \quad , & x=1 \quad \text{عندما} \\ \quad , & x=2 \\ \quad , & x=3 \end{cases}$$

(c)

$$f(x) = \begin{cases} \quad , & x=4 \\ \quad , & x=6 \end{cases}$$

(d)

$$f(x) = \begin{cases} \quad , & 0 \le x \le 2 \\ \quad , & \text{عدا ذلك} \end{cases}$$

(e)

$$f(x) = \begin{cases} \quad , & -\frac{1}{2} \le x \le \frac{1}{2} \\ \quad , & \text{عدا ذلك} \end{cases}$$

(f)

$$f(x) = \begin{cases} \quad , & 0 \le x \le 1 \\ \quad , & \text{عدا ذلك} \end{cases}$$

5 – في تجربة رمي حجر النرد. ليكن X المتغير العشوائي الدال على العدد الأكبر من الوجهين الظاهرين، وعلى أحدهما في حال تساويهما. أكتب جدول التوزيع الاحتمالي للمتغير العشوائي X وأحسب التوقع الرياضي له.

6 – كيس يحوي 5 كرات بيضاء و 3 خضراء. سحبنا 4 كرات عشوائياً فإذا كان المتغير العشوائي Y يدل على عدد الكرات البيضاء المسحوبة، أكتب جدول التوزيع الاحتمالي للمتغير العشوائي Y وأحسب توقعه الرياضي وتباينه.

7 – شركة فيها 11 مهندساً 3 منهم اختصاص طاقة كهربائية. أرادت الشركة تشكيل لجنة مؤلفة من 5 مهندسين بصورة عشوائية. فإذا كان المتغير العشوائي X يدل على عدد مهندسي الطاقة الكهربائية في المجموعة المختارة فالمطلوب:

1- أكتب جدول التوزيع الاحتمالي للمتغير العشوائي X.

2- أحسب التوقع الرياضي والتباين والانحراف المعياري للمتغير العشوائي X.

8 – ليكن X متغيراً عشوائياً منقطعا له جدول التوزيع الاحتمالي التالي:

x	0	1	2	3
$p(X=x)$	0.1	0.35	0.4	0.15

والمطلوب: أحسب التوقع الرياضي للمتغير العشوائي X. ثم أحسب التباين والانحراف المعياري لهذا المتغير.

9 – رمينا 5 قطع نقدية معا و المطلوب:

– احسب احتمال حصولنا على صورتين فقط.

– احسب احتمال حصولنا على صورتين على الأقل.

– احسب توقع عدد الصور الممكنة الظهور و تباينه.

10 – ليكن X متغيراً عشوائياً مستمراً دالة كثافته معرفة بالشكل:

$$f(x) = 6(1-x)x \quad , \quad 0 < X < 1$$

والمطلوب:

1 – أثبت أن $f(x)$ دالة كثافة احتمالية.

2– أوجد دالة التوزيع المجمعة.

3 – أوجد التوقع الرياضي والتباين والانحراف المعياري لهذا المتغير.

11 – إذا كان احتمال إصابة شخص بأعراض جانبية من تناول عقار هو 0.001 فإذا تناول هذا العقار في أحد الأيام 5000 شخص احسب ما يلي:

1 – احتمال إصابة شخصين بأعراض جانبية.

2– احتمال إصابة 5 أشخاص على الأقل .

3– احتمال إصابة شخصين على الأكثر بأعراض جانبية.

4 – توقع عدد المصابين.

12 – يحوي أحد الفصول 20 طالب وطالبة بينهم 5 طالبات ،اخترنا بشكل عشوائي بطريقة القرعة عينة مؤلفة من 6 أشخاص .والمطلوب:

1 –احسب احتمال ان تحوي العينة طالبتين .

2– احسب احتمال ان تحوي العينة طالبتين على الاقل.

13– إذا علمت أن توزيع X هو توزيع معتدل طبيعي متوسطه $\mu = 12$ وانحرافه المعياري $\sigma = 2$ فاحسب مستخدماً الجداول كلاً من الاحتمالات التالية:

$$P(10 < X < 13) \quad , \quad P(X < 10) \quad , \quad P(X > 14)$$

$P(16 < X < 20)$, $P(X < 10.5)$, $P(X > 11)$

14 – أنتج مصنع للأدوية دواءً جديداً لمعالجة أحد الأمراض، معدل نجاحه 80%، أعطي هذا الدواء لـ 15 مريضاً بهذا المرض. ما احتمال شفاء 12 منهم على الأقل.

15 – آلة لإنتاج المسامير لوحظ أنها تنتج مسامير تالفة بنسبة 20%. اخترنا عشوائياً 10 مسامير من المسامير المنتجة من هذه الآلة والمطلوب.

أ– ما احتمال وجود مسمارين تالفين فقط بين هذه المسامير؟

ب– ما احتمال وجود ثلاثة مسامير تالفة على الأكثر؟

16 _ من عينة بحجم 100 دجاجة مأخوذة بشكل عشوائي من إنتاج إحدى المزارع تبين أن متوسط وزن الدجاجة هو 1500 غرام بانحراف معياري 100 غرام و المطلوب حساب احتمال وزن الدجاجة في العينة في كل من الحالات التالية:

1 – $P(X\leq1500)$ 2 – $P(1600\leq X\leq1700)$

3 – $P(X<1400)$ 4 – $P(X>1600)$

5 – $P(1300\leq X\leq1700)$ 6 – $P(X\geq1700)$

مراجع الفصل السابع :

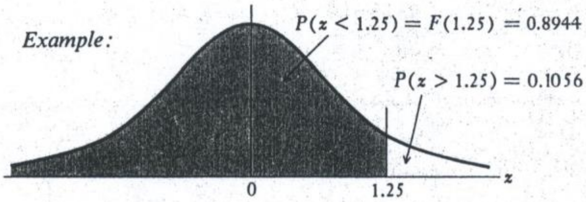
1- حميدان، عدنان؛ الجاعوني، فريد؛ ناصر آغا، عمار؛ العواد، منذر، 2003، مبادئ الإحصاء، منشورات جامعة دمشق.

2- حسام ،كمرجي،2011 - الاحتمال والاحصاء - جامعة دمشق – كلية الهندسة الميكانيكية والكهربائية .

3- Groebner David F., Shannon Patrick W. and Fry Phillip C. (2017)," Business Statistics :A Decision-Making approach", 10 th Edition. person.

Normal distribution

$$F(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz \quad \textit{Example:}$$



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

الفصل الثامن

الارتباط والانحدار

المخرجات والأهداف التعليمية:

بعد الانتهاء من هذا الفصل ،سيكون الطالب قادرا على:

اكتشاف العلاقات بين الظواهر والمتغيرات .

على قياس متانة العلاقة وتحديد نوعها وطبيعتها.

تمثيل العلقة بين المتغيرات بمعادلة رياضية .

التنبؤ بقيم المتغيرات .

تم في هذا الفصل دراسة الظواهر الاحصائية من خلال العلاقات التي تربط بينها. فتم توضيح مفهوم الارتباط بين متغرين أو أكثر وتحديد نوع هذا الإرتباط ودرجة قوته من خلال الشكل الانتشاري الممثل لنقاط المتغيرين أو من خلال بعض المقاييس المهمة كمعامل بيرسون ومعامل سبيرمان ومعامل الإقتران حيث تم التميز بين حساب معامل الارتبط للمتغيرات الكمية وحسابه للمتغيرات النوعية. بعد ذلك تم تحديد المعادلة التقديرية الممثلة للارتباط بين متغرين وهو مايسمى بانحدار أحد المتغيرين على الأخر حيث تم التمييز في الحصول على خط الانحدار بين الارتباط البسيط والمتعدد وبين الارتباط الخطي وغير الخطي. وتم التركيز على حالة الارتباط الخطي البسيط.

1.8- تمهيد:

سنطلع في هذا الفصل على اساليب تحليل جديدة نستطيع باستخدامها دراسة متغير ما من خلال علاقته بمتغير آخر او بعدة متغيرات أخرى، حيث يصادفنا الكثير من الحالات والظواهر التي نحتاج فيها إلى دراسة العلاقة بينها وايجاد نوع هذه العلاقة وتحديد مستوى قوتها. فمثلا قد نحتاج لدراسة العلاقة بين دخل الاسرة و انفاقها ، أو دراسة العلاقة بين دخل الاسرة و مدخراتها أو العلاقة بين أرباح مصنع من جهة و عدد العمال و الألات المتوفرة من جهة أخرى وكذلك العلاقة بين درجات الرياضيات والفيزياء لنفس الطلاب و العلاقة بين لون بشرة الانسان ولون العيون ...إلخ . حيث نستطيع من خلال دراسة العلاقة بين قيم متغيرين أو أكثر وتحليل النتائج، تحديد وجود علاقة أم لا و متانة هذه العلاقة بين المتغيرات في حال وجودها. وهذا الاسلوب يدعى تحليل الارتباط (Correlation) كما نستطيع معرفة تأثير أحد المتغيرين على الآخر أو تأثير عدة متغيرات على متغير واحد وذلك من خلال دراسة الانحدار (Regression) .

بعد التأكد من وجود العلاقة بين المتغيرات يجب تحديد أي المتغيرات يؤثر بالآخر و هو ماسوف ندعوه بالمتغير المستقل (Independent Variable) و نرمز له بالرمز X، وأي المتغيرات الذي يتأثر بالآخر وسوف ندعوه بالمتغير التابع (Dependent Variable) و نرمز له بالرمز Y.

بعد القيام بالإجراءات السابقة يتم حساب الارتباط و الانحدار للظواهر مع العلم بأنه يمكن دراسة علاقة متغير مستقل واحد مع متغير تابع واحد و هو ما يسمى بالارتباط والانحدار البسيط ، كذلك يمكن دراسة علاقة عدة متغيرات مستقلة بمتغير تابع واحد وهو ما يدعى بالارتباط والانحدار المتعدد مثل علاقة انفاق الأسرة كمتغير تابع لدخلها و عدد أفرادها و مكان إقامتها و درجة تعلم أفرادها ...كمتغيرات مستقلة.

2.8- الارتباط (Correlation:)

عند وجود ظاهرتين أو أكثر نبدأ أولا بدراسة وجود علاقة بين هذه الظواهر أم لا؟. فإذا وجد علاقة نحدد نوعها ودرجة قوتها، حيث تتميز العلاقة بين الظواهر أو المتغيرات من حيث كونها طردية (عندما تزداد قيم Y بزيادة قيم X) أو عكسية (عندما تتناقص قيم Y بتزايد قيم X) كما تصنف العلاقات إلى خطية (Linear)

وهي التي يمكن تمثيلها بخط مستقيم وغير خطية (منحنية) (Nonlinear) و تمثل بمنحني غير الخط المستقيم.

وسوف نقوم بدراسة الارتباط حسب الترتيب التالي:

-الارتباط الخطي البسيط للمتغيرات الكمية.

-الارتباط للمتغيرات النوعية .

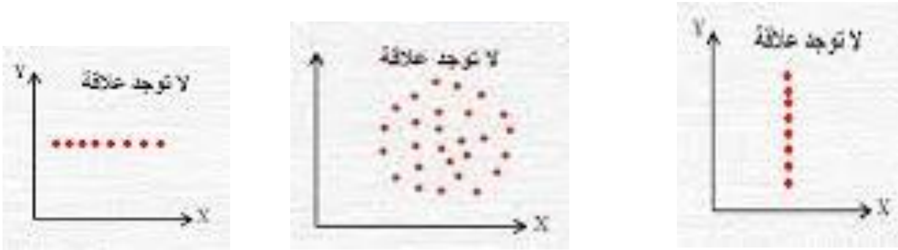
1.2.8 – الارتباط البسيط للمتغيرات الكمية:

بعد التأكد من وجود علاقة بين المتغيرين X و Y و تحديد المتغير المستقل منها وكذلك المتغير التابع ، نقوم بجمع البيانات عن الظاهرتين ومن خلال هذه البيانات نستطيع ان نحدد متانة هذه العلاقة وطبيعتها حيث يوجد طريقتين إما بالاستفادة من الشكل الانتشاري للبيانات أو عن طريق حساب معامل بيرسون للارتباط:

1- الشكل الانتشاري **(Scatter Diagram):**

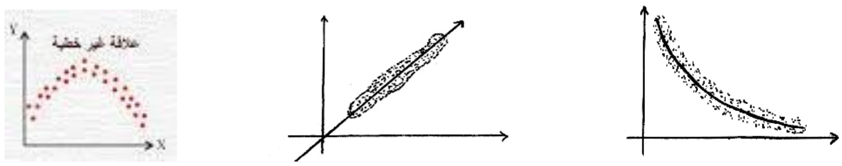
يقصد بالشكل الانتشاري تمثيل البيانات الممثلة للظاهرتين بنقاط على المحاور الإحداثية حيث نمثل قيم المتغير المستقل X على محور السينات وقيم المتغير التابع Y على محور العينات ونمثل كل ثنائية (x, y) بنقطة في المستوى الاحداثي، وينتج عن ذلك عددا كبيرا من النقاط عددها يساوي عدد المشاهدات المأخوذة من المتغيرين، وندعو هذا التمثيل بالشكل الانتشاري.

فمثلاً قد لا يوجد علاقة بين المتغيرينX و Y وذلك عندما تكون النقاط الممثلة للمتغيرين مبعثرة بشكل كبير بحيث لا تمثل أي شكل محدد، أو أن تكون النقاط منظمة بشكل عامودي على أحد المحورين فيكون أحد المتغيرين ثابت دوما وهذا يعني عدم وجود علاقة بين المتغيرين في هذه الحالة أيضاً كما في الأشكال التالية:



الشكل (8.1) – حالات عدم وجود علاقة بين متغيرين

وقد يوجد علاقة بين المتغيرين X و Y حيث تكون إما طردية أو عكسية، وكذلك خطية أو غير خطية كما في الأشكال التالية:



الانتشار يوضح وجود علاقة عكسية و غير خطية بين قيم المتغيرين

الانتشار يوضح وجود علاقة طرية وخطية بين قيم المتغيرين يمكن تمثيلها بمعادلة مستقيم

الشكل (8.2) – بعض أنواع العلاقات بين متغيرين

مما سبق نلاحظ أنه بإمكاننا تحديد نوع العلاقة بين متغيرين (طردية أو عكسية وخطية أو غير خطية) من خلال الشكل الانتشاري كما نستطيع تحديد قوة العلاقة من خلال درجة اقتراب النقاط المبعثرة من المستقيم او المنحني التقريبي الممثل لها.

مثال(1):

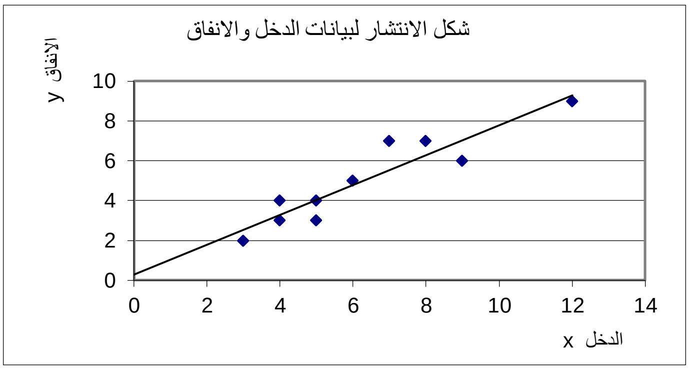
لدينا البيانات التالية عن الدخل الشهري و الإنفاق الشهري لعينة مؤلفة من 10 أسر.

الاسرة	1	2	3	4	5	6	7	8	9	10
الدخل (الف وحدة)	4	5	3	6	7	8	9	4	5	12
الانفاق (الف وحدة)	3	4	2	5	7	7	6	4	3	9

والمطلوب رسم الشكل الانتشاري وتحديد نوع العلاقة بين المتغرين X (الدخل) وy(الانفاق) ودرجة قوتها من خلال الرسم.

الحل:

إن الشكل الانتشاري الموافق للنقاط المبينة بالجدول السابق هو:



الشكل (8.3)

حيث نلاحظ من الشكل بأن القيم الدنيا للدخل تترافق مع القيم الدنيا للإنفاق و القيم العليا مع العليا مما يبين أنه يوجد علاقة طردية بين المتغيرين، كما نلاحظ بأن النقاط متجمعة بحيث يمكن رسم مستقيم يمر من بعضها وقريب من البعض الأخر ومنه يمكن القول بأنه يوجد بين المتغرين علاقة خطية طردية وقوية.

2 _ معامل الارتباط الخطي لبيرسون **(Coefficient of Correlation):**

يساعد معامل الارتباط لبيرسون (Pearson) **في تحديد درجة قوة العلاقة بين المتغيرين** Y , X **وتحديد جهة هذه العلاقة (طردية أم عكسية) عندما يوجد علاقة خطية بين المتغيرين. تعطى قيمة هذا المعامل بالعلاقة:**

$$R = \frac{S_{XY}}{S_X S_Y} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}\ \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i -}} \qquad (8.1)$$

حيث $\bar{y}, \bar{x}$ المتوسطين الحسابيين للمتغيرين المستقل و التابع على التتالي.

S_Y, S_X الانحرافات المعيارية للمتغير المستقل و التابع و n عدد المشاهدات.

و يمكن إصلاح العلاقة السابقة و كتابتها بالأشكال التالية:

$$R = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\ \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \qquad (8.2)$$

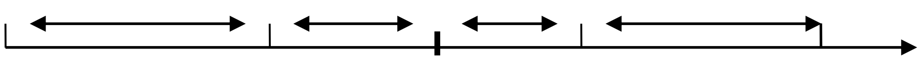
$$R = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{\sqrt{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}(\sum_{i=1}^{n} x_i)^2}\ \sqrt{\sum_{i=1}^{n} y_i^2 - \frac{1}{n}(\sum_{i=1}^{n} y_i)^2}} \qquad (8.3)$$

حيث تعتبر العلاقة الأخيرة (8.3) الأكثر سهولة لحساب معامل الارتباط لبيرسون.

ملاحظة(1):

يسمى المعامل R باسم معامل بيرسون للارتباط البسيط الخطي والذي تنحصر قيمته ضمن المجال: $-1 \le \rho \le +1$ وعلى ضوء قيم هذا المعامل يتم تصنيف الارتباط بين متغيري العينة Y, X ومن ثم بين المتغيرين في المجتمع الإحصائي بكامله كما في الشكل التالي:

ρ=-1



ة تامة قوية موجبا | علاقة ضعيفة موجبة | 0 قة معدومة | علاقة قوية سالبة | تامة قوية سالب

حيث يعبّر عن الارتباط بأنه قوي ايجابي عندما يقترب المعامل R من الواحد، وبأنـه قوي سلبي عندما يقترب المعامل R من 1-، وبأنه ضعيف عندما يقترب المعامل من الصفر.

ملاحظة(2):

يستخدم معامل بيرسون للارتباط الخطي البسيط لحساب قوة أو درجة الارتباط في العينة ومن ثم تقدير قوة أو درجة الارتباط في المجتمع الإحصائي بغض النظر عن معادلة الانحدار بين المتغيرين وحتى من دون اللجوء إلى حسابها.

ملاحظة(3):

إنّ وجود ارتباط بين متغيرين عشوائيين مثل Y,X لا يعني على الإطلاق بأنّ أحدهما سبب أو علة للآخر لأننا وببساطة يمكن أن نرتب بعضاً من الأرقام ليكون معامل الارتباط قوياً في حين أنّ الأرقام المرتبة لا تمثل أي سببية أو عليَّة لأحد المتغيرين على الآخر.

مثال(2):

أحسب معامل بيرسون للارتباط الخطي بين المتغيرين Y,X المبينين بالجدول التالي:

x	1	3	4	6	8	10
y	2	6	7	10	15	20

الحل:

نكوّن الجدول التالي:

x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	2	2	1	4
3	6	18	9	36
4	7	28	16	49
6	10	60	36	100
8	15	120	64	225
10	20	200	100	400

32	60	428	226	814

وبتطبيق العلاقة (8.3) نجد أنّ:

$$R = \frac{428 - \frac{1}{6}(32)(60)}{\sqrt{226 - \frac{1}{6}(32)^2}\sqrt{814 - \frac{1}{6}(60)^2}} = 0.992$$

حيث نجد بأنّ الارتباط قوي و طردي (إيجابي) بين المتغيرين.

– مزايا و مساوئ معامل ارتباط بيرسون:

من مزايا معامل الارتباط لبيرسون أنه يحدد نوع العلاقة طردية أو عكسية من خلال إشارة المعامل موجبة أو سالبة على الترتيب، كما يحدد قوة العلاقة بين المتغرين من خلال القيمة المطلقة للمعامل. ومن المزايا أيضا أن قيمة معامل الارتباط مجردة من وحدات القياس، حيث يتم التخلص من وحدات القياس عند التقسيم على الانحرافات المعيارية .

أما أهم مساوئ معامل الارتباط لبيرسون فهو أنه لايعبر عن متانة العلاقة بشكل صحيح إلا اذا كانت العلاقة خطية، وعندما تكون العلاقة منحنية فإن قيمته لاتعبر عن متانة العلاقة. كما أنه لايمكن استخدام معامل بيرسون في حالة المتغيرات النوعية ويتم عندئذ استخدام مقاييس أخرى.

2.2.8 – معامل سبيرمان لارتباط الرتب (Rank Correlation):

يُستخدم مفهوم الترتيب في الحالات غير الكمية التي تشمل الصفات المعنوية حيث تكون الرتب بديلاً كمياً يمكن التعامل معه. ويسمى معامل الارتباط لهذه الحالات باسم معامل ارتباط الرتب أو معامل سبيرمان والذي يُعرّف بالعلاقة التالية:

$$\boxed{r_s = 1 - \frac{6\sum_{i=1}^{n}(R_{x_i} - R_{y_i})^2}{n(n^2-1)}} \qquad (8.4)$$

حيث: R_{x_i} **يمثل رتبة المتغير** x_i **و** R_{y_i} **يمثل رتبة المتغير** y_i

n **يمثل عدد الأزواج المرتبة** (x, y)

ويتلخص أسلوب حساب معامل ارتباط سبيرمان بترتيب الصفات بصورة متسلسلة حسب معيار معين تصاعدياً أو تنازلياً ومن ثم يتم إعطاء الرتب وفقاً للتسلسل المذكور. ومن الرتب المتناظرة يمكننا حساب الفروق بينها $R_{x_i} - R_{y_i}$ ولكافة الرتب.

ملاحظة**(4): عندما تتكرر صفة من الصفات فإنّ الرتبة التي تُعتمد في مثل هذه الحالة هي متوسط التسلسل لهذه الصفة (أي مجموع التسلسل للصفة المتكررة على عدد تكرارها).**

فلو افترضنا أنّ الصفة المكررة (جـ) أخذت التسلسلات التالية 6 ، 9 ، 12 أي تكررت ثلاث مرّات فإنّ رتبة الصفة (جـ) هي $R = \frac{27}{3} = 9$ للحالات المكررة الثلاث.

مثال**(3):**

في أحد المنافسات تم استعراض 10 لوحات فنية لتقييمها من قبل لجنتي تحكيم فإذا كانت الرتب الممنوحة للوحات العشرة من قبل اللجنتين هي كالتالي:

اللجنة الأولى	1	2	3	4	5	6	7	8	9	10
اللجنة الثانية	4	6	1	2	3	7	8	10	5	9

فأحسب معامل سبيرمان للارتباط وحدد نوعيته.

الحل**: نرتب جدول العناصر لمعامل سبيرمان كالتالي:**

اللوحات	r_1 لجنة أولى	r_2 لجنة ثانية	$r_1 - r_2$	$(r_1 - r_2)^2$
A	1	4	-3	9
B	2	6	-4	16
C	3	1	2	4
D	4	2	2	4
E	5	3	2	4
F	6	7	-1	1
G	7	8	-1	1
H	8	10	-2	4
I	9	5	4	16
J	10	9	1	1

المجموع = 60

ومنه فإنّ:

$$r = 1 - \frac{6\sum_{i=1}^{10}(r_1 - r_2)^2}{n(n^2-1)} = 1 - \frac{6(60)}{10\times 99} = 1 - \frac{60}{165} = \frac{105}{165} = 0.636$$

ويلاحظ بأنّ هنالك ارتباطاً قوياً ايجابياً في التقييم بين اللجنتين في مضمار العينة المأخوذة.

مثال **(4):**

في إحدى الدراسات الاجتماعية لمعرفة مقدار العلاقة بين الحالة العلمية لرب الأسرة وتأثيره على المستوى الاقتصادي لها استُحصلت البيانات التالية:

الة العلمية لرب الأسرة	ابتدائية	أمي	عالية	متوسطة	أمي	يقرأ ويكتب	أمي	يقرأ فقط	ثانوية	جامعية
تتوى الاقتصادي للأسرة	متوسط	معدمة	غنية	فقيرة	فقيرة	فقيرة	متوسط	فقيرة	سط الحال	غنية

فأحسب معامل الارتباط للصفات في الدراسة أعلاه.

الحل:

نُرتّب أولاً الحالات العلمية لرب الأسرة بصورة متسلسلة وتصاعدية وفقاً للدرجة العلمية ومن ثم نرتب المستويات الاقتصادية بصورة متسلسلة ومنفصلة عن الترتيب الأول لتكوين أزواج الرتب الثنائية (x_i, y_i) ومن ثم نوجد مجموع فروقها كما في الجدول التالي حيث تأخذ العينات نفس تسلسلها ما لم تكن مكررة حيث نلاحظ بأنّ رتبة الحالة العلمية (أمي) مكررة ثلاث مرّات وعليه فإنّ رتبتها تساوي مجموع التسلسلات على التكرار أي أنّ:

$$r\ (\text{أمي}) = \frac{1+2+3}{3} = 2$$

$$r\ (\text{فقيرة}) = \frac{2+3+4+5}{4} = \frac{14}{4} = 3.5$$

$$r\ (\text{متوسطة}) = \frac{6+7+8}{3} = \frac{21}{3} = 7$$

حيث نحصل على الجدول التالي:

التسلسل	الحالة العلمية	الرتبة	مستوى الاقتصادي للأسرة	الرتبة	روق d_i للرتب	الفروق للرتب d_i^2
1	أمي	2	معدمة (1)	1	+1	1

2	أمي	2	فقيرة (2)	3.5	-1.5	2.25
3	أمي	2	متوسط الحال (6)	7	-5	25
4	يقرأ فقط	4	فقيرة (3)	3.5	+0.5	0.25
5	يقرأ ويكتب	5	فقيرة (4)	3.5	+1.5	2.25
6	شهادة ابتدائية	6	متوسطة الحال (7)	7	-1	1
7	شهادة متوسطة	7	فقيرة (5)	3.5	3.5	12.25
8	شهادة ثانوية	8	متوسطة الحال (8)	7	+1	1
9	شهادة جامعية	9	غنية (9)	9.5	-0.5	0.25
10	شهادة عالية	10	غنية (10)	9.5	+0.5	0.25

ومنه نجد أنّ معامل سبيرمان للرتب هو:

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)} = 1 - \frac{273}{990} = 0.724$$

والقيمة الأخيرة تدل على وجود علاقة قوية موجبة (طردية) بين الحالة العلمية لرب الأسرة والمستوى الاقتصادي لها.

3.2.8 – الارتباط للمتغيرات النوعية:

راينا أعلاه بعض الأساليب التي نستطيع بمساعدتها الكشف عن متانة العلاقة الارتباطية للمتغيرات الكمية. وراينا ان معامل ارتباط الرتب يصلح لحساب متانة العلاقة الارتباطية للمتغيرات الكمية و النوعية ، سنرى فيما يلي أساليب مختلفة لدراسة ارتباط المتغيرات النوعية ، يستخدم كل منها لحالة معينة ومن هذه الأساليب معامل الاقتران ومعامل التوافق وسنكتفي هنا بتوضيح معامل الاقتران.

– معامل الاقتران:

يستخدم معامل الاقتران عندما يكون لدينا متغيرين نوعيين ولكل متغير قيمتين أو صفتين. كأن ندرس العلاقة بين لون الزهرة ورائحتها أو العلاقة بين الإصابة بالسرطان والتدخين لمجموعة أشخاص مدخنين وغير مدخنين، فالمتغيرين في المثال الثاني هما المرض بالسرطان وله صفتان مصاب وغير مصاب والثاني التدخين وله الصفتين مدخن وغير مدخن.

ولحساب معامل الاقتران نقسم المفردات حسب المتغيرات والصفات كما يلي:

المتغير A / المتغير B	القيمة 1	القيمة 2	المجموع
القيمة 1	n_{11}	n_{12}	$\tilde{n}_1$
القيمة 2	n_{21}	n_{22}	$\tilde{n}_2$
المجموع	n_1	n_2	N

حيث : $n_{11}+n_{12}+n_{21}+n_{22}=N$

ويحسب معامل الاقتران من الصيغة التالية:

$$\boxed{r_c=\frac{n_{11}n_{22}-n_{21}n_{12}}{n_{11}n_{22}+n_{21}n_{12}}} \quad (8.5)$$

حيث تكون العلاقة بين المتغيرات قوية جدا كلما اقتربت قيمة معامل الاقتران من الواحد وبشكل عام نقول انه هناك علاقة عندما تكون قيمة r_c اكبر من 0.5.

مثال**(5):**

قام أحد الباحثين بدراسة العلاقة بين التدخين والاصابة بالتهاب اللثة، فجمع بيانات عن 200 شخص وكان توزيعهم حسب الصفات كما يلي:

المتغير A / المتغير B	مدخن	غير مدخن	المجموع

مصاب	70	10	80
غير مصاب	20	100	120
المجموع	90	110	200

لنحسب معامل الاقتران حسب الصيغة (8.5) فنجد أن:

$$r_c = \frac{n_{11}n_{22} - n_{21}n_{12}}{n_{11}n_{22} + n_{21}n_{12}} = \frac{70(100) - 20(10)}{70(100) + 20(10)} = \frac{6800}{7200} = 0.944$$

وهذا يدل على وجود علاقة قوية جدا بين التدخين ومرض التهاب اللثة.

ونشير أخيرا بأنه ليس هناك أي مدلول لإشارة معامل الاقتران وذلك لأن إشارته تختلف حسب ترتيب تكرار الصفات في الجدول، لذلك نأخذ قيمته المطلقة للدلالة على وجود العلاقة من عدمها (لاحظ بانه لو بدلنا السطر الأول بالثاني بالجدول السابق لأصبحت قيمة معامل الاقتران 0.944 –).

8.3– الانحدار (Regression:)

رأينا في الفقرة السابقة كيف يمكن تحديد وجود علاقة بين المتغيرات من عدمها، وكيف نستطيع معرفة طبيعة هذه العلاقة ومتانتها. ولكن لم نتوصل إلى تحديد العلاقة التي تربط المتغير التابع بالمتغير المستقل والتي نستطيع من خلالها التكهن بقيمة المتغير التابع من أجل قيمة ما للمتغير المستقل. إن عملية ايجاد هذه العلاقة تسمى بالانحدار.

يمكن تقسيم الانحدار الى الأنواع التالية وذلك حسب عدد المتغيرات المستقلة الداخلة بالمعادلة وحسب نوع معادلة التمثيل:

–الانحدار البسيط:

و هو دراسة العلاقة بين متغيرين أحدهما مستقل و الآخر تابع، حيث يقسم إلى قسمين:

1– الانحدار الخطي البسيط (Simple Linear Regression): وذلك عندما يمكن تمثيل العلاقة بين المتغيرين بخط مستقيم.

2- الانحدار غير الخطي البسيط(Simple Non_linear Regression): وذلك عندما لا يمكن تمثيل العلاقة بين المتغيرين بمستقيم وإنما بمنحني من الدرجة ثانية أو الثالثة أو منحني لدالة أسية أو لوغاريتمية ...إلخ وذلك حسب توزع النقاط.

-الانحدار المتعدد(**Multiple Regression**):

وهو دراسة العلاقة بين عدة متغيرات مستقلة و متغير واحد تابع و يقسم بدوره الى قسمين:

-الانحدار المتعدد الخطي(Multiple Linear Regression):

وذلك عندما يمكن تمثيل العلاقة بين المتغيرات المستقلة والمتغير التابع بمعادلة خطية.

2-الانحدار المتعدد اللاخطي(Multiple Nonlinear Regression):

ويكون بتمثيل العلاقة بين المتغيرات المستقلة والتابع بعلاقة غير خطية.

وسوف نكتفي هنا بدراسة الانحدار الخطي البسيط.

1.3.8_الانحدار الخطي البسيط:

من الشكل الانتشاري لقيم المتغيرين يمكننا تمثيل علاقة الانحدار الخطي البسيط بالنموذج الرياضي التالي الذي يمثل معادلة الخط المستقيم:

$$\boxed{\hat{y} = a + bx} \qquad (8.6)$$

حيث يمثل الرمز a نقطة تقاطع المستقيم مع المحور y بينما يمثل الرمز b ميل الخط المستقيم أو معامل انحدار النموذج الخطي.

ويتلخص أسلوب حساب كل من المعاملين a و b باستخدام ما يسمى بطريقة المربعات الصغرى التي تعتمد على فرضية الحصول على أصغر قيمة لمربع الخطأ بين القيمة على النموذج المذكور $\hat{y}$ وبين القيمة الواقعية y ولكافة عناصر العينة ويعبر عن ذلك بعبارة رياضية كالتالي:

$$E_i^2 = (y_i - \hat{y}_i)^2 \qquad , \qquad i = 1,\ldots\ldots n$$

حيث E_i **يمثل مقدار الخطأ الذي تمّ تربيعه للتخلص من القيمة السالبة للخطأ وبالجمع لكافة قيم i نجد أنّ:**

$$G = \sum_{i=1}^{n} E_i^2 = \sum_{i=1}^{n} (y_i - \widehat{y_i})^2 = \sum_{i=1}^{n} (y_i - (a + bx_i))^2$$

وباستخدام التفاضل نجد أنّ القيمة الصغرى تحدث عندما تكون المشتقات الجزئية للدالة G بالنسبة للمتغيرين a **و** b **مساوية للصفر:**

$$\frac{\partial G}{\partial a} = 2\sum_{i=1}^{n} \big(y_i - (a + bx_i)\big)(-1) \equiv 0$$

$$\frac{\partial G}{\partial b} = 2\sum_{i=1}^{n} \big(y_i - (a + bx_i)\big)(-x_i) \equiv 0$$

وبحل المعادلتين الخطيتين الأخيرتين بالنسبة للمجهولين a **و** b **والإصلاح نجد أن:**

$$\boxed{b = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n} (x_i - \overline{x})^2}} \qquad (8.7)$$

$$\boxed{a = \bar{y} - b\bar{x}} \qquad (8.8)$$

حيث نحصل بعد حساب a **و** b **من العلاقتين السابقتين على معادلة الانحدار الخطي** $\hat{y} = a + bx$ **المطلوبة.**

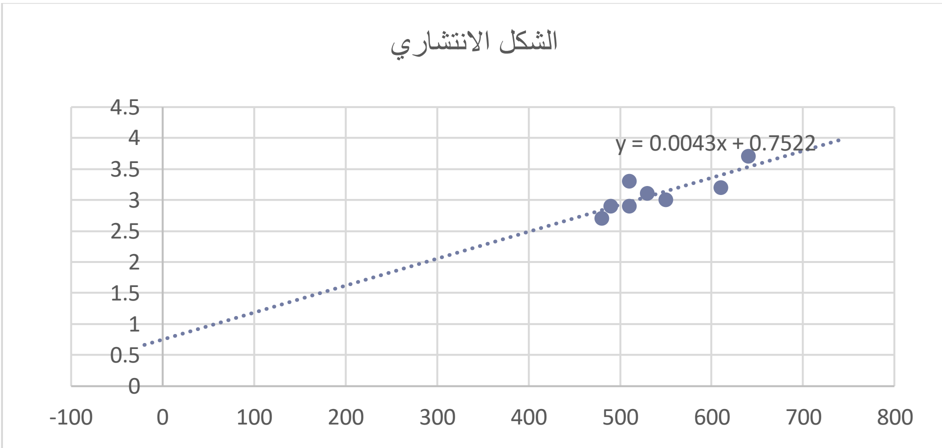
مثال **(6):**

الجدول التالي يوضح مجاميع درجات ثمانية طلاب ونقاط التقدير لهم. أحسب معادلة انحدار نقاط التقدير على مجاميع الطلبة (إن وجدت).

المجاميع x_i	480	490	510	510	530	550	610	640
التقديرات y_i	2.70	2.90	3.30	2.90	3.10	3.00	3.20	3.70

الحل:

نرسم أولاً الشكل الانتشاري على مستوي المحورين x،y بعد تدريجهما.



الشكل (8.4)

يلاحظ من الشكل الانتشاري بأنّ النقاط تتوزع بصورة تقريبية على شكل خط مستقيم لذا نستنتج بأنّ هنالك علاقة خطية بين التقديرات ومجاميع الطلبة يمكن تمثيلها بالنموذج الخطي $\hat{y} = a + bx$ ولحساب كل من المعاملين a و b نحسب أولاً b من العلاقة (8.7) فنكوّن الجدول التالي: حيث $\bar{x} = 540$, $\bar{y} = 3.10$

x_i المجاميع	y_i التقديرات	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$[(x_i - \bar{x})(y_i - \bar{y})]$	$(x_i - \bar{x})^2$
480	2.70	-60	-0.40	24	3600
490	2.90	-50	-0.20	10	2500
510	3.30	-30	+0.20	-6	900
510	2.90	-30	-0.20	+6	900
350	3.10	-10	0.00	0	100
550	3.00	+10	-0.10	-1	100
610	3.20	+70	+0.10	+7	4900
640	3.70	+100	+0.60	+60	10000
4320	24.80	0	0	100	23000

ومن الجدول نجد أنّ:

$$b = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{100}{23000} = 0.00435$$

وكذلك نجد أنّ:

$$a = \bar{y} - b\bar{x} = 3.10 - 0.00435(540) = 0.751$$

وبالتالي فإنّ معادلة الانحدار تأخذ الصورة الخطية التالية: $\hat{y} = 0.751 + 0.00435\,x$

ملاحظة**(5):**

يمكن استخدام نفس الأسلوب عكسياً لإيجاد معادلة انحدار المجاميع على التقديرات أي معادلة انحدار المتغير X على المتغير Y.

ملاحظة**(6):**

تسمى القيم أو النقاط الواقعة على الخط باسم القيم التقديرية $\hat{y}$ (Estimated Value) أو القيم المخمنة بينما تسمى القيم المعطاة باسم القيم المشاهدة .observed V أو القيم الواقعية y. وعلى هذا الأساس يمكن حساب مقدار الخطأ بين القيمة التقديرية والقيمة الواقعية والذي يعرف بالعلاقة التالية:

$$E_i = y_i - \hat{y}_i$$

وبالعودة للمثال السابق يمكن تشكيل جدول الخطأ التالي:

			الخطأ المفسر	خطأ غير المفسر خطأ غير المقبول
x_i	y_i	$\hat{y}_i = 0.751 + 0.00435\ x_i$	$EE_i = (\hat{y}_i - \overline{y})$	$NE_i = y_i - \hat{y}_i$
480	2.70	2.839	-0.261	-0.139
490	2.90	2.883	-0.217	0.017
510	3.30	2.970	-0.120	0.330
510	2.90	2.970	-0.120	-0.070
530	3.10	3.057	-0.043	0.0431
550	3.00	3.144	+0.044	-0.144
610	3.20	3.405	+0.305	-0.205
640	3.70	3.535	+0.435	0.165

4320	24.80	24.80	0.023	-0.003

ويستخدم جدول الخطأ السابق لحساب الخطأ الكلي TE حيث $TE_i = EE_i + NE_i$ ومنها نجد أن مجموع مربعات الخطأ يساوي مجموع مربعات الخطأ المفسر + مجموع مربعات الخطأ غير المفسر أي أن:

أو: $SE_i^2 = SEE_i^2 + SNE_i^2$

$$\sum_{i=1}^{n}(y_i-\bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2 + \sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \qquad (8.9)$$

حيث يكون الانحراف المعياري للخطأ: $\sigma_E = \sqrt{\frac{SE_i^2}{n-2}} = \sqrt{\frac{\sum(y_i-\hat{y}_i)^2}{n-2}}$

وبقسمة المعادلة (8.9) على مجموع مربعات الخطأ الكلي نجد أنّ:

$$1.00 \equiv \frac{SE_i^2}{SE_i^2} = \frac{SEE_i^2}{SE_i^2} + \frac{SNE_i^2}{SE_i^2}$$

حيث يُطلق على النسبة الأولى أسم مربع معامل الارتباط والذي يرمز له بالرمز R^2أي أنّ:

$$R^2 = \frac{SEE_i^2}{SE_i^2} \Rightarrow R = \sqrt{\frac{SEE_i^2}{SE_i^2}}$$

حيث نجد بأنّ: $0 \le R^2 \le 1$.

أخيرا نذكر بأهمية معادلة مستقيم الانحدار المعطى بالعلاقة (8.5) حيث نستطيع من خلال هذه المعادلة التنبؤ بقيم المتغير التابع عندما ياخذ المتغير المستقل قيما محددة، فمثلا في المثال(6) السابق وبعد الوصول لمعادلة خط الانحدار لنفرض بأننا نريد المعدل لطالب مجموع درجاته 540 عندئذ نجد أن:

$$\hat{y} = 0.751 + 0.00435\,x$$

$$\hat{y} = 0.751 + 0.00435(540) = 3.1$$

4.8− معامل التحديد:

يعرف معامل التحديد بأنه القيمة الناتجة عن مربع قيمة معامل الارتباط ويرمز

له بالرمز R^2 حيث R هو معامل الارتباط لبيرسون، و يمثل النسبة المئوية لوثوقية تحقق نموذج الارتباط بين متغيرين.

ولمعامل التحديد أهمية كبيرة حيث يعبر عن مقدار جودة العلاقة بين متغيرين وتكون قيمه محصورة بين الصفر والواحد. فكلما اقتربت قيمته من الواحد كلما زادت جودة النموذج في تفسير العلاقة بين المتغيرين وقلت الحالات الشاذة. وتكون قيمة معامل التحديد مساوية للواحد إذا كانت العلاقة تامه وعندها تقع جميع النقاط على خط الانحدار ويأخذ المعامل القيمة صفر إذا لم يوجد علاقة.

مما سبق نستنتج بأن معامل التحديد هو مقدار موجب دوما. كما نستنتج بأنه يمكن الحصول على معامل الارتباط من معامل التحديد حيث نجد قيمة معامل الارتباط من العلاقة

$$R = \sqrt{R^2}$$

و نحدد جهة الارتباط (إشارة الارتباط) من إشارة الميل في علاقة الانحدار.

مثال(7):

احسب معامل التحديد وحدد مجال الثقة لعلاقة الارتباط بين المتغيرين المبينين بالمثال(2).

الحل:

وجدنا في المثال(2) بأن معامل الارتباط الممثل للعلاقة بين المتغيرين هو:

$$R = 0.992$$

ومنه نجد بأن معامل التحديد هو:

$$R^2 = 0.992^2 = 0.984 \cong 0.98$$

أي أننا على ثقة مقدارها 98 بالمئة بتحقق العلاقة بين المتغيرين وبأن نسبة الحالات الشاذة عن ذلك هي 2 بالمئة.

مثال(8):

احسب معامل الارتباط لعلاقة بين متغيرين إذا علمت أن معامل التحديد بينهما هو:

$$R^2 = 0.868$$

وأن معادلة خط الانحدار بين المتغيرين هو:

$$\hat{y} = a + bx = 0.2819 + 0.7489x$$

الحل:

إن قيمة معامل الارتباط هي:

$$R = \sqrt{R^2} = \sqrt{0.868} = 0.9317$$

كما أن إشارته هي الإشارة الموجبة لأن الميل في معادلة الانحدار موجب، حيث:

$$b = 0.7489$$

ومنه لدينا علاقة ارتباط طردية وقوية.

تمارين الفصل الثامن

1 – بين ماهي طبيعة العلاقة و نوع المتغيرات في العلاقات التالية:

1 – دخل و نفقات المستهلك.

2 – عدد الحوادث وعدد ساعات التدريب على القيادة.

3– درجات مجموعة طلاب في الرياضيات والفيزياء.

4 – دخل ومدخرات الاسرة.

5– نفقات الاسرة و عدد أفرادها.

6 – حجم الانتاج و النفقات الكلية.

7 – حجم الانتاج و كلفة الوحدة الواحدة من الانتاج.

8 – حجم الطلب على سلعة ما و سعرها.

2– البيانات التالية تمثل الوزن x والطول y لاثني عشر شخصاً والمطلوب:

أ– أرسم الشكل الانتشاري واحسب معادلة انحدار Y على X.

ب– أحسب قيمة معامل الارتباط لبيرسون.

3 – قدر متانة 247 | 377 | 334 | 322 | 313 | 354 | 308 | 340 | 301 | 368 | 349 | 293 | T.S(y)

العلاقة بين

المتغيرين X وY و نوعها من خلال رسم الشكل الانتشاري للمسائل التالية

(1) –

x	**2**	**3**	**5**	**8**	**12**	**13**	**15**	**16**	**17**	20
y	5	6	8	11	15	16	18	19	20	23

(2)–

x	**1**	**2**	**4**	**5**	**7**	**8**	**10**	**11**	**13**	**15**
y	3	6	18	27	51	66	105	123	171	227

(3) –

x	**4**	**5**	**7**	**8**	**10**	**11**	**12**	**15**	**17**	**20**
y	**29**	**36**	**44**	**51**	**58**	**66**	**68**	**87**	**93**	**108**

4 – البيانات التالية توضح مقاومة الشد وصلابة الألمنيوم المسبوك. أرسم الشكل الانتشاري واحسب معادلة انحدار مقاومة الشد Y على صلابة الألمنيوم X:

H(x)	53	70	84	55	78	64	71	53	82	67	70	56

5 – لدينا البيانات التالية عن السماد المستخدم و الغلة لعينة عشوائية حجمها 100 قطعة أرض متشابهة الظروف.

الغلة / كمية السماد.	غير وفيرة	وفيرة	المجموع
عالية	15	55	70
قليلة	25	5	30
المجموع	40	60	100

والمطلوب أوجد قوة العلاقة بين كمية السماد والغلة مستخدما المقياس المناسب.

6 – احسب متانة العلاقة بين درجة نضج العنب و المسافة بين الاشجار في البيانات التالية التي تم الحصول عليها من إحدى التجارب مستخدما المقياس المناسب.

المسافة متر	3	4	4.5	5	6	7
درجة النضج	سر صغير (حصرم	ضر كبير (حصرم)	كبير منقط بالاحم	ر كبير منقط بالاحمر	ضر كبير محمر	احمر كبير

7 – البيانات التالية توضح مسافة التوقف وفقاً لسرعات مختلفة في تجربة على موديل جديد للسيارات.

V السرعة كيلومترات بالساعة	35	50	65	80	95	110

d مسافة التوقف بالأمتار	16	26	41	62	88	119

والمطلوب: أحسب معامل الارتباط لبيرسون.

8 – لدينا البيانات التالية عن النفقات الشهرية و الدخل الشهري لعينة عشوائية حجمها 10 مؤسسات انتاجية متشابهة:

المؤسسة	1	2	3	4	5	6	7	8	9	10
الدخل (ألف وحدة)	8	10	10	12	15	16	17	19	20	25
الانفاق(الف وحدة)	20	22	24	25	26	30	30	32	35	36

والمطلوب:

1 – أوجد معادلة مستقيم الانحدار الممثلة للعلاقة و فسر ثوابتها.

2 – احسب معامل التحديد و معامل الارتباط و فسرهما.

3– احسب الخطأ المعياري للتقدير.

4– قدر دخل مؤسسة تبلغ نفقاتها 23 ألف وحدة .

5– قدرة باحتمال 95.5 %دخل مؤسسة تبلغ نفقاتها 23 ألف وحدة.

9 – أحسب الميل ونقطة القطع مع المحور y واكتب معادلة انحدار المتغير Y على المتغير X من البيانات التالية:

y	3	19	18	22	23
x	-2	-1	0	1	2

10 – أحسب معادلة انحدار متوسط التكلفة y لإنتاج عدد من السلع x شهرياً باستخدام الجدول التالي:

متوسط التكلفة	3	9	40	15	33	45	20
كمية الإنتاج شهرياً	15	10	18	10	15	20	13

ثم أحسب جدول الخطأ وأوجد معامل التحديد.

11 – إذا كان خط الانحدار للإنفاق السنوي على المواد الغذائية لأسرة معينة يتمثل بالمعادلة الخطية التالية $y = 300 + 0.2x$ حيث x هو الدخل السنوي للأسرة بينما y مقدار الإنفاق السنوي والمطلوب:

أ – أرسم خط انحدار الإنفاق على الدخل للأسرة.

ب– ما هو إنفاق الأسرة (المصروف السنوي) إذا كان دخلها 30000 ليرة ، 16000 ليرة.

مراجع الفصل الثامن :

1- حميدان، عدنان؛ الجاعوني، فريد؛ ناصر آغا، عمار؛ العواد، منذر، 2003، مبادئ الإحصاء، منشورات جامعة دمشق.

2- حسام ،كمرجي،2011 - الاحتمال والاحصاء - جامعة دمشق – كلية الهندسة الميكانيكية والكهربائية .

3- Groebner David F., Shannon Patrick W. and Fry Phillip C. (2017)," Business Statistics :A Decision-Making approach", 10 th Edition. person.

الفصل التاسع :

السلاسل الزمنية Time series

المخرجات والأهداف التعليمية:

نهاء من هذا الفصل ،سيكون الطالب قادرا على:

عرض بيانات السلاسل الزمنية بيانيا.

التعرف على العوامل المؤثرة ببيانات السلسلة الزمنية.

تمثيل الاتجاه العام للسلسلة الزمنية .

التنبؤ بقيم الظاهرة الممثلة بسلسلة زمنية .

ملخص الفصل:

سنعرض في هذا الفصل تعريفاً للسلاسل الزمنية وأنواعها ونتعرف على العوامل المؤثرة في السلاسل الزمنية وأسبابها وكيفية تحليل السلاسل الزمنية من اجل التنبؤ بقيم السلاسل الزمنية

1.9- تمهيد

تعرف السلسلة الزمنية بانها بيانات عن ظاهرة ما في فترات زمنية متعاقبة ، كل سنة فتدعى السلسلة سنوية او كل شهر فتدعى شهرية او كل اسبوع ...... الخ . فعدد طلاب الجامعة خلال عدة سنوات او عدد سكان سورية تشكل سلسلة زمنية سنوية ،والانتاج الشهري لمعمل البرادات يشكل سلسلة زمنية شهرية ،والمبيعات اليومية لمتجر هي سلسلة يومية .... . تدعى كل قيمة للظاهرة والمقابلة لزمن معين بحد السلسلة .

2.9 – عرض السلاسل الزمنية

1.2.9 –العرض الجدولي للسلاسل الزمنية :

تعرض السلاسل الزمنية بجداول احصائية بسيطة حيث يعرض بالجدول سلسلة واحدة وعندها يتالف الجدول من سطرين اوعمودين احد الاسطر او الاعمدة للزمن والاخر لقيم الظاهرة . ويمكن عرض عدة سلاسل زمنية في الجدول الواحد، نضع في احد الاسطر او الاعمدة الزمن وفي الباقي حدود السلاسل الزمنية ويتضح من هذا بان السلاسل من الممكن ان تكون بالاسطر او في الاعمدة ونختار ذلك حسب حجم الجدول وشكل الورقة ،انظر الجدول (1.9) .اما ترتيب البيانات داخل الجدول فيجب ان يكون ترتيبا تاريخيا وعادة مايكون من الاقدم الى الاحدث حسب تطور التاريخ، وهناك اتجاه اخر بحيث نبدا من الاحدث ،ولانصار هذا الاتجاه حججهم باننا نهتم بالقيم الحديثة للظاهرة اكثر .

مثال:فيما يلي حجم الانتاج لاحد المعامل خلال الاعوام (2010-2017)

العام	2010	2011	2012	2013	2014	2015	2016	2017
حجم الانتاج(طن)	**20**	**25**	**17**	**18**	**25**	**30**	**30**	**40**

جدول (9-1) المصدر : فرضي

2.2.9 – العرض البياني للسلاسل الزمنية :

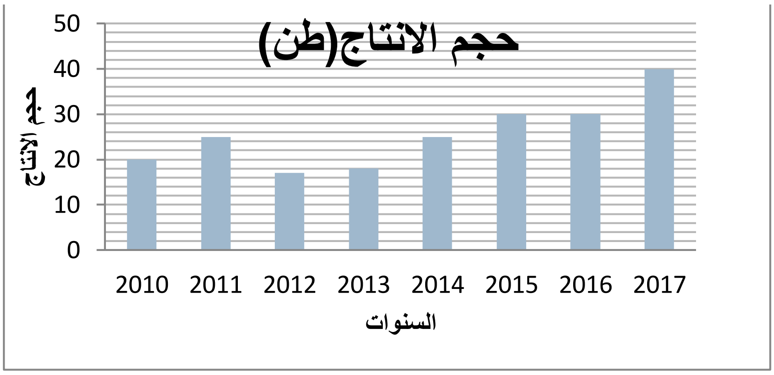
تعرض السلاسل الزمنية بيانيا باستخدام :

-الاعمدة البيانية

-منحنى السلسلة الزمنية .

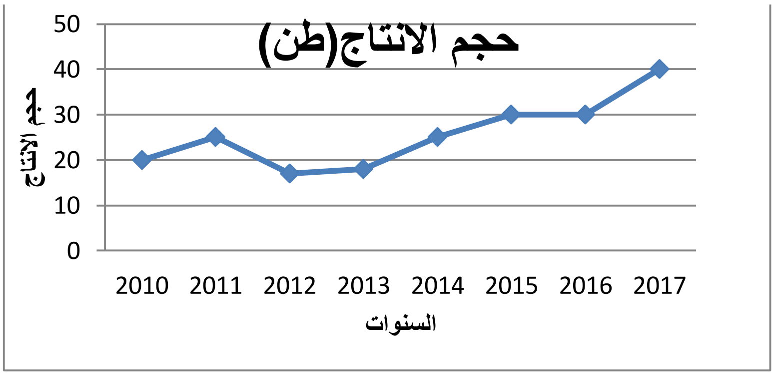
يتم العرض على المحاور الاحداثية حيث يخصص محور السينات للزمن و محور العينات لحدود السلسلة الزمنية .

تعرفنا على اسلوب الاعمدة البيانية في الفصل الاول من هذا الكتاب ، ولن نكرر شرح هذا الاسلوب هنا ، وسنعرض بيانات الجدول (9-1) في الشكل (1.9).



الشكل (9-1).

سنبين فيما يلي كيفية رسم منحنى السلسلة الزمنية و القواعد و الملاحظات الواجب مراعاتها عند رسمها. يقسم محور السينات الى اقسام متساوية بحيث يعبر كل قسم عن فترة زمنية واحدة ،يوم اسبوع او شهر او سنة حسب حدود السلسلة الزمنية و من ثم نمثل كل ثنائية (الزمن ،الظاهرة) بنقطة على المستوى الاحداثي توضع عند نهاية الفترة الزمنية على اعتبار ان قيمة الظاهرة حصلت نهاية العام او يمكن ان توضع النقطة منتصف الفترة الزمنية باعتبار انها تشير الى الفترة كلها، بعد تمثيل كل الثنائيات بنقاط نقوم بوصلها بخط منكسر و يكون الشكل هو منحنى السلسلة الزمنية انظر الشكل (9-2) .



الشكل (9-2)

يساعد هذا الاسلوب على ملاحظة تطور الظاهرة خلال مرور الزمن و معرفة التغيرات التي طرأت عليها ، كما يمكن باستخدام هذا الاسلوب مقارنة عدة ظواهر و ذلك بتمثيل كل ظاهرة بمنحنى و من ثم مقارنة هذه المنحنيات .و يجب عند تمثيل اكثر من منحنى على نفس الشكل تميز هذه المنحنيات عن بعضها و ذلك بألوان مختلفة او برسمها باشكال مختلفة كما يجب وضع وسيلة ايضاح على الشكل لكي نشير الى عنوان كل منحنى.

3.9- العوامل المؤثرة بالسلسلة الزمنية :

عند تحليل بيانات السلسلة الزمنية لابد من معرفة العوامل المؤثرة بها والتي ندعوها ايضا بمركبات السلسلة الزمنية وهذه العوامل هي :

1- الاتجاه العام Trend :وهو الحركة المستمرة على المدى البعيد في بيانات السلسلة ، فاذا كانت متزايدة (خط السلسلة صاعد) قلنا للسلسلة اتجاه عام متزايد ،واذا كانت متناقصة (منحنى السلسة هابط) قلنا

للسلسلة اتجاه عام هابط .ان العوامل المسببة للاتجاه العام هي التقدم التقني والتزايد السكاني .من الامثلة على الاتجاه العام :تزايد مبيعات اجهزة الحاسوب والتلفون الجوال ،انخفاض مبيعات اقراص CD ......

2- الحركات الدورية **Cycle:عندما يكون المنحنى للسلسلة الزمنية السنوية، على شكل موجات متلاحقة فهنا تتاثر السلسلة بعوامل دورية مثل الدورات الاقتصادية او المناخية او انتاج الزيتون (المعاومة).**

3- العوامل الموسمية **Season :هو الحركة المتكررة بنفس الاتجاه الاوقات ولاقل من سنة ،مثل زيادة مبيعات الثلاجات صيفا والمدافئ شتاء ،وزيادة مبيعات الملابس بالاعياد...ان سبب هذه العوامل المناخ والعادات والتقاليد.**

4- العوامل غير المنتظمة **Irregular :وهي الحركة الفجائية في منحنى السلسلة صعودا او هبوطا وبشكل غير منتظم .سبب هذه العوامل الكوارث الطبيعية الحروب الازمات .**

4.9- دراسة الاتجاه العام :

تتم دراسة الاتجاه العام بتمثيل بيانات السلسلة بمعادلة انحدار[3] تدعى معادلة الاتجاه العام ،وذلك بتمثيل العلاقة بين حدود السلسلة Y_t (متغير تابع)والزمن t(متغير مستقل). ونشير هنا ان الزمن لا يؤثر بشكل مباشر كمتغير مستقل في حدود السلسلة الزمنية ، ولكنه يعكس تأثير كل المتغيرات المستقلة التي تؤثر في حدود السلسلة .

عند ايجاد معادلة الاتجاه العام نستخدم عدد من الطرق ، كما مر معنا ،وسنركز اهتمامنا هنا على طريقة واحدة هي المربعات الصغرى واتجاه عام خطي .

حسب هذه الطريقة لتمثيل العلاقة بمعادلة خط مستقيم من الشكل : $\hat{y}_t = a + b\,t$

حيث t هي الزمن ويمثل بالقيم 0 1 2 .... بعام الاساس تكون قيمته الصفر. وهنا الباحث يحدد ذلك .اما ثوابت المعادلة تحسب من العلاقات التالية :

$$b = \frac{\sum yt - \bar{t}\sum y}{\sum t^2 - \bar{t}\sum t} \quad \text{و} \quad a = \bar{y} - b\,\bar{t}$$

مثال :فيما يلي الانتاج السنوي لمعمل للأعوام 2010-2016 (طن)

العام	2010	2011	2012	2013	2014	2015	2016	المجموع
الانتاج	110	120	130	70	150	160	170	910

المطلوب :ايجاد معادلة الاتجاه العام .

[3] راجع الفصل الثامن

يجب اولا ايجاد القيم اللازمة في العلاقات اعلاه $\sum yt, \sum y, \sum t, \sum t^2, \bar{y}, \bar{t}$ كما يلي في الجدول (2.9):

جدول حساب القيم اللازمة لحساب معادلة الاتجاه العام

العام	الانتاج y_t	t	$y*t$	t^2	القيم الاتجاهية T
2010	110	1	110	1	100
2011	120	2	240	4	110
2012	130	3	390	9	120
2013	70	4	280	16	130
2014	150	5	750	25	140
2015	160	6	960	36	150
2016	170	7	1190	49	160
المجموع	910	28	3920	140	910

جدول (2.9)

$$b = \frac{\sum yt - \bar{t}\sum y}{\sum t^2 - \bar{t}\sum t} = \frac{3920 - 4*910}{140 - 4*28} = \frac{280}{28} = 10$$

حيث: $\bar{y} = \frac{\sum y}{n} = \frac{910}{7} = 130$ **و** $\bar{t} = \frac{\sum t}{n} = \frac{28}{7} = 4$

$$a = \bar{y} - b\,\bar{t} = 130 - 10*4 = 90$$

ومنه معادلة الاتجاه العام هي : $\hat{y}_t = 90 + 10\,t$

4.9. 1- استخدامات معادلة الاتجاه العام :

تستخدم معادلة الاتجاه العام في :

1- تقدير قيم الظاهرة متأثرة في الاتجاه العام فقط ،ويكون ذلك بتعويض الزمن t بقيمه 1و2و3......حيث t=0 هي سنة الاساس .تدعى القيم الناتجة بالقيم الاتجاهية ويرمز لها ب T .تم ايجادها في الجدول .

2- التنبؤ بقيم الظاهرة في المستقبل بافتراض ثبات العوامل المؤثرة بالظاهرة ،وذلك بتعويض قيمة t حسب العام المقابل .فتوقع حجم الانتاج لعام 2018 تكون قيمة t=9 وحجم الانتاج المتوقع هو 180.

-3 التوصل الى متوسط التغير السنوي بقيم الظاهرة ،وهي قيمة معامل الانحدار b ،في مثالنا 10 .اي ان حجم الانتاج يزداد سنويا بمقدار 10 طن .

-4 ان قيمة ثابت الانحدار هي متوسط قيمة الظاهرة عام الاساس عندما t=0 .

5.9. – دراسة وتحليل العوامل الموسمية (الرقم القياسي الموسمي) :

تعرفنا اعلاه على العوامل الموسمية ومسبباتها. سنرى الان كيف نتوصل الى معرفة تأثير الموسم على حدود السلسلة الزمنية وكيف يمكن استبعاده او التنبؤ به ، وذلك من خلال ايجاد الرقم القياسي الموسمي S .

يتم ايجاد الرقم القياسي الموسمي بعدة حالات :

1.5.9 – الرقم القياسي الموسمي اذا لم تتأثر السلسلة الموسمية بالاتجاه العام :

– اذا لم يكن في السلسلة اتجاه عام فيتم ايجاد الرقم القياسي الموسمي بطريقة النسب الى المتوسط العام .سنرى هذا الاسلوب من خلال المثال التالي :

فيما يلي المبيعات الفصلية لمتجر خلال 5 سنوات (جدول 3.9):

المبيعات السنة	الفصل الأول	الفصل الثاني	الفصل الثالث	الفصل الرابع	المجموع
2012	10	20	30	50	110
2013	15	25	35	45	120
2014	10	20	30	45	105
2015	20	30	40	50	140
2016	15	25	35	50	125
المجموع	70	120	170	240	600
متوسط الفصل	14	24	34	48	
م القياسي الموسمي S	46.66%	80%	113.33%	160%	400

جدول (3.9)

سنوجد اولا متوسط الفصل من خلال تقسيم مجموع الفصل على عدد السنوات 5 ،ثم نوجد المتوسط العام بتقسيم المجموع الكلي 600 على عدد الفصول 20 ويساوي $30 = \frac{600}{20}$ ،بعد ذلك نوجد الرقم القياسي الموسمي $100 * \frac{\text{متوسط الفصل}}{\text{المتوسط العام}} = s$. (انظر الجدول 3.9).

لا حظ ان البيانات لا تتأثر بالاتجاه العام من خلال قراءة المجموع السنوي .كما ان مجموع الارقام القياسية الموسمية هو عدد المواسم مضروبا ب 100 .

2.5.9 – الرقم القياسي الموسمي اذا كانت السلسلة الموسمية تتأثر بالاتجاه العام :

لإيجاد الرقم القياسي الموسمي يتم اولا ايجاد معادلة الاتجاه العام ثم ايجاد القيم الاتجاهية ومنها يحسب الرقم القياسي الموسمي كما يلي في المثال:

الجدول التالي (4.9) يعطي المبيعات الفصلية من السكر لمحل تجاري (طن)

المبيعات السنة	الربع الأول	الربع الثاني	الربع الثالث	الربع الرابع	المجموع
2012	10	20	30	40	
2013	20	30	40	50	
2014	58	40	50	60	
2015	40	50	60	70	
2016	60	70	80	90	

الجدول (4.9)

من اجل حساب ثوابت معادل الاتجاه العام الفصلية فان :

اما $\sum yt = 12292, \sum y = 968, \sum t = 210, \sum t^2 = 2870, \bar{y} = \frac{968}{20} = 48.4, \bar{t} = \frac{210}{20} = 10.5$

$$b = \frac{\sum yt - \bar{t}\sum y}{\sum t^2 - \bar{t}\sum t} = \frac{12292 - 10.5 * 968}{2870 - 10.5 * 210} = \frac{2128}{665} = 3.2$$

$$a = \bar{y} - b\,\bar{t} = 48.4 - 3.2 * 10.5 = 14.8$$

ومنه معادلة الاتجاه العام الفصلية : $\hat{y}_t = 14.8 + 3.2\,t$

بعد ذلك نوجد من هذه المعادلة القيم الاتجاهية ،وهي عبارة عن المبيعات متأثرة بالاتجاه العام فقط ، الجدول (5.9):

المبيعات متأثرة بالاتجاه العام فقط

المبيعات السنة	الفصل الأول	الفصل الثاني	الفصل الثالث	الفصل الرابع
2012	18	21.2	24.4	27.6
2013	30.8	34	37.2	40.4
2014	43.6	46.8	50	53.2

2015	56.4	59.6	62.8	66
2016	69.2	72.4	75.6	78.8

الجدول (5.9)

نسبت الان القيم الفعلية الى الاتجاهية لكل القيم $\boldsymbol{s = \frac{y_t}{T} * 100}$، الجدول (6.9)

جدول حساب القيم $\boldsymbol{s = \frac{y_t}{T} * 100}$

المبيعات السنة	الربع الأول	الربع الثاني	الربع الثالث	الربع الرابع	المجموع
2012	55.55556	94.33962	122.9508	144.9275	417.7735
2013	64.93506	88.23529	107.5269	123.7624	384.4596
2014	133.0275	85.47009	100	112.782	431.2796
2015	70.92199	83.89262	95.5414	106.0606	356.4166
2016	86.7052	96.68508	105.8201	114.2132	403.4236
المجموع	411.1453	448.6227	531.8392	601.7457	1993.353
قم القياسي الموسمي الخام	82.22907	89.72454	106.3678	120.3491	398.6706
القياسي الموسمي المعدل	82.50327	90.02373	106.7225	120.7504	400

الجدول (6.9)

الرقم القياسي الموسمي الخام هو الوسط الحسابي للقيم في كل عمود. اما الرقم القياسي الموسمي المعدل ينتج عن تصحيح الرقم القياسي الخام بعد تصحيحه بضربه بالعامل المصحح والذي يعطى بالعلاقة التالية :

$$s_{\text{المعدل}} = \frac{s_{\text{الخام}}}{\text{المجموع الفعلي}} * \text{المجموع النظري}$$

حيث : المجموع الفعلي هو 398.6706.

المجموع النظري هو عدد الفصول ضرب 100 ،بمثالنا 400.

3.5.9– مدلول الرقم القياسي الموسمي واستخداماته:

– مدلول الرقم القياسي الموسمي :

1- اذا كان $S = 100$ لا يوجد تأثير للموسم .

2- اذا كان $S > 100$ يوجد تأثير موجب للموسم .

3- اذا كان $S<100$ يوجد تأثير سالب للموسم .

– استخدامات الرقم القياسي الموسمي :

1- يستخدم باستبعاد تأثير الموسم عن حدود السلسلة الزمنية ،اي الحصول على قيم الظاهرة بدون اثر الموسم .يتم ذلك بالتقسيم على الرقم القياسي الموسمي .

2- كما يستخدم بأخذ تأثير الموسم في حدود السلسلة الزمنية ،اي الحصول على قيم الظاهرة مع اثر الموسم .يتم ذلك بالضرب في الرقم القياسي الموسمي .

مثال : بالعودة لمثالنا السابق احسب ما يلي:

1- توقع مبيعات الفصل الاول لعام 2017 متاثرة بالاتجاه العام فقط :

في هذا الفصل تكون $t = 21$ **وبالتالي المبيعات المتوقعة هي :**

$\hat{y}_t = 14.8 + 3.2\,t$ **ومنه** $\hat{y}_t = 14.8 + 3.2 * 21 = 82$ **طن .**

2- توقع مبيعات الفصل الاول لعام 2017 متأثرة بالاتجاه العام والموسم :

لأخذ اثر الموسم بالحسبان ،نضرب القيمة الاتجاهية بالرقم الموسمي للفصل الاول 82.5%.

اي $82 * 82.5\% = 67.65$ **طن**

3- ما هي المبيعات المتوقعة للفصل الثالث للعام 2016 لو لم يكن هناك تأثيرا للموسم :

هنا نقسم القيمة الفعلية على الرقم القياسي الموسمي للفصل الثالث 106.72%،

اي المبيعات المتوقعة لو لم يكن هنا تاثيرا للموسم هي $\frac{80}{106.72\%} = 74.96$ **طن .**

6.9- تحليل وتركيب العوامل المؤثرة بالسلسلة الزمنية :

ان القيم الفعلية للسلسلة هي ناتجة عن اربع عوامل :

1- الاتجاه العام T **.**

2- العوامل الدورية C **.**

3- العوامل الفصلية S **.**

4- العوامل الغير منتظمة I **.**

وبناء عليه فان القيم الفعلية يمكن تحليلها باحد اسلوبين :

1- هي ناتجة عن جداء العوامل السابقة :

$$y_t = T * C * S * I$$

2- هي ناتجة عن مجموع العوامل السابقة:

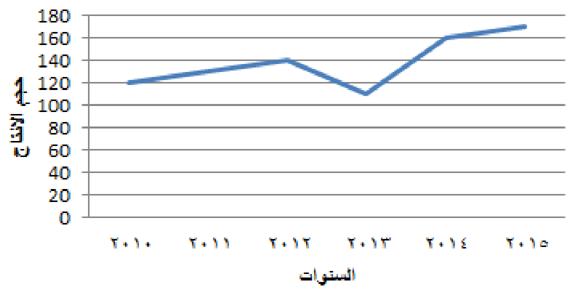
$$y_t = T + C + S + I$$

هناك قواعد لاختيار احد الاسلوبين ،لن نتطرق لذكرها هنا ونكتفي باستخدام الاسلوب الاول وهو الاكثر شيوعا .ومنه نرى:

- **انه اذا علمنا قيمة** T **فان :**$\frac{y_t}{T} = C * S * I$
- **انه اذا علمنا قيمة** T **ولم يكن هناك عوامل موسمية فان :**$\frac{y_t}{T} = C * I$ **وهذا ندعوه الرقم القياسي الدوري .لا حظ ان العوامل الغير منتظمة تبقى مرفقة مع الدورية لأنه لا يمكن التنبؤ بها ودراستها .**

تمارين الفصل التاسع

1- ان العوامل المؤثرة في السلسلة الزمنية في الشكل التالي، هي:



A) الاتجاه العام والموسمية   B) العوامل الدورية والعوامل غير المنتظمة

C) العوامل الدورية والعوامل الموسمية   D)الاتجاه العام والعوامل غير المنتظمة

2- ان سبب الاتجاه العام هو :

منا(D التقدم التقني والتزايد السكاني C) الحروب B) العادات والتقالي

3- من اسباب الموسم هو :

ب B) العادات والتقاليد و المناخ

D)غير ذلك   C) التقدم التقني والتزايد السكاني

4- لديك البيانات التالية عن حجم الإنتاج السنوي لمعمل صناعات غذائية (طن):

العام	1995	1996	1997	1998	1999	2000	2001	2002	2003	مجموع
الإنتاج	38	40	45	48	52	63	67	75	84	512

المطلوب :

1- أوجد معادلة الاتجاه العام معتبر عام 1994 عام أساس.

2- ما هو متوسط التغير السنوي بحجم الانتاج.

3- ما هو متوسط حجم الانتاج عام 1994.

4- تنبأ بحجم الانتاج عام 2004.

5- فيما يلي المبيعات الفصلية لمعمل برادات (المبيعات براد)

السنة	بيعات	لفصل الأول	لفصل الثاني	لفصل الثالث	لفصل الرابع
2012		100	120	150	160
2013		110	108	140	170
2014		90	100	110	120
2015		80	90	120	130
2016		100	110	120	150

المطلوب:

1. **أوجد الرقم القياسي الموسمي بطريقة النسب الى المتوسط العام.**
2. **ما هو تأثير الموسم في الفصل الرابع.**
3. **تنبأ بمبيعات الفصل الرابع عام 2013 لو لم يكن هناك تأثير للموسم.**

5- **لدينا فيما يلي الانتاج السنوي لمعمل بملايين الوحدات النقدية :**

العام	2000	2001	2002	2003	2004	2005	2006	المجموع
الانتاج	100	110	120	74	140	150	160	854

المطلوب :

-1	**ان السلسلة الزمنية المعروضة بالجدول تتأثر بالعوامل التالية :** **بالاتجاه العام** T **والموسمية**S. B - **بالعوامل الموسمية** S **والعشوائية** I . C **جاه العام والعوامل الموسمية والعشوائية** S+I+T . D- **بكل العوام** **المذكورة** T+S+C+I . E- **غير ذلك** .
-2	ان معادلة الاتجاه العام السنوية بطريقة المربعات الصغرى معتبرا عام 1999 عام أساس هي : $\hat{y}=110+12.96t$ **B)** $\hat{y}=100+10t$ **C)** $\hat{y}=110.48+10.88t$ $\hat{y}=82+10t$ **E)**غير ذلك
-3	ان كمية الانتاج المتوقعة لعام 2007 متأثرة بالاتجاه العام فقط هي : **A)**162 **B)** 168 **C)** 142.88 **D)**182.2 **E)**غير كل ما ذكر
-4	ان كمية الانتاج المقدرة لعام 1999 متأثرة بالاتجاه العام فقط هي : **A)**82 **B)** 130 **C)** 100 **D)**110 **E)**غير كل ما ذك
-5	ان الرقم القياسي الدوري لعام 2001 هو : **A)** % 100 **B)**%110 **C)**%120.5 **D)** %113.72 **E)**غير ذلك

وإذا علمت ان معادلة الاتجاه العام الفصلية المحسوبة لهذا المعمل باعتبار الفصل الرابع عام 1999 هو فصل الاساس هي
$\hat{y} = 21.75 + 0.625t$ ، وان الرقم القياسي الموسمي هو كما يلي مع العلم انه لا يوجد تأثير للموسم في الفصل الرابع :

	الفصل 1	الفصل 2	الفصل 3	الفصل 4	المجموع
الرقم القياسي الموسمي المعدل	% 110	% ؟	% 120	% ؟	؟

6-	يبلغ الرقم القياسي الموسمي المعدل للفصل الثاني : **A)** % 70 **B)**%110 **C)**%120.5 **D)** %113.72 **E)**غير ذلك
7-	يبلغ الرقم القياسي الموسمي المعدل للفصل الرابع : **A)** % 100 **B)**%100 **C)**%120.5 **D)** %113.72 **E)**غير ذلك
8-	ان كمية الانتاج المقدرة للفصل الثاني لعام 2007 متأثرة بالاتجاه العام فقط هي : **A)**22.3 **B)** 40.5 **C)** 40.5 **D)**24.325 **E)**غير كل ما ذكر
9-	ان كمية الانتاج المقدرة للفصل الثاني لعام 2007 متأثرة بالاتجاه العام والموسم هي : **A)**22.3 **B)** 40.5 **C)** 41.875 **D)**28.35 **E)**غير كل ما ذكر
10-	**ان المبيعات الفعلية للفصل الثالث عام 2000 هي 18 مليون ،ماهي المبيعات المتوقعة الثالث عام 2000 بعد عزل تأثير الموسم :** **A)** 15 **B)** 22.5 **C)** 21.875 **D)**16.0625 **E)**غير كل ما ذكر

مراجع الفصل :

1- حميدان، عدنان؛ الجاعوني، فريد؛ ناصر آغا، عمار؛ العواد، منذر، 2003، مبادئ الإحصاء، منشورات جامعة دمشق.

1- **Groebner David F., Shannon Patrick W. and Fry Phillip C. (2017)," Business Statistics :A Decision-Making approach", 10 th Edition. person**.

الفصل العاشر : الارقام القياسية
Index numbers

المخرجات والأهداف التعليمية:

بعد الانتهاء من هذا الفصل ،سيكون الطالب قادرا على:

1- حساب الرقم القياسي للأسعار.

2- المقارنة الكمية لظاهرة في زمنين مختلفين او مكانين مختلفين ، وكذلك مقارنة الاسعار بين فترتين او مكانين.

3- استبعاد اثر التغير في قيمة النقود من البيانات النقدية والوصول الى التغير الفيزيائي في الظاهرة .

4- التمييز بين اسعار ثابتة واسعار جارية .

ملخص الفصل:

يتناول هذا الفصل الارقام القياسية : المقصود بالرقم القياسي ومشاكل انشاؤه وطرق حسابه البسيطة والتجميعية المرجحة ، وتطبيقات عن حساب الرقم القياسي بطرقه المختلفة، وتطبيقات على استخدام الرقم القياسي .

. تعريف الرقم القياسي:1.10

الرقم القياسي هو أداة لمقارنة ظاهرة أو عدة ظواهر تعود لفترات زمنية مختلفة لمعرفة مقدار التغير في الظواهر أو الفرق بينهما كأن نقارن سعر سلعة عام 2019 مع سعرها عام 2010 أو مقارنة اسعار مجموعة من السلع في عام ما بأسعارها في عام أخر، أو مقارنة سعر سلعة في مكانين مختلفين ، مثلا نقارن سعر سلعة في دمشق مع سعرها في اللاذقية ،أو مقارنة انتاج سورية من القمح عام 2015 بإنتاجها عام 2010 .... .

تتعدد الأرقام القياسية و ذلك بسبب اختلاف طريقة إنشاؤها(كما سنرى) او بسبب اختلاف الظواهر التي يقارنها، فحسب الظواهر التي يقارنها نجد عدد كبير من الأرقام القياسية على سبيل المثال: الرقم القياسي للأجور و الرقم القياسي الزراعي والرقم القياسي الصناعي والرقم القياسي لأسعار الجملة والرقم القياسي لأسعار المستهلك وسنهتم هنا بشكل عام بطرق انشاء الرقم القياسي للأسعار.

عند انشاء الرقم القياسي للأسعار فإننا نقارن أسعار عام ما وسنطلق عليه عام الدراسة او عام الهدف ونرمز لأسعاره بالرمز p_I ،بأسعار عام آخر وهو عام الأساس ونرمز لأسعاره ب p_0، وتتم هذه المقارنة بطرق مختلفة تختلف باختلاف إنشاء الرقم القياسي للأسعار .

قبل البدء بإنشاء الرقم القياسي وتطبيقه لابد من حل بعض المشكلات المتعلقة بهذا الرقم وهي:

- تحديد سنة الاساس :

سنة الاساس هي السنة التي تتم المقارنة بها يجب أن تكون خالية من الظروف غير الطبيعية ، مثل الحروب والمجاعات والفيضانات وعدم الاستقرار السياسي ...... إلخ. يمكن اختيار سنة الأساس بطريقتين:

– باستخدام قاعدة ثابتة تظل فيها سنة الأساس ثابتة ، او من خلال قاعدة السلسلة التي تتغير سنة الأساس فيها على سبيل المثال ، لعام 2019 ستكون سنة الأساس 2018 ، لعام 2018 ستكون 2017 ، وهكذا.

- اختيار السلع التي ستدخل بحساب الرقم القياسي:

المشكلة الثانية في بناء الأرقام القياسية هي اختيار السلع. ان عدد السلع في السوق كبير جدا ، بسبب عدم إمكانية إدراج جميع السلع ، يجب اختيار عينة من السلع ، عند اختيار هذه السلع يجب اخذ النقاط التالية في الاعتبار:

- **يجب أن تكون السلع ممثلة لأذواق الناس وعاداتهم .**
- **يجب أن تكون السلع ثابتة في الجودة على مدى فترتين او مكانين مختلفتين.**
- **ينبغي إدراج جميع أصناف السلعة التي تكون شائعة الاستخدام او كما نقول ذات طابع شعبي.**

* تحديد الأسعار :

بعد اختيار السلع ، تكمن المشكلة التالية في معرفة أسعارها، من اين سنحصل على الاسعار ،هل سيتم اختيار أسعار الجملة أم أسعار التجزئة وهل يجب تضمين الضرائب في .

أثناء جمع أسعار السلع التي ستدخل في حساب الرقم القياسي ، تجدر الإشارة إلى النقاط التالية:

– يتم جمع الأسعار من تلك الأماكن التي يتم فيها تداول سلعة معينة بكميات كبيرة الاسواق .

– يجب أيضًا استخدام المعلومات المنشورة من جهات رسمية والمتعلقة بالأسعار .

– عند اختيار الأفراد والمؤسسات الذين يقدمون عروض أسعار ، ينبغي توخي الحذر لعدم تحيزهم.

– يعتمد اختيار أسعار الجملة أو التجزئة على نوع رقم القياسي الذي يتعين إعداده. تستخدم أسعار الجملة في بناء الرقم القياسي العام للأسعار وتستخدم أسعار التجزئة في بناء الرقم القياسي لتكلفة المعيشة.

– يمكن استخدام متوسط الأسعار التي يتم جمعها من أماكن مختلفة.

* اختيار المتوسط:

نظرًا لأن الارقام القياسية هي وسط الاسعار ، فإن المشكلة الرابعة هي اختيار متوسطاً مناسباً، نظرياً الوسط الهندسي[4] هو الأفضل لهذا الغرض كما راينا. ولكن في الممارسة العملية ، يتم استخدام الوسط الحسابي لأنه الأسهل استخداما .

* اختيار الأوزان (الترجيح) :

بشكل عام ليس لجميع السلع المدرجة في بناء الأرقام القياسية نفس الأهمية او نفس الثقل ، لذلك عند انشاء الأرقام القياسية يجب أن يؤخذ ذلك بعين الاعتبار ، يجب الترجيح بأوزان مناسبة للسلع وفقًا لأهميتها النسبية .على سبيل المثال ، يجب إعطاء أسعار مستلزمات الاطفال أهمية أكبر أثناء إعداد مؤشر تكلفة المعيشة للأسرة في المجتمعات الفتية (معدل نمو مرتفع) مقارنةً بإعداد مؤشر تكلفة المعيشة للأسرة في المجتمعات الكهلة . يجب أن تكون الأوزان غير منحازة وأن يتم اختيارها بشكل علمي و عقلاني.

* الغرض من الرقم القياسي:

من أهم الاعتبارات عند انشاء الأرقام القياسية هو الهدف من الرقم القياسي . يجب حل جميع المشكلات أو الخطوات الأخرى في ضوء الغرض من إعداد رقم قياسي معين. نظرًا لأنه يتم إعداد أرقام قياسية مختلفة لأغراض محددة ولا يوجد رقم قياسي واحد هو "الرقم القياسي لكل شيء" ، فمن المهم أن يكون واضحًا الغرض من الرقم القياسي قبل إنشائه. إن الغاية من انشاء الرقم القياسي للأسعار هو تحديد مدى التغير في الأسعار وبالتالي دراسة المستوى المعاشي في البلد من أجل تخطيط الأجور والضرائب والانفاق العام ...واعتمادا على ذلك يجب وضع الحلول المناسبة لكل المشاكل لخدمة هذا الهدف.

*اختيار طريقة انشاء الرقم القياسي:

[4] انظر الفصل الثالث ،الوسط الهندسي

الخطوة الأخيرة هي اختيار طريقة مناسبة لبناء الأرقام القياسية، يؤثر الأسلوب المستخدم في انشاء الرقم القياسي على قيمته فكما سنرى أدناه، أن كل طريقة تعطي قيمة مختلفة للرقم القياسي ،حيث يمكن اختيار الأسلوب المناسب اعتماداً على مزاياه وبشكل عام فإنه يجب اختيار الأسلوب الذي يجتاز أكبر عدد من اختبارات الأرقام القياسية.

1 – الأرقام القياسية البسيطة (المناسيب).

2 – الأرقام القياسية التجميعية البسيطة .

3 – الأرقام القياسية التجميعية المرجحة .

2.10 الأرقام القياسية البسيطة (المناسيب) :

تعتبر الأرقام القياسية البسيطة سهلة للغاية وهي عبارة عن نسبة قيمة ظاهرة واحدة في عام المقارنة إلى عام الأساس، او نسبة قيمة ظاهرة واحدة في مكان المقارنة إلى مكان الأساس، وعليه يكون الرقم القياسي لسعر سلعة ما هو إلا عبارة عن نسبة سعرها في عام الدراسة (المقارنة) إلى سعرها في عام الأساس وضرب الناتج بمائة للتعبير عن الناتج بشكل مئوي:

مثال: **لنفرض أن سعر كيلو التفاح في عام 2010 كان 40 ليرة ، وأصبح في عام 2016 يساوي 60 ليرة ، فعليه يكون الرقم القياسي للسعر عام 2016 مقارنةً بعام 2010 :**

$$I = \frac{p_I}{p_0} * 100$$

$$I = \frac{60}{40} * 100 = 150\%$$

أي أن سعر التفاح في عام 2016 يشكل 150% من سعره عام 2010 أو بمعنى آخر نقول أن سعر التفاح ارتفع بنسبة 50% من سعره عام 2010.

أما إذا اعتبرنا أن عام 2016 عام الأساس 2010 عام الدراسة فان الرقم القياسي سيكون :

$$I = \frac{40}{60} * 100 = 66.66\%$$

أي أن سعر التفاح بعام 2010 كان أقل من سعره عام 2016 بنسبة 33.333%.

توضيح :من هذا المثال يمكن ان نستنتج كيفية قراءة الرقم القياسي

- **اذا كان I=100 اي ان قيمة الظاهرة لم تتغير مقارنة بعام او مكان الاساس.**
- **اذا كان I>100 اي ان قيمة الظاهرة ارتفعت مقارنة بعام الاساس او مكان الاساس ، ونسبة الارتفاع هي الفرق بين الرقم القياسي و 100.**
- **اذا كان I<100 اي ان قيمة الظاهرة انخفضت مقارنة بعام الاساس او مكان الاساس ، ونسبة الانخفاض هي الفرق بين الرقم القياسي و 100.**

رغم سهولة الأرقام القياسية البسيطة فإنها لا تفي بالغرض، والذي هو مقارنة تغير مستوى الأسعار ككل وليس سلعة واحدة ، فإذا ما أردنا مقارنة أسعار 160 سلعة مثلاً فإن المقارنة ستكون باستخدام 160 رقم، لذلك في مثل هذه الحالات تفضل الارقام القياسية التجمعية البسيطة أو المرجحة.

3.10 الأرقام القياسية التجميعية البسيطة :

الأرقام القياسية التجميعية تعطي الرقم القياسي لعدد من السلع وذلك حسب الطرق التالية:

1- الرقم التجميعي البسيط: حسب هذا الأسلوب فإننا نعبر عن الرقم القياسي للأسعار بأنه نسبة مجموع أسعار السلع عام المقارنة (الدراسة) إلى مجموع أسعار السلع عام الأساس مضروباً بمائة:

$$I_s = \frac{\sum p_I}{\sum p_0} * 100$$

حيث : $\sum p_I$مجموع أسعار السلع سنة المقارنة.

$\sum p_0$ مجموع أسعار السلع سنة الأساس.

مثال: **لتكن لدينا أسعار ثلاث سلع في عام الأساس و عام المقارنة كما يلي:**

السلعة	السعر عام الاساس p_0	السعر عام الدراسة p_I
التفاح	4	8
اللحم	10	12
الخبز	2	2
لمجموع	16	22

المطلوب : **أوجد الرقم القياسي التجميعي البسيط:**

$$I_s = \frac{\sum p_I}{\sum p_0} * 100$$

$$I_s = \frac{22}{16} * 100 = 137.5\%$$

أي أن الأسعار عام المقارنة ارتفعت بنسبة 37.5%.

2- الأرقام القياسية كمتوسط:

بهذا الأسلوب يتم حساب الأرقام القياسية باستخدام مقاييس النزعة المركزية التي مرت معنا :

- المتوسط البسيط للمناسيب:

للتخلص من عيوب المناسيب عند استخدامها لمقارنة أسعار عدداً كبيراً من السلع فإننا نأخذ متوسط عدداً كبيراً من المناسيب للسلع ، ويمكن أن نستخدم في هذه الحالة الوسط الحسابي أو الوسيط أو الوسط الهندسي.

1. الرقم القياسي للأسعار باستخدام الوسط الحسابي للمناسيب: لتكن لدينا المناسيب $I_1, I_2, \ldots\ldots I_n$ حيث n عدد السلع ،فأن الرقم القياسي لأسعار هذه السلع في سنة المقارنة بالنسبة لأسعار سنة الأساس هو :

$$\mathrm{I} = \frac{\sum \frac{p_{Ii}}{p_{0i}}}{n} * 100$$

2. الرقم القياسي للأسعار باستخدام الوسط الهندسي:

$$I = \sqrt[n]{\frac{p_{I1}}{p_{01}} * \frac{p_{I2}}{p_{02}} ****** \frac{p_{In}}{p_{0n}}} * 100$$

مثال: لتكن لدينا البيانات التالية عن سعر السلعتين A , B في سنة الأساس والمقارنة فإن الرقم القياسي كوسط حسابي وكوسط هندسي يتم حسابهما كما في الجدول:

السلعة	السـعر		$\frac{p_I}{p_0}$
	سنة الاساس p_0	سنة المقارنة p_I	
A	8	4	0.5
B	4	8	2
المجموع	12	12	2.5
الوسط الحسابي للمناسيب	$I = \frac{2.5}{2} * 100 = 125\%$		
الوسط الهندسي للمناسيب	$I = \sqrt[2]{0.5 * 2} * 100 = 100\%$		

من عيوب الارقام القياسية البسيطة رغم بساطته ، أنه يعطي لكل السلع نفس الأهمية النسبية كما أنه لا يأخذ بالحسبان الوحدات التي تباع بها السلع كيلو ، ليتر...

4.10 الأرقام القياسية التجميعية المرجحة بالكميات :

للتغلب على عيوب الرقم القياسي البسيط فإننا نقوم بترجيح الأسعار بالكميات من هذه السلع بسنة الأساس أو سنة المقارنة وهنا يمكن حساب الأرقام القياسية التالية:

1. رقم لاسبير القياسي

Laspeyre: **لحساب هذا الرقم فإننا نقوم بترجيح أسعار سنة الأساس وسنة المقارنة بكميات سنة الأساس q_0 :**

$$I_l = \frac{\sum p_I \, q_0}{\sum p_0 \, q_0} * 100$$

.2 **رقم باش القياسي**

Paasche: **لحساب هذا الرقم فإننا نقوم بترجيح أسعار سنة المقارنة وسنة الأساس بكميات سنة المقارنة** q_I:

$$I_p = \frac{\sum p_I \, q_I}{\sum p_0 \, q_I} * 100$$

.3 **الرقم المثالي**

(فيشر) Fisher: **وهو عبارة عن الوسط الهندسي للرقمين لاسبير وباش:**

$$I_f = \sqrt{\frac{\sum p_I \, q_0}{\sum p_0 \, q_0} * \frac{\sum p_I \, q_I}{\sum p_0 \, q_I}} * 100$$

.4 **رقم مارشال –**

ادجورث القياسي Marshall – Edgeworth : **حسب هذا الرقم فإننا نقوم بترجيح الأسعار بالوسط الحسابي لكميات سنة الأساس والمقارنة:**

$$I_m = \sqrt{\frac{\sum p_I (q_0 + q_I)}{\sum p_0 (q_0 + q_I)}} * 100$$

مثال:

لدينا البيانات التالية عن أسعار السلع وكمياتها المنتجة في عامي الأساس والمقارنة جدول (1.10)

السلعة	عام الاساس		عام المقارنة		$p_I \, q_0$	$p_0 \, q_0$	$p_I \, q_1$	$p_0 \, q_I$	$q_0 + q_I$	$p_I(q_0 + q_I)$	$p_0(q_0 + q_I)$
	السعر p_0	الكمية q_0	السعر p_I	الكمية q_I							
تفاح	**5**	**2**	**6**	**4**	**12**	**10**	**24**	**20**	**6**	**36**	30
دجاج	**7**	**6**	**8**	**8**	**48**	**42**	**64**	**56**	**14**	**112**	98
البطاطا	**10**	**10**	**12**	**20**	**120**	**100**	**240**	**200**	**30**	**360**	300
لمجموع	**22**	**18**	**26**	**32**	**180**	**152**	**328**	**252**	**50**	**508**	429

جدول (1-10)

ولحساب الأرقام القياسية التجميعية والبسيطة المرجحة فقد قمنا بحساب المجاميع اللازمة في الجدول.

.1 **الرقم التجميعي**

البسيط:

$$I_s = \frac{\sum p_I}{\sum p_0} * 100$$

$$I_s = \frac{26}{22} * 100 = 118.18\%$$

أي أن الأسعار ارتفعت بنسبة 18.18%.

2. رقم لاسبير القياسي:

$$I_l = \frac{\sum p_I \, q_0}{\sum p_0 \, q_0} * 100$$

$$I_l = \frac{180}{152} * 100 = 118.42\%$$

أي أن الأسعار ارتفعت بنسبة 18.42%.

3. رقم باش القياسي:

$$I_p = \frac{\sum p_I \, q_I}{\sum p_0 \, q_I} * 100$$

$$I_p = \frac{328}{252} * 100 = 130.15\%$$

أي أن الأسعار ارتفعت بنسبة 30.15%.

4. رقم فيشر القياسي المثالي:

$$I_f = \sqrt{\frac{\sum p_I \, q_0}{\sum p_0 \, q_0} * \frac{\sum p_I \, q_I}{\sum p_0 \, q_I}} * 100$$

$$I_f = \sqrt{\frac{180}{152} * \frac{328}{252}} * 100 = 124.15$$

أي أن الأسعار ارتفعت بنسبة 24.15%.

5. رقم مارشال – ادجورث القياسي:

$$I_m = \frac{\sum p_I \, (q_0 \, + q_I)}{\sum p_0 \, (q_0 \, + q_I)} * 100$$

$$I_m = \frac{508}{429} * 100 = 118.41\%$$

أي أن الأسعار ارتفعت بنسبة 18.41%.

4.10 الأرقام القياسية التجميعية للمناسيب المرجحة بالقيم :

يتم ترجيح المناسيب حسب هذا الأسلوب بقيم السلع المنتجة أو المستهلكة ، والمقصود بالقيمة هو جداء الكميات q بالأسعار p : وعليه يكون الرقم القياسي حسب القيم المأخوذة للترجيح بأحد الأشكال التالية:

1- الترجيح بقيم كميات عام الأساس محسوبة بأسعار عام الأساس:

$$I=\frac{\sum\left(\frac{p_I}{p_0}*p_0q_0\right)}{\sum p_0q_0}*100=\frac{\sum p_Iq_0}{\sum p_0q_0}*100$$

لاحظ أن هذا الرقم بعد الاختصار ينتج الرقم القياسي لاسبير.

2- الترجيح بقيم كميات عام المقارنة مأخوذة بأسعار عام الأساس:

$$I=\frac{\sum\left(\frac{p_I}{p_0}*p_0q_I\right)}{\sum p_0q_I}*100=\frac{\sum p_Iq_I}{\sum p_0q_I}*100$$

لاحظ أن هذا الرقم بعد الاختصار ينتج الرقم القياسي باش.

3- الترجيح بقيم كميات عام الأساس مأخوذة بأسعار عام المقارنة:

$$I=\frac{\sum\left(\frac{p_I}{p_0}*p_Iq_0\right)}{\sum p_Iq_0}*100$$

4- الترجيح بقيم كميات عام المقارنة مأخوذة بأسعار عام المقارنة:

$$I=\frac{\sum\left(\frac{p_I}{p_0}*p_Iq_I\right)}{\sum p_Iq_I}*100$$

مثال: لدينا البيانات التالية عن أسعار السلع X, Y ,Z وكمياتها المنتجة في عامي الأساس والمقارنة جدول (10-2):

السلعة	سنة الاساس		سنة المقارنة		$\frac{p_1}{p_0}$
	p_0	q_0	p_I	q_I	
X	5	6	6	8	1.2
Y	4	8	5	10	1.25
Z	6	20	9	22	1.5
المجموع	15	34	20	44	

جدول (10-2)

المطلوب: احسب الرقم القياسي للاسعار كمتوسط مرجح للمناسيب وفق ما يلي:

- متوسط المناسيب المرجح بقيم كميات عام الأساس محسوبة بأسعار عام الأساس.
- متوسط المناسيب المرجح بقيم كميات عام المقارنة مأخوذة بأسعار عام الأساس.
- متوسط المناسيب المرجح بقيم كميات عام الأساس مأخوذة بأسعار عام المقارنة.
- متوسط المناسيب المرجح بقيم كميات عام المقارنة مأخوذة بأسعار عام المقارنة.

الحل: لقد حسبنا القيم اللازمة لتطبيق العلاقات اللازمة في الجدول(9-3):

48	36	40	30	57.6	43.2	48	36
50	40	40	32	62.5	50	50	40
198	180	132	120	297	270	198	180
296	256	212	182	417.1	363.2	296	256

(3-10)

1. متوسط المناسيب المرجح بقيم سنة الأساس:

$$I = \frac{\sum\left(\frac{p_I}{p_0} * p_0 q_0\right)}{\sum p_0 q_0} * 100 = \frac{256}{182} * 100 = 140.66\%$$

2. متوسط المناسيب المرجح بكميات عام المقارنة مأخوذة بأسعار عام الأساس:

$$I = \frac{\sum\left(\frac{p_I}{p_0} * p_0 q_I\right)}{\sum p_0 q_I} * 100 = \frac{296}{212} * 100 = 139.62\%$$

3. متوسط المناسيب المرجح بكميات عام الأساس مأخوذة بأسعار عام المقارنة:

$$I = \frac{\sum\left(\frac{p_I}{p_0} * p_I q_0\right)}{\sum p_I q_0} * 100 = \frac{363.2}{256} * 100 = 141.875\%$$

4. متوسط المناسيب المرجح بكميات عام المقارنة مأخوذة بأسعار عام المقارنة:

$$I = \frac{\sum\left(\frac{p_I}{p_0} * p_I q_I\right)}{\sum p_I q_I} * 100 = \frac{417.1}{296} * 100 = 140.9\%$$

6.10 استخدام الأرقام القياسية في عزل تأثير ارتفاع الأسعار:

في كثير من الأحيان ،عند مقارنة تطور الظواهر بالقيم النقدية تصادفنا مشكلة تأثير ارتفاع الأسعار أو بتعبير آخر مشكلة انخفاض قيمة النقود، الذي بدوره يؤدي أحياناً إلى ظهور تعاريف مثل زيادة حقيقية

وزيادة ظاهرية، وبالمثل في الحسابات القومية ظهور تسميات بالأسعار الثابتة وبالأسعار الجارية. في مثل هذه الحالات حتى تكون مقارنتنا صحيحة معبرة عن الزيادة الحقيقية في قيمة الظاهرة فإننا نستخدم الأرقام القياسية للأسعار لعزل تأثير انخفاض قيمة النقود وتتم المقارنة على أساس أسعار عام الأساس حيث تكون قيمة الرقم القياسي لهذا العام 100%.

سنبين فيما يلي كيفية استخدام الرقم القياسي للأسعار لعزل تأثير ارتفاع الأسعار في الظاهرة ، وبالتالي الوصول الى التغير بقيمة الظاهرة بغض النظر عن التغير في الاسعار او ما يعرف بالتغير الفيزيائي بقيمة الظاهرة .

مثال: لدينا البيانات التالية في الجدول (10-4)عن متوسط الدخل الشهري للعامل في إحدى البلدان وكذلك الرقم القياسي للأسعار (حيث 2010=100).

العام	2010	2011	2012	2013	2014	2015	2016	2017
دخل الشهري	5500	6000	6500	6600	6900	7000	7100	9250
لرقم القياسي %100=201	100	130	135	140	160	170	180	185

(4-10)

المطلوب:

- أوجد القيمة الحقيقية للدخل الشهري العامل في عام 2017 اعتماداً على أسعار عام 2010 كعام أساس .
- احسب الزيادة الحقيقية في متوسط الدخل في عام 2015 اعتماداً على عام 2010 كعام أساس .

الحل: لإيجاد القيمة الحقيقية للدخل اعتماداً على أسعار 2010 نقسم القيم الاسمية للدخل عام 2017 على الرقم القياسي للأسعار عام 2010 حيث 2000 هو عام أساس (100 %) ويساوي :

$$y_{adj} = \frac{9250}{185} * 100 = 5000$$

حيث y_{adj} القيمة الحقيقية للظاهرة بعد استبعاد اثر ارتفاع الاسعار (الدخل الحقيقي).

واعتماداً على ذلك تكون الزيادة الحقيقية للدخل في عام 2017 بالنسبة لعام 2010 هي:

$5000 - 5500 = -500$ وحدة نقدية. ماذا يعني هذا؟

ان الزيادة في قيمة الدخل عام 2017 مقارنة بعام 2010 والبالغة $9250 - 5500 = 3750$ وحدة نقدية هي زيادة ظاهرية. او يمكن القول ان الدخل في عام 2017 يعادل ما قيمته 5000 باسعار 2010

ولا يكفي للعيش او لشراء السلع التي كان ثمنها **5500** في عام **2010** . وان ما نحتاجه من النقود حتى نشتري السلع ذاتها التي كنا نشتريها عام **2010** هو :

$$5500 * \frac{185}{100} = 10175$$

مثال:

فيما يلي ارباح متجر خلال الاعوام 2012-2018 (مليون)

العام	2012	2013	2014	2015	2016	2017	2018
بح بالأسعار الجارية (مليون)	**60**	**65**	**70**	**78**	**95**	**95**	**98**
رقم القياسي **2012=100**	**100**	**120**	**140**	**160**	**190**	**190**	**200**
بح بأسعار الاساس (مليون)	**60**	**54.16**	**50**	**48.75**	**50**	**50**	**49**

الجدول (5.10)

المطلوب:

- **احسب الربح الحقيقي عام 2018 مقاسا بأسعار 2012 .**

$$y_{adj} = \frac{98}{200} * 100 = 49\text{مليون}$$

- **ايهما اكبر ربح المتجر عام 2018 ام 2012 .**

ان ارباحه عام 2012 بأسعار 2012 هو 60 مليون ،اما ارباحه عام 2018 بأسعار 2012 هي 49 مليون .اي ارباه عام 2012 هي اكبر.

- **اوجد سلسلة ارباح المتجر بأسعار عام 2012 (بالأسعار الثابتة) .ثم قارن الارباح بالأسعار الجارية مع الارباح بالأسعار الثابتة.**

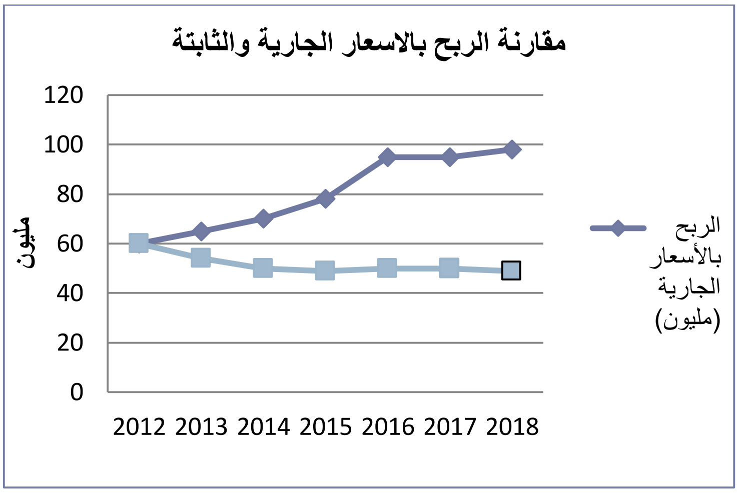
تم ايجادها بالسطر الاخير من الجدول (*5.9*) **باستخدام العلاقة السابقة :**

$$y_{adj} = \frac{Y_t}{I_t} * 100$$

حيث Y_t: **الربح في العام** t .

I_t **الرقم القياسي عام** t .

من مقارنة الارباح بالأسعار الجارية مع الارباح بالأسعار الثابتة ،اسعار عام 2012 .نرى ان الارباح بالأسعار الجارية تتزايد بينما بالأسعار الثابتة تتناقص(انظر الجدول وكذلك الشكل1.9) هذا يدل الى ان نشاط هذا المتجر لم يتوسع ولكن الاسعار هي التي ارتفعت ،واذا دققنا الشكل نرى بوضوح انخفاض ارباحه الحقيقية .


مقارنة الربح بالاسعار الجارية والثابتة
120
100
80
60
40
20
0
مليون
2012 2013 2014 2015 2016 2017 2018
الربح بالأسعار الجارية (مليون)

الشكل(1.10)

مسائل وتمارين

1 – لدينا البيانات التالية عن أسعار بعض السلع في مدينة دمشق للعامين 2010و 2018:

السلعة	عام 2010	عام 2018
	p_0	p_I
اللحمة	**300**	**4000**
البيض	**100**	**1000**
الحليب	**25**	**200**
بطاطا	**30**	**150**
المجموع	**455**	**5350**

المطلوب:

1. **احسب مناسيب الأسعار للسلع الواردة في الجدول معتبراً عام 2010 عام الأساس.**
2. **احسب الرقم القياسي التجميعي البسيط لهذه السلع معتبراً عام 2010عام الأساس.**
3. **احسب الرقم القياسي لهذه السلع باستخدام الوسط الحسابي للمناسيب معتبراً عام 2010عام الأساس.**
4. **احسب الرقم القياسي لهذه السلع باستخدام الوسط الهندسي للمناسيب معتبراً عام 2010عام الأساس.**

2- الجدول التالي يبين أسعار ثلاث سلع في عامي الأساس والمقارنة والكميات المستهلكة منها.

السلعة	سنة الاساس		سنة المقارنة	
	p_0	q_0	p_I	q_I
A	**6**	**5**	**10**	**12**
B	**10**	**10**	**12**	**20**
C	**20**	**20**	**30**	**30**
المجموع	**36**	**35**	**52**	**62**

المطلوب:

1. **احسب الرقم القياسي التجميعي المرجح بكميات سنة الأساس (لاسبير).**
2. **احسب الرقم القياسي التجميعي المرجح بكميات سنة المقارنة (باش)**
3. **احسب الرقم القياسي الأمثل (فيشر)**
4. **احسب الرقم القياسي المرجح بمتوسط كميات سنة الأساس والمقارنة ،مارشال.**

5. **احسب الرقم القياسي التجميعي البسيط.**
6. **احسب متوسط المناسيب المرجح بقيم سنة الأساس.**
7. **احسب متوسط المناسيب المرجح بقيم سنة المقارنة.**
8. **احسب متوسط المناسيب المرجح بقيم سنة الأساس محسوبة بأسعار سنة المقارنة.**
9. **احسب متوسط المناسيب المرجح بقيم سنة الدراسة محسوبة بأسعار سنة الاساس.**

3- الجدول التالي يبين الرقم القياسي العام للأسعار في احدى الدول العربية ودخل الاسرة الشهري(الف) :

السنة	2011	2012	2013	2014	2015	2016	2017
الرقم القياسي %	**100**	**120**	**125**	**130**	**150**	**160**	**200**
الدخل الشهري للأسرة	**20**	**21**	**24**	**25**	**30**	**32**	**36**

المطلوب:

1. **أوجد الدخل الحقيقي الشهري للأسرة عام 2017 بأسعار 2011.**
2. **إذا علمت أن سعر البرتقال في هذه الدولة في عام 2017مساويا 100 وحدة نقدية، احسب سعر كيلو البرتقال بأسعار عام 2011 .**
3. **احسب قيمة الدخل الواجب الحصول عليه عام 2017 حتى تعيش الاسرة كما كانت تعيش بعام 2011.**

مراجع الفصل :

1- حميدان، عدنان؛ الجاعوني، فريد؛ ناصر آغا، عمار؛ العواد، منذر، 2003، مبادئ الإحصاء، منشورات جامعة دمشق.

2- **Groebner David F., Shannon Patrick W. and Fry Phillip C. (2017)," Business Statistics :A Decision-Making approach", 10 th Edition. person**.

المراجع باللغة العربية :

أبو صالح، محمد صبحي؛ عوض، عدنان محمد ، 1990 ـ مقدمة في الإحصاء ـ مركز الكتب الأردني.

أبو عمة، عبد الرحمن، راضي، الحسني عبد البر، الهندي، محمود إبراهيم، 1995، الإحصاء والاحتمالات، الر
ملك سعود.

الصياد،جلال،سمرة،عادل،2001، مبادئ الإحصاء، دار حافظ.

حميدان، عدنان؛ الجاعوني، فريد؛ ناصر آغا، عمار؛ العواد، منذر، 2003، مبادئ الإحصاء، منشورات جامعة دم

حسام ،كمرجي،2011 - الاحتمال والاحصاء - جامعة دمشق – كلية الهندسة الميكانيكية والكهربائية .

كنجو، انيس، 2000 ،الاحصاء والاحتمالات – مكتبة العبيكان .

مخول، مطانيوس، غانم، عدنان،2006 مبادئ الإحصاء، كلية الآداب والعلوم الإنسانية، قسم المكتبات، جامعة دم

طبيه، احمد عبد السميع،2008-مبادئ الاحصاء- دار البداية .

المراجع باللغة الانكليزية :

Deborah Rumsey,(2010)" Statistics Essentials For Dummies" Wiley.

Freund, J.E. ,Modern elementary statistics. Prentice-Hall, 2001.

Johnson, R.A. & Bhattacharyya, G.K.(2009), Statistics: Principles and Methods,
ı. Wiley.

Groebner David F., Shannon Patrick W. and Fry Phillip C. (2017)," Business Statistic
ɔn-Making approach", 10 th Edition. person.

Weiss, N.A.(1999), Introductory Statistics. Addison Wesley.