جامعة دمشق

كلية السياحة

مقرر مبادئ الإحصاء

للسنة الأولى

إعداد د. غزوة حسن الصرن

العام الدراسي2018–2019

مفاهيم أساسية:

علم الاحصاء: هو العلم الذي يبحث في تصميم أساليب جمع البيانات والتقنيات المختلفة لتنظيم وتصنيف وعرض هذه البيانات وتلخيصها في صورة مؤشرات رقمية لوصف وقياس خصائصها الأساسية وتحليلها بغرض اتخاذ قرارات مناسبة.

جمع البيانات ← **تنظيم وعرض البيانات** ← **تلخيص البيانات** ← **تحليل البيانات واتخاذ القرارات**

يقسم علم الاحصاء إلى قسمين هما : الاحصاء الوصفي والاحصاء الاستدلالي.

الاحصاء الوصفي: مجموعة الطرائق والأساليب التي تستخدم في تنظيم وعرض وتلخيص البيانات واستكشاف خصائصها الأساسية وتلخيصها في صورة مؤشرات رقمية.

الاحصاء الاستدلالي: مجموعة الطرائق والأساليب التي تستخدم في تعميم نتائج العينة على خصائص المجتمع الذي سحبت منه وقياس العلاقات بين الخصائص المختلفة للمجتمع والتنبؤ بالقيم المستقبلية لهذه الخصائص.

أهداف علم الاحصاء: يهدف إلى الآتي:

- تبسيط البيانات الاحصائية بعرضها في جداول ورسوم بيانية وذلك لتسهيل فهمها وتحليلها.
- التعبير عن الحقائق بصورة عددية واضحة ودقيقة بدلاً من عرضها والتعبير عنها بصورة إنشائية.
- مقارنة المجموعات المختلفة وإيجاد العلاقات القائمة بينها.
- التنبؤ بالقيم المستقبلية مما يساعد على عملية التخطيط.
- استخلاص النتائج واتخاذ القرارات المناسبة بقدر كبير من الصحة وذلك بعد قيام الباحث في أي فرع من فروع العلوم المختلفة بتحليل البيانات المتوفرة لديه.

أهمية علم الاحصاء في مجال الادارة والاقتصاد والبحوث العلمية:

-الاحصاءات هي القاعدة الأساسية التي تبنى عليها سياسة الدولة في كافة المجالات (النمو الاقتصادي وتطور قطاعات الاقتصاد المختلفة).

-التعرف على المجالات الحيوية التي تعتمد على الأساليب الإحصائية في البحث والتحليل.

-يساعد الاحصاء في البحوث العلمية على تقديم أدق نوع ممكن من الوصف للمعطيات فالوصف والموضوعية من سمات العلم الحديث.

-الاسلوب الاحصائي هو الوسيلة الأساسية في دراسة الظواهر الاقتصادية وقياس العلاقات بينها وهو وسيلة للتنبؤ بالقيم المستقبلية لهذه الظواهر، يعتمد الاقتصاد القياسي والكمي على النماذج الاحصائية الاحتمالية كنموذج الانحدار بين الكمية المطلوبة والسعر، والذي يمكن من تقدير مرونة الطلب السعرية وغيرها من العلاقات.

-تستخدم الأساليب الإحصائية في إدارة جودة الإنتاج والمقارنة بين السياسات التسويقية والادارية.

-يستخدم الاحصاء في قياس تغيرات الظواهر الاقتصادية المختلفة وذلك باستخدام الأرقام القياسية، كالرقم القياسي للأسعار.

علاقة علم الاحصاء بالعلوم الأخرى:

- علم الاقتصاد: تفسير الظواهر الاقتصادية كنظريات الطلب والعرض، والعلاقة بين مستوى الدخل والإنفاق الاستهلاكي ونوع العلاقات الاقتصادية وكيفية قياسها، وفي مراقبة الانتاج في الشركة الصناعية.

- علم النفس: قياس درجات ذكاء الأشخاص.

- علم الفلك: تحديد مدارات الكواكب والنجوم.

- علم الجغرافية: جغرافيا المدن، علم الخرائط، جغرافيا المناخ.

- علم الطب والصيدلة: مقارنة الامراض وسبل علاجها، قياس كفاءة دواء معين.

- البحث العلمي: الإحصاء أداة من أدوات البحث العلمي، حيث يستخدم لمعالجة البيانات في معظم الدراسات العلمية الحديثة.

- علم الاجتماع: دراسة وتفسير الظواهر الاجتماعية، ميول ورغبات الأفراد في المجتمع وغير ذلك.
- علم السكان: تطور السكان، والتراكيب المختلفة لهم.

وظائف علم الاحصاء: له ثلاثة وظائف هي: وصف البيانات، الاستدلال الاحصائي، التنبؤ.

المجتمع: المجموعة الكلية لمفردات الدراسة سواء كانت أفراد أو أشياء، واستخلاص خصائص هذا المجتمع هو الهدف النهائي للدراسة الاحصائية.

العينة: هي مجموعة جزئية من مفردات المجتمع محل الدراسة يتم اختيارها بحيث تكون ممثلة للمجتمع تمثيلاً صحيحاً. ولصعوبة دراسة جميع أفراد مجتمع الدراسة يلجأ الباحث إلى عينة من هذا المجتمع ويتم تعميم النتائج على باقي المجتمع.

البيانات: مجموعة المشاهدات التي يتم جمعها من مفردات المجتمع أو العينة لخاصية (متغير) معينة. وقد تكون البيانات أرقام أو أسماء أو رموز أو أحرف أو كلمات أو جمل.....الخ. ومن خلال معالجة البيانات نحصل على المعلومات.

جمع البيانات: هناك عدة أساليب لجمعها منها:

1- **الاسلوب التجريبي:** يتم الحصول على البيانات عن طريق تصميم التجربة، يتم فيها قياس تأثير العامل محل الدراسة مع ثبات العوامل الأخرى، حيث نحصل على البيانات في هذه الحالة عن طريق المشاهدة.
2- **أسلوب المسح** الإحصائي: نحصل على البيانات في هذه الحالة من السجلات والتقارير وقواعد البيانات والانترنت أو عن طريق الاستبيانات والمقابلات الشخصية ويقسم أسلوب المسح إلى نوعين هما:

✓ المسح الشامل: يتم جمع البيانات من كل مفردات المجتمع محل الدراسة، ودراسة آراء طلاب السنة الأولى (سياحة) عن اسلوب الاختبارات.

✓ اسلوب المسح بالعينة: نجمع البيانات من بعض مفردات المجتمع محل الدراسة كدراسة آراء طلاب السنة الأولى في جامعة دمشق عن اسلوب الاختبارات (بعض الطلاب بأخذ عينة) ومن ثم تعميم النتائج على طلاب الجامعة في جميع الكليات.

والعينات نوعان عينات احتمالية عشوائية وعينات غير عشوائية.

أنواع العينات العشوائية:

A. **العينة العشوائية البسيطة:** تعطى كل مفردة من مفردات المجتمع نفس الفرصة (الاحتمال) في الاختيار، وتستخدم عندما تكون وحدات المجتمع متجانسة، ويتم السحب بواسطة الأرقام العشوائية، أو البطاقات المرقمة، أو دولاب الحظ....الخ. أي بالقرعة. مثال أخذ عينة عشوائية بسيطة من طلاب سنة أولى سياحة يمكن الاستعانة بأرقام الطلاب والسحب العشوائي لهذه الأرقام.

B. **العينة العشوائية الطبقية:** تستخدم عندما تكون وحدات المجتمع غير متجانسة، حيث يتم تقسيم المجتمع محل الدراسة إلى مجموعات متجانسة وغير متداخلة تسمى كل مجموعة بطبقة، ويتم اختيار عينة عشوائية بسيطة من كل طبقة حجمها يتناسب مع حجم تلك الطبقة، وتشكل هذه العينات العشوائية البسيطة بمجموعها العينة الطبقية. مثال: نريد دراسة مستوى الذكاء لطلاب الجامعة نقسم الكليات إلى قسمين كليات علمية وكليات أدبية أي طبقتين ونأخذ عينة عشوائية بسيطة من طلاب كل طبقة يتناسب حجمها مع حجم الطلاب الكلي في الطبقة.

C. **العينة العشوائية المنتظمة:** تقسم مفردات المجتمع إلى مجموعات عددها مساو لعدد مفردات العينة التي نريد اختيارها ثم نختار مفردة من المجموعة الأولى وبشكل عشوائي ولتكن المفردة الثالثة على سبيل المثال فإننا نختار المفردة الثالثة من كل مجموعة حتى يكتمل حجم العينة التي نريدها.

D. **العينة العشوائية العنقودية:** ويستخدم هذا النوع من العينات في حالة المجتمعات التي تتكون من عدة مجموعات تشكل كل مجموعة عنقوداً يتفرع منه أيضاً العديد من المجموعات. ومنها العينة العنقودية البسيطة، والعينة العنقودية ذات المرحلتين، والعينة العنقودية المتعددة المراحل. مثال: تقدير حجم الدخل في سورية: يستلزم تقسيم سورية إلى محافظات ونختار عينة عشوائية من

المحافظات كمرحلة أولى ومن ثم نقسم المحافظات المختارة إلى مجموعات من المدن ونختار عينة عشوائية بسيطة من المدن كمرحلة ثانية وذلك داخل كل محافظة تم اختيارها في المرحلة الأولى ومن ثم نقسم المدن المختارة في المرحلة الثانية إلى أحياء ونختار عينة عشوائية بسيطة من الأحياء داخل كل مدينة تم اختيارها في المرحلة الثانية.

أنواع العينات غير العشوائية: العينة العمدية (القصدية)، العينة الحصصية.

3- **أسلوب السلاسل الزمنية:** يتم الحصول على البيانات عن طريق رصد البيانات التي تعبر عن ظاهرة ما عند نقاط زمنية متتالية مثل: عدد السياح سنوياً خلال الفترة 2000-2018

انواع المتغيرات وفق طبيعة البيانات:

- المتغير: هو مقدار الشيء أو الخاصة في العنصر أو الفرد مثلاً. ويوصف المتغير بأنه متغيراً كمياً عندما يشير إلى مقدار ما لدى الفرد مثال الطول، الوزن....الخ. اما إذا كانت القيمة لا تعبر عن مقدار الخاصة بل عن وجودها أو عدمه أي إذا كان يمتلك الخاصة أم لا فإن المتغير نوعي(وصفي) حيث يمكن ترتيب الأفراد.

ومنه فإن البيانات تقسم إلى نوعين:

- **بيانات نوعية (وصفية):** هي البيانات التي يمكن حصرها في عدة أوجه وصفية(حالات) ولا يمكن إجراء عمليات حسابية عليها مثال الجنس: له وجهان أو حالتان ذكر – أنثى.
- **بيانات كمية:** نحصل عليها في شكل أعداد ويمكن ترتيبها مثل: عدد السياح، عدد الفنادق، عدد العاملين، عدد الأدلاء السياحيين...الخ. وتقسم إلى بيانات كمية منفصلة(منقطعة): وهي التي يمكن عدها مثال: عدد أفراد الأسرة وإلى كمية متصلة(مستمرة): وهي البيانات التي لا يتم عدها وإنما يتم الحصول عليها عن طريق القياس وتأخذ أي قيمة داخل مدى معين سواء كانت صحيحة أو كسرية (الطول، الوزن، درجة الحرارة...الخ.).

قياس البيانات: للبيانات مقاييس حسب نوعها منها:

1. المقياس الاسمي: مجموعة الأوجه أو الصفات التي يأخذها المتغير الوصفي(النوعي) مع عدم إمكانية ترتيبها مثال: زمرة الدم، جنسية السائح، الجنس، الحالة الاجتماعية، نتيجة الطالب في الامتحان...الخ.
2. المقياس الترتيبي: مجموعة الأوجه والصفات التي يأخذها المتغير الوصفي ويمكن ترتيبها، مثال: متغير المؤهل العلمي، تقدير الطالب في الامتحان، مستوى رضا العميل عن الخدمة...الخ.
3. مقياس الفترة: مجموعة من الأعداد أو القيم التي يأخذها المتغير الكمي وليس للصفر معنى حقيقي أي لا يعني انعدام الخاصية محل الدراسة، مثال: درجة الطالب في الامتحان، فإن الدرجة صفر لا تعني انعدام قدرة الطالب على التحصيل. أي بيانات الفترة تقاس بمقدار بعدها عن الصفر ولا تخضع للعمليات الحسابية.
4. مقياس النسبة: مجموعة الأعداد أو القيم التي يأخذها المتغير الكمي والصفر له معنى حقيقي، أي يعني انعدام الخاصية محل الدراسة، مثال: الطول، الوزن، إنتاجية الهكتار، المساحة المزروعة ، عدد الوحدات المعيبة في الإنتاج.

تمرين: حدد نوع البيانات التالية ومقياسها

لون العيون: أخضر، أزرق، أسود، بني. (نوعي أو وصفي مقياسه اسمي)

الجنسية: سوري، أردني، لبناني، امريكي، روسي (نوعي أو وصفي مقياسه اسمي)

منطقة السكن: حضر، ريف (نوعي أو وصفي مقياسه اسمي)

مستوى الطالب(تقديره): ضعيف، متوسط، جيد، ممتاز(نوعي أو وصفي مقياسه ترتيبي)

مستوى الثقافة: منخفض جدا، منخفض، متوسط، مرتفع (نوعي أو وصفي مقياسه ترتيبي)

المستوى التعليمي: أمي، ملم، ابتدائية، إعدادية، ثانوية، معهد، إجازة جامعية، دراسات عليا(نوعي أو وصفي مقياسه ترتيبي)

لون الشعر: أسود، بني، غير ذلك (نوعي أو وصفي مقياسه اسمي)

التخصص: أدبي، علمي، فندقي(نوعي أو وصفي مقياسه اسمي)

حجم السلعة: كبيرجدا، كبير، متوسط، صغير(نوعي أو وصفي مقياسه ترتيبي)

الانحراف عند الشباب: سرقة، جنس، تعاطي، (نوعي أو وصفي مقياسه اسمي)

نتيجة الطالب في الامتحان: راسب، ناجح (نوعي أو وصفي مقياسه اسمي)

الاتجاه نحو مهنة معينة: نعم، لا (نوعي أو وصفي مقياسه اسمي)

كمية الهطول المطري في فترة معينة(كمي مقياسه نسبة)

درجات الحرارة:(كمي مقياسه فترة)

وصف البيانات وتحليلها:

إذا كانت البيانات التي يراد تحليلها إحصائياً في صيغة قيم رقمية فالإحصاء يساعد الباحث في أربع صور هي:

1. يستطيع الاحصاء أن يحدد النقطة المركزية التي تجتمع حولها البيانات عن طريق استخدام مقاييس النزعة المركزية.
2. يشير الإحصاء إلى كيفية انتشار البيانات عن طريق حساب التشتت.
3. يوضح الاحصاء العلاقة التي تربط بين نوع ما من البيانات وبيانات أخرى كما هو الحال في قياس الارتباط بين المتغيرات.
4. يساعد الاحصاء على توفير بعض الاجراءات الاحصائية لاختيار الدرجة التي تتطابق أو تبعد عن تلك القيم المتوقعة أو مدى قربها من المقاييس كما هو الحال عند استخدام المقاييس الاستدلالية.

عرض البيانات: وهناك نوعان/ العرض الجدولي للبيانات. والعرض البياني للبيانات.

العرض البياني للبيانات: وهو كل ما يتعلق بتمثيل البيانات بشكل بياني بحسب ما يتطلبه نوع البيان وسوف نبين ذلك وفق طريقة علمية.

الأشكال والرسوم البيانية: تعد من أكثر الطرائق الإحصائية استخداماً في وصف وتلخيص البيانات، وذلك لبساطتها وسهولتها ووضوحها.

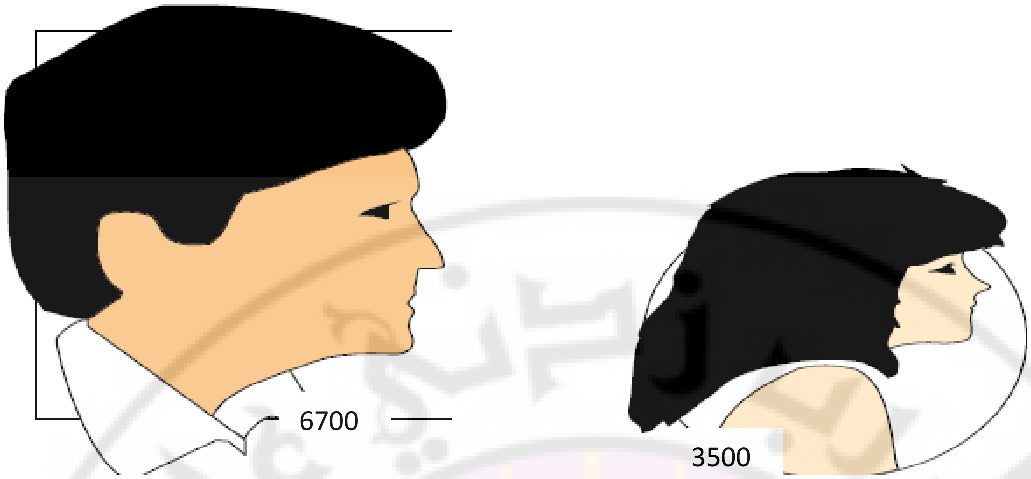
هناك أنواع عديدة من الأشكال والرسوم البيانية يمكن استخدامها لوصف البيانات لكل منها ميزاته وعيوبه من حيث درجة البساطة والوضوح ومن حيث إظهار معالم البيانات وتبعاً لذلك يختلف استخدامها حسب نوع البيانات وأحجامها وعدد المتغيرات، والغرض من وصف البيانات.

أنواع الأشكال والرسوم البيانية:

1. **الأشكال المصورة:** يعطي انطباعاً بصرياً عن مجموعة البيانات المبحوثة ولكن يعيبها عدم الدقة وإخفاء التفاصيل.
2. **الدوائر المجزأة:** pie Chart: تستخدم عندما يكون الهدف مقارنة الأجزاء المختلفة بالنسبة للمجموع الكلي وعدد الأجزاء المقارنة قليل نسبياً.
3. **الأعمدة** Bar Chart: قد تكون أعمدة بسيطة، متلاصقة مزدوجة، مركبة، مجزأة. تستخدم عندما تكون أجزاء الظاهرة المقارنة كثيرة العدد نسبياً، وعندما نرغب في توضيح قيم الأجزاء المقارنة.
4. **الخطوط البيانية:** تستخدم عندما يكون عدد المفردات كبير نسبياً أو عندما يكون الغرض توضيح العلاقة بين المتغيرات لفترات زمنية متعاقبة كما في حالة السلاسل الزمنية.

أمثلة على الأشكال والرسوم البيانية:

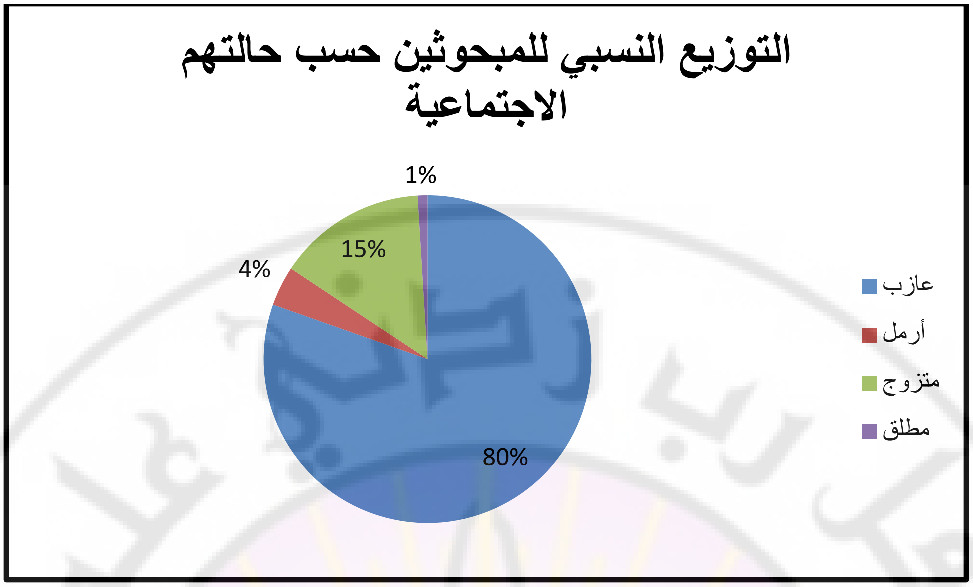
1- الأشكال المصورة: مثلاً التمثيل البياني للمبحوثين حسب متغير الجنس حيث يوجد في مجتمع الدراسة 3500 أنثى، 6700 ذكر.



2- الدوائر المجزأة: لدينا توزيع المبحوثين حسب الحالة الاجتماعية كما يلي: 48 مطلق، 732 متزوج، 195 أرمل، 4000 عازب والمطلوب عرض البيانات بيانياً. بما أن الغرض هو مقارنة الأجزاء المختلفة بالنسبة للمجموع الكلي وعدد الحالات(الأوجه) قليل نسبياً، فإننا نستخدم الدوائر المجزأة ولرسم الدائرة نقوم بالآتي:

- رسم دائرة بمقياس رسم مناسب.
- نحسب نسبة كل مجموعة إلى المجموع الكلي.
- تقسيم 360 درجة على المجموعات حسب نسبة كل مجموعة فنحصل على تقسيم الدائرة إلى قطاعات تتناسب كل منها مع حجمها من المجموع الكلي.
- نقوم بتقسيم مساحة الدائرة إلى القطاعات المذكورة حسب قيمة زوايا القطاع كما في الشكل أدناه، ويتم تمييز كل قطاع من هذه القطاعات بلون معين وكتابة اسم الجزء الذي يخص كل قطاع إلى جانبه أو داخله.

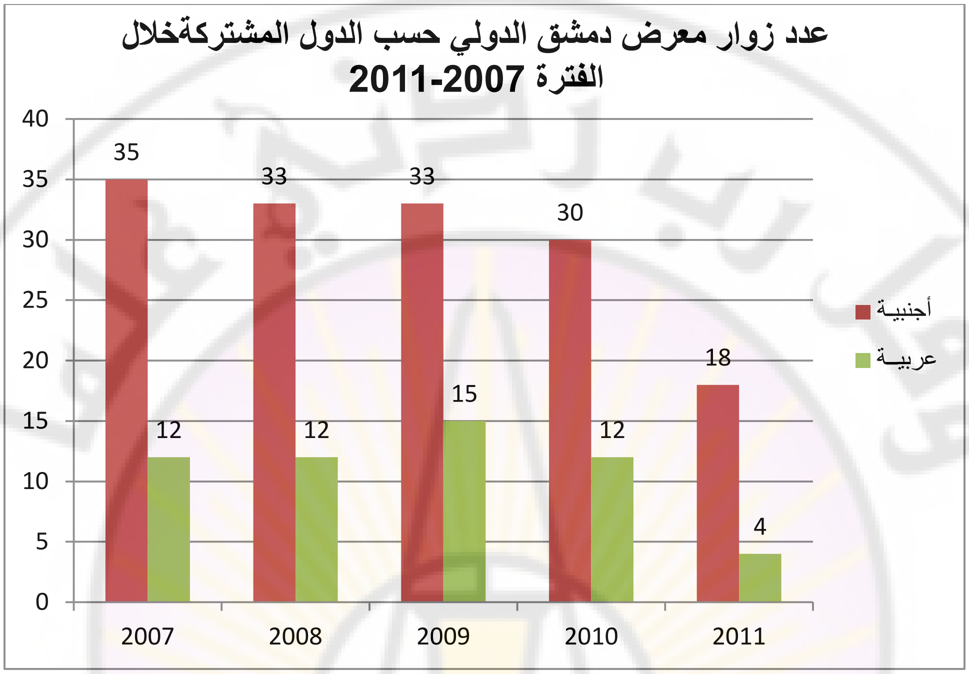
مثال: التوزيع النسبي للمبحوثين حسب الحالة الاجتماعية: لمتغير الحالة الاجتماعية أربعة حالات فقط الشكل البياني المناسب هو الدوائر المجزأة كما يلي:



3- الأعمدة البسيطة: تستخدم هذه الرسوم لعرض تقسيمات ظاهرة واحدة في نفس الفترة الزمنية أو لعرض الظاهرة في عدة فترات زمنية حيث يخصص لكل فترة زمنية عمود مستقل مع ترك مسافات متساوية بين كل عمود وآخر.

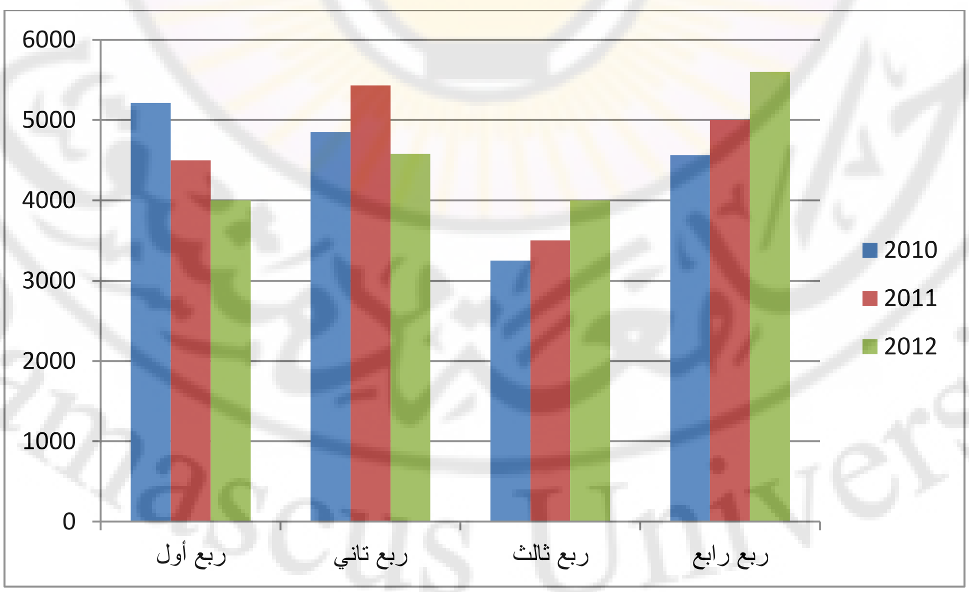
يبين الشكل التالي التوزيع التكراري المطلق للقادمين الأجانب خلال الفترة2007–2011 كما يلي:

4- الأعمدة المتلاصقة: تستخدم عندما يكون المطلوب عرض ظاهرة معينة في عدة فترات زمنية أو عدة أماكن جغرافية او التقسيم خلال فترة زمنية واحدة. ففي هذه الحال تمثل كل فترة بعدد من الأعمدة المتلاصقة مساوية لعدد أوجه الظاهرة مع تمييز كل عمود بلون او تظليل

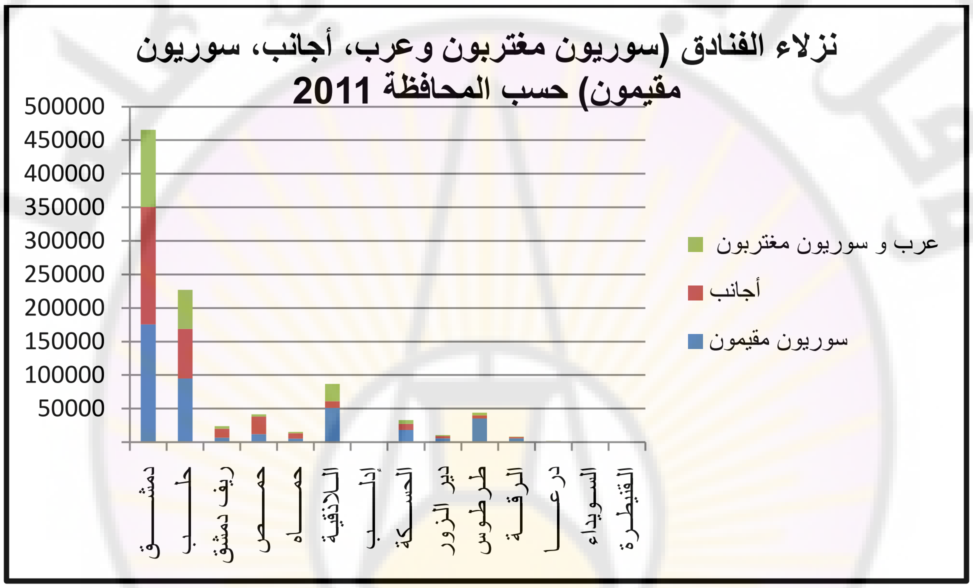


أعمدة متلاصقة مزدوجة

أو قد تكون غير مزدوجة كما يلي:

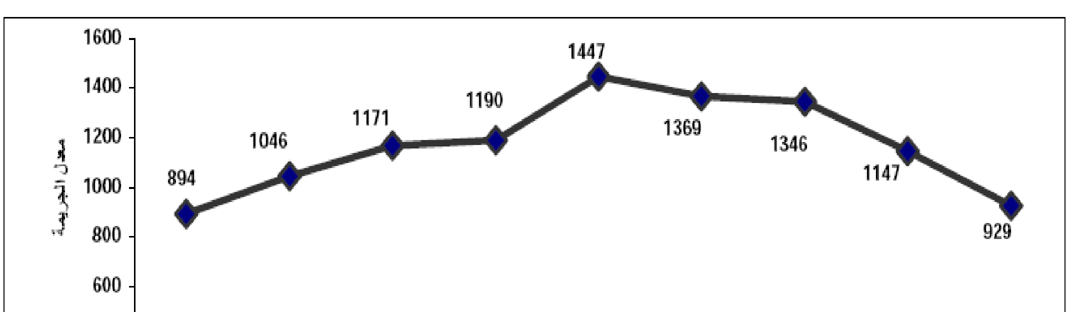


5- الأعمدة المجزأة: تستخدم هذه الرسوم البيانية عندما تكون القيمة الاجمالية للظاهرة موزعة على مجموعات فرعية مميزة ويشمل البيان عدة فترات زمنية أو تقسيمات. حيث يتم اللجوء إلى رسم عمود يمثل قيمة الظاهرة الكلية لكل فترة زمنية، ويقطع هذا العمود إلى أجزاء تتناسب مع قيم المجموعات الجزئية ويميز كل جزء من الأعمدة بألوان معينة أو بأشكال وتظليلات مختلفة لكل مجموعة فرعية على حدة. ولإجراء عملية المقارنة بسهولة يفضل استخدام النسب المئوية.



6- الأعمدة المركبة: تستخدم عندما نريد عرض قيم الظاهرة الاجمالية على مجموعات فرعية حيث كل مجموعة فرعية موزعة بحد ذاتها إلى مجموعات جزئية. كأن تمثل كل مجموعة فرعية بأعمدة متلاصقة ومجزأة بآن واحد.

7- الخطوط البيانية: تستخدم للتعبير عن العلاقة بين ظاهرتين بخط بياني، يتم رسم محورين متقاطعين، يمثل المحور الأفقي أحد الظاهرتين والمحور العمودي الظاهرة الأخرى، وقد يتعلق الأمر بتمثيل قيم ظاهرة ما خلال فترة زمنية معينة، وفي هذه الحالة يمثل المحور الأفقي الفترات الزمنية المتعاقبة، ويمثل المحور العمودي قيم الظاهرة في تلك الفترة. ونمثل أزواج القيم على الرسم ويوصل بين تلك النقاط فنحصل على الخط البياني كما في الشكل التالي:



يصعب استخدام الأشكال والرسوم البيانية لتوضيح المقارنات البيانية عندما يزيد عدد المتغيرات عن حد معين وفي تلك الحالة تستخدم الجداول الاحصائية لعرض البيانات ومقاييس النزعة المركزية والتشتت لتلخيصها.

العرض الجدولي للبيانات: بعد عملية تبويب البيانات وتصنيف البيانات أو الصفات التي تميز المفردات ترصد النتائج في جداول مناسبة توضح الشكل النهائي للمجموعات وفق الآتي:

- تصنيف جغرافي: حيث تجمع الوحدات التي تشترك بصفة مكانية واحدة، مثل: توزع الطلاب حسب مناطق سكنهم.
- تصنيف تاريخي أو زمني: حيث تجمع الوحدات التي تشترك بصفة زمانية واحدة مثلاً عدد الطلاب في كلية السياحة في كل سنة دراسية منذ افتتاحها ولغاية الآن فنحصل على ما يسمى السلاسل الزمنية.
- تصنيف نوعي: حيث تجمع الوحدات التي تشترك بصفة معينة واحدة كتوزيع الموظفين في فندق معين حسب المؤهل العلمي.
- تصنيف كمي: يتم تجميع الوحدات التي تشترك بصفة معينة تأخذ شكلاً رقمياً في مجموعة واحدة مثلاً: توزيع السكان حسب فئات الأعمار.

وتسمى الجداول وفق التصنيفات السابقة بجداول التوزيع التكرارية.

الشروط الواجب مراعاتها عند إعداد الجدول:

1. وجود عنوان واضح له.
2. ذكر المصدر الذي أخذت منه البيانات.
3. تسجيل الملاحظات الخاصة في أسفل الجدول وتحديدها بعلامات خاصة.
4. توضيح عناوين الأعمدة والصفوف.
5. يفضل فصل الأعمدة بخطوط رأسية عندما تكون متعددة.

يوجد شكلين للجداول:

1- الجدول البسيط: جيث يحتوي على عمود واحد لبيان موضوع الدراسة ويقابله عمود آخر بعدد التكرارات مثال: توزع السياح حسب الجنسية

الجدول(1): توزع السياح حسب جنسياتهم لعام 2008 الوحدة: سائح*

الجنسية	عدد السياح (التكرارات)
روسية	2451
بريطانية	645
فرنسية	245
عربية	5646
ماليزية	241
صينية	579
يابانية	122
هندية	254
إيطالية	345

المصدر: سجلات الهجرة والجوازات

- الارقام افتراضية والغاية بيان شروط إعداد الجدول.

2- الجدول المتعدد: أي يتم تصنيف البيانات حسب ظاهرتين أو أكثر مثل الجدول الذي يبين توزع السياح حسب الجنسية والجنس كما يلي:

الجدول(2): توزع السياح حسب جنسياتهم لعام 2008 الوحدة: سائح*

الجنسية	عدد السياح الذكور (التكرارات)	عدد السياح الإناث (التكرارات)
روسية	1151	1300
بريطانية	500	145
فرنسية	100	145
عربية	3000	2646
ماليزية	141	100
صينية	279	300

يابانية	32	90
هندية	200	54
إيطالية	200	145

المصدر: سجلات الهجرة والجوازات

- الارقام افتراضية والغاية بيان شروط إعداد الجدول.

وقد يكون الجدول على الشكل التالي:

الجدول(3): توزع السياح حسب القارة التي ينتمون إليها والجنس لعام 2008 الوحدة: سائح*

القارة	عدد السياح 2008		عدد السياح2009	
	ذكور	إناث	ذكور	إناث
أوروبية	254	245	121	451
آسيوية	658	1455	1215	1545
افريقية	2458	548	3551	2545
أمريكية	121	141	124	135
استرالية	210	142	342	215

المصدر: سجلات الهجرة والجوازات

- الارقام افتراضية والغاية بيان شروط إعداد الجدول.

التوزيعات التكرارية: هي عبارة عن جداول لجميع الأوجه(الحالات) أو القيم التي يمكن أن يأخذها المتغير موضوع الدراسة، وعدد المفردات التي تمثل تكرارات مقابلة لكل وجه(حالة) أو قيمة.

مثال: لدينا عدد من السياح الأجانب والبالغ عددهم/30/ سائحاً، كانت جنسياتهم كما يلي:

فرنسي، ايطالي، اسباني، فرنسي، ايطالي، الماني، روسي، صيني، روسي، ايراني، فرنسي، ايراني، ايطالي، اسباني، روسي، ايراني، صيني، ايراني، الماني، الماني، روسي، ايراني، روسي، بريطاني، روسي، ايراني، روسي، صيني، روسي، صيني.

المطلوب: وضع البيانات السابقة في جدول توزيع تكراري؟

الجنسية	علامات احصائية(التفريغ)	التكرار المطلق
فرنسي	III	3
ايطالي	III	3
اسباني	II	2
ألماني	III	3
روسي	𝍸 III	8
صيني	IIII	4
ايراني	𝍸 I	6
بريطاني	I	1
المجموع		30

ملاحظة: عندما نريد حساب نسبة السياح من كل جنسية نقوم بتطبيق العلاقة التالية:

$$p = \frac{f_i}{\sum_{i=1}^{n} f_i} \times 100$$

أي بمعنى آخر نسبة خاصة معينة هي ناتج قسمة عدد الأفراد الذين يتمتعون بالخاصة على المجموع الكلي للأفراد مضروباً بـــ 100

مثال(1): نسبة السياح من الجنسية الايرانية = (عدد السياح الايرانيين/ عدد السياح الاجمالي)×100=(6/30)×100=20%.

مثال(2): أخذت عينة عشوائية بسيطة بحجم n=10 فوجد فيها أن 4 أشخاص مدخنين أوجد نسبة المدخنين ونسبة غير المدخنين في العينة؟

نسبة المدخنين= (عدد المدخنين/ حجم العينة)×100=(4/10)×100=40%.

نسبة غير المدخنين = (عدد غير المدخنين/ حجم العينة)×100=(6/10)×100=60%.

او بطريقة أخرى نسبة غير المدخنين=100−40%=60%.

تبويب البيانات في جدول توزيع تكراري:

لتبويب البيانات نقوم بالخطوات التالية:

1− نحسب المدى الكلي R حيث: $R = X_{MAX} - X_{MIN}$ أي المدى الكلي هو الفرق بين أكبر قيمة وأصغر قيمة بين البيانات(القيم أو المشاهدات).

2− نحسب عدد المجالات (الفئات)m من خلال العلاقة الرياضية التالية:

$m = 1 + 3.33log(n)$ ، حيث:

n حجم العينة أي عدد القيم أو المشاهدات الكلي،

بينما $log(n)$ هو اللوغاريتم العشري لــn.

3− حساب طول(مدى) الفئة(المجال)h وذلك من خلال إيجاد ناتج قسمة المدى الكلي على عدد المجالات(الفئات) أي: $h = \frac{R}{m}$

4− نبدأ بكتابة المجالات(الفئات) بأن نضع أصغر قيمة في البيانات حداً أدنى للفئة الأولى ونضيف h إلى الحد الأدنى فنحصل على الحد الثاني للفئة الأولى ونضع المجالات نصف مفتوحة بحيث يكون المجال مغلق عند الحد الأدنى ومفتوح عند الحد الثاني، أما المجال الثاني فيكون حده الأدنى هو الحد الثاني للمجال الأول ونضيف لحده الأدنى h فنحصل على الحد الثاني للمجال الثاني ويكون المجال الثاني مغلق عند حده الأدنى ومفتوح عند الحد الثاني له، وهكذا حتى نحصل على عدد المجالاتm,

ملاحظة: إذا حددت قيمة m في نص المسألة فليس هناك ضرورة لتطبيق العلاقة الرياضية لحساب قيمتها فيتم متابعة الحل على القيمة المعطاة في نص المسألة.

5− نحسب التكرار المطلق f_i لكل فئة(ومجال من المجالات السابقة وذلك بأن نحسب تكرارات كل قيمة من القيم التي تنتمي للمجال ونقوم بجمع هذه التكرارات للقيم فنحصل على التكرار المقابل للمجال، ويجب أن يكون مجموع التكرارات المطلقة لكافة المجالات يساوي n عدد القيم الكلي.

مثال:

لدينا درجات 20 طالب في مادة الاحصاء كما يلي:

20,40,60,65,70,40,50,70,75,61,82,70, 60,35,40,46,50,60,70,75

بوب البيانات السابقة في جدول توزيع تكراري.

الحل:

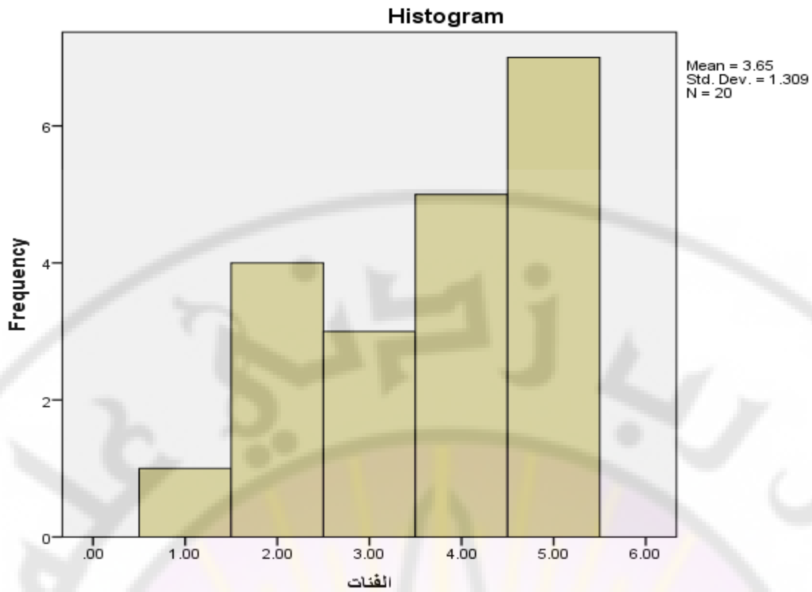
المدى الكلي: $R = X_{MAX} - X_{MIN} = 82 - 20 = 62$

عدد المجالاتm: $m = 1 + 3.33log(n) = 1 + 3.3log(20) = 5.293 \approx 5$

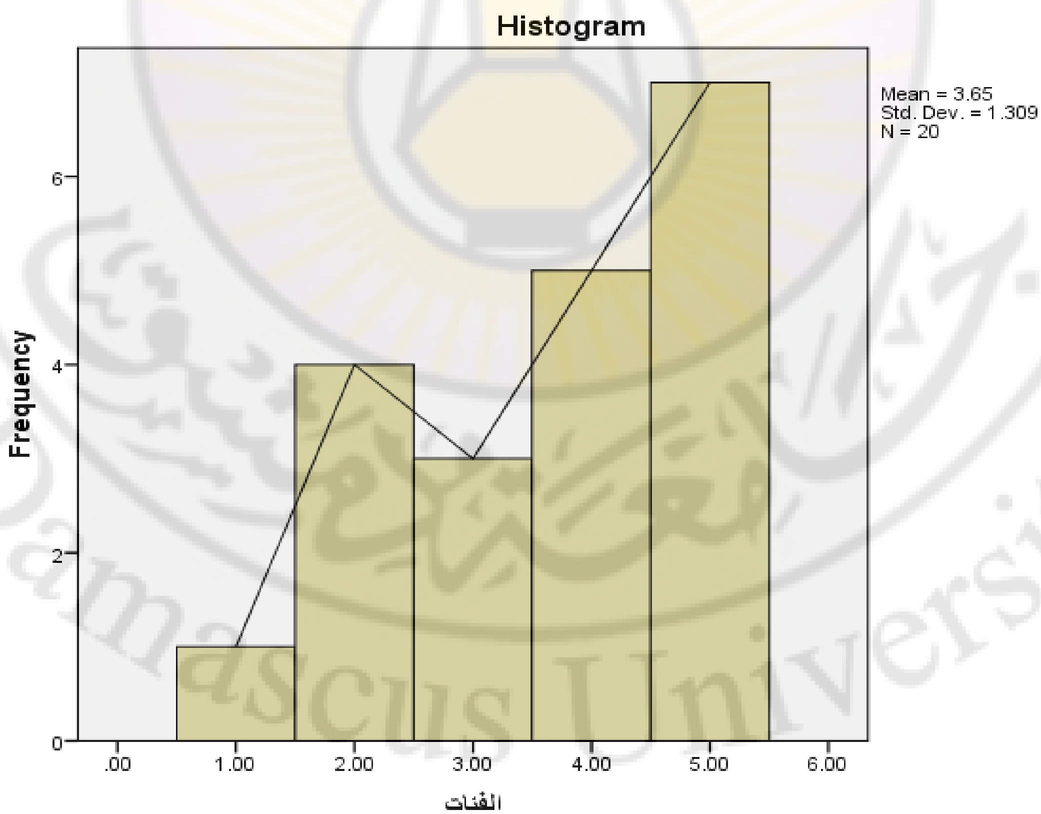
مدى(طول) المجال(الفئة)= $h = \frac{R}{m} = \frac{62}{5} = 12.4 \approx 12$

الفئات(المجالات)	التكرارات المطلقة$\boldsymbol{f_i}$
$[20 - 32[$	1
$[32 - 44[$	4
$[44 - 56[$	3
$[56 - 68[$	5
68 فاكثر	7
المجموع	$\sum_{i=1}^{m=5} f_i = 20$

المدرج التكراري: هو عبارة عن مستطيلات متلاصقة طول قاعدتها(عرضها)هو طول الفئة وارتفاعها(طولها) هو التكرار المقابل لكل فئة. كما في الشكل التالي:

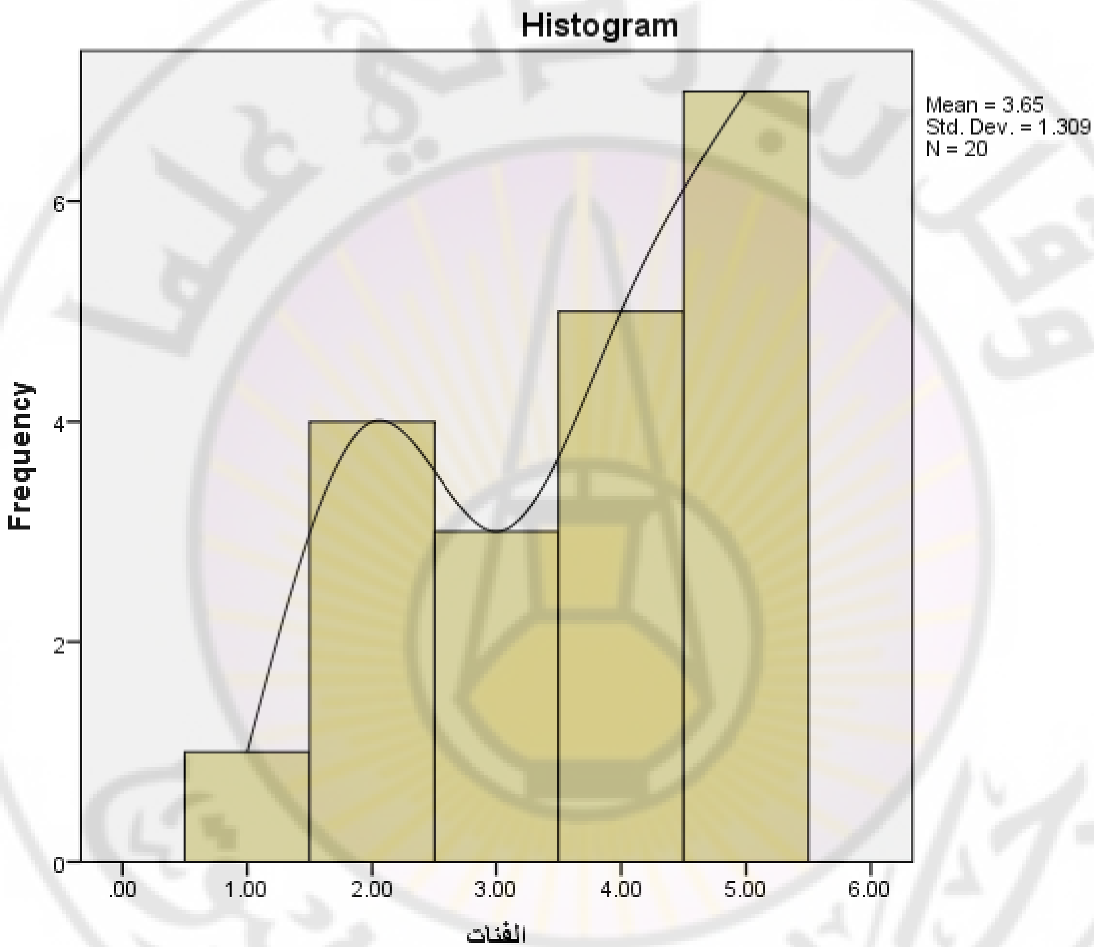


المضلع التكراري: لرسم المضلع التكراري نحسب مراكز المجالات (الفئات)، ومن ثم نرسم مراكز الفئات مقابل التكرارات المطلقة المقابلة لها نصل بين مراكز الفئات بقطع مستقيمة فنحصل على المضلع التكراري. كما في الشكل التالي الذي يوضح المضلع التكراري مع المدرج التكراري:



ملاحظة: في البرمجيات يضع على المحور الأفقي رقم كل فئة بحيث تمثل القيمة على المحور الأفقي عند كل مستطيل حده الأدنى للفئة التالية والأعلى للفئة السابقة.

المنحنى التكراري: نحسب مراكز الفئات ونرسمها مقابل التكرارات المطلقة المقابلة لها ونصل بين مراكز الفئات بخط منحني فنحصل على المنحنى التكراري. كما في الشكل التالي:



التكرار التجميعي الصاعد: نجمع التكرارات المطلقة المقابلة لكل فئة من البداية حتى نحصل على المجموع الكلي للتكرارات المطلقة بمعنى التكرار التجميعي الصاعد المقابل للفئة الأخيرة يجب أن يساوي مجموع التكرارات المطلقة. ويكون العمود الناتج معبراً عن عدد الأفراد الذين حصلوا على درجة أقل من الحد الأعلى للفئة.

- التكرار التجميعي الهابط: نبدأ بالمجموع الكلي للبيانات ونضعه كتكرار تجميعي هابط للفئة الأولى، ونطرح منه التكرارات المطلقة المقابلة لكل فئة من البداية حتى نحصل على التكرار

التجميعي الهابط للفئة الأخيرة، بمعنى التكرار التجميعي الهابط الناتج معبراً عن عدد الأفراد الذين حصلوا على درجة هي الحد الأدنى للفئة فأكثر.

- التكرار النسبي: ونحصل عليه بإيجاد ناتج قسمة كل تكرار مطلق لكل فئة على مجموع التكرارات المطلقة مضروباً بـــ100 أي: التكرار النسبي للفئة = التكرار المطلق للفئة / مجموع التكرارات المطلقة)×100

التكرار النسبي= $100 \times \frac{f_i}{\sum_{i=1}^{m} f_i}$

مجموع االتكرارات النسبية = 100

- التكرار النسبي الصاعد: نقوم بجمع التكرارات النسبية من البداية حتى نحصل على مجموع التكرارات النسبية للفئة الأخيرة والذي يساوي100 ونحصل من خلال هذه التكرارات على نسبة الأفراد الذين حصلوا على درجة أقل من الحد الأعلى للفئة.
- التكرار النسبي الهابط: نبدا بمجموع التكرارت النسبية وهو 100 ونطرح منه التكرار النسبي المقابل لكل فئة حتى نحصل على التكرار النسبي للفئة الأخيرة ونحصل من خلال هذه التكرارات على نسبة الأفراد الذين حصلوا على درجة الحد الأدنى للفئة فأكثر.

مثال:

الفئات(المجالات)	التكرارات المطلقةf_i	التكرارالتجميعي الصاعد	التكرار التجميعي الهابط	التكرار النسبي%	التكرار النسبي الصاعد	التكرار النسبي الهابط
$[20-32[$	1	1	20	5	5	100
$[32-44[$	4	5	19	20	25	95
$[44-56[$	3	8	15	15	40	75
$[56-68[$	5	13	12	25	65	60
68 فاكثر	7	20	7	35	100	35
المجموع	$\sum_{i=1}^{m=5} f_i = 20$			100		

عدد الطلاب الذين درجاتهم أقل من 56 هو 8 بينما نسبة الطلاب الذين درجاتهم أقل من 56 هو 40%.

عدد الطلاب الذين درجاتهم 32 فأكثر هو 19 ونسبتهم هي 95%

مقاييس النزعة المركزيةMeasures of central tendency

هي الطرائق الاحصائية التي تقوم بحساب القيمة التي تتمركز حولها معظم القيم(المشاهدات) تسمى مقاييس النزعة المركزية وسندرسها حسب أنواع البيانات(مفردة، مرتبة، مبوبة) وهذه المقاييس هي: المتوسط الحسابي، الوسيط، المنوال، المتوسط الهندسي.

المقياس الأول: المتوسط الحسابي Mean:

سنميز الحالات التالية:

أولاً :حالة البيانات المفردة: أي كل قيمة تتكرر مرة واحدة فقط. يعرف المتوسط الحسابي بأنه ناتج قسمة مجموع القيم(المشاهدات) على عددها أي أن $\bar{x}=\frac{\sum_{i=1}^{n}x_i}{n}$

حيث: x_i المشاهدات او القيمة ذات الترتيبi. وi=1,2,3,.....n، وn: عدد المشاهدات (القيم). $\bar{x}$: المتوسط الحسابي.

مثال: احسب المتوسط الحسابي للمفردات التالية بالطريقة العامة: 10,2,4,7,5,8

$$\bar{x}=\frac{\sum_{i=1}^{n}x_i}{n}=\frac{10+2+4+7+5+8}{6}=\frac{36}{6}=6$$

التفسير بفرض أن البيانات السابقة هي درجات طلاب الدرجة العظمى من/10/ فيكون تفسير المتوسط الحسابي بأنه لو تساوت درجات الطلاب لحصل كل واحد منهم على 6 درجات.

مثال(2): إذا كان مجموع درجات /10/ طلاب هو 230 فأوجد المتوسط الحسابي لدرجاتهم؟

$$\bar{x}=\frac{\sum_{i=1}^{n}x_i}{n}=\frac{230}{10}=23$$

التفسير: لوتساوت درجات الطلاب لحصل كل واحد منهم على 23 درجة.

مثال(3): إذا كان المتوسط الحسابي لدرجات عدد من الطلاب هو 56 ومجموع درجاتهم هو 2800 اوجد عدد هؤلاء الطلاب؟

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{2800}{n} = 56$$

ومنه طالباً n=2800/56=50

ثانياً البيانات مرتبة: (قيم مع تكرارات): يعطى المتوسط الحسابي بالعلاقة التالية:

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

مثال: لتكن لدينا المشاهدات التالية مع تكرارتها المطلقة كما يلي:

القيم x_i	التكرارات المطلقة f_i	جداء القيم بتكراراتها المطلقة $f_i x_i$
2	1	2
4	2	8
5	1	5
3	2	9
8	1	8
المجموع	$\sum_{i=1}^{5} f_i$=8	$\sum_{i=1}^{5} f_i x_i = 32$

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{32}{8} = 4$$

يفسر بالطريقة نفسها.

مثال(2): مجموعة من المشاهدات المتكررة وسطها الحسابي 14 ومجموع تكرارتها 30 اوجد مجموع هذه الشماهدات.

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} \Rightarrow 14 = \frac{\sum_{i=1}^{n} f_i x_i}{30} \Rightarrow \sum_{i=1}^{n} f_i x_i = 30 \times 14 = 420$$

ثالثاً: المتوسط الحسابي للبيانات المبوبة: يعطى بالعلاقة التالية:

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i \acute{x}_i}{\sum_{i=1}^{n} f_i}$$

حيث $\acute{x}_i$ هي مركز المجال(الفئة) والتي تحل محل القيم وتحسب بغيجاد المتوسط الحسابي لحدي الفئة أي مركز المجال(الفئة)= (الحد الأعلى للفئة+الحد الأدنى لفئة)/2

مثال: لدينا البيانات المبوبة التالية:

الفئات	التكرارات المطلقة f_i	مركز المجال(الفئة) $\acute{x}_i$	$f_i\acute{x}_i$
$[0-20[$	5	(0+20)/2=10	50
$[20-40[$	2	30	60
$[40-60[$	7	50	350
$[60-80[$	8	70	560
$[80-100]$	4	90	360
المجموع	26		1380

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i\acute{x}_i}{\sum_{i=1}^{n} f_i} = \frac{1380}{26} = 53.077$$

المتوسط الحسابي المرجح: إذا كان لدينا أكثر من مجموعة من البيانات بحيث يكون لكل مجموعة خصائص مشتركة فإن المتوسط الحسابي للبيانات ككل هو المتوسط المرجح الذي يعطى بالعلاقة التالية:

$$\bar{x} = \frac{n_1\overline{x_1} + n_2\overline{x_2} + \cdots + n_n\overline{x_n}}{n_1 + n_2 + \cdots \ldots + n_n}$$

مثال: لدينا درجات الذكور كما يلي: 50,25,40,70,65 ودرجات الاناث كما يلي: 50,80,75,70 أوجد المتوسط الحسابي لدرجات الطلاب؟

بما ان الطلاب مقسمين إلى مجموعتين فإن المتوسط الحسابي للطلاب هو متوسط مرجح كما يلي:

$$\bar{x} = \frac{n_1\overline{x_1} + n_2\overline{x_2}}{n_1 + n_2}$$

لدينا $n_1 = 5$ و $n_2 = 4$

المنتوسط الحسابي لدرجات الذكور $\overline{x_1} = \frac{\sum x_{i1}}{n_1} = \frac{50+25+40+70+65}{5} = \frac{250}{5} = 50$

المتوسط الحسابي لدرجات الاناث$\overline{x_2} = \frac{\sum x_{i2}}{n_2} = \frac{50+80+75+70}{4} = \frac{275}{4} = 68.75$

$$\bar{x} = \frac{n_1\overline{x_1} + n_2\overline{x_2}}{n_1 + n_2} = \frac{5 \times 50 + 4 \times 68.75}{5 + 4} = 58.333$$

خصائص المتوسط الحسابي:

1- مجموع انحرافات القيم عن متوسطها الحسابي يساوي الصفر.

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

بيانات مرتبة: $\sum f_i(x_i - \bar{x}) = 0$

بيانات مبوبة: $\sum f_i(\acute{x}_i - \bar{x}) = 0$

2- المتوسط الحسابي يتأثر بالقيم المتطرفة والشاذة.

3- مجموع مربعات انحرافات القيم عن متوسطها الحسابي أقل من مجموع مربعات انحرافاتها عن أي قيمة أخرى لتكن a. أي أن:

البيانات مفردة$\sum(x_i - \bar{x})^2 < \sum(x_i - a)^2$

البيانات مرتبة$\sum f_i(x_i - \bar{x})^2 < \sum f_i(x_i - a)^2$

البيانات مبوبة$\sum f_i(\acute{x}_i - \bar{x})^2 < \sum f_i(\acute{x}_i - a)^2$

4- إذا كان هناك مجموعة من المفردات ووسطها الحسابي $\bar{x}$ وقمنا بتبديل المفردات بضربها بعدد a وجمع العدد b إلى ناتج الضرب فإن المتوسط الحسابي الجديدللبيانات بعد التعديل يكون كما يلي:

المتوسط الحسابي الجديد= المتوسط الحسابي القديم مضروباً بالعدد a وجمع العدد b إلى الناتج أي: $\bar{x}$ الجديد=($\overline{x}$ القديم×a) +b حيث a,b اعداد حقيقية.

5- إذا أضفنا إلى كل قيمة العدد a فإن المتوسط الحسابي الجديد= المتوسط الحسابي القديم +a

6- إذا ضربنا كل مفردة بالعدد a فيكون المتوسط الحسابي الجديد يساوي المتوسط الحسابي القديم مضروباً بالعدد a.

7- إذا طرحنا من كل مفردة القيمة a فإن المتوسط الحسابي الجديد يساوي المتوسط الحسابي القديم مطروحاً منه العددa.

مزايا المتوسط الحسابي:

- يعتمد في حسابه على كل القيم والمشاهدات.
- سهل الحساب والفهم.
- يتوفر فيه القابلية للتعامل الجبري.

عيوب المتوسط الحسابي: - يتأثر بالقيم الشاذة والمتطرفة.

- عدم إمكانية حسابه في البيانات الوصفية.
- يصعب حسابه في جالة الجداول التكرارية المفتوحة.

المقياس الثاني: الوسيط Median (Me):

يعرف على أنه القيمة التي تتوسط البيانات بعد ترتيبها ترتيباً تصاعدياً أو تنازلياً بحيث يكون عدد المفردات التي قبلها يساوي عدد المفردات التي بعدها بمعنى 50% من البيانات أقل من قيمة الوسيط و50% منها أكبر من قيمة الوسيط. وسنميز الحالات التالية للبيانات في حسابه:

أولاً: البيانات مفردة: نميز حالتين

الحالة الأولى: عدد المفردات عدد فردي: لحساب الوسيط نقوم بترتيب البيانات ترتيباً تصاعدياً أو تنازلياً، نوجد القيمة التي ترتيبها $\frac{n+1}{2}$ أي أن $Me = X_{\frac{N+1}{2}}$

مثال: لدينا المفردات التالية: 50,60,80,70,40,100,90 اوجد الوسيط؟

ترتيب البيانات تصاعدياً: 40,50,60,70,80,90,100

عدد المفردات n=7 عدد فردي فالوسيط هو القيمة التي ترتيبها $4 = \frac{7+1}{2} = \frac{n+1}{2}$ أي القيمة الرابعة وتساوي 70 أي أن $Me = 70$ التفسير: 50% من المبحوثين حصلوا على درجات أقل من 70 و50% منهم حصلوا على درجات أعلى من 70.

الحالة الثانية: عدد المفردات عدد زوجي: لحساب الوسيط نقوم بترتيب البيانات ترتيباً تصاعدياً أو تنازلياً، نوجد القيمتين الوسيطيتين اللتين ترتيبهما $\frac{n}{2}$ ، $\frac{n}{2}+1$ أي أن $Me = \frac{X_{\frac{n}{2}+1} + X_{\frac{n}{2}}}{2}$

أي أن الوسيط هو المتوسط الحسابي للقيمتين الوسيطيتين $X_{\frac{n}{2}+1}$ و $X_{\frac{n}{2}}$.

مثال: 40,50,60,70,100,80 اوجد Me؟

n=6 عدد البيانات عدد زوجي.

لحساب الوسيط نرتب البيانات تصاعدياً كما يلي: 40, 50, 60,70, 80,100

القيمة الوسيطية الأولى التي ترتيبها n/2 أي: $X_{\frac{n}{2}}$ هي $x_{6/2} = x_3$ القيمة الثالثة وهي 60

القيمة الوسيطية الثانية والتي ترتيبها $\frac{n}{2}$+1 أي $X_{\frac{n}{2}+1}$ وهي $x_{\frac{6}{2}+1} = x_{3+1} = x_4$ أي القيمة الرابعة وهي 70 فيكون الوسيط المتوسط الحسابي للقيمتين 60 و70 أي أن $Me = \frac{X_{\frac{n}{2}+1} + X_{\frac{n}{2}}}{2} = \frac{x_{3+x_4}}{2} =$ $\frac{60+70}{2} = 65$ التفسير 50% من المبحوثين درجاتهم أعلى من 65 و50% منهم درجاتهم أقل من 65.

ثانياً: البيانات مرتبة:

لحساب الوسيط نقوم بالاتي:

- نرتب البيانات ترتيباً تصاعدياً أو تنازلياً.
- نوجد ترتيب الوسيط $\frac{\sum f_i}{2}$
- نوجد التكرار التجميعي الصاعد (ت ت ص↑).
- نبحث في عمود التكرار التجميعي الصاعد عن **أول** عدد يساوي أو أكبر من ترتيب الوسيط فتكون القيمة المقابلة له هي الوسيط.

مثال:

X_i	f_i	ت ت ص ↑
2	1	1
4	3	4
10	2	6
14	4	10

17	1	11
المجموع	11	

ترتيب الوسيط = n/2=11/2=5.5

اول تكرار تجميعي صاعد يساوي أو أكبر من ترتيب الوسيط5.5 هو 6 القيمة المقابلة له هي 10 ومنه Me=10 نفس التفسير.

ثالثاً: الوسيط لبيانات مبوبة: لحساب الوسيط لبيانات مبوبة نقوم بالاتي:

1- نوجد تكرار تجميعي صاعد.

2- نوجد ترتيب الوسيط $\frac{\sum f_i}{2}$.

3- نحدد الفئة الوسيطية وذلك بأن نبحث في عمود التكرار التجميعي الصاعد عن أول عدد يساوي أو أكبر من ترتيب الوسيط فتكون الفئة المقابلة له هي فئة الوسيط.

4- نطبق العلاقة الرياضية التالية:

$Me = L_k + \frac{\frac{\sum f_i}{2} - F_{k-1}}{F_{k+1} - F_{k-1}} h_k$ حيث Me الوسيط، L_k الحد الأدنى للفئة الوسيطية، $\frac{\sum f_i}{2}$ ترتيب الوسيط، F_{k+1} التكرار التجميعي الصاعد للفئة التي تلي الفئة الوسيطية، F_{k-1} التكرار التجميعي الصاعد للفئة التي تسبق الفئة الوسيطية، h_k طول(مدى) الفئة الوسيطية.

مثال:

الفئات	fi	ت ت ص
$[50 - 60[$	8	8
$[60 - 70[$	10	18
$[70 - 80[$	16	34
$[80 - 90[$	14	48
$[90 - 100[$	10	58
$[100 - 110[$	5	63
$[110 - 120[$	2	65
المجموع	65	

نوجد ترتيب الوسيط = 65/2=32.5 ، نبحث في عمود التكرار التجميعي الصاعد عن أول عدد يساوي أو يتجاوز ترتيب الوسيط وهو 34 يقابل الفئة 80-70 فتكون هي الفئة الوسيطية. ومنه يكون:

$L_k = 70$ ، $F_{k+1} = 48$ ، $F_{k-1} = 18$ ، h_k طول(مدى) الفئة الوسيطية = الحد الأعلى للفئة الوسيطية – الحد الأدنى للفئة الوسيطية = 80–70=10.

$$Me = L_k + \frac{\frac{\sum f_i}{2} - F_{k-1}}{F_{k+1} - F_{k-1}} h_k = 70 + \frac{32.5 - 18}{48 - 18} \times 10 = 74.833$$

ويفسر بنفس الطريقة.

ملاحظة: يمكن إيجاد الوسيط في حالة البيانات المبوبة ضمن جداول تكرارية مفتوحة، ويمكن إيجاده إذا كانت أطوال الفئات غير متساوية دون الحاجة لتعديل التكرارت.

مزايا وعيوب الوسيط:

المزايا: – لا يتأثر بالقيم الشاذة والمتطرفة.

- سهل الحساب.
- مجموع قيم الانحرافات المطلقة عن الوسيط أقل من مجموع الانحرافات المطلقة عن أي قيمة أخرى*a*.

عيوبه: إنه لا يأخذ كل القيم عند حسابه يعتمد على قيمة او قيمتين فقط.

- يصعب حسابه في حالة البيانات الوصفية المقاسة بمقياس اسمي.

المقياس الثالث: المنوال***Mode***: ***(Mo)***

هو أحد مقاييس النزعة المركزية لمستويات القياس الاسمي ويعرف المنوال بأنه القيمة أو الحالة الأكثر شيوعاً والأكثر تكراراً. وقد يكون التوزيع وحيد المنوال، أو بمنوالين أو عدة مناويل.

في مجموعة البيانات الصغيرة حيث لا تتكرر القيم فلا يوجد منوال. وعندما يكون للبيانات أكثر من منوال فلا يجوز حساب متوسطها لأن ذلك يتنافى مع مفهومه. ونميز الحالات التالية:

الحالة الأولى: بيانات مفردة: كل قيمة تتكررر مرة واحدة فقط. فلا يوجد لها منوال.

مثال(1): *3,4,5,6,2,* ليس لها منوال.

مثال(2): *3,4,5,6,2,3* المنوال*Mod=3* لن تكرارها يساوي 2 اكبر من تكرار بقية القيم.

ويفسر بأن غالبية المبحوثين أخذوا الدرجة 3.

الحالة الثانية: البيانات مرتبة:

مثال: لدينا البيانات التالية: 2,4,8,10,12,2,5,4,6

نرتب البيانات تصاعدياً ونحسب تكراراتها كما يلي:

x_i	f_i
2	2
4	2
5	1
6	1
8	1
10	1
12	1

للتوزيع منوالين هما *Mod1=2*، و*Mod2=4* لأن تكرارهما =2 واكبر من تكرار بقية القيم.

الحالة الثالثة: المنوال لبيانات مبوبة: لحسابه نحدد الفئة المنوالية وهي الفئة الأكثر تكراراً أي تكرارها المطلق أكبر من تكرار بقية الفئات.

نحدد الحد الأدنى للفئة المنوالية L_k، طول(مدى) الفئة المنوالية h_k، نحدد التكرار المطلق للفئة المنواليةf_k ، نحدد التكرار المطلق للفئة التي تسبق الفئة المنواليةf_{k-1}، نحدد التكرار المطلق للفئة التي تلي الفئة المنواليةf_{k+1}، ومن ثم نطبق العلاقة الرياضية التالية:

$$Mod = L_K + \frac{(f_k - f_{k-1})}{(f_k - f_{k-1}) + (f_k - f_{k+1})} h_k$$

مثال:

الفئات	fi
$[50 - 60[$	8
$[60 - 70[$	10
$[70 - 80[$	16
$[80 - 90[$	14
$[90 - 100[$	10
$[100 - 110[$	5
$[110 - 120[$	2
المجموع	65

الفئة المنوالية هي الفئة 80−70 لأن تكرارها المطلق هو 16 اكبر من تكرار بقية الفئات، وبالتالي حدها الأدنى هو 70 وطولها هو 80−70=10 ، وتكرارها المطلق 16، التكرار المطلق للفئة التي تسبقها هو 10، والتكرار المطلق للفئة التي تليها هو 14، نعوض في العلاقة:

$$Mod = L_K + \frac{(f_k - f_{k-1})}{(f_k - f_{k-1}) + (f_k - f_{k+1})} h_k$$

$$Mod = 70 + \frac{(16 - 10)}{(16 - 10) + (16 - 14)} 10 = 77.5$$

أي أن غالبية المبحوثين درجاتهم هي 77.5.

مزايا وعيوب المنوال:

مزاياه: سهل الحساب، لا يتأثر بالقيم الشاذة أو المتطرفة، يمكن إيجاده للبيانات الوصفية وفي حالة التوزيعات التكرارية المفتوحة.

عيوبه: لا يأخذ بالاعتبار جميع القيم عند حسابه، قد يكون لبعض البيانات أكثر من منوال، وبالتالي لا يمكن تحديد قيمة واحدة له للتمركز حولها البيانات.

العلاقة بين $\bar{x}$,Me, Mod: المتوسط الحسابي−المنوال=3(المتوسط الحسابي−الوسيط)

$\bar{x} = Me = Mod$ التوزيع متماثل

$Mod < Me < \bar{x}$ التوزيع موجب الالتواء

$Mod > Me > \bar{x}$ *التوزيع سالب الالتواء*

المقياس الرابع: المتوسط الهندسي:

هو الجذر النوني لجداءات القياسات(القيم) الأرقام موجبة، القيم عبارة عن معدلات ونسب وأرقام قياسية

بفرض لدينا القيم التالية: $x_1, x_2, \ldots\ldots x_n$ فإن المتوسط الهندسي لها يعطى بالعلاقتين الاتيتين:

$$G = \sqrt[n]{x_1, x_2, \ldots\ldots x_n}$$

$$\mathbf{G = (x_1, x_2, \ldots\ldots x_n)^{\frac{1}{n}}}$$

مثال: كانت لدينا معدلات تخرج 5 طلاب كما يلي: 70,75,72,65,60, فإن متوسط معدلات التخرج هو

$$G = \sqrt[5]{60 \times 65 \times 72 \times 75 \times 70} = 68.189$$

أو

$$G = (60 \times 65 \times 72 \times 75 \times 70)^{\frac{1}{5}} = 68.189$$

مقاييس التشتت

يمثل التشتت مدى انحراف (تقارب، أو تباعد) البيانات بعضها عن البعض، وهناك عدة مقاييس للتشتت منها:

1- دليل التشتت للبيانات الوصفية.

2- المدى.

3- التباين والانحراف المعياري ومعامل الاختلاف.

4- معامل التفرطح ومعامل التطاول.

1- **دليل التشتت للبيانات الوصفية(النوعية):** هو مقياس نسبي يقيس تشتت البيانات الوصفية سواء الاسمية منها او الوصفية ويرمز له بالرمز DI ويعطى بالعلاقة:

$$DI = \frac{C\left(n^2 - \sum n_i^2\right)}{n^2(C-1)} \times 100$$

C: عدد الأصناف(الحالات) أو الأوجه لكل متغير وصفي.

n: المجموع الكلي لمشاهدات الحالات جميعاً.

$\sum n_i$ مجموع المشاهدات لكل الحالات الخاصة بالمتغير الوصفي.

تتراوح قيمة دليل التشتت بين صفر (تجانس كامل) ومائة (تشتت كامل).

مثال: لدينا الجدول التالي الذي يوضح عدد الطلاب في أقسام كلية الاقتصاد بجامعة دمشق.

القسم	علاقات عامة	تسويق	موارد بشرية	مالية	محاسبة	قانون	سياسة	ادارة مالية	مجموع
عدد الطلاب	125	111	77	53	45	34	33	26	504

المطلوب: قياس مدى التشتت بين أعداد طلاب كلية الاقتصاد حسب أقسام الكلية.

n=504 ،C=8

$$DI=\frac{C\left(n^2-\sum n_i^2\right)}{n^2(C-1)}\times 100$$

$$=\frac{8\left(504^2-(125^2+111^2+77^2+53^2+45^2+34^2+33^2+26^2)\right)}{504^2(8-1)}$$

$$\times 100=\frac{1699088}{1778112}\times 100=95.56\%$$

التشتت كبير جداً لأن قيمة المقياس قريبة من المائة.

مثال: المستوى التعليمي للعاملين في إحدى الجهات العامة قارن التشتت في المستوى التعليمي بين الذكور والإناث.

المستوى التعليمي	ثانوي	جامعة	ماجستير	دكتوراه	مجموع
ذكور	5	10	6	2	23
إناث	3	7	4	1	15

بحساب دليل التشتت للذكورنجد ***DI1=91.75%***

دليل التشتت للإناث يساوي ***DI2=88.89%***

المستوى التعليمي للإناث أقل تشتت من المستوى التعليمي للذكور

2- **المدى**: كما عرفنا سابقاً المدى هو الفرق بين أكبر قيمة وأصغر قيمة في البيانات. والمدى في حالة البيانات المبوبة في جداول تكرارية مغلقة هو الفرق بين الحد الأعلى للفئة الأخيرة والحد الأدنى للفئة الأولى.

مثال: ليكن لدينا سعر الخدمة السياحية في فندقين:

تقلبات سعر الخدمة في الفندق الأول: *62,60,55,58,65,59*

تقلبات سعر الخدمة في الفندق الثاني: *55,59,60,61,59,65*

أوجد قيمة المدى لسعر الخدمة في الفندقين؟

R1=65-55=10

R2=65-55=10

وهذا لا يعني أن التقلبات في سعر الخدمة السياحية للفندقين متشابهين وذلك بالنظر إلى الجدول السابق يتضح خلاف ذلك، وبناء عليه لا يمكن الاعتماد على المدى كثيراً ويفضل استخدام الانحراف المعياري لأن جميع القيم تدخل في حسابه.

مزايا المدى: سهولة حسابه، مقياس يعطي فكرة سريعة عن تفاوت البيانات.

عيوب المدى:

- لا يدخل في حسابه كافة القيم يعتمد فقط على قيمتين، وقد تكون إحداهما شاذة لذلك لا نعتمد عليه كثيراً.
- يصعب حسابه في البيانات الوصفية أو الجداول التكرارية المفتوحة.

3- التباين σ^2 والانحراف المعياري σ :

من أهم مقاييس التشتت ويقيس مدى تشتت البيانات عن متوسطها الحسابي. حيث أن الانحراف المعياري هو الجذر التربيعي الموجب للتباين. ومنه فإن التباين هو مربع الانحراف المعياري والتباين بالتعريف هو المتوسط الحسابي لمربعات انحرافات القيم عن متوسطها الحسابي. ويعطى بالعلاقات الرياضية التالية حسب حالة البيانات(مفردة، مرتبة، مبوبة).

أولاً: البيانات المفردة: الطريقة العامة:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}$$

$$\sigma = +\sqrt{\sigma^2}$$

بالطريقة المختصرة:

$$\sigma^2 = \frac{\sum_{i=1}^{n} {x_i}^2}{n} - (\overline{x})^2$$

$$\sigma = +\sqrt{\sigma^2}$$

ثانياً: البيانات المرتبة: الطريقة العامة:

$$\sigma^2 = \frac{\sum_{i=1}^{n} f_i (x_i - \bar{x})^2}{\sum_{i=1}^{n} f_i}$$

$$\sigma = +\sqrt{\sigma^2}$$

حيث: $\sum_{i=1}^{n} f_i = n$

بالطريقة المختصرة:

$$\sigma^2 = \frac{\sum_{i=1}^{n} f_i {x_i}^2}{\sum_{i=1}^{n} f_i} - (\bar{x})^2$$

$$\sigma = +\sqrt{\sigma^2}$$

ثانياً: البيانات المبوبة: الطريقة العامة:

$$\sigma^2 = \frac{\sum_{i=1}^{n} f_i (\acute{x}_i - \bar{x})^2}{\sum_{i=1}^{n} f_i}$$

$$\sigma = +\sqrt{\sigma^2}$$

حيث: $\sum_{i=1}^{n} f_i = n$

بالطريقة المختصرة:

$$\sigma^2 = \frac{\sum_{i=1}^{n} f_i {\acute{x}_i}^2}{\sum_{i=1}^{n} f_i} - (\bar{x})^2$$

$$\sigma = +\sqrt{\sigma^2}$$

مزايا الانحراف المعياري:

1- سهولة الحساب والتعامل معه جبرياً.

2- تدخل جميع القيم في حسابه لذا يعد أدق مقاييس التشتت.

3- له نفس وحدة القياس للظاهرة محل الدراسة.

عيوب الانحراف المعياري:

1- يتأثر بالقيم الشاذة.

2- لا يمكن حسابه للبيانات الوصفية.

3- يصعب حسابه للجداول التكرارية المفتوحة.

معامل الاختلاف (مقياس التشتت النسبي):

يستخدم معامل الاختلاف في المقارنة بين تشتت مجموعتين أو أكثر من البيانات حيث لا يمكننا استخدام أحد مقاييس التشتت لعمل هذه المقارنة مباشرة لسببين:

- اختلاف وحدات القياس المستخدمة في المجموعتين كما لو كنا نقارن تشتت درجات مجموعة من الطلاب وتشتت أطوالهم.
- وجود فرق كبير بين المتوسطين الحسابيين للمجموعتين المراد المقارنة بين تشتتهما.

يعطى معامل الاختلاف أو مقياس التشتت النسبي بالعلاقة التالية: $C.V = \frac{\sigma}{\bar{x}} \times 100$

أمثلة عن التباين والانحراف المعياري ومعامل الاختلاف:

مثال(1) لدينا المعطيات التالية: *15,20,10,25,30* احسب معامل الاختلاف؟

يعطى معامل الاختلاف بالعلاقة التالية: $C.V = \frac{\sigma}{\bar{x}} \times 100$ ولحسابه يجب حساب المتوسط الحسابي $\bar{X}$ والانحراف المعياريσ

حساب المتوسط الحسابي$\bar{X}$: البيانات مفردة كل قيمة تتكرر مرة واحدة فقط. يعطى المتوسط الحسابي بالعلاقة التالية: حيث $\boldsymbol{n = 5}$ عدد القيم

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{15 + 20 + 10 + 25 + 30}{5} = \frac{100}{5} = 20$$

حساب الانحراف المعياريσ : يحسب بالعلاقة

$$\boldsymbol{\sigma = +\sqrt{\sigma^2}}$$

لحسابه نحسب التباين$\boldsymbol{\sigma^2}$ بالطريقة العامة كما في العلاقة التالية:

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} = \frac{250}{5} = 50$$

ومنه

$$\boldsymbol{\sigma} = +\sqrt{\sigma^2} = +\sqrt{50} = 7.071$$

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	${x_i}^2$
10	-10	100	100
15	-5	25	225
20	0	0	400
25	5	25	625
30	10	100	900
$\sum_{i=1}^{n} x_i = 100$		$\sum_{i=1}^{n} (x_i - \bar{x})^2 = 250$	$\sum_{i=1}^{n} {x_i}^2 = 2250$

حساب التباين بالطريقة المختصرة:

$$\sigma^2 = \frac{\sum_{i=1}^{n} {x_i}^2}{n} - (\bar{x})^2 = \frac{2250}{5} - (20)^2 = 50$$

وبالتالي قيمة الانحراف المعياري هو 7.071

مقياس التشتت النسبي(معامل الاختلاف) $\boldsymbol{c.v = \frac{7.071}{20} \times 100 = \%35.4}$ تشتت قليل نسبياً.

مثال(2) لدينا البيانات المرتبة كما في الجدول التالي: احسب معامل الاختلاف؟

$\boldsymbol{x_i}$	**2**	**4**	**5**	**7**	**8**	**9**	**10**
$\boldsymbol{f_i}$	**1**	**1**	**1**	**1**	**3**	**2**	**1**

يعطى معامل الاختلاف بالعلاقة التالية $C.V = \frac{\sigma}{\bar{x}} \times 100$ ولحسابه يجب حساب $\bar{x}$، $\boldsymbol{\sigma^2}$ و $\boldsymbol{\sigma}$ من خلال العلاقات التالية:

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{70}{10} = 7$$

بالطريقة العامة $\sigma^2 = \frac{\sum_{i=1}^{n} f_i (x_i - \bar{x})^2}{\sum_{i=1}^{n} f_i} = \frac{58}{10} = 5.8$

$$\sigma = +\sqrt{\sigma^2} = +\sqrt{5.8} = 2.408$$

حيث: $\sum_{i=1}^{n} f_i = n = 10$

بالطريقة المختصرة: $\sigma^2 = \frac{\sum_{i=1}^{n} f_i x_i^2}{\sum_{i=1}^{n} f_i} - (\bar{x})^2 = \frac{548}{10} - (7)^2 = 5.8$

$$C.V = \frac{\sigma}{\bar{X}} \times 100 = \frac{2.408}{7} \times 100 = 34.4\%$$

ونحتاج إلى الجدول المساعد التالي:

x_i	f_i	$f_i x_i$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$f_i (x_i - \bar{x})^2$	${x_i}^2$	$f_i {x_i}^2$
2	1	2	5−	25	25	4	4
4	1	4	3−	9	9	16	16
5	1	5	2−	4	4	25	25
7	1	7	0	0	0	49	49
8	3	24	1	1	3	64	192
9	2	18	2	4	8	81	162
10	1	10	3	9	9	100	100
المجموع	$\sum_{i=1}^{n} f_i = n = 10$	$\sum_{i=1}^{n} f_i x_i = 70$			$\sum_{i=1}^{n} f_i (x_i - \bar{x})^2 = 58$		$\sum_{i=1}^{n} f_i {x_i}^2 = 548$

مثال(3): لدينا البيانات المبوبة كما في الجدول التالي والمطلوب حساب معامل الاختلاف؟

الفئات(المجالات)	$[2-4[$	$[4-6[$	$[6-8[$	$[\![8-10]\!]$
f_i	2	1	3	4

الحل:

يعطى معامل الاختلاف بالعلاقة التالية $C.V = \frac{\sigma}{\bar{X}} \times 100$ ولحسابه يجب حساب $\bar{x}$، σ^2 و σ من خلال العلاقات التالية:

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i \acute{x}_i}{\sum_{i=1}^{n} f_i} = \frac{68}{10} = 6.8$$

$$\sigma^2 = \frac{\sum_{i=1}^{n} f_i (\acute{x}_i - \bar{x})^2}{\sum_{i=1}^{n} f_i} = \frac{51.6}{10} = 5.16$$ بالطريقة العامة

$$\sigma = +\sqrt{\sigma^2} = +\sqrt{5.16} = 2.272$$

حيث: $\sum_{i=1}^{n} f_i = n = 10$

بالطريقـــــــــة المختصـــــــــرة: $\sigma^2 = \frac{\sum_{i=1}^{n} f_i {\acute{x}_i}^2}{\sum_{i=1}^{n} f_i} - (\bar{x})^2 = \frac{514}{10} - (6.8)^2 = 5.16$

$$C.V = \frac{\sigma}{\bar{X}} \times 100 = \frac{2.272}{6.8} \times 100 = 33.4\%$$

ونحتاج إلى الجدول المساعد التالي:

الفئات(المجالات)	f_i	$\acute{x}_i$	$f_i \acute{x}_i$	$(\acute{x}_i - \bar{x})$	$(\acute{x}_i - \bar{x})^2$	$f_i (\acute{x}_i - \bar{x})^2$	${\acute{x}_i}^2$	$f_i {\acute{x}_i}^2$
$[2-4[$	2	3	6	3.8−	14.44	28.88	9	18
$[4-6[$	1	5	5	1.8−	3.24	3.24	25	25
$[6-8[$	3	7	21	0.2	0.04	0.12	49	147
$[\![8-10]\!]$	4	9	36	2.2	4.84	19.36	81	324
المجموع	$\sum_{i=1}^{n} f_i = n = 10$		$\sum_{i=1}^{n} f_i \acute{x}_i = 68$			$\sum_{i=1}^{n} f_i (\acute{x}_i - \bar{x})^2 = 51.6$		$\sum_{i=1}^{n} f_i {\acute{x}_i}^2 = 514$

معامل الالتواء: (أحد مقاييس عدم التماثل):

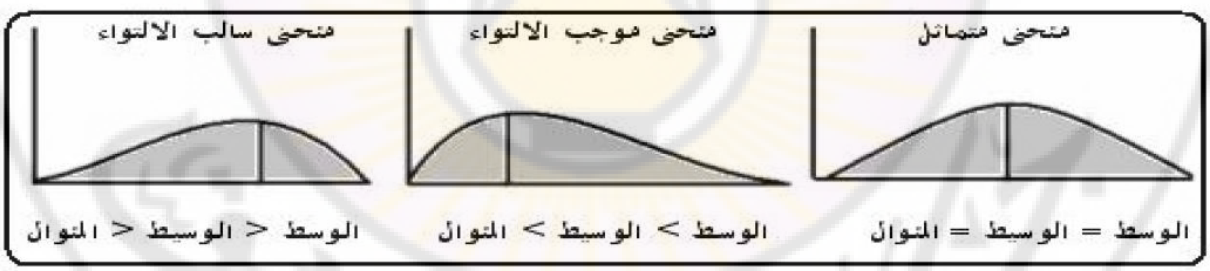
الالتواء: هو بعد المنحني التكراري للظاهرة عن التماثل ويقاس بمعامل الالتواء فإما:

- أن يكون المنحنى التكراري متماثلاً عندما تكون قيمة معامل الالتواء صفراً أي أن: $\bar{x} = Me = Mod$
- أو أن يكون المنحنى التكراري ملتوٍ نحو اليمين عندما تكون قيمة معامل الالتواء موجبة أي أن: $\bar{x} > Mod$
- أو أن يكون المنحنى التكراري ملتوٍ نحو اليسار عندما تكون قيمة معامل الالتواء سالبة أي أن: $\bar{x} < Mod$

يعطى معامل الالتواء بالعلاقة التالية:

$$SK = \frac{\bar{X} - Mod}{\sigma}$$

ويظهر ذلك على الشكل التالي:



مثال (1): إذا علمت أن قيمة الانحراف المعياري هي 2.19، وقيمة المتوسط الحسابي 8، وقيمة المنوال 8 أوجد معامل الالتواء؟

$$SK = \frac{\bar{X} - Mod}{\sigma} = \frac{8-8}{2.19} = \frac{0}{2.19} = 0$$

التوزيع متماثل(طبيعي)(معتدل)

مثال (1): إذا علمت أن قيمة الانحراف المعياري هي 1.48، وقيمة المتوسط الحسابي 7، وقيمة المنوال 6 أوجد معامل الالتواء؟

$$SK = \frac{\bar{X} - Mod}{\sigma} = \frac{7-6}{1.48} = \frac{1}{1.48} = 0.67 > 0$$

قيمة معامل الالتواء موجبة وبالتالي المنحنى التكراري ملتو نحو اليمين.

مقياس التفرطح: وهو مقياس يصف ارتفاع قمة المنحنى التكراري من حيث الاعتدال، أو التدبب، أو التفرطح.

عند تمثيل التوزيع التكراري في شكل منحنى تكراري قد يكون هذا المنحنى منبسط، مدبب، معتدل.

فعندما يتركز عدد أكبر من القيم بالقرب من منتصف المنحنى ويقل في طرفيه يكون المنحنى مدبباً، وعندما يتركز عدد أكبر على طرفي المنحنى ويقل عدد القيم في المنتصف يكون المنحنى مفرطحاً، أو منبسطاً، ويكون المنحنى متماثلاً عندما تتوزع القيم بشكل متماثل على طرفي المتوسط الحسابي.

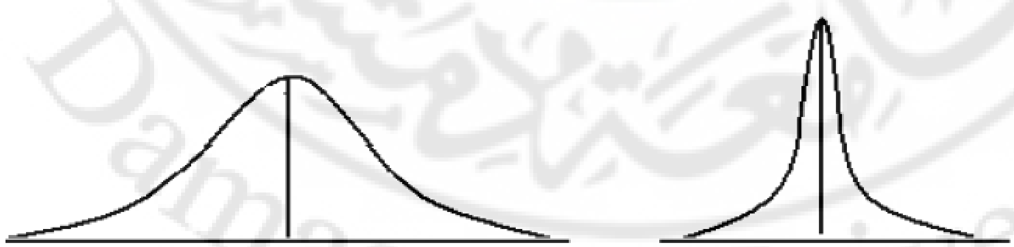
ويقاس التفرطح بالعلاقة التالية:

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

معامل التفرطح في التوزيع الطبيعي يساوي/3/. أي: $K=3$ التوزيع متماثل.

$k>3$: التوزيع مدبب.

$K<3$: التوزيع مفرطح(منبسط).



منحنى مفرطح　　　　منحنى مدبب

مثال: لدينا البيانات التالية: *66,85, 52,78,80,91,74,58* إيجاد شكل التفرطح.

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^4$
66	7−	49	2401
85	12	144	20736
52	21−	441	194481
78	5	25	625
80	7	49	2401
91	18	324	104976
74	1	1	1
58	15−	285	50625
$\sum_{i=1}^{n} x_i = 584$		$\sum_{i=1}^{n} (x_i - \bar{x})^2 = 1258$	$\sum_{i=1}^{n} (x_i - \bar{x})^4 = 376246$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{66 + 85 + 52 + 78 + 80 + 91 + 74 + 58}{8} = \frac{584}{8} = 73$$

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} = \frac{1258}{8} = 157.25$$

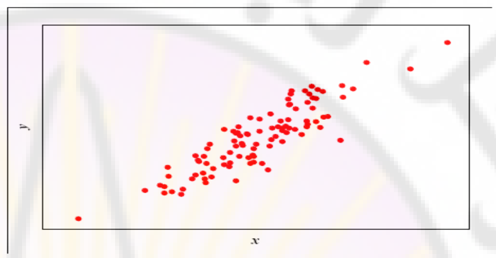
$$\sigma = +\sqrt{\sigma^2} = +\sqrt{157.25} = 12.540$$

$$k = \frac{\frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^4}{\sigma^4} = \frac{376246/8}{12.540^2} = 1.99 > 3$$

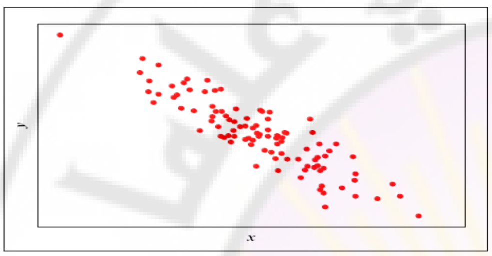
شكل توزيع البيانات مفرطح

الارتباط البسيط

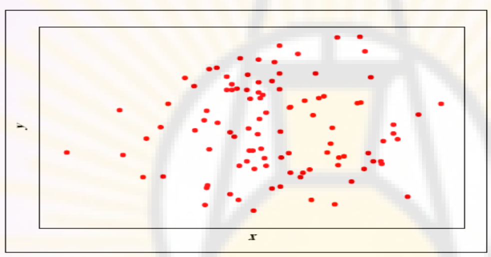
هو علاقة بين متغيرين x,y بمعنى إذا تغير أحد المتغيرين فإن المتغير الآخر يتبعه إما في نفس الاتجاه فيكون الارتباط طردي كعلاقة الصادرات بالميزان التجاري، الدخل والانفاق، أو في الاتجاه المضاد فيكون الارتباط عكسي كعلاقة الواردات بالميزان التجاري. ويقال أن المتغيرين مستقلين عندما ينعدم الارتباط بينهما كعلاقة دخل الفرد بوزنه كما هو موضح في أشكال الانتشار التالية:



رسم توضيحي 6: الارتباط الطردي



رسم توضيحي 7: الارتباط العكسي



رسم توضيحي 8: لا يوجد ارتباط (استقلال)

ملاحظة: الارتباط لا يدل على السببية، حيث ليس شرطاً أن يتغير أحد المتغيرين دائماً بدلالة الآخر.

يرمز لمعامل الارتباط بـــ r وهو عبارة عن مقياس كمي نسبي يقيس قوة العلاقة بين متغيرين تتراوح قيمته بين +1 و −1 أي أن: $-1 \le r \le +1$ يقال أن الارتباط تام وطردي إذا كان $r=+1$ ويكون الارتباط تام وعكسي إذا كان $r=-1$، ويقال أن المتغيرين مستقلين $r=0$

كلما اقتربت قيمة معامل الارتباط من الواحد الصحيح كلما زادت العلاقة الارتباطية بين المتغيرين قوة ومتانة.

تفسير قيمة معامل الارتباط: نقوم بتفسير الاشارة أولاً فإذا كانت سالبة الارتباط عكسي وإذا كانت موجبة الارتباط طردي. ومن ثم نقوم بتفسير القيمة على النحو التالي:

$r < 50\%$ *الارتباط ضعيف*

$50\% \leq r < 60\%$ الارتباط مقبول

$60\% \leq r < 70\%$ الارتباط جيد

$70\% \leq r < 80\%$ الارتباط جيد جداً

$80\% \leq r < 90\%$ الارتباط قوي

$90\% \leq r < 100\% = 1$ الارتباط قوي جداً

معامل الارتباط الخطي(بيرسون): يقيس العلاقة الارتباطية بين متغيرين كميين ويسمى بمعامل ارتباط بيرسون ويعطى بالعلاقة التالي: $r_p = \frac{\frac{\sum x.y}{n} - \bar{x}\bar{y}}{\sigma_x \sigma_y}$

$\sum x.y$: مجموع حاصل جداء قيم *x* بقيم *y* المقابلة لها.

$\bar{x}$ متوسط قيم *x*

$\bar{y}$ متوسط قيم *y*.

σ_x: الانحراف المعياري لقيم *x*

σ_y: الانحراف المعياري لقيم *y*

ملاحظة: قيمة معامل الارتباط الخطي تساوي الصفر تدل على عدم وجود علاقة ارتباط خطية فقط بين المتغيرين قيد الدراسة.

مثال: لدينا الصادرات والميزان التجاري لإحدى الدول هل يوجد علاقة ارتباط خطية بين المتغيرين وما نوعها وما مدى قوتها؟

الحل: الصادرات هي المتغير المستقلx والميزان التجاري هو المتغير التابع y والمتغيرين كميين فمعامل الارتباط المناسب لقياس قوة وشدة العلاقة بينهما هو معامل ارتباط بيرسون الخطي الذي يعطى بالعلاقة التالية:

$$r_p = \frac{\frac{\sum x.y}{n} - \bar{x}\bar{y}}{\sigma_x \sigma_y}$$

لحسابه نحتاج على الجدول المساعد التالي:

الصادراتx	الميزان التجاريy	$x.y$	X^2	Y^2
9	1	9	81	1
11	2	33	121	9
17	8	136	289	64
18	7	126	324	49
19	6	114	361	36
16	5	80	256	25
16	7	112	256	49
19	8	152	361	64
23	12	276	529	144
23	12	276	529	144
$\sum x = 171$	$\sum y = 69$	$\sum x.y = 1314$	$\sum x^2 = 3107$	$\sum y^2 = 585$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{171}{10} = 17.1$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{69}{10} = 6.9$$

$$\sigma_x^2 = \frac{\sum x^2}{n} - (\bar{x})^2 = \frac{3107}{10} - (17.1)^2 = 18.29$$

$$\text{ومنه } \sigma_x = +\sqrt{\sigma_x^2} = +\sqrt{18.29} = 4.277$$

$$\sigma_y^2 = \frac{\sum y^2}{n} - (\bar{y})^2 = \frac{585}{10} - (6.9)^2 = 10.89$$

ومنه $\sigma_y = +\sqrt{\sigma_y^2} = +\sqrt{10.89} = 3.3$

معامل ارتباط بيرسون:

$$r_p = \frac{\frac{1314}{10} - (17.1)(6.9)}{(4.277)(3.3)} = \frac{131.4 - 114.99}{14.11} = 0.95 = 95\%$$

الاشارة موجبة فالعلاقة طردية بين متغيري الصادرات والميزان التجاري. وبما أن قيمة معامل الارتباط تنتمي للمجال $1 = 100\% > r \leq 90\%$ فالارتباط قوي جداً بين المتغيرين.

معامل ارتباط الرتب (معامل ارتباط سبيرمان): يستخدم عندما يكون المتغيرين رتبيين أو أحدهما كمي والآخر رتبي ويحسب بالعلاقة: $r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$

d: الفرق بين رتب x ورتبy، $\sum d = 0$

مثال: لدراسة العلاقة بين المستوى التعليمي لرب الاسرة وبين عدد الأولاد أخذت عينة عشوائية بسيطة من الأسر بحجم n=5 أسرة وكان عدد أفرادها والمستوى التعليمي لرب الاسرة كما يلي:

المستوى التعليمي لرب الاسرة	تعليم أساسي	ثانوية	جامعة	معهد	أمي
عدد الأولاد	4	6	7	4	3

الحل: المستوى التعليمي لرب الاسرة يؤثر في عدد الأولاد فيكون المستوى التعليمي هو المتغير المستقل x، وعدد الأولاد هو المتغير التابع y. متغير المستوى التعليمي رتبي، بينما متغير عدد الاولاد هو كمي وفي هذه الحالة معامل الارتباط المناسب هو ارتباط سبيرمان الذي يعطى بالعلاقة التالية:

$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$ ولحسابه نحتاج إلى حساب الرتب والجدول المساعد التالي:

المستوى التعليمي لرب الاسرةx	عدد الأولادy	رتبx	رتبy	d	d^2

تعليم أساسي	4	2	2.5	-0.5	0.25
ثانوية	6	3	4	-1	1
جامعة	7	5	5	0	0
معهد	4	4	2.5	1.5	2.25
أمي	3	1	1	0	0
المجموع					$\sum d^2 = 3.5$

ويكون معامل الارتباط

$$r_s = 1 - \frac{6(3.5)}{5(25-1)} = 0.825 = 82.5\%$$

الاشارة موجبة الارتباط طردي بين المتغيرين

قيمة معامل الارتباط تنتمي للمجال $80\% \le r \le 90\%$ فالارتباط قوي بين المتغيرين.

معامل الاقتران: يستخدم عندما يكون المتغيرين نوعيين وبحالتين(وجهين) فقط لكل منهما ويعطى بالعلاقة التالية:

معامل الاقتران= $\frac{A.D-B.C}{A.D+B.C}$ حيث

المتغير1/المتغير 2	الحالة الأولى للمتغير الثاني	الحالة الثانية للمتغير الثاني
الحالة الأولى للمتغير الأول	A	B
الحالة الثانية للمتغير الأول	C	D

حيث A,B,C,D هي التكرارات.

مثال: لدراسة فيما إذا كان هناك علاقة ما بين الجنس والتدخين أخذت عينة عشوائية من 100 مبحوث وتم الحصول على البيانات كما في الجدول التالي:

التدخين/الجنس	ذكر	أنثى	المجموع
مدخن	25	15	40

غير مدخن	5	55	60
المجموع	30	70	100

المطلوب:هل هناك علاقة بين التدخين والجنس.

الحل: المتغيرين نوعيين وبحالتين فقط لكل منهما فمعامل الارتباط المناسب هو معامل الاقتران الذي يعطى بالعلاقة التالية:

معامل الاقتران= $\frac{A.D-B.C}{A.D+B.C}$ حيث A=25 و B=15 و C=5، D=55

معامل الاقتران: $\frac{25\times55-15\times5}{25\times55+15\times5}=\frac{1300}{1450}=0.897=89.7\%$

الاشارة موجبة الارتباط طردي بين المتغيرين.

قيمة معامل الارتباط تنتمي للمجال $80\% \le r \le 90\%$ فالارتباط قوي بين الجنس والتدخين.

تحليل الانحدار الخطي البسيط

عبارة عن اسلوب إحصائي يقوم بصياغة دالة رياضية لعملية ذات عوامل مؤثرة عدة $x_1, x_2, ... x_n$ لوصف متغير ناتج y عن هذه العملية والتحكم به، وتوقع قيم غير معروفة له، ويكون شكل الدالة الرياضية على الصورة التالية: $Y = f(x_1, x_2, \dots \dots x_n)$ تسمى المتغيرات $x_1, x_2, \dots \dots x_n$ المتغيرات المستقلة والمتغير y المتغير الناتج أوالتابع،

و دالة الانحدار الخطي البسيط هي أبسط حالة حيث يوجد متغير مستقل واحد يرتبط مع المتغير التابع بعلاقة خط مستقيم وتأخذ الدالة الشكل التالي: $y = a + bx$

حيث: y المتغير التابع، x المتغير المستقل.

b_0 نقطة تقاطع المستقيم مع المحور العمودي y

b_1 ميل المستقيم أو معامل انحدار x على y

b_1 ، b_0 : ثوابت تحسبان من خلال العلاقتين التاليتين:

$$b_1 = \frac{\frac{\sum x.y}{n} - \bar{x}.\bar{y}}{\sigma_x^2}$$

$$b_0 = \bar{y} - b_1\bar{x}$$

ولإيجاد أي قيمة مقدرة جديدة لــy والتي يرمز لها $\hat{Y}_h$ نعوض بقيمة معلومة للمتغير المستقل ولتكن x_h في المعادلة. حيث تكون تشكل x_h و$\hat{Y}_h$ زوج من القيم وتكون قيمة $\hat{Y}_h$ مقدرة في نفس الزمن والمكان اللذين علمت فيهما قيمة x_h.

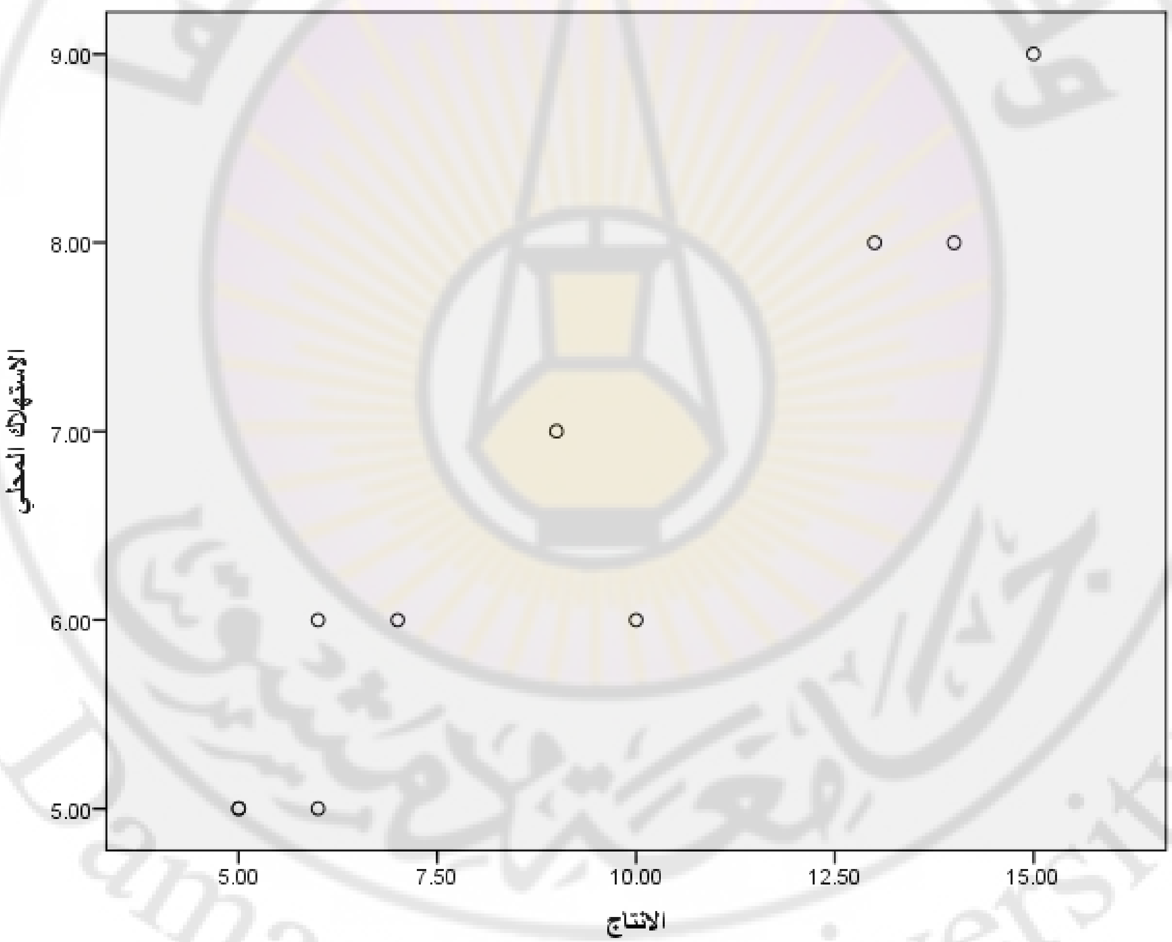
مثال: أوجد معادلة الانحدار الخطي البسيط بين للاستهلاك المحلي والانتاج من سلعة معينة، وتوقع قيمة الاستهلاك المحلي إذا كان الإنتاج المخطط هو 16000000 وحدة منتجة. أوجد القيم المقدرة للمتغير التابع. والبيانات الخاصة بالمتغيرين هي كما في الجدول التالي:

الوحدة: مليون ل.س.

الاستهلاك المحلي	6	8	9	8	7	6	5	6	5	5
الانتاج	10	13	15	14	9	7	6	6	5	5

الحل:

الاستهلاك المحلي تابع للإنتاج لذلك يكون الاستهلاك المحلي هو المتغير التابع y والانتاج هو المتغير المستقل x لإيجاد العلاقة الخطية (معادلة الانحدار الخطي البسيط) نوجد شكل الانتشار فنحصل على الشكل التالي:



حيث يتضح من الشكل بأن العلاقة خطية من الدرجة الأولى والمعادلة هي معادلة مستقيم بين المتغيرين تأخذ الشكل التالي: y=a+bx لتقدير الثوابت لهذه المعادلة من خلال العلاقتين:

$$b_1 = \frac{\frac{\sum x.y}{n} - \bar{x}.\bar{y}}{\sigma_x^2}$$

$$b_0 = \bar{y} - b_1\bar{x}$$

ونحتاج إلى الجدول المساعد التالي:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{90}{10} = 9$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{65}{10} = 6.5$$

$$\sigma_x^2 = \frac{\sum x^2}{n} - (\bar{x})^2 = \frac{942}{10} - (9)^2 = 13.2$$

xالانتاج	الاستهلاك yالمحلي	x.y	X2	Y2	ŷ
10	6	60	100	36	6.86
13	8	104	169	64	7.94
15	9	135	225	81	8.66
14	8	112	196	64	8.3
9	7	63	81	49	6.5
7	6	42	49	36	5.78
6	5	30	36	25	5.42
6	6	36	36	36	5.42
5	5	25	25	25	5.06
5	5	25	25	25	5.06
$\sum x = 90$	$\sum y = 65$	$\sum x.y = 632$	$\sum x^2 = 942$	$\sum y^2 = 441$	

$$b_1 = \frac{\frac{\sum x.y}{n} - \bar{x}.\bar{y}}{\sigma_x^2} = \frac{\frac{632}{10} - (9)(6.5)}{13.2} = 0.36$$

المعادلة المقدرة $b_0 = \bar{y} - b_1\bar{x} = 6.5 - (0.36)(9) = 3.26$

$$\hat{y} = 3.26 + 0.36x$$

القيمة المتوقعة للاستهلاك عندما يكون الانتاج 16 أي x=16 هي:

$$\hat{y} = 3.26 + 0.36(16) = 9.02$$

أي أن الاستهلاك المحلي سيصل إلى 9.02 مليون وحدة منتجة عندما يصبح الانتاج 16 مليون وحدة.

القيم المقدرة للاستهلاك(القيم النظرية) من المعادلة هي $\hat{y}$ حيث نحصل على كل قيمة من خلال تعويض قيم x في المعادلة المقدرة فكل قيمة لـــ x يقابلها قيمة مقدرة لـــ y

فمثلاً القيمة المقدرة الأولى لـــy هي

$$\hat{y} = 3.26 + 0.36(10) = 6.86$$

التوزيع الطبيعي

Noraml Distribution

تمهيد:

التوزيع الطبيعي من أهم التوزيعات الإحتمالية المستمرة وأكثرها استخداماً في مجال البحث العلمي على وجه الخصوص. فهو أداة إحصائية هامة في اختبار فرضيات البحث العلمي عندما يكون حجم العينة كبيراً (أكبر من 30 مفردة). يستخدم في إدارة وضبط الجودة للسلع والخدمات في مختلف المنظمات، حيث يعد الأساس الذي تبنى عليه خرائط الرقابة الإحصائية للجودة على اختلاف أنواعها. إضافة لاستخدامه كأداة لتقدير قيم الظاهرة المعبر عنها بمتغير مستمر كالوزن، الطول، درجة الحرارة، درجة الاختبار......الخ. وهذا التقدير يكون بشكليه النقطي والمجالي، وتقدير لنسبة أو حجم الظاهرة المدروسة.

إضافة إلى أن التوزيع الطبيعي هو أسلوب لاختبار الفرضيات، فإنه يعد شرطاً أساسياً يجب تحققه عند إجراء اختبار الفرضيات باستخدام الأساليب المعلمية. لذا كان من الضروري أن نتناول هذا التوزيع الاحتمالي المستمر في هذا الفصل نظراً لأهميته العلمية والتطبيقية.

يقدم هذا الفصل توضيحاً لمفهوم التوزيع الطبيعي العام وخصائصه المختلفة إضافة إلى التوزيع الطبيعي المعياري وخصائصه أيضاً، ليتضح الفرق بين التوزيعين وخصائصهما. كما يتطرق هذا الفصل لمفهوم الدرجة المعيارية واستخدامها في عملية التقييم أو للتحويل من توزيع طبيعي عام إلى توزيع طبيعي معياري عند حساب الاحتمالات من جدول التوزيع الطبيعي المعياري. مع تعريف بالنسب الشهيرة للاحتمالات تحت مساحة التوزيع واستخدامها في عملية التقدير الإحصائي، مع تطبيق عملي حول ذلك.

التوزيع الطبيعي العام:

التوزيع الطبيعي أو توزيع غاوس، وهو من أفضل وأكثر التوزيعات الاحتمالية المتصلة استخداماً، ويعزى ذلك إلى أن توزيعات كثيرة لمتغيرات عشوائية مستمرة كالطول، الوزن، درجة الطالب في الامتحان..الخ. تتبع التوزيع الطبيعي.

والتوزيع الطبيعي لمتغير عشوائي متصل X مداء لفترة المفتوحة $]-\infty,+\infty[$ ودالة كثافته الاحتمالية دالة أسية تعتمد على القيمتين μ (التوقع) σ (الانحراف المعياري) تأخذ الشكل العام التالي:

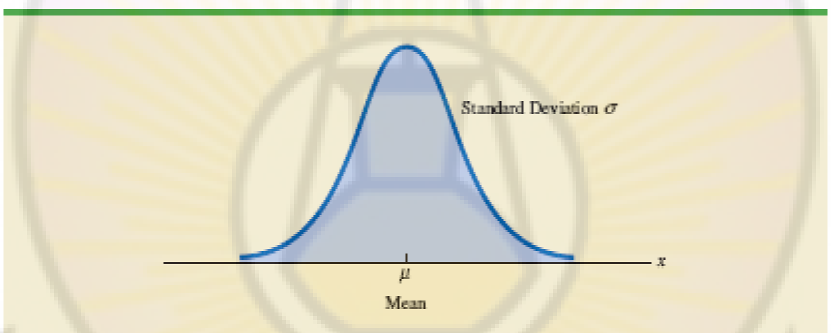
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(X-\mu)^2/2\sigma^2} \ , \ -\infty < x,\mu < \infty \ , \ \sigma > 0$$

e: ثابت قيمته 2.71828

π: هي النسبة التقريبية وتساوي 3.14159

ومن أهم صفات هذه الدالة أن: $\int_{-\infty}^{+\infty} f(x)dx = 1$ أي أن السماحة تحت المنحنى تساوي الواحد وهو مجموع الاحتمالات.

ومنحنى دالة الاحتمال للتوزيع الطبيعي له خاصية شكل الجرس المقلوب، ويتحدد شكل الجرس تماماً لأي توزيع طبيعي خاصة إذا علمنا المتوسط الحسابي μ والانحراف المعياري σ لهذا التوزيع. حيث تدل قيمة μ على مكان مركز الجرس، كما تدل σ على كيفية الانتشار. ويوضح الشكل التالي منحنى هذا التوزيع (Anderson, Sweeney, & Williams, 2011)[1]:



رسم توضيحي 1 منحنى توزيع طبيعي عام

بعض خواص التوزيع الطبيعي:

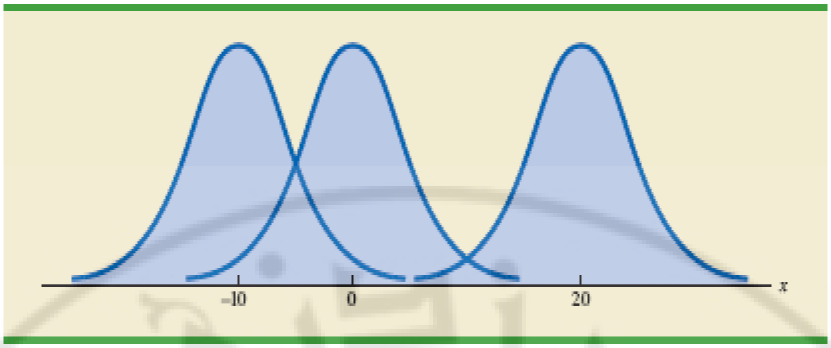
1- المنحنى متصل ويقع بالكامل فوق المحور الأفقي.

2- متماثل بالنسبة للمستقيم $x = \mu$ أي أن هذا المستقيم يقسم المساحة تحت المنحنى وفوق المحورالأفقي إلى قسمين متناظرين ومتساويين.

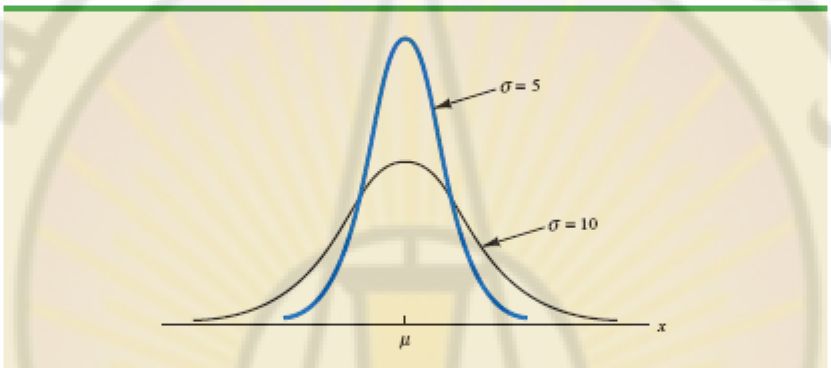
3- قيم المعلم μ تحدد مركز التوزيع، بينما قيم المعلم σ تحدد التشتت كما في الشكلين الآتيين:

---

[1] الأشكال البيانية مصدرها (Anderson, Sweeney, & Williams, 2011)



رسم توضيحي 2 منحنيات توزيع طبيعي لها نفس الانحراف مع اختلاف المتوسط.



رسم توضيحي 3منحنيات توزيع طبيعي لها نفس التوقع مع اختلاف الانحراف المعياري.

4- منحى التوزيع الطبيعي يقترب طرفاه من المحور الأفقي ولكن لا يمسه.

5- المساحة الكلية تحت المنحنى تساوي الواحد.

6- المتوسط الحسابي =المنوال= الوسيط لهذا التوزيع.

7- معامل التطاول لهذا التوزيع يساوي /3/.

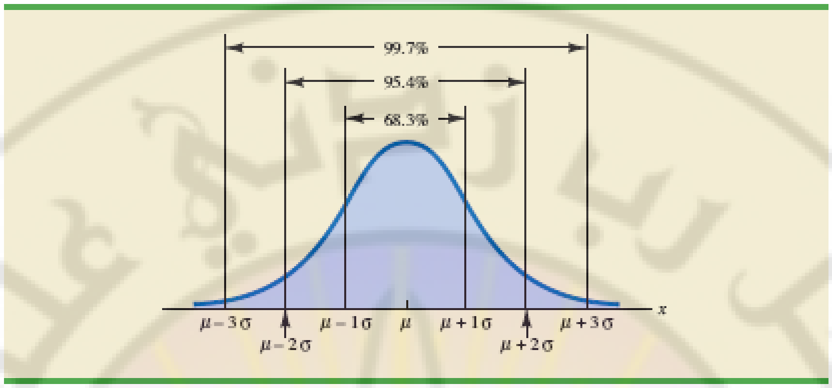
8- معامل الالتواء لهذا التوزيع يساوي صفر.

9- 68.27% من مساحة المنحنى الطبيعي تتحصر بين $\mu + \sigma, \mu - \sigma$

10- 95.45% من مساحة المنحنى الطبيعي تتحصر بين$\mu + 2\sigma, \mu - 2\sigma$

11- 99.75% من مساحة المنحنى الطبيعي تتحصر بين $\mu + 3\sigma, \mu - 3\sigma$

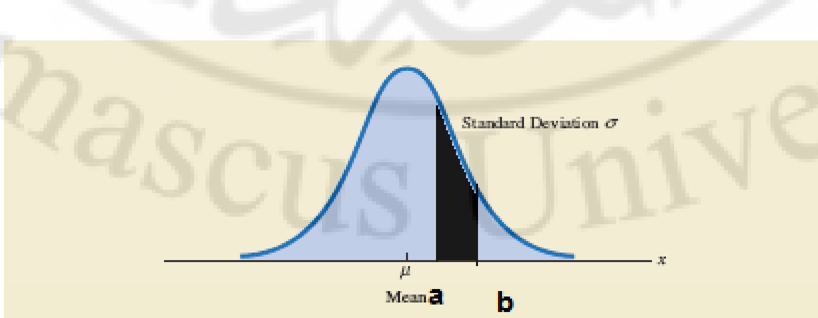
تدعى هذه النسب بالنسب الشهيرة وتستخدم في عملية التقدير الإحصائي لقيمة المتغير العشوائي كما هو موضح في الشكل الآتي:



رسم توضيحي 4النسب الشهيرة لتوزيع المساحة تحت منحنى التوزيع الطبيعي

نرمز للمتغير العشوائي X الذي يتبع توزيع طبيعي بتوقع μ وانحراف معياري σ بالرمز $X\sim N(\mu,\sigma)$ فمثلاً $X\sim N(5,4)$ تعني أن المتغير العشوائي الطبيعي له توقع (متوسط حسابي) يساوي 5 وتباين يساوي 4 (انحراف معياري يساوي 2).

وحيث أن $f(X)$ دالة كثافة إحتمالية فإن المساحة الكلية تحت منحنى هذه الدالة يساوي الواحد، لذلك فإن إحتمال أن متغيراً عشوائياً X متوسطه μ وتباينه σ^2 يقع بين قيمتين محددتين a, b هو $P(a\leq X\leq b)$ يساوي المساحة المحصورة بين منحنى الدالة والقيمتين $X=a, X=b$ كما في الشكل الآتي(Nelson, 2020) :



رسم توضيحي 5 احتمال أن يأخ المتغير قيمة ضمن قيمتين.

التوزيع الطبيعي المعياري:

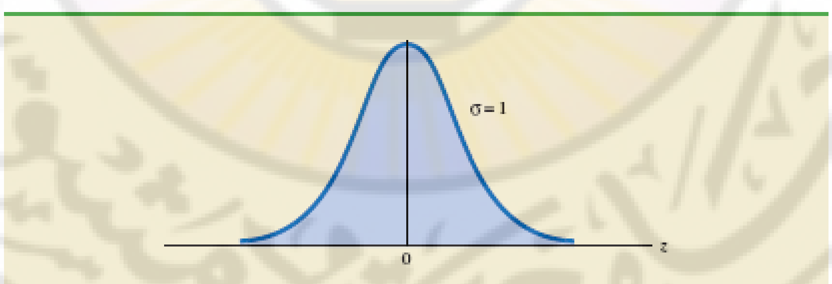
التوزيع الطبيعي للمتغير العشوائي X وانحرافه المعياري σ وهاتان القيمتان تؤثران على وضع التوزيع ومقدار تشتته، ولكن توجد صعوبة في تحديد كل قيم σ, μ ، لذا يتم تحويل قيم المتغير العشوائي X إلى قيم معيارية كما يلي:

$$Z = \frac{X - \mu}{\sigma}$$

ويتبع المتغير العشوائي Z توزيعاً جديداً يسمى التوزيع الطبيعي المعياري توقعه $\mu = 0$ وتباينه $\sigma^2 = 1$ أي أن: $Z \sim N(0,1)$. وتكون دالة الكثافة الاحتمالية للمتغير العشوائي المعياري Z هي:

$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} \qquad\qquad -\infty < z < +\infty$$

ويرسم على المحاور الإحداثية منحنياً متناظراً بالنسبة للمحور الشاقولي كما في الشكل الآتي:



رسم توضيحي 6 منحنى توزيع طبيعي معياري.

بعض خواص التوزيع الطبيعي المعياري:

1- المنحنى متصل ويقع بالكامل فوق المحور الأفقي.

2- متماثل بالنسبة للمستقيم Z=0، أي أن المستقيم Z=0 يقسم المساحة تحت المنحنى وفوق المحور الأفقي إلى قسمين متساويين ومتناظرين.

3- مركزه ينطبق على مبدأ الإحداثيات أي أن متوسطه صفر.

4- متناظر بالنسبة للمجور العمودي.

5- المساحة الكلية تحت المنحنى تساوي الواحد.

6- 68.27% من مساحة المنحنى الطبيعي تتحصر بين 1 - و1.

7- 95.45% من مساحة المنحنى الطبيعي تتحصر بين 2 - و2.

8- 99.75% من مساحة المنحنى الطبيعي تتحصر بين 3- و3.

9- المنحنى يصل إلى نهايته العظمى عندما Z=0 وقيمة دالة الكثافة الاحتمالية عند Z=0 (نهايتها العظمى) تساوي: $f(x) = \frac{1}{\sqrt{2\pi}}$ وله جداول خاصة سيتم حساب الاحتمالات المختلفة بناء عليه كما سنرى لاحقاً.

الدرجة المعيارية:

"مقياس يقيس الانحرافات عن المتوسط الحسابي بوحدات من الانحراف المعياري ويرمز لها بالرمز z فإذا كان لدينا المتغير X وله القيم $(X_1, X_2 \ldots . X_n)$ والتي لها المتوسط $\bar{X}$ والانحراف المعياري σ فإن Z تحسب من العلاقة الآتية" (أبو ظبي، دليل رقم/10/):

$$Z = \frac{X_i - \mu}{\sigma}$$

نورد المثال الآتي:

خضع العاملين في إحدى الفنادق لدورتي تدريب الأولى في مجال القيادة الدولية للحاسوب ICDL، والثانية في اللغة الانكليزية وأجرى اختبار في نهاية كل دورة تدريبية للمتدربين. فإذا كانت درجة أحد العاملين المتدربين في اختبار القيادة الدولية للحاسوب 86 وكان متوسط درجات الاختبار هو 77 بانحراف معياري 11 درجة. وكانت درجته في اختبار اللغة الانكليزية 96 درجة ومتوسط درجات العاملين المتدربين هو 84 درجة بانحراف معياري 17 درجة، ففي أي الاختبارين كان أداء العامل المتدرب أفضل؟

باستخدام الدرجة المعيارية لدرجة العامل في كلا موضوعي التدريب كما يلي:

الدرجة المعيارية لدرجة العامل في اختبار قيادة الحاسوب:

$$Z = \frac{X_i - \mu}{\sigma} = \frac{86 - 77}{11} = 0.82$$

الدرجة المعيارية لدرجة العامل في اختبار اللغة الانكليزية:

$$Z = \frac{X_i - \mu}{\sigma} = \frac{96 - 84}{17} = 0.7$$

وتدل هذه النتائج على أن أداء العامل المتدرب في قيادة الحاسوب أفضل من أدائه في اللغة الانكليزية وإن كانت درجته في الحاسوب أقل.

حساب الاحتمالات: نميز بين الحالتين:

الاحتمالات للمتحولX الخاضع للتوزيع الطبيعي المعياري:

تحسب هذه الاحتمالات من جدول التوزيع الطبيعي المعياري، حيث تمثل هذه الاحتمالات المساحة تحت المنحني من $-\infty$ وحتى القيمة المتداولة Z_1 ويرمز لهذه السماحة بالرمز $\emptyset(Z)$ أي أن:

$$P(Z < Z_1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Z_1} e^{\frac{-z^2}{2}}\, dz \qquad = \emptyset(z_1)$$

نحسب الاحتمالات الآتية بناء على جدول التوزيع الطبيعي المعياري:

$$P(z < 0) = \emptyset(0) = 0.5$$

$$P(z < 0.53) = \emptyset(0.53) = 0.7019$$

$$P(z < 2.82) = \emptyset(2.82) = 0.9976$$

$$P(z < 1) = \emptyset(1) = 0.8413$$

الاحتمالات للمتحولX الخاضع للتوزيع الطبيعي العام:

ينبغي عند حساب الاحتمالات لمتغير عشوائي X يخضع للتوزيع الطبيعي العام أن نحول التوزيع من توزيع طبيعي عام إلى توزيع طبيعي معياري باستخدام الدرجة المعيارية. نحصل من جداول التوزيع الطبيعي المعياري Z على الاحتمالات المطلوبة. ونميز الآتي (العلي و دريباتي، 2021):

- حساب الاحتمال العام من الشكل $P(X \le X_1)$

باستخدام رمز التحويلية المعيارية يكون الاحتمال المطلوب هو:

$$P(X \le X_1) = \emptyset\left(\frac{X_1 - \mu}{\sigma}\right)$$

- حساب الاحتمال العام في مجال محدد من الشكل $P(X_1 \le X \le X_2)$:

$$P(X_1 \le X \le X_2) = \emptyset\left(\frac{X_2 - \mu}{\sigma}\right) - \emptyset\left(\frac{X_1 - \mu}{\sigma}\right)$$

- حساب الاحتمال العام من الشكل $P(X \ge X_1)$:

$$P(X \ge X_1) = 1 - P(X \le X_1) = 1 - \emptyset\left(\frac{X_1 - \mu}{\sigma}\right)$$

نورد المثال الآتي: بفرض أن X هي درجة الحرارة ومتوسطها 25 درجة وانحرافها المعياري 5 درجات أوجد احتمال أن تكون درجة الحرارة أقل من 30. واحتمال أن تكون درجة الحرارة بين 35–20، واحتمال أن تكون درجة الحرارة أعلى من 40.

$$P(X \leq 30) = \emptyset\left(\frac{30-25}{5}\right) = \emptyset(1) = 0.8413$$

$$P(20 \leq X \leq 35) = \emptyset\left(\frac{35-25}{5}\right) - \emptyset\left(\frac{20-25}{5}\right) = \emptyset(2) - \emptyset(1)$$

$$= 0.9772 - 0.8413 = 0.1359$$

$$P(X \geq 40) = 1 - P(X \leq 40) = 1 - \emptyset\left(\frac{40-25}{5}\right) = \emptyset(3) = 0.9987$$

- حساب الاحتمالات المقابلة لمجالات الثقة الأساسية (النسب الشهيرة):

✓ المجال الأول: $P(\mu - \sigma \leq X \leq \mu + \sigma)$

نحسب:

$$Z_1 = \frac{(\mu - \sigma) - \mu}{\sigma} = -1$$

$$Z_1 = \frac{(\mu + \sigma) - \mu}{\sigma} = +1$$

ومن النسب الشهيرة فإن الاحتمال المطلوب قيمته هي 68.26% أي 0.6826

✓ المجال الثاني: $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma)$

نحسب:

$$Z_1 = \frac{(\mu - 2\sigma) - \mu}{\sigma} = -2$$

$$Z_1 = \frac{(\mu + 2\sigma) - \mu}{\sigma} = +2$$

ومن النسب الشهيرة فإن الاحتمال المطلوب قيمته هي 95.45% أي 0.9545

✓ المجال الثالث: $P(\mu - 3\sigma \leq X \leq \mu + 3\sigma)$

$$Z_1 = \frac{(\mu - 3\sigma) - \mu}{\sigma} = -3$$

$$Z_1 = \frac{(\mu + 3\sigma) - \mu}{\sigma} = +3$$

ومن النسب الشهيرة فإن الاحتمال المطلوب قيمته هي 99.73% أي 0.9973

إذا فترضنا أن متوسط الطول العام للسياح الذكور في مجتمع ما هو 176 CM بانحراف معياري قدره 7CM المطلوب تقدير عدد السياح الذكور الذين يقع طولهم في المجالات الآتية، مع العلم أن عدد أفراد المجتمع هو 3500 سائح.

المجال الأول: [183 – 169]، المجال الثاني: [190 – 162]، المجال الثالث: [197 – 155]،

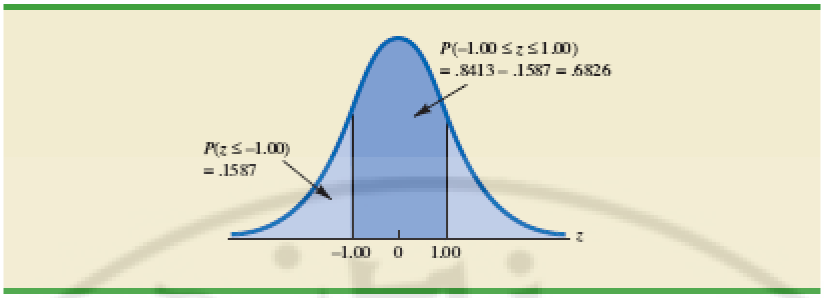
تحسب الاحتمالات كما يلي: المجال الأول:

$$Z_1 = \frac{169 - 176}{7} = -1$$

$$Z_1 = \frac{183 - 176}{7} = +1$$

إذا الاحتمال هو 0.6826 ومنه يكون عدد السائحين الذكور الذين أطوالهم تنتمي للمجال الأول:

3500*0.6826=2389.1=2389 سائح. ويوضح ذلك بالشكل:



رسم توضيحي 7 الاحتمال المقابل لمجال الثقة الأول

المجال الثاني:

$$Z_1 = \frac{162 - 176}{7} = -2$$

$$Z_1 = \frac{190 - 176}{7} = +2$$

إذا الاحتمال هو 0.9545 ومنه يكون عدد السائحين الذكور الذين أطوالهم تنتمي للمجال الثاني:

3500*0.9545=3340.75=3341 سائح. ويوضح بيانياً بنفس الطريقة المتبعة في المجال الأول.

المجال الثالث:

$$Z_1 = \frac{155 - 176}{7} = -3$$

$$Z_1 = \frac{197 - 176}{7} = +3$$

إذا الاحتمال هو 0.9973 ومنه يكون عدد السائحين الذكور الذين أطوالهم تنتمي للمجال الثاني:

3500*0.9973=3490.55=3491 سائح. ويوضح بيانياً بنفس الطريقة المتبعة في المجال الأول.

أسئلة للمناقشة:

1- مالفرق بين التوزيع الطبيعي والتوزيع الطبيعي المعياري من حيث الخصائص ودالة التوزيع؟

2- اكتب الصيغة الرياضية لكل من التوزيع الطبيعي، التوزيع الطبيعي المعياري، الدرجة المعيارية.

3- اختر الإجابة الصحيحة لكل مما يأتي:

- منحنى التوزيع الطبيهي هو:

A. منحنى متماثل.

B. منحنى معامل التوائه يساوي الصفر.

C. منحنى معامل تطاوله يساوي /3/.

D. كل ما ذكرصحيح.

- لدينا توزيع بمتوسط يساوي /1/ وانحراف معياري يساوي الصفر يمكن وصفه بــ

A. توزيع طبيعي معياري.

B. توزيع طبيعي عادي.

C. توزيع غير طبيعي .

D. غير ذلك.

- المساحة تحت منحنى التوزيع الطبيعي تساوي:

A. الواحد.

B. مجموع الاحتمالات $\sum P_i$

C. $\int_{-\infty}^{+\infty} f(x)dx$

D. كل ما ورد صحيح.

- لدينا توزيع طبيعي بمتوسط $\mu = 3,\ \sigma^2 = 2$ فإن الدرجة المعيارية لقيمة المتغير x المساوية لـــ 4 هي:

A. $\frac{4-3}{2}$
B. $\frac{4-3}{\sqrt{2}}$
C. $\frac{4-2}{3}$
D. غير ذلك

- احسب الدرجة المعيارية لدخل أحد العاملين في شركة سياحية ما البالغ /120 ألف وحدة نقدية إذا علمت بأن متوسط دخول العاملين في هذه الشركة 90 ألف وحدة نقدية بانحراف معياري 10 آلاف وحدة نقدية. وإذا علمت بأن عامل آخر في شركة سياحية أخرى دخله 150 ألف وحدة نقدية ومتوسط دخول العاملين في الشركة الثانية هو 125 ألف وحدة نقدية بانحراف معياري قدره 25 ألف وحدة نقدية، أي العاملين دخله هو الأفضل.

- إذا علمت بأن درجات الطلاب في إحدى المقررات في برنامج السياحة في الجامعة الافتراضية السورية تخضع للتوزيع الطبيعي بمتوسط يساوي64 درجة وانحراف معياري قدره 8 درجات والمطلوب:

  ✓ احتمال الحصول على طالب درجته أقل من 72.
  ✓ احتمال الحصول على طالب درجته أكثر من 80.
  ✓ احتمال الحصول على طالب درجته تتراوح بين60-70.

- إذا افترضنا متوسط الأعمار العام للسياح الوافدين لسورية هو 50 سنة بانحراف معياري /10/ سنوات والمطلوب تقدير عدد السياح الذين تقع أعمارهم في المجالات الآتية مع العلم بأن عدد السياح الوافدين إلى سورية هو 3000 سائح موضحاً بالرسم:

- ✓ $[40 - 60]$
- ✓ $[30 - 70]$

$[20 - 80]$