● علم الإحصاء

الإحصاء كعلم، هو الذي يهتم بطرق جمع البيانات، وتبويبها، وتلخيصها بشكل يمكن الاستفادة منها في وصف البيانات وتحليلها للوصول إلى قرارات سليمة في ظل ظروف عدم التأكد.

● وظائف علم الإحصاء

من التعريف السابق يمكن تحديد أهم وظائف علم الإحصاء في الآتي:

1. وصف البيانات (Data Description).
2. الاستدلال الإحصائي (Statistical Inference).
3. التنبؤ (Forecasting).

● أولا: وصف البيانات (Data Description)

تعتبر طريقة جمع البيانات وتبويبها وتلخيصها من أهم وظائف علم الإحصاء، إذ لا يمكن الاستفادة من البيانات الخام، وصف الظواهر المختلفة محل الاهتمام ، إلا إذا تم جمع البيانات وعرضها في شكل جداول، أو بياني من ناحية ، وحساب بعض المؤشرات الإحصائية البسيطة التي تدلنا على طبيعة البيانات من ناحية أخرى.

● ثانيا: الاستدلال الإحصائى (Statistical Inference)

ويستند الاستدلال الإحصائي على فكرة اختيار جزء من المجتمع (Population) يسمى عينة (Sample) بطريقة علمية مناسبة ، لغرض استخدام بيانات هذه العينة في التوصل إلى نتائج ، يمكن تعميمها على مجتمع الدراسة ، وعليه يهتم الاستدلال الإحصائي بموضوعين هما:

التقدير (Estimation) : وفيه يتم حساب مؤشرات من بيانات العينة تسمى إحصاء (statistics) تستخدم كتقدير لمؤشرات المجتمع وتسمى معالم (Parameters).

إختبارات الفروض (Tests of hypothesis) : وفيه يتم استخدام بيانات العينة للوصول إلى قرار علمي سليم بخصوص الفروض المحددة حول معالم المجتمع.

● ثالثا : التنبؤ (Forecasting)

وفيه يتم استخدام نتائج الاستدلال الإحصائي ، والتي تدلنا على سلوك الظاهرة في الماضي في معرفة ما يمكن أن يحدث لها في الحاضر والمستقبل.

● أهمية علم الاحصاء

يحتل الاحصاء مكاناً بين العلوم لما له من استعمالات واسعة للوصول الى قرارات صائبة لوصف او تفسير الظواهر المختلفة في جميع العلوم وهو المستعمل من قبل للافراد والجماعات المختلفة والدول على حد السواء وفي الحقيقة ان الانتصار العظيم في نزول الانسان على القمر ماكان يحدث لولا مساعدة علم الاحصاء ، واستخدم الاحصاء في مجالات كثيرة ونركز على اهمية علم الاحصاء في العلوم البايلوجية والطبية والصحة العامة والكيمياء .

1. **في علم الاحياء (البايولوجي) :-** تستخدم الطرق الاحصائية في دراسة الاجناس والفصائل المختلفة للحيوان والنبات ومعرفة خواص كل جنس بما يتميز عن غيره واختلاف مفردات الجنس الواحد في اي خاصية معينة من الناحية الاحصائية ، فمثلاً نرى الذكور في الجنس البشري اطول قامة من الاناث مع ان الذكور فيما بينهم يختلفون في الطول الى درجة ما وكذلك الاناث ، كل ذلك يتم عن طريق جمع البيانات وتبويبها ودراستها دراسة احصائية والخروج بنتائج من هذه الصفات .
2. **في الطب :-** يستخدم الاحصاء لدراسة العلاقة بين متغيرات كثيررة منها على سبيل المثال العلاقة بين العمر وضغط الدم وكذلك العلاقة بين الوراثة والبيئة وتأثيراتهما على تكوين الفرد .
3. **في الصحة العامة :-** يستخدم الاحصاء لدراسة الامراض السارية ونسبة زيادتها ونقصها في المجتمع وكذلك دراسة حالة المعوقين والوفيات ونسبة الزيادة في السكان .
4. **في الكيمياء :-** يستخدم الاحصاء لتحليل البيانات المتعلقة بتكرير النفط ومعرفة نسبة مكوناته وكذلك دراسة العلاقة بين الغازات او الفلزات او العمليات الكيميائية من ناحية تحليل البيانات المتعلقة بها وكذلك التجارب الكيمياوية في اعداد بحوث الماجستير والدكتوراه والبحوث العلمية الاخرى وغيرها من التجارب في مجال النفط والمعادن وجمع البيانات المتعلقة بها ودراستها دراسة احصائية لغرض الاستفادة منها في اعداد خطط التنمية الصناعية والبتروكيمياوية .

● أنواع البيانات

يمكن تقسيم البيانات إلى مجموعتين هما:
1. البيانات الوصفية (Qualitative Data).
2. البيانات الكمية (Quantitative Data).

أولا : البيانات الوصفية (Qualitative Data)

هي بيانات غير رقمية ، أو بيانات رقمية مرتبة في شكل مستويات أو في شكل فئات رقمية ، ومن ثم تقاس البيانات الوصفية بمعيارين هما:

1. بيانات وصفية مقاسة بمعيار اسمي (Nominal or classificatory data) : وهي بيانات غير رقمية تتكون من مجموعات متنافية ، كل مجموعة لها خصائص تميزها عن المجموعة الأخرى، كما أن هذه المجموعات لا يمكن المفاضلة بينها، ومن الأمثلة على ذلك:
- نوع الجنس : متغير وصفي تقاس بياناته بمعيار اسمي (ذكر – أنثى).
- الحالة الاجتماعية : متغير وصفي تقاس بياناته بمعيار اسمي(متزوج، أعزب ، أرمل ، مطلق). الجنسية.
- فصيلة الدم : متغير وصفي تقاس بياناته بمعيار اسمي (A,A+, A-,B,B+,…).
- حالة المادة (سائل ، صلب ، غاز).

2. بيانات وصفية مقاسة بمعيار ترتيبي (Ordinal Scale) (ترتيب تصاعدي أو تنازلي): وتتكون من مستويات، أو فئات يمكن ترتيبها تصاعديا أو تنازليا، ومن الأمثلة على ذلك:
- تقدير الطالب : متغير وصفي تقاس بياناته بمعيار ترتيبي (A,A+,B,B+,C,C+).
- حالات المرضى : متغير وصفي تقاس بياناته بمعيار ترتيبي (جيدة جداً ، جيدة ، مقبولة ، مستقرة ، ...).
- المستوى التعليمي : متغير وصفي تقاس بياناته بمعيار ترتيبي (أمي، يقرأ ويكتب،ابتدائية، متوسطة،ثانوية،جامعية،أعلى من جامعية).
- تركيز خلات الصوديوم المستخدم في حفظ لحوم الدجاج من البكتريا : متغير وصفي ترتيبي يقاس بياناته بمعيار ترتيبي(10%، 15%،20%،25%).

ثانيا : البيانات الكمية (Quantitative Data)

وهي البيانات التي يمكن قياسها مباشرة بأرقام عددية كالاختلاف بين الافراد في الطول والوزن ومستوى الهيموكلوبين والهرمونات وعدد خلايا الدم الحمراء ومستوى الدهون في مصل الدم Lipid profile (VLDL ، LDL-C ، HDL-C، TG ، TC) ويمكن قياسها بوحدات القياس المختلفة كالسنتيمتر والكيلوغرام (mg , pg ,g) وتنقسم المتغيرات الكمية الى :-

1. متغيرات متصلة او مستمرة Continuous variable

المتغير المتصل هو المتغير الذي تأخذ كل مفردة قيمة رقمية او كسر بين حدي المتغير الكلي فلو فرضنا اطوال الطلبة يتراوح بين (130.5 و 170 سم) ،كمية الهيموكلوبين (12.5 – 14 ملغم لكل لتر من الدم).

2. متغيرات غير متصلة او مستمرة Discontinuous Variable

هي المتغيرات التي تأخذ المشاهدة او المفردة فيها قيم متباعدة او متقطعة غير مستمرة اي هو الذي لا تأخذ كل مفردة فيه قيمة كسرية بل لا تزيد قيمة المتغير او تنقص بأقل من واحد فعدد الطلاب عدد الكتب كلها متغيرات غير متصلة او مستمرة .

● طرق جمع البيانات (Methods of Data Collections)

تعتبر طريقة جمع البيانات من أهم المراحل التي يعتمد عليها البحث الإحصائي، كما أن جمع البيانات بأسلوب علمي صحيح ، يترتب عليه الوصول إلى نتائج دقيقة في التحليل.

ولدراسة طرق جمع البيانات، يجب الإلمام بالنقاط التالية:

1. **مصادر البيانات (Data Sources).**
2. **أسلوب جمع البيانات (Data collection Methods).**
3. **أنواع العينات (Samples Types).**

أولاً : مصادر جمع البيانات

هناك مصدرين للحصول منها على البيانات هما:

1. **المصادر الأولية:** وهي المصادر التي نحصل منها على البيانات بشكل مباشر، حيث يقوم الباحث نفسه بجمع البيانات من المفردة محل البحث مباشرة ،

- ويتميز هذا النوع من المصادر بالدقة و الثقة في البيانات، لأن الباحث هو الذي يقوم بنفسه بجمع البيانات من المفردة محل البحث مباشرة.
- ولكن أهم ما يعاب عليها إنها تحتاج إلى وقت ومجهود كبير، وكذلك مكلفة من الناحية المادية.

2. **المصادر الثانوية :** وهي المصادر التي نحصل منها على البيانات بشكل غير مباشر (إي يتم الحصول على البينات جاهزة ولا يقوم الباحث بجمعها من المصدر قيد البحث).

- ومن مزايا هذا النوع من المصادر، توفير الوقت والجهد والمال،
- إلا أن درجة ثقة الباحث فيها ليست بنفس الدرجة في حالة المصادر الأولية.

ثانياً : أسلوب جمع البيانات (Data Collection Procedure)

يتحدد الأسلوب المستخدم في جمع البيانات ، حسب الهدف من البحث ، وحجم المجتمع البحث، ويوجد أسلوبين لجمع البيانات هما:

1. **أسلوب المسح الشامل:** يستخدم هذا الأسلوب إذا كان الغرض من البحث هو حصر جميع مفردات المجتمع ، وفي هذه الحالة يتم جمع بيانات عن كل مفردة من مفردات المجتمع بلا استثناء،
   - ويتميز أسلوب الحصر الشامل بالشمول وعدم التحيز، ودقة النتائج
   - ولكن يعاب عليه أنه يحتاج إلى الوقت والمجهود ، والتكلفة العالية.
2. **أسلوب المعاينة :** يعتم هذا الأسلوب على معاينة جزء من المجتمع (محل الدراسة) يتم اختياره بطريقة علمية سليمة ، ودراسته ثم تعميم نتائج العينة على المجتمع ككل .

- ومن ثم يتميز هذا الأسلوب بالآتي:
  - تقليل الوقت والجهد.
  - تقليل التكلفة.
  - الحصول على بيانات أكثر تفصيلا ، وخاصة إذا جمعت البيانات من خلال استمارة استبيان.
  - كما أن أسلوب المعاينة يفضل في بعض الحالات التي يصعب فيها إجراء حصر شامل، مثل معاينة دم المريض ، أو إجراء تعداد لعدد الأسماك في البحر...
- و يعاب على أساوب المعاينة : أن النتائج التي تعتمد على هذا الأسلوب أقل دقة من نتائج أسلوب المسح الشامل، وخاصة إذا كانت العينة المختارة لا تمثل المجتمع تمثيلا جيدا.

ثالثاً : أنواع العينات (Samples Types)

لمعرفة أنواع العينات ، يتم أولا تحديد الفرق بين مجتمع (Population) الدراسة، والعينة (Sample) المأخوذة من هذا المجتمع :

- **المجتمع (Population) :** هو مجموعة من المفردات التي تشترك في صفات، وخصائص محددة، ومجتمع الدراسة هو الذي يشمل جميع مفردات الدراسة ، أي هو الكل الذي نرغب دراسته.

والمجتمع اما ان يكون :-

- **مجتمع محدود Finite Population :** وهو المجتمع الذي يمكن حصر مفرداته كما هو الحال في مستوى الهيموكلوبين في دم طلبة جامعة كربلاء او عدد ردهات المرضى في مستشفى الحسين .
- **مجتمع غير محدود Infinite Population :** هو المجتمع الذي من الصعب او المستحيل حصر مفرداته مثل عدد البكتريا في مستعمرة بكترية او حقل معين .

- **العينة (Sample) :** هو جزء من المجتمع يتم اختيارها بطرق مختلفة لغرض دراسة هذا المجتمع.

ويتوقف نجاح استخدام أسلوب المعاينة على عدة عوامل هي:

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

أقسام العينات

ويمكن تقسيم العينات وفقا لأسلوب اختيارها إلى نوعين هما:

1. **العينات الاحتمالية :** هي العينات التي يتم اختيار مفرداتها وفقا لقواعد الاحتمالات ، بمعنى آخر هي التي يتم اختيار مفرداتها من مجتمع الدراسة بطريقة عشوائية ، بهدف تجنب التحيز الناتج عن اختيار المفردات، ومن أهم أنواع العينات الاحتمالية ، ما يلي:

- العينة العشوائية البسيطة (Simple Random Sample).
- العينة العشوائية الطبقية (Stratified Random Sample).
- العينة العشوائية المنتظمة (Systematic Random Sample).
- العينة العنقودية أو المتعددة المراحل (Cluster Sample).

العينة العشوائية البسيطة Simple Random Sample

وهي تلك العينة التي تسحب من مجتمع الدراسة بحيث يكون احتمال فرض ظهور اية مفردة من مفردات المجتمع الاحصائي في العينة متساوياً وبمعنى اخر تعني اعطاء كل فرد من المجتمع نفس الفرصة للظهور في العينة ويتم اختيارها كما يلي : مثل استخدام طريقة البطاقات او القرعة

العينة العشوائية الطبقية (Stratified Random Sample).

يتم في هذا النوع من العينة تقسيم المجتمع الاحصائي اولاً الى مجموعات فرعية تسمى كل منها (طبقة Strate) ومن ثم تتم عملية المعاينة من كل طبقة ، وعادة تكون جميع عناصر الطبقة الواحدة متجانسة فيما يتعلق بالخصائص موضوع الدراسة .

العينة العشوائية المنتظمة (Systematic Random Sample).

وهي اختيار العينات بشكل منتظم من قائمة المجتمع حيث يتم اختيارها من خلال ترقيم عناصر المجتمع الاحصائي بحيث يتم تحديد قاعدة للاختيار تستند على تحديد اختيار العنصر الاول ولتبسيط الشرح لو كان مجتمع الاصل (100 مريض) وتريد اختيار (10 مرضى) لأجراء بعض الفحوصات عليهم فمثلاً تأخذ الارقام العشرة الاولى وتوضع في صندوق ويتم السحب.

العينة العنقودية متعددة المراحل Multi-stage cluster sample

تعتبر المعاينة العنقودية احد الآليات التي يمكن استخدامها لاختيار العينات من خلال تقسيم المجتمع الى مجموعات او عناقيد على سبيل المثال نريد التعرف على مستوى التعليم الطبي في العراق فنختار 5 كليات طب ومن كل كلية فرعين ومن كل فرع 5 طلاب

2. العينات غير الاحتمالية

هي التي يتم اختيار مفرداتها بطريقة غير عشوائية ، حيث يقوم الباحث باختيار مفردات العينة بالصورة التي تحقق الهدف من المعاينة ، مثل اختيار عينة ، وأهم أنواع العينات غير الاحتمالية:

1. العينة العمدية (Judgmental Sample).
2. العينة الحصصية (Quota Sample).

● طرق عرض البيانات

الخطوة التالية بعد جمع البيانات ، هو تبويب البيانات وعرضها بصورة يمكن الاستفادة منها في وصف الظاهرة محل الدراسة، من حيث تمركز البيانات، ودرجة تجانسها.

وهناك طريقتين لعرض البيانات هما:

1. عرض البيانات جدوليا.
2. عرض البيانات بيانيا.

أولاً : عرض البيانات جدوليا

يمكن عرض البيانات في صورة **جدول تكراري (Frequency Tables)**، ويختلف شكل الجدول طبقا لنوع البيانات ، وحسب عدد المتغيرات، وفيما يلي عرض بيانات متغير(وصفي أو كمي) في شكل جدول تكراري بسيط.

- **عرض بيانات المتغير الوصفي في شكل جدول تكراري بسيط**

في حال دراسة ظاهرة ما تحتوي على متغير وصفي واحد ، فإنه يمكن عرض بياناتها في شكل **جدول تكراري بسيط (Simple frequency table)**، وهو جدول يتكون من عمودين، أحدهما به مستويات (مجموعات) المتغير، والثاني به عدد المفردات (التكرارات (لكل مستوى) مجموعة.

التكرار (Frequency) (التكرار المطلق): وهو عدد مرار حصول الحدث (event) لعينة واحدة (sample) ضمن نفس المجتمع (population) ويوجد نوعين من التكرارات الإحصائية هما :

1. **التكرار المطلق (Absolute Frequency)**.
2. **التكرار النسبي (Relative Frequency) :**وهو حاصل قسمة التكرار المطلق على عدد المحاولات (القراءات) الكلي (n).

◈ مثال

لو فرضنا إنه قمنا بـ (60) رمية عشوائية لقطعة زهر ، فإن إحتمال ظهور أي وجه هي (10) مرات ، وعليه يكون عدد مرات التكرار المطلق لظهور اي وجه هي (10) ، وعدد التكرارات الكلي (n) هو (60) و التكرار النسبي لظهور أي وجه هو ($\frac{10}{60}=\frac{1}{6}$) .

مثال

فيما يلي بيانات عينة من(40) مزرعة عن نوع التمر الذي تنتجه المزرعة.(مرمزة بحروف)

B	**A**	**C**	**D**	**C**	**E**	**C**	**B**	**D**	**C**	**A**	**C**	**B**	**D**	**B**	**C**	**B**	**C**	**B**	**A**
B	**A**	**C**	**A**	**E**	**D**	**C**	**B**	**D**	**E**	**D**	**E**	**A**	**C**	**B**	**C**	**B**	**D**	**C**	**C**

والمطلوب:

1. ما هو نوع المتغير؟، وما هو المعيار المستخدم في قياس البيانات؟.
2. اعرض البيانات في شكل جدول تكراري.
3. كون جدول التوزيع التكراري النسبي.

الحل

1. (نوع التمر) هو متغير وصفي، تقاس بياناته بمعيار اسمي (A ، B ، C ، D ، E).
2. لعرض البيانات في شكل جدول تكراري ، يتم إتباع الآتي:

نوع التمر	عدد المزارع(عدد التكرارات) (f)	التكرار النسبي = $\frac{f}{\Sigma}$
A	**6**	**6/40 = 0.150**
B	**10**	**10/40 = 0.250**
C	**13**	**13/40 = 0.325**
D	**7**	**7/40 = 0.175**
E	**4**	**4/40 = 0.100**
المجموع Σ	**40**	**40/40 = 1**

حيث نلاحظ أنه يمكن حساب التكرار النسبي (Relative frequency) من خلال قسمة تكرار (f) كل نوع على مجموع التكرارات الكلي (n)

التكرار النسبي = (تكرار المجموعة) ÷ (مجموع التكرارات n) = $\frac{f}{\Sigma f}$

مثال

فيما يلي المستوى التعليمي ل (50) شخص

A	**C**	**B**	**D**	**B**	**D**	**E**	**B**	**F**	**B**
D	**A**	**E**	**C**	**A**	**F**	**C**	**E**	**C**	**E**
E	**B**	**D**	**D**	**A**	**C**	**D**	**A**	**D**	**D**
C	**D**	**B**	**C**	**B**	**B**	**D**	**D**	**C**	**B**
D	**C**	**B**	**C**	**D**	**D**	**C**	**A**	**D**	**C**

والمطلوب:

1. اعرض البيانات في شكل جدول تكراري.
2. كون التوزيع التكراري النسبي.

الحل

المستوى	*f*	*f*/Σ
A	**6**	**0.12**
B	**10**	**0.20**
C	**12**	**0.24**
D	**15**	**0.30**
E	**5**	**0.10**
F	**2**	**0.04**
Σ	**50**	**1.00**

● عرض بيانات المتغير الكمي في شكل جدول تكراري بسيط

بنفس الأسلوب السابق المتبع في تكوين جدول تكراري، يمكن أيضا عرض بيانات المتغير الكمي في شكل جدول تكراري بسيط ، ويتكون هذا الجدول من عمودين ، الأول يحتوي على فئات تصاعدية للقراءات التي يأخذها المتغير، والثاني يشمل التكرارات أو عدد المفردات التي تنتمي قراءات الفئة المناسبة لها، والمثال التالي يبين كيف يمكن عرض البيانات الكمية بيانيا.

◈ مثال

فيما يلي بيانات درجات (70) طالب في الاختبار النهائي لمقرر مادة الإحصاء التطبيقي.

56	65	70	65	55	60	66	75	56	70	76
60	70	61	67	61	71	67	71	66	62	82
68	72	57	68	72	69	57	69	75	71	77
72	62	67	73	58	63	66	63	65	73	83
58	73	74	76	74	80	81	74	58	60	77
76	82	77	83	77	85	91	94	72	78	85
79	64	57	79	55	87	64	78	62	88	91

والمطلوب:

1. كون التوزيع التكراري لدرجات الطلاب.
2. كون التوزيع التكراري النسبي.
3. ما هو نسبة الطلاب الحاصلين على درجة ما بين 70 إلى أقل من80 .
4. ما هو نسبة الطلاب الحاصلين على درجة أقل من 70 درجة.
5. ما هو نسبة الطلاب الحاصلين على درجة 80 أو أكثر .

☑ الحل

1. تكوين التوزيع التكراري:

إن درجة الطالب في الاختبار متغير كمي مستمر، ولكي يتم تبويب البيانات في شكل جدول تكراري، يتم اتباع الآتي:

– **تقسيم الدرجات إلى فئات (Classes "K").**

– يتم تحديد عدد الفئات (K) وفقاً لإعتبارات عدة، منها:

رأي الباحث ، الهدف من البحث ، حجم البيانات.

وإن أفضل عدد للفئات يتراوح بين (5 – 15).

في هذا المثال سيتم إتخاذ عدد الفئات (k=8).

– **يتم حساب عدد الفئات (K) عادة بإستخدام صيغة (Sturges Formula) :**

$$\boxed{K \approx 1 + 3.3(\log_{10} n)}$$

حيث (K) : عدد الفئات ، (n) عدد المشاهدات (القراءات) الكلي.

$K \approx 1 + 3.3 (\log_{10} 70) \approx 8$ (where $\log_{10} 70 \approx 2$)

– **لحساب المدى (Range "R"): المدى = الفرق بين أكبر درجة و أوطأ درجة**

$$\boxed{\text{Range } (R) = X_{max} - X_{min}}$$

حيث (X_{max}) أكبر قيمة عددية للمشاهدات ،(X_{min}) أصغر قيمة عددية للمشاهدات.

$(\therefore\ R = 94 - 55 = 39)$

– **حساب طول الفئة الواحدة (Length "L") يتم قسمة المدى (R) على عدد الفئات (k) :**

$$\boxed{L = R / K}$$

$$(\therefore L = \frac{R}{K} = \frac{39}{8} \approx 5)$$

∴ طول الفئة الواحدة هو (5).

– تحديد الفئات: الفئة تبدأ بقيمة دنيا تسمى الحد الأدنى و تنتهي بقيمة عليا تسمى الحد الأعلى، وعليه فإن الحد الأدنى للفئة الأولى = أقل قراءة = 55 ، و الحد الأعلى للفئة الأولى = الحد الأدنى + طول الفئة (L) = 55 + 5 = 60 ، و عليه فإن الفئة الأولى هي : (60 – 55) (55 – less than 60)

C1 = 55 – less than 60
C2 = 60 – less than 65
C3 = 65 – less than 70
C4 = 70 – less than 75
C5 = 75 – less than 80
C6 = 80 – less than 85
C7 = 85 – less than 90
C8 = 90 – less than 95

k	*f*	*f*/Σ
55 – 60	10	0.143
60 – 65	12	0.171
65 – 70	13	0,186
70 – 75	16	0.229
75 – 80	10	0.143
80 – 85	4	0.057
85 – 90	3	0.043
90 - 95	2	0.028
Σ	70	1.00

● العرض البياني للبيانات الكمية

العرض البياني للبيانات ، هو أحد طرق التي يمكن استخدامها في وصف البيانات ، من حيث شكل التوزيع ومدى تمركز البيانات ، وفي كثير من النواحي التطبيقية يكون العرض البياني أسهل وأسرع في وصف الظاهرة محل الدراسة ، وتختلف طرق عرض البيانات بيانيا حسب نوع البيانات المبوبة في شكل جدول تكراري ، وفيما يلي عرض للأشكال البيانية المختلفة.

أولاً : المدرج التكراري (Histogram)

المدرج التكراري هو التمثيل البياني للجد ول التكراري البسيط الخاص بالبيانات الكمية المتصلة، وهو عبارة عن أعمدة بيانية متلاصقة ، حيث تمثل التكرارات على المحور الرأسي ، بينما تمثل قيم المتغير(حدود الفئات) على المحور الأفقي ، ويتم تمثيل كل فئة بعمود ، ارتفاعه هو تكرار الفئة ، وطول قاعدته هو طول الفئة.

◈ مثال

فيما يلي التوزيع التكراري لأوزان عينة من الدواجن بالغرام ، حجمها 100 اختيرت من أحد المزارع بعمر 45 يوم.

الوزن (غم)	**600 -**	**620 -**	**640 -**	**660 -**	**680 -**	**700 - 720**	**sum**
عدد الدجاج	**10**	**15**	**20**	**25**	**20**	**10**	**100**

أوجد :

1. طول الفئة.
2. المدرج التكراري.
3. المدرج التكراري النسبي.

☑ الحل

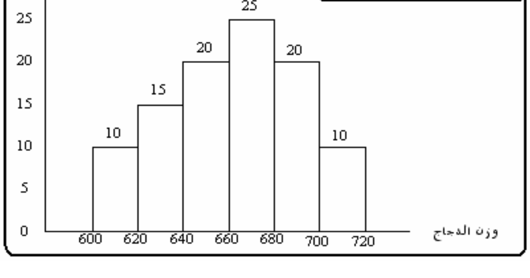
1. طول الفئة (L) = الحد الأعلى للفئة – الحد الأدنى للفئة

L = 620-600 = 640-620 = 660-640 = ….. = 720-700= 20

∴ طول الفئة = 20 .

2. لرسم المدرج التكراري :

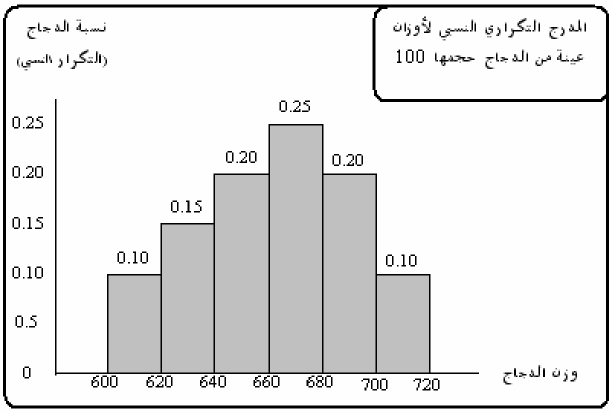
- يتم تمثيل المحور الأفقي للأوزان ، و المحور الرأسي يمثل التكرارات .
- كل فئة تمثل بعمود إرتفاعه = تكرار الفئة و طول قاعدته = طول الفئة (L).
- كل عمود يبدأ من حيث انتهى به عمود الفئة السابقة .
- الشكل التالي يمثل المدرج التكراري لأوزان الدجاج:



3. لرسم المدرج التكراري النسبي : يتم حساب التكرارات النسبية

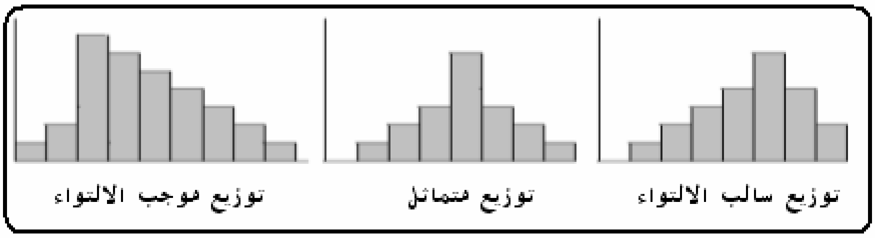
الوزن	600 -	620 -	640 -	660 -	680 -	700 - 720	sum
عدد الدجاج	10	15	20	25	20	10	100
التكرار النسبي	0.10	0.15	0.20	0.25	0.20	0.10	1.00

و بإتباع نفس الخطوات السابقة عند رسم المدرج التكراري، يتم رسم المدرج التكراري النسبي، بإحلال التكرارات النسبية محل التكرارات المطلقة على المحور الرأسي، كما هو مبين في الشكل التالي:



- **ملاحظات على شكل المدرج التكراري**

1. أن المساحة أسفل المدرج التكراري تساوي مجموع التكرارات (n) .
2. أما المساحة أسفل المدرج التكراري النسبي ، فهي تعبر عن مجموع التكرارات النسبية، وهي تساوي الواحد الصحيح.
3. يمكن تقدير القيم الشائعة ، وهي القيم التي يناظرها أكبر ارتفاع.
4. يمكن معرفة شكل توزيع البيانات ، كما هو مبين بالأشكال الثلاث التالية:



ثانياً : المضلع التكراري

هو تمثيل بياني أيضا للجدول التكراري البسيط ، حيث تمثل التكرارات على المحور الرأسي ، ومراكز الفئات على المحور الأفقي ، ثم التوصيل بين الإحداثيات بخطوط منكسرة ، وبعد ذلك يتم توصيل طرفي المضلع بالمحور الأفقي.

ومركز الفئة هي القيمة التي تقع في منتصف الفئة ، وتحسب بتطبيق المعادلة التالية:

مركز الفئة = (الحد الأدنى للفئة + الحد الأعلى للفئة) ÷ 2

$$\text{Midpoint} = \frac{\text{Lower} + \text{Upper}}{2}$$

ونظرا لعدم معرفة القيم الفعلية لتكرار كل فئة، يعتبر مركز الفئة هو التقدير المناسب لقيمة كل مفردة من مفردات الفئة.

◈ مثال

لرسم المضلع التكراري في المثال السابق:

الوزن	600 -	620 -	640 -	660 -	680 -	700 - 720	sum
عدد الدجاج	10	15	20	25	20	10	100
التكرار النسبي	0.10	0.15	0.20	0.25	0.20	0.10	1.00

☑ الحل

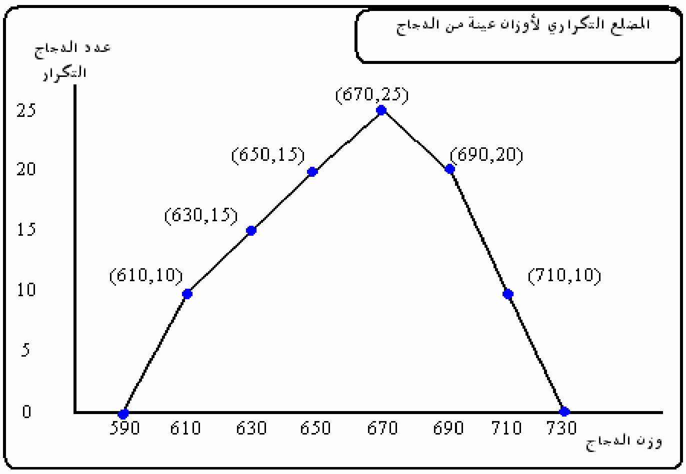
لرسم المضلع التكراري يتبع الآتي:

يتم حساب مراكز الفئات بتطبيق المعادلة الخاصة بحساب مراكز الفئات:

الوزن	*f*	Midpoint(x)
600-	10	$\frac{600+620}{2}=610$
620-	15	$\frac{620+640}{2}=630$
640-	20	650
660-	25	670
680-	20	690
700-720	10	710
Sum	100	

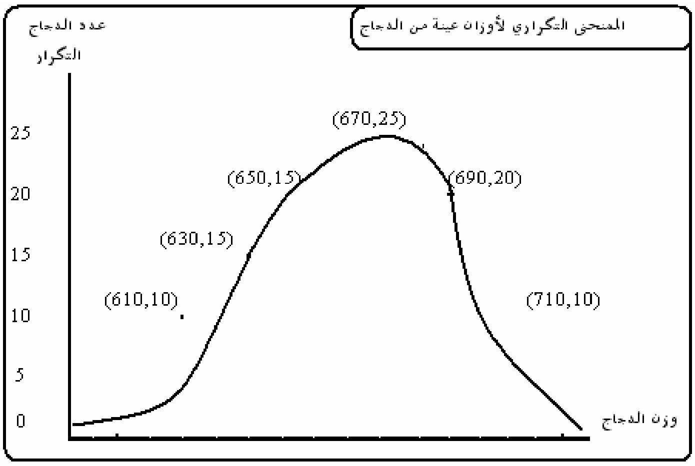
∴ نقط الإحداثيات هي :

Midpoint (x)	590	610	630	650	670	690	710	730
f (y)	0	10	15	20	25	20	10	0

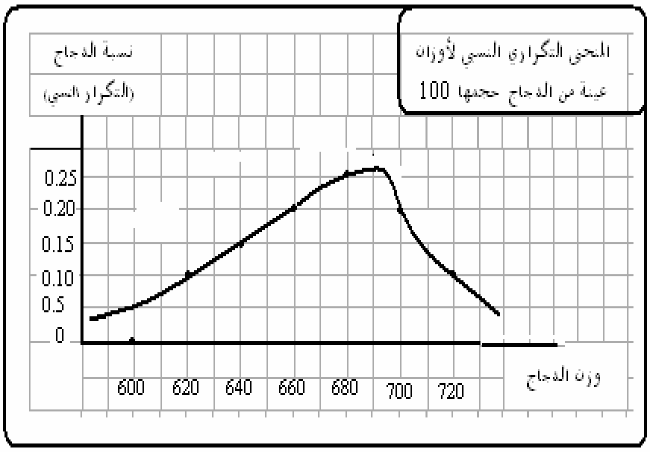


ثالثاً : المنحنى التكراري (Frequency Diagram)

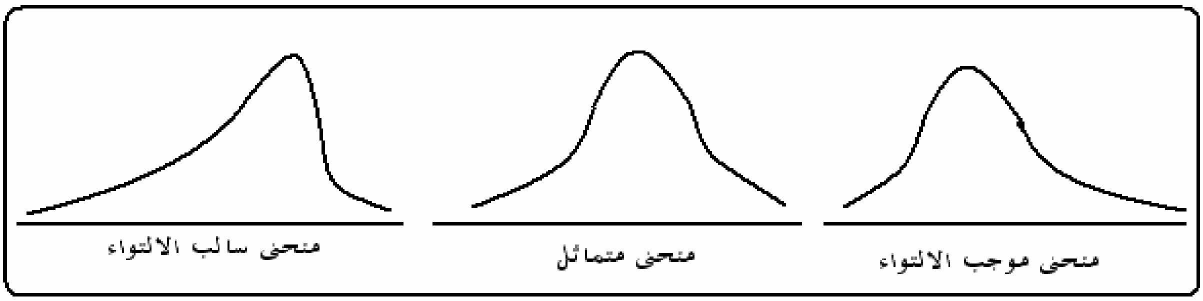
بإتباع نفس الخطوات السابقة في رسم المضلع يمكن رسم المنحنى التكراري ، ولكن يتم تمهيد الخطوط المنكسرة في شكل منحنى بحيث يمر بأكثر عدد من النقاط ، وفي المثال السابق يمكن رسم المنحنى



كما يمكن رسم المنحنى التكراري النسبي بتمثيل التكرارات النسبية على المحور الرأسي بدلا من التكرارات المطلقة، ومن ثم يأخذ هذا المنحنى الشكل التالي :



والمنحنى التكراري أعلاه موجب الالتواء، كما أن المساحة أسفل هذا المنحنى تعبر عن مجموع التكرارات النسبية، أي إنها ا تساوي الواحد الصحيح ، وهناك أشكال مختلفة للمنحنى التكراري النسبي، تدل على أشكال توزيع البيانات، ومن أهمها ما يلي:



مقاييس النزعة المركزية (Central Tendency)

في كثير من النواحي التطبيقية يكون الباحث في حاجة إلى حساب بعض المؤشرات التي يمكن الاعتماد عليها في وصف الظاهرة من حيث القيمة التي تتوسط القيم أو تنزع إليها القيم ، ومن حيث التعرف على مدى تجانس القيم التي يأخذها المتغير، وأيضا ما إذا كان هناك قيم شاذة أم لا .

والاعتماد على العرض البياني وحدة لا يكفى ، ولذا سيتم تناول بعض المقاييس الإحصائية التي يمكن من خلالها التعرف على خصائص الظاهرة محل البحث ، وكذلك إمكانية مقارنة ظاهرتين أو أكثر.

تسمى مقاييس الترعة المركزية **بمقاييس الموضع (Location)** أو **المتوسطات** ، ومن هذه المقاييس ، **الوسط الحسابي (Mean)** ، **والمنوال (Mode)** ، **والوسيط (Median)** .

الوسط الحسابي (Mean)

من أهم مقاييس الترعة المركزية ، وأكثرها استخداما في النواحي التطبيقية ، ويمكن حسابه للبيانات المبوبة وغير المبوبة ، كما يلي:

أولا : الوسط الحسابي للبيانات غير المبوبة

يعرف الوسط الحسابي بشكل عام على أنه مجموع القيم مقسوما على عددها . فإذا كان لدينا عدد (n) من القيم وهي $(x_1,x_2,x_3,\ldots\ldots,x_n)$ ، وعليه فإن **الوسط الحسابي " $\bar{X}$ "(Mean)** لهذه القيم يحسب كالأتي:

$$\boxed{\bar{X}=\frac{\sum_{n=1}^{n}x_n}{n}=\frac{x_1+x_2+x_3+\ldots..+x_n}{n}}$$

حيث يدل الرمز (Σ): على مجموع كل قيم (x).

مثال

أوجد الوسط الحسابي لدرجات (8) طلاب في إختبار مادة معينة.

34	32	42	37	35	40	36	40

الحل

$$\bar{X}=\frac{\sum_{n=1}^{8}x_n}{8}=\frac{40+36+40+35+37+42+32+34}{8}=\frac{296}{8}=37$$

أي أن الوسط الحسابي لدرجة الطالب يساوي 37 درجة .

ثانيا :الوسط الحسابي للبيانات المبوبة

من المعلوم أن القيم الأصلية ، لا يمكن معرفتها من جدول التوزيع التكراري ، حيث أن هذه القيم موضوعة في شكل فئات ، ولذا يتم التعبير عن كل قيمة من القيم التي تقع داخل حدود الفئة بمركز هذه الفئة ، ومن ثم يؤخذ في الاعتبار أن مركز الفئة هو القيمة التقديرية لكل مفردة تقع في هذه الفئة.

فإذا كانت (K) هي عدد الفئات ، وكانت $(\chi_1,\chi_2,\chi_3,\ldots.\chi_k)$ هي مراكز الفئات، وإن $(f_1, f_2, f_3,\ldots.., f_k)$ هي التكرارات ، فإن الوسط الحسابي (Mean) يحسب بالمعادلة التالية :

$$\overline{\chi} = \frac{\chi_1 f_1 + \chi_2 f_2 + \ldots + \chi_k f_k}{f_1 + f_2 + \ldots + f_k} = \frac{\sum_{i=1}^{k} \chi_i f_i}{\sum_{i=1}^{k} f_i}$$

مثال

الجدول التالي يبين توزيع (40) تلميذ حسب أوزانهم، و المطلوب إيجاد الوسط الحسابي.

فئات الوزن	32-34	34-36	36-38	38-40	40-42	42-44
عدد التلاميذ	4	7	13	10	5	1

الحل

لحساب الوسط الحسابي في هذه الحالة ، يتوجدب علينا حساب مايلي:

1. مجموع التكرارت.
2. مراكز الفئات (χ).
3. حاصل ضرب مركز الفئة في تكرارها(χf).
4. مجموع حاصل ضرب مركز الفئة في تكرارها $(\Sigma\chi f)$.

فئات الوزن (C)	التكرارات (f)	مراكز الفئات (χ)	$\chi \cdot f$
32-34	4	$\frac{32+34}{2} = 33$	132
34-36	7	35	245
36-38	13	37	481
38-40	10	39	390
40-42	5	41	205
42-44	1	43	43
Σ	40		1496

$$\therefore \ \overline{\chi} = \frac{\sum_{i=1}^{k} \chi_i f_i}{\sum_{i=1}^{k} f_i} = \frac{1496}{40} = 37.4 \text{k.g}$$

● خصائص الوسط الحسابي

1. الوسط الحسابي للمقدار الثابت يساوي الثابت نفسه ، اي إذا كانت قيم (χ) هي : $(\chi = a,a,a,..,a)$ فإن الوسط الحسابي لهذه القيم هو (a).

2. مجموع إنحرافات القيم عن وسطها الحسابي يساوي صفر: $\boxed{\sum\left(x-\overline{x}\right)=0}$.

ويمكن التحقق من الخاصية أعلاه بإستخدام البيانات في مثال (درجات (8) الطلبة المذكور سابقا)، حيث نجد إن الوسط الحسابي لدرجاتهم هو (37) و عليه :

χ	34	32	42	37	35	40	36	40	296
$(x-\overline{x}) = (x-37)$	34-37= -3	32-37= -5	42-37= 5	37-37= 0	35-37= -2	40-37= 3	36-37= -1	40-37= 3	0

3. إذا أضيف مقدار ثابت إلى كل قيمة من القيم ، فإن الوسط الحسابي للقيم المعدلة (بعد الإضافة) يساوى الوسط الحسابي للقيم الأصلية (قبل الإضافة) مضافا إليها هذا المقدار الثابت.

أي إنه :

$$\boxed{\bar{y} = \bar{x} + a}$$

4. إذا ضرب مقدار ثابت (a) في كل قيمة من القيم ، فإن الوسط الحسابي للقيمة المعدلة (القيمة الناتجة بعد الضرب)= الوسط الحسابي الأصلي (للقيم قبل الضرب) مضروباً في هذا المقدار الثابت (a) ، أي إن :

$$\boxed{\bar{y} = \bar{x} \times a}$$

5. مجموع مربعات انحرافات القيم عن وسطها الحسابي أقل ما يمكن ، أي أن:

$$\boxed{\sum(x-\bar{x})^2 < \sum(x-a)^2 \quad \textit{if } a \neq \bar{x}}$$

ثالثا :الوسط الحسابي المرجح (Weighted Mean)

في بعض الأحيان يكون لكل قيمة من قيم المتغير أهمية نسبية تسمى الوزن (weight "*w*") ، أو الترجيح ، وعند عدم أخذ هذه الأوزان في الاعتبار عند حساب الوسط الحسابي ، تكون القيمة المعبرة عن الوسط الحسابي غير دقيقة ، فمثلا لو أخذنا خمسة طلاب ، وسجلنا درجات هؤلاء الطلاب في مادة معينة ، وعدد ساعات المذاكرة في الأسبوع.

الطالب	1	2	3	4	5	Sum
الدرجة (χ)	23	40	36	28	46	173
عدد ساعات المذاكرة (*w*)	1	3	3	2	4	

نجد إن الوسط الحسابي غير المرجح للدرجة التي حصل عليها الطالب هي :

$$\bar{\chi} = \frac{\sum \chi}{n} = \frac{23+40+36+28+46}{5} = 34.6$$

وإذا أردنا أن نحسب الوسط الحسابي للدرجات (χ) المرجحة بعدد ساعات المذاكرة (w) يتم تطبيق المعادلة الأتية :

$$\bar{w} = \frac{\sum(\chi w)}{\sum w} = \frac{(23\times1)+(40\times3)+(36\times3)+(28\times2)+(46\times4)}{1+3+3+2+4} = \frac{491}{13} = 37.769$$

وهذا الوسط الحسابي يكون أكثر دقة من الوسط الحسابي غير المرجح.

وعليه فإن الوسط الحسابي المرجح (weighted mean) يتم حسابه من المعادلة الأتية:

$$\boxed{\bar{w} = \frac{\sum(\chi w)}{\sum w}}$$

- **مزايا وعيوب الوسط الحسابي**

1. أنه سهل الحساب.
2. يأخذ في الاعتبار كل القيم.
3. أنه أكثر المقاييس استخداما وفهما.

ومن عيوبه :

1. أنه يتأثر بالقيم الشاذة والمتطرفة .
2. يصعب حسابه في حالة البيانات الوصفية .
3. يصعب حسابه في حالة الجداول التكرارية المفتوحة .

● الوسيط (Median)

هو أحد مقاييس النزعة المركزية ، والذي يأخذ في الاعتبار رتب القيم ، ويعرف الوسيط بأنه القيمة التي يقل عنها نصف عدد القيم ($\frac{n}{2}$) ، ويزيد عنها النصف الأخر ($\frac{n}{2}$) أي إن (50%) من القيم أقل منه و (50%) من القيم اعلى منه ، وفيما يلي كيفية حساب الوسيط في حالة البيانات غير مبوبة ، والبيانات المبوبة.

أولاً : الوسيط للبيانات غير المبوبة

لبيان كيف يمكن حساب الوسيط للبيانات غير المبوبة ، نتبع الخطوات التالية:

1. ترتب القيم تصاعديا .
2. تحديد رتبة الوسيط ، وهي :

- ففي حالة كون عدد القيم (n) فردياً فإن الوسيط يتمثل بالقيمة ذات الرتبة $\left(\frac{n}{2}+1\right)$.
- وفي حالة كون عدد القيم (n) زوجياً فإن الوسيط يقع بين القيمة ذات الرتبة ($\frac{n}{2}$) و القيمة ذات الرتبة $\left(\frac{n}{2}+1\right)$، وعليه فإن الوسيط يتم حسابه بأخذ معدل القيمتين ذواتا المراتب ($\frac{n}{2}$) و $\left(\frac{n+1}{2}\right)$ و كما يلي :

$$\text{الوسيط} = \frac{\text{القيمة رقم}\left(\frac{n}{2}\right) + \text{القيمة رقم}\left(\frac{n}{2}+1\right)}{2}$$

◈ مثال

تم تقسيم قطعة أرض زراعية إلى 17 وحدة تجريبيةمتشابهة ، وتم زراعتها بمحصول القمح ، وتم استخدام نوعين من التسميد هما : النوع (a) و جرب على (7) وحدات تجريبية ، و النوع (b) وجرب على (10) وحدات تجريبية ، وبعد إنتهاء الموسم الزراعي ، تم تسجيل إنتاجية الوحدة بالطن/ هكتار ، وكانت على النحو التالي:

النوع (a)	**1.2**	**2.75**	**3.25**	**2**	**3**	**2.3**	**1.5**			
النوع (b)	**4.5**	**1.8**	**3.5**	**3.75**	**2**	**2.5**	**1.5**	**4**	**2.5**	**3**

والمطلوب حساب وسيط الإنتاج لكل نوع من السماد المستخدم ، ثم قارن بينها.

☑ الحل

1. حساب وسيط الإنتاج للنوع الأول (a) :

- ترتيب القيم تصاعديا (حيث إن عدد القيم هنا (n = 7) وهو عدد فردي ، وإن الوسيط يتمثل بالعدد ذي المرتبة ($4 = \frac{7+1}{2} = \frac{n+1}{2}$)

				قيمة الوسيط			
الإنتاج	1.2	1.5	2	**2.3**	2.75	3	3.25
الرتبة	1	2	3	**4**	5	6	7
				رتبة الوسيط			

$\therefore\ Med_a = 2.3$ Ton/Hectare

2. حساب الوسيط لإنتاج النوع الثاني (b) :

- ترتيب القيم تصاعديا (حيث إن عدد القيم هنا (n = 10) وهو عدد زوجي ، وإن الوسيط يتمثل بالعدد الواقع بين المرتبتين $(\frac{n}{2})$ و $\left(\frac{n}{2}+1\right)$ أي بين المرتبتين (5) و (6) وقيمة الوسيط تمثل معدل القيمتين الواقعتين في هاتين المرتبتين.
- رتبة الوسيط = $5.5 = \frac{5+6}{2}$
- قيمة الوسيط = $2.75 = \frac{2.5+3}{2}$

					2.75					
الإنتاج	1.5	1.8	2	2.5	**2.5**	**3**	3.5	3.75	4	4.5
الرتبة	1	2	3	4	**5**	**6**	7	8	9	10
					5.5					

$\therefore$ Med_b=2.75 Ton/Hectare

وبمقارنة النوعين من السماد ، نجد إن وسيط إنتاجية النوع (a) أقل من وسيط إنتاجية النوع (b).

ثانيا :الوسيط للبيانات المبوبة

لحساب الوسيط من بيانات مبوبة في جدول توزيع تكراري ، يتم إتباع الخطوات التالية:

1. تكوين الجدول التكراري المتجمع الصاعد.

2. تحديد رتبة الوسيط : $\left(\frac{n}{2}\right)=\left(\frac{\sum f}{2}\right)$

3. تحديد فئة الوسيط كما يلي :

تكرار متجمع صاعد سابق f_1 ← الحد الأدنى لفئة الوسيط (A)

رتبة الوسيط (n/2) ← الوسيط Med

تكرار متجمع صاعد لاحق f_2 ← الحد الأعلى لفئة الوسيط

ويحسب الوسيط بتطبيق المعادلة :

$$\boxed{Med = A + \frac{\frac{n}{2}-f_1}{f_2-f_1}\times L}$$

حيث (L) طول فئة الوسيط = الحد الأعلى – الحد الأدنى.

- **ملاحظة**

يتم تكوين التكرار المتجمع الصاعد من خلال حساب مجموع التكرارات (عدد القيم) التي تقل عن كل حد من حدود الفئات.

◈ مثال

فيما يلي توزيع (50) عجل متوسط الحجم، حسب إحتياجاته اليومية من الغذاء الجاف(بالكيلوغرام):

فئات الإحتياجات اليومية	1.5 -	4.5 -	7.5 -	10.5 -	13.5 – 16.5
عدد العجول *f*	4	12	19	10	5

أوجد الوسيط حسابيا.

☑ الحل

حساب الوسيط

- **رتبة الوسيط :** $\frac{n}{2}=\frac{\sum f}{2}=\frac{50}{2}=25$
- **تكوين الجدول التكراري المتجمع الصاعد:**

	أقل من	تكرار متجمع صاعد	
	1.5	0	
	4.5	0+4=4	
A	7.5	0+4+12=16	f_1
	Med.	25	
	10.5	0+4+12+19=35	f_2
	13.5	45	
	16.5	50	

- تحديد فئة الوسيط : وهى الفئة التي تشمل قيمة الوسيط ، وهي قيمة أكبر من (n/2) من القيم و أقل من (n/2) من القيم ، ويمكن معرفتها بتحديد التكرارين المتجمعين الصاعدين الذين يقع بينهما رتبة الوسيط.

ومن خلال الجدول المتجمع الصاعد أعلاه نلاحظ مايلي:

$f_1=16$, $f_2=35$, $A=7.5$, $L=10.5-7.5=3$

$$\therefore\ Med=A+\frac{\frac{n}{2}-f_1}{f_2-f_1}\times L=7.5+\frac{25-16}{35-16}\times 3=8.921\ \text{K.g}$$

● مزايا و عيوب الوسيط

من مزايا الوسيط:

1. لا يتأثر بالقيم الشاذة أو المتطرفة.
2. سهل في الحساب.
3. مجموع قيم الانحرافات المطلقة عن الوسيط أقل من مجموع الانحرافات المطلقة عن أي قيم أخرى ، أي إن $\sum|x-Med|\le\sum|x-a|\ ,a\neq Med$

ومن عيوب الوسيط:

1. أنه لا يأخذ عند حسابه كل القيم في الاعتبار، فهو يعتمد على قيمة أو قيمتين فقط .
2. يصعب حسابه في حالة البيانات الوصفية المقاسة بمعيار اسمي (Nominal) .

● المنوال (Mode)

ويعرف المنوال على إنه القيمة الأكثر شيوعاً أو تكراراً، ويمكن إستخدامه في حالة البيانات الوصفية لمعرفة النمط (المستوى) الشائع ، ويتم حسابه كما يلي:

أولا : حساب المنوال في حالة البيانات غير المبوبة

المنوال (Mode) = القيمة (المستوى) الأكثر تكراراً

ثانيا : حساب المنوال في حالة البيانات المبوبة (طريقة الفروق)

$$\boxed{\text{Mod} = A + \frac{d_1}{d_1 + d_2} \times L}$$

حيث أن :
(A): الحد الأدنى لفئة المنوال (الفئة المناظرة لأكبر تكرار).
(d_1): الفرق الأول = تكرار فئة المنوال – التكرار السابق لها.
(d_2): الفرق الثاني = تكرار فئة المنوال – التكرار اللاحق لها.
(L): طول فئة المنوال.
فئة المنوال = الفئة المناظرة لأكبر تكرار.

◈ مثال

أدناه درجات عينة عشوائية من الطلبة في أربعة مواد دراسية:

المادة 1	**80**	**77**	**75**	**77**	**77**	**77**	**65**	**70**	**58**	**67**
المادة 2	**88**	**68**	**60**	**75**	**93**	**65**	**77**	**85**	**95**	**90**
المادة 3	**80**	**65**	**69**	**80**	**65**	**88**	**76**	**65**	**86**	**80**
المادة 4	**85**	**73**	**69**	**85**	**73**	**69**	**69**	**73**	**72**	**85**

المطلوب حساب منوال الدرجات لكل مادة من الموال الدراسية.

☑ الحل

إن هذه البيانات غير مبوبة لذا فإن المنوال = القيمة الأكثر تكراراً:

المادة	الدرجة الأكثر تكرارا ⇐ عدد مرات التكرار	القيمة المنوالية
1	(77) ⇐ 4 مرات	المنوال = (77)
2	لا تحتوي تكرار لأي قيمة	لايوجد منوال
3	(65) ⇐ 3 مرات (80) ⇐ 3 مرات	المنوال الأول = (65) المنوال الثاني = (80)
4	(69) ⇐ 3 مرات (73) ⇐ 3 مرات (85) ⇐ 3 مرات	المنوال الأول = (69) المنوال الثاني = (73) المنوال الثالث = (85)

مثال

فيما يلي توزيع (30) أسرة حسب الإنفاق الشهري

فئات الإنفاق	2 -	5 -	8 -	11 -	14 – 17
عدد الأسر *f*	4	7	10	5	4

المطلوب حساب منوال الإنفاق الشهري للأسرة ، بإستخدام طريقة الفروق.

الحل

1. تحديد الفئة المنوالية وهي الفئة المناظرة لأكبر تكرار.

	الفئات	التكرارات	
	2 -	4	
	5 -	7	$d_1=10-7=3$
فئة المنوال A	8 -	10	أكبر تكرار
	11 -	5	$d_2=10-5=5$
	14 - 17	4	

$\therefore\ A = 8\ ,\ d_1 = 3\ ,\ d_2 = 5\ ,\ L = 3$

$$\therefore\ Mod = A + \frac{d_1}{d_1 + d_2} \times L = 8 + \frac{3}{3+5} \times 3 = 8 + 1.125 = 9.125$$

مقاييس التشتت (Dispersion Measurements)

عند مقارنة مجموعتين من البيانات ، يمكن استخدام شكل التوزيع التكراري ، أوالمنحنى التكراري ، وكذلك بعض مقاييس النزعة المركزية ، مثل الوسط الحسابي والوسيط والمنوال ، والإحصاءات الترتيبية ، ولكن استخدام هذه الطرق وحدها لا يكفي عند المقارنة ، فقد يكون مقياس النزعة المركزية للمجموعتين متساوي ، وربما يوجد اختلاف كبير بين المجموعتين من حيث مدى تقارب وتباعد البيانات من بعضها البعض ، أو مدى تباعد أو تقارب القيم عن مقياس الترعة المركزية.
ومثال على ذلك ، إذا كان لدينا مجموعتين من الطلاب ، وكان درجات المجموعتين كالتالي:

Group 1	63	70	78	81	85	67	88
Group 2	73	78	77	78	75	74	77

لو قمنا بحساب الوسط الحسابي لكل مجموعة ، نجد أن الوسط الحسابي لكل منهما يساوي (76) درجة ، ومع ذلك درجات المجموعة الثانية أكثر تجانسا من درجات المجموعة الأولى .
من أجل ذلك لجأ الإحصائيون إلى استخدام مقاييس أخرى لقياس مدى تجانس البيانات ، أو مدى انتشار البيانات حول مقياس الترعة المركزية ، ويمكن استخدامها في المقارنة بين مجموعتين أو أكثر من البيانات ، ومن هذه المقاييس ، مقاييس التشتت (Dispersion Measurements).
من هذه المقاييس : المدى (Range) ، و الإنحراف المتوسط (Mean Deviation) والتباين (Variance)، والانحراف المعياري (Standard Deviation).

● المدى (Range)

هو أبسط مقاييس التشتت ، ويحسب المدى في حالة البيانات غير المبوبة بتطبيق المعادلة التالية:

(المدى)$_{\text{البيانات غير المبوبة}}$ = أكبر قراءة – أقل قراءة

$$\boxed{\mathbf{R = Max. - Min.}}$$

وأما المدى في حالة البيانات المبوبة له أكثر من صيغة، ومنها المعادلة التالية:

(المدى)$_{\text{البيانات المبوبة}}$ = مركز الفئة الأخيرة – مركز الفئة الأولى

$$\boxed{\mathbf{Range = (Mid)_{last\ faction} - (Mid)_{first\ faction}}}$$

◈ مثال

تم زراعة (9) وحدات تجريبية بمحصول القمح ، وتم تسميدها بنوع معين من الأسمدة الفسفورية ، وفيما يلي بيانات كمية الإنتاج من القمح بالطن/دونم:

4.8	6.21	5.4	5.18	5.29	5.18	5.08	4.63	5.03

والمطلوب حساب المدى.

☑ الحل

Range = max – min

$\therefore$ **Range = 6.21 – 4.63 = 1.58 طن/دونم**

◈

مثال

الجدول التكراري التالي يبين توزيع (60) مزرعة حسب المساحة المزروعة بالذرة بالألف دونم ،و المطلوب حساب مدى المساحة المزروعة بالذرة.

Area	15-20	20-25	25-30	30-35	35-40	40-45
No. Of Farms	3	9	15	18	12	3

☑ **الحل**

Range = $(\text{Mid})_{\text{last faction}} - (\text{Mid})_{\text{first faction}}$

$$\therefore \text{Range} = \frac{45+40}{2} - \frac{20+15}{2} = \frac{85}{2} - \frac{35}{2} = \frac{50}{2} = 25$$ دونم

- **مزايا وعيوب المدى**

المزايا	**1.** أنه بسيط وسهل الحساب. **2.** يكثر استخدامه عند الإعلان عن حالات الطقس، و المناخ الجوي، مثل درجات الحرارة، والرطوبة، والضغط الجوي. **3.** يستخدم في مراقبة الجودة.
العيوب	**1.** أنه يعتمد على قيمتين فقط ، ولا يأخذ جميع القيم في الحسبان. **2.** يتأثر بالقيم الشاذة .

الإنحراف المتوسط (Mean Deviation "*MD*")

هو أحد مقاييس التشتت ، و يعبر عنه **بمتوسط الانحرافات المطلقة للقيم عن وسطها الحسابي ،**

فإذا كانت $(x_1, x_2, x_3, \ldots, x_n)$ هي القراءات التي تم أخذها عن ظاهرة معينة ، وكان $(\bar{x} = \frac{\sum x}{n})$ هو الوسط الحسابي (Mean) لهذه القراءات ، فإن الإنحراف المتوسط (*MD*) يحسب بتطبيق المعادلة الأتية : **(بالنسبة للبيانات غير المبوبة)**

$$\boxed{MD = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}}$$

أما في حالة البيانات المبوبة فيتم إستخدام المعادلة الأتية:

$$\boxed{MD = \frac{\sum_{i=1}^{n} |x_i - \bar{x}| \times f}{n}}$$

حيث إن (***f***) هو تكرار الفئة ، (***x***) هو مركز الفئة ، $(\bar{x})$ هو الوسط الحسابي.

مثال

إذا كانت الطاقة التصديرية لخمس محطات لتحلية المياه بالمليون متر مكعب كما يلي:

4	5	2	10	7

أوجد قيمة الانحراف المتوسط للطاقة التصديرية.

الحل

1. يتم حساب الوسط الحسابي $(\bar{x})$:

$$\therefore \bar{x} = \frac{\sum x_i}{n} = \frac{4+5+2+10+7}{5} = \frac{28}{5} = 5.6$$

x	$(x-\bar{x}) = (x-5.6)$	|x-5.6|
4	4 – 5.6 = -1.6	1.6
5	-0.6	0.6
2	-3.6	3.6
10	4.4	4.4
7	1.4	1.4
sum	0	11.6

$$MD = \frac{11.6}{5} = 2.32m^3$$

مثال

يبين الجدول التالي التوزيع التكراري ل(40) أسرة حسب الإنفاق الشهري و المطلوب إيجاد الإنحراف المتوسط لهذا الإنفاق:

الإنفاق	2-5	5-8	8-11	11-14	14-17
عدد الأسر	1	8	13	10	8

الحل

حدود الإنفاق	عدد الأسر(f)	مركز الفئة(x)	$x.f$	$\bar{x}$	$\|x-\bar{x}\|$	$\|x-\bar{x}\| \times f$
2-5	1	3.5	3.5	$\bar{x} = \frac{\sum x \times f}{n} = \frac{428}{40} = 10.7$	7.2	7.2
5-8	8	6.5	52		4.2	33.6
8-11	13	9.5	123.5		1.2	15.6
11-14	10	12.5	125		1.8	18
14-17	8	15.5	124		4.8	38.4
sum	40		428			112.8

$$\therefore MD = \frac{112.8}{40} = 2.82$$

- **مزايا وعيوب الانحراف المتوسط**

المزايا	**يأخذ كل القيم بنظر الإعتبار**
العيوب	**1. يتأثر بالقيم الشاذة.** **2. يصعب التعامل معه رياضياً.**

التباين (Variance)

هو أحد مقاييس التشتت ، وأكثرها استخداماً في النواحي التطبيقية ، ويعبر عن **متوسط مربعات انحرافات (بعد) القيم عن وسطها الحسابي.**

• أولا : التباين في المجتمع (σ^2) (Pobulation Variance)

إذا توافر لدينا قراءات عن كل مفردات المجتمع ، ولتكن (x_1 , x_2 , ..., x_N) فإن التباين في المجتمع ويرمز له بالرمز (σ^2) يحسب بإستخدام المعادلة التالية:

$$\boxed{\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}}$$

حيث (μ) هو الوسط الحسابي في المجتمع = $(\frac{\sum x}{N})$

مثال

مصنع لتعبئة المواد الغذائية ، يعمل به (15) عامل ، وكانت عدد سنوات الخبرة لهؤلاء العمال كما يلي :

5	**13**	**7**	**14**	**12**	**9**	**6**	**8**	**10**	**13**	**14**	**6**	**11**	**12**	**10**

بفرض أن هذه البيانات تم جمعها عن كل مفردات المجتمع ، فأوجد التباين لعدد سنوات الخبرة .

الحل ☑

لحساب تباين سنوات الخبرة في المجتمع ، يتم استخدام المعادلة أعلاه :

$$\mu = \frac{\sum x}{N} = \frac{5+13+7+14+12+9+6+8+10+13+14+6+11+12+10}{15} = \frac{150}{15} = 10$$

	x	$(x-\mu)$	$(x-\mu)^2$
	5	-5	25
	13	3	9
	7	-3	9
	14	4	16
	12	2	4
	9	-1	1
	6	-4	16
	8	-2	4
	10	0	0
	13	3	9
	14	4	16
	6	-4	16
	11	1	1
	12	2	4
	10	0	0
Sum Σ	**150**	**0**	**130**

$$\therefore \sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N} = \frac{130}{15} = 8.67$$ تباين سنوات الخبرة في المصنع هو

- **ثانيا: التباين في العينة (S^2) (Sample Variance)**

في كثير من الحالات يكون تباين المجتمع (σ^2) غير معلوم ، وعندئذ يتم سحب عينة من هذا المجتمع ، ويحسب التباين من بيانات العينة كتقدير لتباين المجتمع ، فإذا كانت قراءات عينة عشوائية حجمها (n) هي $(x_1 , x_2 , \ldots, x_n)$ فإن تباين العينة ويرمز له بالرمز (S^2) يحسب كما يلي:

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

حيث $(\bar{x})$ هو الوسط الحسابي في المجتمع = $(\frac{\sum x}{n})$.

مثال

في المثال السابق ، إذا تم سحب عينة من عمال المصنع حجمها (5) عمال ، وسجل عدد سنوات الخبرة وكانت كالتالي :

8	13	10	5	9

احسب تباين سنوات الخبرة في العينة.

الحل ☑

$$\bar{x} = \frac{8+13+10+5+9}{5} = \frac{45}{5} = 9$$

x	$(x-\bar{x})$	$(x-\bar{x})^2$
8	-1	1
13	4	16
10	1	1
5	-4	16
9	0	0
45	**0**	**34**

$$\therefore S^2 = \frac{34}{5-1} = \frac{34}{4} = 8.5$$ تباين سنوات الخبرة في العينة

- **ملاحظة هامة :**

(درجات الحرية df - Degrees of Freedom) هي عدد القيم القابلة للتغير في حساب خاصية إحصائية ما.

- ✓ يعتمد حساب الخصائص الإحصائية المختلفة على مجموعة من المعلومات أو البيانات.
- ✓ يسمى عدد المعلومات المستقلة عن بعضها والتي تدخل في حساب خاصية إحصائية معينة) كالتباين Variance ، والارتباط Correlation ، ...الخ) بدرجات الحرية (df) .
- ✓ بشكل عام، عدد درجات الحرية في تقييم خاصية إحصائية معينة يساوي عدد القراءات المستقلة التي تدخل في حساب الخاصية الإحصائية (تباين، ارتباط، ...) ناقص عدد الخصائص الإحصائية المستخدمة في حساب الخاصية الإحصائية المطلوبة (مثل استخدام قيمة المتوسط الحسابي في حساب التباين مثلا).

على سبيل المثال: لدينا عينة إحصائية مكونة من 100 قراءة لنتائج امتحان ذكاء أجري على 100 شخص. عدد درجات الحرية في حساب التباين لهذه العينة يساوي عدد القراءات المستقلة "100" ناقص عدد الخصائص الإحصائية المستخدمة في حساب التباين وهي هنا تساوي "1" لأننا فقط نستخدم المتوسط الحسابي في حساب التباين.

رياضيا، درجات الحرية تمثل بعد نطاق متجه عشوائي، أو بشكل أوضح عدد مركبات المتجه "الحرة": أي هي عدد مركبات المتجه التي تجب معرفتها للتمكن من تحديده بشكل كامل.

يستخدم مصطلح درجات الحرية بكثرة ضمن سياق الحديث عن نموذج الانحدار الخطي وتحليل التباين، حيث يكون لدينا مجموعة من المتجهات العشوائية مقيدة بالوقوع ضمن فضاء جزئي خطي، و يكون عدد درجات الحرية هو عدد أبعاد ذلك الفضاء الجزئي الخطي. درجات الحرية تستخدم أيضا في حساب مربعات أطوال تلك المتجهات، وفي حساب خصائص توزيع كاي تربيع) بالإنكليزية (Chi-Squared : وأنواع التوزيعات الإحصائية الأخرى.

● الانحراف المعياري (Standard Deviation)

عند استخدام التباين كمقياس من مقاييس التشتت ، نجد أنه يعتمد علي مجموع مربعات الانحرافات ، ومن ثم لا يتمشى هذا المقياس مع وحدات قياس المتغير محل الدراسة ، ففي المثال السابق ، نجد أن تباين سنوات الخبرة في العينة(8.5) ، فليس من المنطق عند تفسير هذه النتيجة أن نقول ، إن تباين سنوات الخبرة هو" **8.5 سنة تربيع**" ، لأن وحدات قياس المتغير هو عدد السنوات ، من أجل ذلك لجأ الإحصائيين إلى مقياس منطقي يأخذ في الاعتبار **الجذر التربيعي للتباين** ، لكي يناسب وحدات قياس المتغير ، وهذا المقياس هو **الانحراف المعياري (Standard Deviation)**.

إذا **الانحراف المعياري ، هو الجذر التربيعي الموجب للتباين** ، أي أن:

$$\boxed{Standard\ \ Deviation = \sqrt{Variance}}$$

وعليه يكون الإنحراف المعياري للمجتمع (σ) و للعينة (S) ويحسبان كما يلي:

$$\boxed{\begin{array}{l} \sigma = \sqrt{\dfrac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}} \\ where\ \mu = \dfrac{\sum_{i=1}^{N} x_i}{N} \end{array}}$$

الإنحراف المعياري للمجتمع(البيانات غير مبوبة)

$$\boxed{\begin{array}{l} S = \sqrt{\dfrac{\sum_{i=1}^{N}(x_i-\bar{x})^2}{n-1}} \\ where\ \bar{x} = \dfrac{\sum_{i=1}^{N} x_i}{n-1} \end{array}}$$

الإنحراف المعياري للعينة(البيانات غير مبوبة)

ومثال على ذلك:

نجد أن **الانحراف المعياري** لسنوات الخبرة لعمال المصنع (المجتمع) هو :

$\sigma = \sqrt{8.67} = 2.94$

و **الإنحراف المعياري** لعينة من العمال هو :

$S = \sqrt{8.5} = 2.92$

● الانحراف المعياري في حالة البيانات المبوبة

إذا كانت بيانات الظاهرة ، مبوبة في جدول توزيع تكراري ، فإن الانحراف المعياري يحسب بتطبيق المعادلة التالية :

$$S=\sqrt{\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2 . f_i}{n-1}} \quad or \quad S=\sqrt{\frac{\sum_{i=1}^{n}\left(x_i^2 . f_i\right)-\frac{\left(\sum_{i=1}^{n}\left(x_i . f_i\right)\right)^2}{n}}{n-1}}$$

حيث إن (f) هو تكرار الفئة ، (x) هو مركز الفئة ، ($\bar{x}=\frac{\sum x.f}{n}$) هو الوسط الحسابي ، ($n=\sum f$) هو مجموع التكرارات ، و المقدار الذي تحت الجذر يعبر عن التباين.

◈ مثال

احسب الانحراف المعياري للإنفاق الشهري للأسرة :

الإنفاق	2-5	5-8	8-11	11-14	14-17
عدد الأسر	1	8	13	10	8

☑ الحل

لحساب الانحراف المعياري للإنفاق الشهري ، تستخدم المعادلة أعلاه

$$\bar{x}=\frac{\sum xf}{n}=\frac{428}{40}=10.7$$

حدود الإنفاق	عدد الأسر(f)	مركز الفئة(x)	$x . f$	$(x-\bar{x})$	$(x-\bar{x})^2$	$(x-\bar{x})^2 . f$
2-5	1	3.5	3.5	3.5-10.7= -7.2	51.84	51.84
5-8	8	6.5	52	6.5-10.7= - 4.2	17.64	141.12
8-11	13	9.5	123.5	9.5-10.7= - 1.2	1.44	18.72
11-14	10	12.5	125	12.5-10.7= 1.8	3.24	32.4
14-17	8	15.5	124	15.5-10.7= 4.8	23.04	184.32
sum	40		428		97.2	428.4

$$\therefore S=\sqrt{\frac{428.4}{39-1}}=\sqrt{\frac{482.4}{39}}=\sqrt{10.984615}=3.314$$

يترك للطالب تطبيق الصيغة الثانية للإنحراف المعياري للتوصل إلى قيمة الإنحراف المعياري.

● **خصائص الانحراف المعياري**

1. **الانحراف المعياري للمقدار الثابت يساوي صفرا** ، أي أنه إذا كان لدينا القراءات التالية : (x : a , a , a , …a) حيث (a) مقدار ثابت ، فإن $(S_x=0)$ حيث (S_x) هو الإنحراف المعياري لقيم (x) .
2. إذا أضيف مقدار ثابت إلى كل قيمة من قيم المفردات ، فإن الانحراف المعياري للقيم الجديدة (القيم بعد الإضافة)= الانحراف المعياري للقيم الأصلية(القيم قبل الإضافة) ، فإذا كانت القيم الأصلية هي $(x_1 , x_2 ,….,x_n)$ وتم إضافة مقدار ثابت (a) إلى كل قيمة من قيم (x) فإن الإنحراف المعياري للقيم الجديدة $(x_1+a ,x_2+a ,…., x_n+a)$ يساوي الإنحراف المعياري للقيم الأصلية ، أي $(S_x = S_{x+a})$.
3. *ونفس الحالة أعلاه تكون في حال ضرب القيم بثابت عددي.*

◈ **مثال**

إذا كان من المعلوم أن تطبيق برنامج غذائي معين للتسمين لفترة زمنية محددة سوف يزيد من وزن الدجاجة (0.5) كيلوغرام ، سحبت عينة عشوائيا من مزرعة دجاج حجمها (5) دجاجات ، وكانت أوزانها :

1	**1.75**	**2**	**1.25**	**2.5**

1. احسب الانحراف المعياري لوزن الدجاجة.
2. إذا طبق البرنامج الغذائي المشار إليه ، ما هو الانحراف المعياري لوزن الدجاجة في هذه العينة.

☑ **الحل**

1. حساب الانحراف المعياري للوزن قبل تطبيق البرنامج.

$$S = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{n-1}} , \quad \bar{x} = \frac{\sum_{i=1}^{N} x_i}{n} = \frac{1+1.75+2+1.25+2.5}{5} = \frac{8.5}{5} = 1.7$$

x	$(x-\bar{x})$	$(x-\bar{x})^2$
1	**-0.7**	**0.49**
1.75	**0.05**	**0.0025**
2	**0.3**	**0.09**
1.25	**-0.45**	**0.2025**
2.5	**0.8**	**0.64**
8.5	**0**	**1.425**

$$\therefore S = \sqrt{\frac{1.425}{5-1}} = \sqrt{\frac{1.425}{4}} = \sqrt{0.35625} = 0.597$$ **الإنحراف المعياري للوزن قبل البرنامج في العينة**

2. الإنحراف المعياري للوزن بعد تطبيق البرنامج الذي من المتوقع فيه حصول زيادة في الوزن بمقدار (0.5)كيلوغرام وعليه يكون الإنحراف المعياري بعد الزيادة مساوي لماهو عليه قبل الزيادة.

- **مزايا وعيوب الانحراف المعياري**

المزايا	**1.** أنه أكثر مقاييس التشتت استخداما . **2.** يسهل التعامل معه رياضيا . **3.** يأخذ كل القيم في الاعتبار.
العيوب	أنه يتأثر بالقيم الشاذة.

● الإرتباط و الإنحدار (Correlation & Regression)

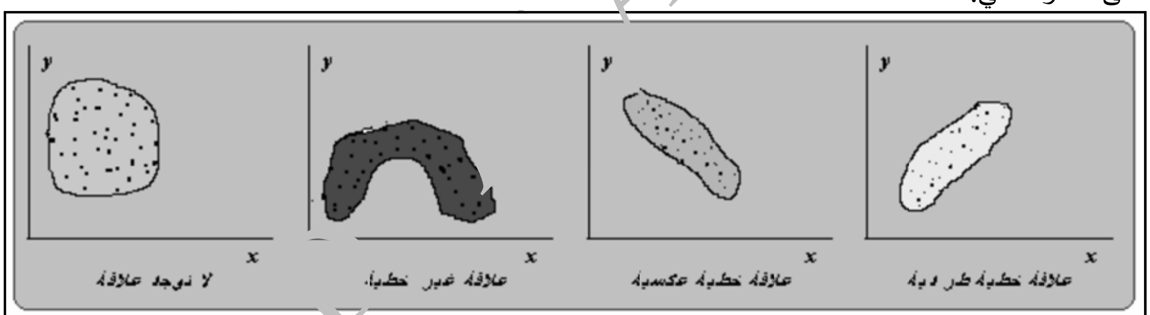
فيما سبق من محاظرات تم عرض بعض المقاييس الوصفية ، مثل مقاييس النزعة المركزية ،والتشتت، وغيرها من المقاييس الأخرى والتي يمكن من خلالها وصف شكل توزيع البيانات التي تم جمعها عن متغير واحد، وننتقل من التعامل مع متغير واحد إلى التعامل مع متغيرين أو أكثر، ويتناول هذا الفصل دراسة وتحليل العلاقة بين متغيرين ، وذلك باستخدام بعض طرق التحليل الإحصائي مثل تحليل الارتباط (Correlation)، والانحدار الخطي البسيط (Simple regression)،

- **<u>فإذا كان اهتمام الباحث هو دراسة العلاقة بين متغيرين استخدم لذلك أسلوب تحليل الارتباط ،</u>**
- **<u>وإذا كان الاهتمام بدراسة أثر أحد المتغيرين على الآخر استخدم لذلك أسلوب تحليل الانحدار ،</u>**

ومن الأمثلة على ذلك:

1. الإنفاق، والدخل العائلي.
2. سعر السلعة ، والكمية المطلوبة منها.
3. تقديرات الطلاب في مقرر الإحصاء ، وتقديراتهم في مقرر الرياضيات.
4. كميات السماد المستخدمة ، وكمية الإنتاج من محصول معين تم تسميده بهذا النوع من السماد.
5. عدد مرات ممارسة نوع معين من الرياضة البدنية ، ومستوى الكلسترول في الدم.
6. وزن الجسم ، وضغط الدم.

والأمثلة على ذلك في المجال التطبيقي كثيرة ، فإذا كان لدينا المتغيرين $(y\ ,x)$ وتم جمع بيانات ، عن أزواج قيم هذين المتغيرين ، وتم تمثيلها بيانيا فيما يسمى بشكل الانتشار ، فإن العلاقة بينها تأخذ أشكالا مختلفة على النحو التالي:



● الإرتباط الخطي البسيط (Simple Correlation)

إذا كان الغرض من التحليل هو تحديد نوع وقوة العلاقة بين متغيرين ، يستخدم تحليل الارتباط ، وأما إذا كان الغرض هو دراسة وتحليل أثر أحد المتغيرين على الآخر ، يستخدم تحليل الانحدار، وفي هذا الفصل يتم عرض أسلوب تحليل الارتباط الخطي البسيط ، أي في حالة افتراض أن العلاقة بين المتغيرين تأخذ الشكل الخطي ، وسوف يجرى حسابه في حالة البيانات الكمية ، والبيانات الوصفية المقاسة بمعيار ترتيبي.

● الغرض من تحليل الارتباط الخطى البسيط

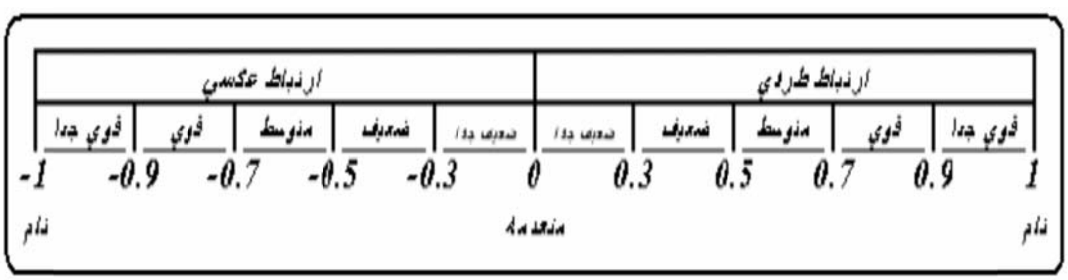
الغرض من تحليل الارتباط الخطي البسيط هو تحديد نوع وقوة العلاقة بين متغيرين ، ويرمز له في حالة المجتمع بالرمز (ρ) وفي حالة العينة بالرمز (r) وحيث أننا في كثير من النواحي التطبيقية نتعامل مع بيانات عينة مسحوبة من المجتمع ، سوف يتم حساب معامل الارتباط في العينة (r) كتقدير لمعامل الإرتباط في المجتمع ، ومن خلال تحديد معامل الارتباط ، يمكن الإستدلال على :

أولا : نوع العلاقة : وتأخذ ثلاث أنواع حسب إشارة معامل الارتباط كما يلي:

إشارة معامل الإرتباط	العلاقة بين المتغيرين
سالبة ($r < 0$)	عكسية
موجبة ($r > 0$)	طردية
صفر ($r = 0$)	لاتوجد علاقة بين المتغيرين

ثانياً : قوة العلاقة : ويمكن الحكم على قوة العلاقة من حيث درجة قرابتها أو بعدها عن (± 1) حيث إن قيمة معامل الإرتباط تقع في المدى ($-1 < r < 1$) وقد صنف بعض الإحصائيين درجات لقوة العلاقة يمكن تمثيلها كما يلي:

قيمة معامل الإرتباط	القوة
$0 < r \le \lvert 0.3\rvert$	ضعيف جدا
$\lvert 0.3\rvert < r \le \lvert 0.5\rvert$	ضعيف
$\lvert 0.5\rvert < r \le \lvert 0.7\rvert$	متوسط
$\lvert 0.7\rvert < r \le \lvert 0.9\rvert$	قوي
$\lvert 0.9\rvert < r < \lvert 1\rvert$	قوي جداً
$r = 1$	تام



●معامل الإرتباط الخطي "لبيرسون" (Pearson Correlation Coefficient)

في حالة جمع البيانات عن متغيرين كميين (x , y) يمكن قياس الإرتباط بينهما ، بإستخدام طريقة (بيرسون Pearson) ، ومن الأمثلة على ذلك قياس العلاقة بين الوزن و الطول ، و العلاقة بين الإنتاج و الكلفة ، و العلاقة بين الإنفاق الإستهلاكي و الدخل ، و العلاقة بين الدرجة التي حصل عليها الطالب وعدد ساعات المذاكرة ...إلخ

ولحساب معامل بيرسون للإرتباط يتم إستخدام صيغة بيرسون الأتية:

$$\boxed{r=\frac{S_{xy}}{S_x.S_y}=\frac{\dfrac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)\times\left(y_i-\bar{y}\right)}{(n-1)}}{\sqrt{\dfrac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}{(n-1)}}\times\sqrt{\dfrac{\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2}{(n-1)}}}}$$

حيث إن :

$S_{xy}=\dfrac{\sum\left(x-\bar{x}\right)\times\left(y-\bar{y}\right)}{(n-1)}$ = التباين (Variance) بين كل من (x) و (y) ،

$S_x=\sqrt{\dfrac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}{(n-1)}}$ = الإنحراف المعياري (Standard Deviation) لقيم (x) ،

$S_y=\sqrt{\dfrac{\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2}{(n-1)}}$ = الإنحراف المعياري (Standard Deviation) لقيم (y) .

ويمكن إختصار الصيغة أعلاه لتأخذ الشكل :

$$\boxed{r=\frac{\sum\left(x-\bar{x}\right)\times\left(y-\bar{y}\right)}{\sqrt{\sum\left(x-\bar{x}\right)^2\times\sum\left(y-\bar{y}\right)^2}}}$$

ملاحظة هامة

يمكن تبسيط معادلة إحتساب معامل بيرسون للإرتباط لتأخذ الصيغة الأتية:

$$\boxed{r=\frac{\left(\sum_{i=1}^{n}x_iy_i\right)-\dfrac{\left(\sum_{i=1}^{n}x_i\right)\times\left(\sum_{i=1}^{n}y_i\right)}{n}}{\sqrt{\left(\left(\sum_{i=1}^{n}x_i^2\right)-\dfrac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}\right).\left(\left(\sum_{i=1}^{n}y_i^2\right)-\dfrac{\left(\sum_{i=1}^{n}y_i\right)^2}{n}\right)}}}$$

$$\boxed{\therefore r=\frac{n.\left(\sum_{i=1}^{n}x_iy_i\right)-\left(\sum_{i=1}^{n}x_i\right)\times\left(\sum_{i=1}^{n}y_i\right)}{\sqrt{\left(n.\left(\sum_{i=1}^{n}x_i^2\right)-\left(\sum_{i=1}^{n}x_i\right)^2\right)\times\left(n.\left(\sum_{i=1}^{n}y_i^2\right)-\left(\sum_{i=1}^{n}y_i\right)^2\right)}}}$$

حيث نجد إن الصيغة أصبحت تعتمد في حساباتها على مجاميع القيم بدلا من إعتمادها على إنحرافات القيم عن وسطها الحسابي.

◈ مثال

في ما يلي المساحة المزروعة بالأعلاف الخضراء (بالألف دونم) ، و إجمالي اللحوم المنتجة (بالألف طن) خلال السنوات من (1995) إلى (2002):

السنة	1995	1996	1997	1998	1999	2000	2001	2002
المساحة	305	313	397	289	233	214	240	217
الكمية	592	603	662	607	635	699	719	747

أوجد معامل الإرتباط بين المساحة المزروعة وكمية اللحوم المنتجة و ما مدلول ذلك؟

☑ الحل

بفرض إن المساحة المزروعة هي (x) ، (y) هي كمية اللحوم المنتجة ، أكمل الحل، علماً أن:
معامل الإرتباط الناتج هو (0.798 -) أي يوجد إرتباط عكسي قوي بين المساحة المزروعة وكمية إنتاج اللحوم.

● معامل إرتباط الرتب (سبيرمان Sperman)

إذا كانت الظاهرة محل الدراسة تحتوي على متغيرين وصفيين ترتيبين، ومثال على ذلك قياس العلاقة بين تقديرات الطلبة في مادتين ، أو العلاقة بين درجة تفضيل المستهلك لسلعة معينة ، ومستوى الدخل، فإنه يمكن استخدام طريقة" بيرسون "السابقة في حساب معامل ارتباط يعتمد على رتب مستويات المتغيرين كبديل للقيم الأصلية ، ويطلق على هذا المعامل " معامل ارتباط سبيرمان Sperman "
ويعبر عنه بالمعادلة التالية:

$$\boxed{r_s = 1 - \frac{6\sum_{i=1}^{n} d^2}{n\left(n^2 - 1\right)}}$$

حيث إن (d) هو الفرق بين رتب مستويات المتغير الأول (x) ورتب مستويات المتغير الثاني (y) ، اي إن ($d=R_x - R_y$) .

◈ مثال

فيما يلي درجات (10) طلاب في مادتين دراسيتين ،إحسب معامل الإرتباط بين درجات الطلبة في المادتين.

المادة 1	A	C+	D	D+	B+	C+	A+	B	B+	B+
المادة 2	A+	D	C	C	A	B	B+	B	C	B

☑ الحل

بفرض إن (x) هي درجات (المادة 1) و (y) هي درجات (المادة 2) ، يتم ترتيب التقديرات أعلاه في رتب تنازليا كما يلي:

<table>
<tr><th>الرتب</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th>8</th><th>9</th><th>10</th></tr>
<tr><td>تقديرات المادة 1</td><td>A+</td><td>A</td><td>B+</td><td>B+</td><td>B+</td><td>B</td><td>C+</td><td>C+</td><td>D+</td><td>D</td></tr>
<tr><td>رتب (x)</td><td>1</td><td>2</td><td colspan="3">(3+4+5)/3=4</td><td>6</td><td colspan="2">(7+8)/2=7.5</td><td>9</td><td>10</td></tr>
<tr><td>تقديرات المادة 2</td><td>A+</td><td>A</td><td>B+</td><td>B</td><td>B</td><td>B</td><td>C</td><td>C</td><td>C</td><td>D</td></tr>
<tr><td>رتب (y)</td><td>1</td><td>2</td><td>3</td><td colspan="3">(4+5+6)/3=5</td><td colspan="3">(7+8+9)/3=8</td><td>10</td></tr>
</table>

ولأجل حساب قيمة (d) يتم العودة إلى الجدول الأصلي:

x	رتب (x)	y	رتب (y)	فرق الرتب d=x-y	$(d)^2$
A	2	A+	1	1	1
C+	7.5	D	10	-2.5	6.25
D	10	C	8	2	4
D+	9	C	8	1	1
B+	4	A	2	2	4
C+	7.5	B	5	2.5	6.25
A+	1	B+	3	-2	4
B	6	B	5	1	1
B+	4	C	8	-4	16
B+	4	B	5	-1	1
sum					44.5

$$\therefore r = 1 - \frac{6.\sum d^2}{n(n^2-1)} = 1 - \frac{6 \times 44.5}{10(100-1)} = 1 - \frac{267}{990} = 1 - 0.2697 = 0.7303$$

$\therefore$
نجد إن معامل الإرتباط موجب (علاقة طردية) وقيمته أكبر من (0.7) أي إنه (قوي).
أي إن الإرتباط طردي قوي بين تقديرات الطلبة في المادتين 1 و 2 .

● الانحدار الخطي البسيط (Simple Regresion)

إن الغرض من استخدام أسلوب تحليل الانحدار الخطي البسيط، هو دراسة وتحليل أثر متغير كمي على متغير كمي آخر ، ومن الأمثلة على ذلك ما يلي:

1. دراسة أثر كمية السماد على إنتاجية الدونم.
2. دراسة أثر الإنتاج على التكلفة.
3. دراسة أثر كمية البروتين التي يتناولها الأبقار على الزيادة في الوزن.

● نموذج الانحدار الخطي

في تحليل الانحدار البسيط ، نجد أن الباحث يهتم بدراسة أثر أحد المتغيرين ويسمى بالمتغير **المستقل أو المتنبأ منه (independent variable)**، على المتغير الثاني و يسمى **بالمتغير التابع أو المتنبأ به (dependent variable)** ، ومن ثم يمكن عرض نموذج الانحدار الخطي في شكل معادلة خطية من الدرجة الأولى ، تعكس المتغير التابع كدالة في المتغير المستقل كما يلي:

$$\boxed{y = Ax + B}$$

حيث أن:

(y) : هو المتغير التابع (الذي يتغير)،
(x): هو المتغير المستقل (الذي يؤثر)،
(B): هو الجزء المقطوع من المحور الرأسي (y) ، وهو يمثل قيمة المتغير التابع في حالة إنعدام المتغير المستقل (x) ،أي في حالة (x = 0).
(A) : ميل الخط المستقيم (الذي معادلته أعلاه) ويعكس مقدار التغير في (y) إذا تغيرت قيمة (x) وحدة واحدة.

● تقدير نموذج الانحدار الخطي البسيط

يمكن تقدير معاملات الإنحدار (A) ، (B) بإستخدام طريقة المربعات الصغرى (least squares) وكما يلي:

$$\boxed{\begin{array}{l} A = \dfrac{n.\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i . \sum_{i=1}^{n} y_i}{n.\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}, \\ B = \bar{y} - A\bar{x} \end{array}}$$

حيث إن ($\bar{x}$) هو الوسط الحسابي لقيم (x) وكذلك ($\bar{y}$) هو الوسط الحسابي لقيم (y).

◈ مثال

فيما يلي بيانات عن كمية البروتين اليومي بالغرام التي يحتاجهاالعجل ومقدار الزيادة في وزن العجل بالكيلوغرام وذلك لعينة من العجول حجمها(10)،قدر معادلة إنحدار الوزن على كمية البروتين، وما هو مقدار الزيادة في الوزن عند إعطاء العجل (90) غرام من البروتين.

Protein	10	11	14	15	20	25	46	50	59	70
Weight	10	10	12	12	13	13	19	15	16	20

☑ الحل

$\therefore$ n= 10

$\sum x = 320,$

$\sum y = 140,$

$\sum xy = 5111,$

$\sum x^2 = 14664$

$$\therefore \bar{x} = \frac{\sum x}{n} = \frac{320}{10} = 32,$$

$$\bar{y} = \frac{\sum y}{n} = \frac{140}{10} = 14,$$

$$\therefore A = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - \left(\sum x\right)^2} = \frac{(10)(5111) - (320)(140)}{(10)(14664) - (320)^2} = 0.1426,$$

$$\therefore B = \bar{y} - A\bar{x} = 14 - (0.1326)(32) = 9.4368$$

$$y = 9.44 + 0.143x$$

من خلال ما تقدم أعلاه ومن معادلة الإنحدار نجد إنه في حال إنعدام إستخدام البروتين فإن الوزن يزيد (9.44)كغم.

أما معامل الإنحدار (B) يدل على إنه كلما زادت كميةالبروتين غرام واحد حدثت زيادة في الوزن بمقدار (0.143)كغم.

وحساب مقدار الزيادة في الوزن عند (x=90) يترك واجب للطالب.

- **ملاحظات مهمة**

في بعض الأدبيات تكون درجات الإرتباط كالأتي:

الطردي	درجة الإرتباط	العكسي
0.4 > ر > 0	**ضعيف**	**0 > ر > 0.4 -**
0.6 ≥ ر ≥ 0.4	**متوسط**	**0.6 - ≤ ر ≤ 0.4 -**
1 > ر > 0.6	**قوي**	**1 - > ر > 0.6 -**

◈ مثال

من بيانات الجدول الآتى:

x	7	6	8	3	10	5
y	5	4	6	2	8	4

1. أوجد معامل ارتباط بيرسون.
2. أوجد معامل ارتباط الرتب لسبيرمان.
3. قارن بين معامل الارتباط فى الحالتين.

☑ الحل

1. لإيجاد معامل إرتباط بيرسون :

	x	x^2	y	y^2	x.y
	7	49	5	25	35
	6	36	4	16	24
	8	64	6	36	48
	3	9	2	4	6
	10	100	8	64	80

	5	25	4	16	20
SUM	39	283	29	161	213

n = 6

$$\therefore \; r = \frac{n.\left(\sum_{i=1}^{n} x_i y_i\right) - \left(\sum_{i=1}^{n} x_i\right).\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{\left(n.\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2\right).\left(n.\left(\sum_{i=1}^{n} y_i^2\right) - \left(\sum_{i=1}^{n} y_i\right)^2\right)}}$$

$$r = \frac{6(213) - (39)(29)}{\sqrt{\left(6(283) - (39)^2\right)\left(6(161) - (29)^2\right)}} = \frac{1278 - 1131}{\sqrt{177 \times 125}} = \frac{147}{\sqrt{22125}} = 0.988$$

∴ معامل إرتباط بيرسون = (0.988)

2 لإيجاد معامل إرتباط رتب سبيرمان:

● في حالة الترتيب التنازلي للبيانات :

x↓	10	8	7	6	5	3
الرتب	1	2	3	4	5	6
y↓	8	6	5	4	4	2
الرتب	1	2	3	4.5		6

	x	الرتبة	y	الرتبة	d	d^2
	7	3	5	3	0	0
	6	4	4	4.5	-0.5	0.25
	8	2	6	2	0	0
	3	6	2	6	0	0
	10	1	8	1	0	0
	5	5	4	4.5	0.5	0.25
sum						0.5

n = 6

وبتطبيق معادلة سبيرمان لحساب معامل إرتباط الرتب ، نجد إن معامل سبيرمان لإرتباط الرتب = (0.986)

∴ قيم معاملي الإرتباط (لبيرسون و سبيرمان) متقاربتين جدا.

∴ الإرتباط طردي قوي.

● **ملاحظات**

1. إذا كان الارتباط قوياً قربت قيم المتغيرين من خط مستقيم يمثل العلاقة بينهما يسمى **خط الانحدار.**
2. معادلة خط انحدار (y) على (x) (لتقدير قيمة (y) عند أى قيمة لـ(x) هي: (y=Ax+B).

3. معادلة خط انحدار (x) على (y) (لتقدير قيمة (x) عند أي قيمة لـ(y) هي: (x=Cy+D) ، حيث :

$$C=\frac{n.\sum_{i=1}^{n}x_iy_i-\sum_{i=1}^{n}x_i.\sum_{i=1}^{n}y_i}{n.\sum_{i=1}^{n}y_i^2-\left(\sum_{i=1}^{n}y_i\right)^2},$$

$$D=\bar{x}-C\bar{y}$$

4. العلاقة بين معامل الإرتباط (r) و معاملي الإنحدار (A) و (C) هي : ($r^2 = A.C$) حيث يأخذ (r) نفس إشارة معامل الإرتباط.

مقاييس أخرى لوصف البيانات

هناك مقاييس أخرى يمكن استخدامها في وصف البيانات ، من حيث درجة تشتت البيانات، ومدى انتشارها، ومن هذه المقاييس ، ما يلي :

أولا : معامل الاختلاف (Variation Coefficient)

أحد المقاييس المستخدمة لقياس درجة التشتت ، وفيه تحسب قيمة التشتت كنسبة مئوية من قيمة مقياس النزعة المركزية ، و يفضل استخدام معامل الاختلاف عند مقارنة درجة تشتت بيانات مجموعتين مختلفتين أو أكثر لها وحدات قياس مختلفة ، بدلا من الانحراف المعياري ، لأن معامل الاختلاف يعتمد على التغيرات النسبية في القيم عن مقياس الترعة المركزية ، بينما يعتمد الانحراف المعياري على التغيرات المطلقة للقيم ، فعند مقارنة درجة تشتت بيانات الأطوال بالسنتمتر ، وببيانات الأوزان بالكيلوغرام ، لا يمكن الاعتماد على الانحراف المعياري في هذه المقارنة ، وإنما يستخدم معامل الاختلاف ، ومن ثم يطلق عليه بمعامل الاختلاف النسبي .

- ### معامل الاختلاف النسبي $(v.c)$

ويحسب معامل الاختلاف النسبي بتطبيق المعادلة التالية:

$$\boxed{v.c\% = \frac{S}{\bar{x}} \times 100}$$

❖ **مثال**

تم اختيار مجموعتين من الدواجن في أحد المزارع ، وتم استخدام عليقة معينة لتسمين المجموعة الأولى، بينما تم استخدام عليقة أخرى لتسمين المجموعة الثانية ، وبعد فترة زمنية تم جمع بيانات عن أوزان المجموعتين بالكيلوغرام ، وتم الحصول على المقاييس التالية :

المقياس	G1	G2
$\bar{x}$	173	198
S	23	25

والمطلوب مقارنة درجة تشتت المجموعتين.

☑ **الحل**

1. معامل الإختلاف النسبي للمجموعة الأولى:

$$v.c_{G1}\% = \frac{S}{\bar{x}} \times 100 = \frac{23}{173} \times 100 = 13.295\%$$

2. معامل الإختلاف النسبي للمجموعة الثانية:

$$v.c_{G2}\% = \frac{S}{\bar{x}} \times 100 = \frac{25}{198} \times 100 = 12.626\%$$

حيث نلاحظ إن درجة تشتت أوزان المجموعة الثانية أقل من درجة تشتت المجموعة الأولى.

● تقدير مدى الانحراف المعياري

يمكن قياس درجة تشتت البيانات من خلال تقدير المدى (R) الذي يقع داخله الانحراف المعياري (S) وهو:

$$\boxed{\frac{R}{4} > S > \frac{R}{6}}$$

وإذا كان المدى الذي يقع فيه الانحراف المعياري صغير دل ذلك على أن تشتت البيانات صغير، أما إذا كان المدى كبير دل ذلك على وجود تشتت كبير في البيانات، وإذا وقع الانحراف المعياري خارج المدى دل ذلك على وجود قيم شاذة.
اي يجب ان تقع قيمة الانحراف المعياري للبيانات بين قيمتي (R/6) و (R/4).

● الدرجة المعيارية (Standarized Degree)

تقيس الدرجة المعيارية (لقيمة معينة) عدد وحدات الانحراف المعياري التي تزيد بها أو تقل بها هذه القيمة عن الوسط الحسابي ، فإذا كانت ($x_1 , x_2 , \ldots , x_n$) هي قيم المشاهدات ، وعددها (n) ، وكان ($\bar{x}$) هو الوسط الحسابي لهذه القيم ، و (S) هو الإنحراف المعياري لها ، فإن الدرجة المعيارية للقيمة (x) و يرمز لها بالرمز (z) تحسب بإستخدام المعادلة التالية :

$$\boxed{z = \frac{x - \bar{x}}{S}}$$

ويمكن إستخدام هذه الدرجة في مقارنة قيمتين أو أكثر مختلفة من حيث وحدات القياس.

◈ مثال

في المثال السابق تم إختيار عينة من المجموعة الأولى ووجد إن وزنها (178 كغم) و عينة من المجموعة الثانية ووجد إن وزنها (180 كغم) ، قارن بين هاتين القيمتين من حيث أهمية كل منهما في المجموعة المنتمية إليها.

☑ الحل

من خلال معطيات المجموعتين في المثال السابق :

المقياس	G1	G2
$\bar{x}$	173	198
S	23	25
x	178	180

للمقارنة بين الوحدتين من حيث أهمية وزن كل منها في المجموعة التي تنتمي إليها ، يتم حساب الدرجة المعيارية لوزن كل منها :

$$z_1 = \frac{x-\bar{x}}{S} = \frac{178-173}{23} = 0.22$$

$$z_2 = \frac{x-\bar{x}}{S} = \frac{180-198}{25} = -0.75$$

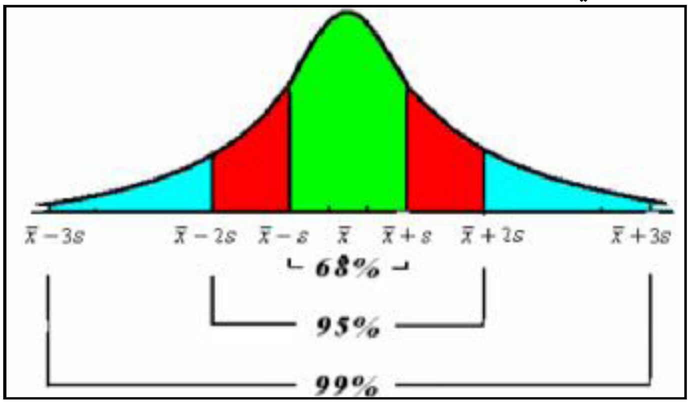
نجد أن الوزن (178 كغم) يزيد عن الوسط الحسابي ب(0.22)انحراف معياري ، بينما نجد أن الوزن (180 كغم) يقل عن الوسط الحسابي ب(0.75) انحراف معياري . ومن ثم الوزن الأول أهميته النسبية أعلى من الوزن الثاني.

● القاعدة العملية

إذا كان لدينا المشاهدات التالية ($x_1 , x_2 , ...,x_n$) وكان ($\bar{x}$) هو الوسط الحسابي لهذه القيم ، و (S) هو الإنحراف المعياري لها ، يكون منحنى توزيع هذه المشاهدات متماثل، إذا تحقق الآتي:

% 68	تقريبا من قيم المشاهدات تقع قيمها بين	$(\bar{x}+S)$	و	$(\bar{x}-S)$
% 95	تقريبا من قيم المشاهدات تقع قيمها بين	$(\bar{x}+2S)$	و	$(\bar{x}-2S)$
% 99	تقريبا من قيم المشاهدات تقع قيمها بين	$(\bar{x}+3S)$	و	$(\bar{x}-3S)$

ويمكن بيان ذلك من الشكل التالي:



● القاعدة النظرية

تسمى هذه القاعدة بقاعدة ***(تشيبشيف)*** ، وفكرة هذه القاعدة تنص على ما يلي :

في أي توزيع من التوزيعات فإنه على الأقل ($1-\frac{1}{k^2}$) من قيم المشاهدات تقع في المدى ($\bar{x} \pm kS$)

حيث $(k > 1)$ ، وطبقا لهذه القاعدة، فإنه على الأقل (75%) من قيم المشاهدات تقع في المدى ($\bar{x} \pm 2S$) و على الأقل (89%) من قيم المشاهدات تقع في المدى ($\bar{x} \pm 3S$).

الإحتمالات (Probabilities)

يقصد **بالإحتمال** فرصة حدوث أو وقوع حادثة معينة ، وتستخدم الاحتمالات في كثير من النواحي التطبيقية، مثل المجالات الاقتصادية ، والتجارية ، والزراعية ، والطبية ، والسلوكية ، وغيرها ، خاصة عند اتخاذ القرار في دراسات الجدوى، والتنبؤ بسلوك الظواهر المختلفة ، ولكي يمكن فهم موضوع الاحتمال ، وأهميته في النواحي التطبيقية ، سيتم عرض بعض المفاهيم الخاصة بالاحتمالات.

بعض المفاهيم الخاصة بالاحتمال

• التجربة العشوائية (Randomized Experiment)

هي أي عملية (يمكن إتمامها) يمكن تحديد كل النتائج الممكنة لها ، ولكن لا يمكن مسبقا تحديد النتيجة التي ستظهر أو تحدث ، ومثال على ذلك عند إلقاء قطعة عملة معدنية مرة واحدة ، فإن النتائج الممكنة لها نتيجتان هما : (ظهور الصورة) ويرمز لها بالرمز (H) أو (ظهور الكتابة) و يرمز لها بالرمز (T) ، أي إن النتائج الممكنة هي : **{H , T}** ، وقبل إلقاء القطعة ، لا يمكن تحديد أي من النتيجتين سوف تظهر.

• فراغ العينة (Sample Space)

هي مجموعة النتائج الممكنة للتجربة ، ويرمز لها بالرمز (S) ، ويرمز لعدد النتائج المكونة لفراغ العينة بالرمز **(n(S))** ،ومن الأمثلة على ذلك :

1. عند إلقاء قطعة عملة غير متحيزة مرة واحدة ، نجد أن فراغ العينة هو (**S:{H , T}**) وعدد النتائج هي : (**n(S)=2**) .
2. عند إلقاء عملة معدنية غير متحيزة مرتين ، فإن فراغ العينة يمكن الحصول عليه من خلال شجرة الإحتمالات كما يلي:

الرمية الاولى	الرمية الثانية	S
H	H	HH
	T	HT
T	H	TH
	T	TT

$\therefore$ n(S)= 4

3. عند رمي النرد مرة واحدة غير متحيزة ،فإن فراغ العينة هو مجموعة النقاط التي تظهر على الوجه ، وهي :

S = {1, 2, 3, 4, 5, 6}

$\therefore$ n(S) = 6

4. عند إلقاء عملة معدنية غير متحيزة عدد من المرات حتى نحصل على الصورة مرة واحدة ، نجد إن التجربة هي عدد من المحاولات يتم إيقافها عندما نحصل على الصورة مرة واحدة، إذا فراغ العينة هو : $S = \{H, TH, TTH, TTTH \ldots.\}$ ويكون $N(S) = \infty$.

5. عند سحب ثلاث كرات بدون إرجاع من كيس به خمس كرات حمراء (Red) ، وثلاث زرقاء (Blue) ، وكرتان خضراء (Green) ، نجد إن فراغ العينة هو :

السحبة الأولى	السحبة الثانية	السحبة الثلثة	S
R	R	R	RRR
		B	RRB
		G	RRG
	B	R	RBR
		B	RBB
		G	RBG
	G	R	RGR
		B	RGB
		G	RGG
B	R	R	BRR
		B	BRB
		G	BRG
	B	R	BBR
		B	BBB
		G	BBG
	G	R	BGR
		B	BGB
		G	BGG
G	R	R	GRR
		B	GRB
		G	GRG
	B	R	GBR
		B	GBB
		G	GBG
	G	R	GGR
		B	GGB
		G	GGG

• الحادث (Event)

هو فئة جزئية من النتائج المكونة لفراغ العينة ، ويرمز للحالات بالحروف الهجائية (A, B ,C)

ويقسم الحادث إلى نوعين هما:

1. الحادث البسيط (Simple Event) : وهو الذي يحتوي على نتيجة واحدة من النتائج المكونة لفراغ العينة.
2. حادث مركب (Complex Event) :ويشمل نتيجتين أو أكثر من النتائج المكونة لفراغ العينة ، أي إن الحادث المركب يمكن تقسيمه إلى حوادث بسيطة . ويرمز لعدد النتائج المكونة للحادث بالرمز n(A) , n(B) ..... وهكذا. فعند

إلقاء عملة معدنية غير متحيزة ، وتعريق الحادث (A) بأنه ظهور الصورة مرتين، و الحادث (B) ظهور الصورة مرة واحدة على الأقل ، نجد إن فراغ العينة في هذه الحالة هو : S:{HH, HT ,TH, TT} وبالنسبة للحادث (A) فهو حادث بسيط ، يشمل نتيجة واحدة هي : A:{HH} أي إن : n(A)=1 ، أما الحادث (B) فهو حادث مركب يشمل ثلاث نتائج هي B:{HT, TH, HH} ، أي إن n(B)=3 ، وهذا الحادث يمكن تقسيمه إلى أحداث بسيطة.

• الإتحاد ("∪" Union)

يعبر عن إتحاد الحادثان (A) ، (B) عن وقوع أحدهما على الأقل أي وقوع الأول أو الثاني أو كلاهما ويعبر عن ذلك رياضيا ($A \cup B$) .

◈ مثال

عند إلقاء قطعة النرد مرة واحدة متزنة ، وتعريف الحادث (A) بإنه ظهور وجه يقبل القسمة على (3) ، والحادث (B) بأنه ظهور عدد فردي فإن :

$A: \{3, 6\},$
$B: \{1, 3, 5\}$
$\therefore (A \cup B) = \{1, 3, 5, 6\}$

• التقاطع ("∩" Intersection)

يعبر عن تقاطع الحادثان (A) ، (B) عن وقوع الإثنان في أن واحد ، ويشمل كل النتائج المشتركة بين الحادثين ، ويعبر عن ذلك رياضيا ب ($A \cap B$).

وفي المثال السابق فإن : ($(A \cap B):\{3\}$).

• الأحداث المتنافية أو المتعارضة (Mutually Exclusive events)

يقال عن الحادثين (A) ، (B) بأنهمامتنافيان إذاكان وقوع أحدهما يلغي الأخر أي إستحالة وقوعهما في أن واحد ومن ثم تكون نتيجة تقاطع الحادثان المتنافيان هي المجموعة الخالية و يرمز لهما (ϕ) أي إن ($A \cup B = \phi$) .

• الحادث المكمل (Compliment Event)

الحادث المكمل للحادث (A) هو الذي ينفي وقوعه ، ويرمز للحادث المكمل للحادث (A) بالرمز ($\overline{A}$) وعليه فإن :

$A \cap \overline{A} = \phi,$

$A \cup \overline{A} = S$

◈ مثال

ألقيت عملة معدنية غير متحيزة ثلاث مرات ، وعرفت الأحداث التالية:

الحادث (A) ظهور الصورة مرتين.

الحادث (B) ظهور الصورة مرة واحدة.

الحادث (C) ظهور الصورة في الرمية الأولى.

أوجد :

1. الأحداث الخاصة بالإتحاد ($A\cup B$, $A\cup C$, $A\cup B\cup C$).
2. الأحداث الخاصة بالتقاطعات ($A\cap B$, $A\cap C$, $A\cap B\cap C$).
3. الحادث ($\overline{B}$).

☑ الحل

1st	2nd	3rd	S
H	H	H	HHH
		T	HHT
	T	H	HTH
		T	HTT
T	H	H	THH
		T	THT
	T	H	TTH
		T	TTT

$\therefore\ n(S)= 8$

$A= \{HHT, HTH, THH\} \Rightarrow n(A)= 3$

$B= \{HTT, THT, TTH\} \Rightarrow n(B) =3$

C= {HHH, HHT, HTH, HTT} $\Rightarrow$ n(C) =4

1. الأحداث الخاصة بالإتحاد:

$(A\cup B)$: {HHT, HTH, THH, HTT, THT, TTH} $\Rightarrow n(A\cup B)=6$
$(A\cup C)$: {HHT, HTH, THH, HHH, HTT} $\Rightarrow n(A\cup C)=5$
$(A\cup B\cup C)$: {HHH, HHT, HTH, HTT, THT, TTH, THH} $\Rightarrow n(A\cup B\cup C)=7$

2. الأحداث الخاصة بالتقاطع:

$(A\cap B)$: $\phi \Rightarrow n(A\cap B)=0$
$(A\cap C)$:{ HHT, HTH} $\Rightarrow n(A\cap C)=2$
$(A\cap B\cap C)$: $\phi \Rightarrow n(A\cap B\cap C)=0$

3. إيجاد $(\bar{B})$

$(\bar{B})$:{ HHH, HHT, HTH, THH, TTT} $\Rightarrow n(\bar{B})=5$

طرق حساب الإحتمالات

يعتمد حساب الاحتمال من الناحية النظرية على أسس وقواعد الرياضيات ، ويعتبر هذا النوع من الاحتمال هو العنصر الأساسي في الاستدلال الإحصائي، ولكن في المجال التجريبي تعتمد الاحتمالات على النتائج الفعلية لمشاهدات التجربة ، وعلى تكرار الحادث محل الاهتمام ، فإذا رمزنا لإحتمال وقوع الحادث (A) بالرمز ($P(A)$)فإن طريقة حساب هذا الاحتمال تتحدد وفقا لنوع الإحتمال، وهما نوعان:

- **الإحتمال التجريبي (Emperical Probability)**

ويعبر عنه بالتكرار النسبي ، ويحسب بالمعادلة التالية:

$$\boxed{P(A) = \frac{f(A)}{n}}$$

حيث (n) هومجموع التكرارات (العدد الكلي للمشاهدات)،

($f(A)$) هو تكرار الحادث

❖ مثال

إذا تم إلقاء قطعة معدنية غير متحيزة (500) مرة وتم ملاحظة عدد مرات ظهور كل وجه و لخصت كما يلي:

face	H	T	Sum
عدد مرات ظهور الوجه	260	240	500

إحسب إحتمال ظهور كل وجه.

☑ الحل

$$P(H) = \frac{f(H)}{n} = \frac{260}{500} = 0.52$$

$$P(T) = \frac{f(T)}{n} = \frac{240}{500} = 0.48$$

- **الإحتمال النظري (Theoretical Probability)**

وهو الذي يعتمد في حسابه على أسس وقواعد الرياضيات ، والتي تستخدم في تحديد عدد النتائج الممكنة للتجربة ، وعدد النتائج الممكنة لوقوع الحادث ، ومن ثم يحسب هذا النوع من الاحتمال ، بتطبيق المعادلة التالية :

$$\boxed{P(H) = \frac{n(A)}{n(S)}}$$

حيث ($n(S)$) : هو عدد النتائج الممكنة للتجربة.

($n(A)$) : هو عدد النتائج الممكنة لوقوع الحادث (A).

فلو تم إلقاء عملة معدنية غير متحيزة مرة واحدة ، فإن فراغ العينة هو $S:\{H,T\}$ أي إن عدد النتائج الممكنة هي : $n(S)=(2)^1=2$ ، و إذا كان الحادث (A) يمثل ظهور الصورة :

$A:\{H\} \Rightarrow n(A)= 1$

إحتمال وقوع الحادث (A)

$$P(A) = \frac{n(A)}{n(S)} = \frac{1}{2} = 0.5$$

العلاقة بين الاحتمال التجريبي و الاحتمال النظري

عند زيادة عدد المحاولات (n) يقترب الإحتمال التجريبي من الإحتمال النظري .

- **النتائج المتشابهة**

إذا أجريت تجربة ، وكانت كل نتيجة من النتائج الممكنة للتجربة لها نفس الفرصة من الظهور ، أي إن كل نتيجة لها إحتمال هو $\left(\frac{1}{n(S)}\right)$ ، تسمى هذه النتئج بالنتائج المتشابهة.

<u>مثلا :</u> عند إلقاء قطعة النرد مرة واحدة متزنة ، نجد إن فراغ العينة هو $S:\{1, 2, 3, 4, 5,6\}$ ، وإحتمال ظهور كل وجه هو $\left(\frac{1}{6}\right)$ وعند إلقاءه مرتيننجد إن فراغ العينة هو $(n(S) = 6^2 = 36)$ نتيجة وهي:

	1	2	3	4	5	6
1	1,1	1,2	1,3	1,4	1,5	1,6
2	2,1	2,2	2,3	2,4	2,5	2,6
3	3,1	3,2	3,3	3,4	3,5	3,6
4	4,1	4,1	4,3	4,4	4,5	4,6
5	5,1	5,2	5,3	5,4	5,5	5,6
6	6,1	6,2	6,3	6,4	6,5	6,6

وهذه النتائج متماثلة و إحتمال كل نتيجة هو (1/36).

- **النتائج غير المتماثلة**

هي النتائج التي تحدث عند تكرار المحاولة بحيث إن إحتمالات نتائج كل محاولة غي متساوي، وبالتالي لا تتساوى إحتمالات نتائج التجربة.

مثلا : كيس به ثلاث كرات حمراء (R) و كرتان بيضاء (W) ، فإن في كل سحب يكون إحتمال ظهور كرة حمراء هو $(\frac{3}{5})$ و إحتمال ظهور كرة بيضاء هو $(\frac{2}{5})$ وعليه تكون نتائج فراغ العينة، وإحتمال كل نتيجة في حالة سحب كرتين هو:

1^{st}	2^{nd}	S = إحتمالات النتائج
R (3/5)	**R** (3/5)	**RR** (3/5)(3/5) = **(9/25)**
	W (2/5)	**RW** (3/5)(2/5) = **(6/25)**
W (2/5)	**R** (3/5)	**WR** (2/5)(3/5) = **(6/25)**
	W (2/5)	**WW** (2/5)(2/5) = **(4/25)**

بعض قوانين الإحتمالات (Propability Laws)

إذا كان لدينا الحوادث (A) و (B) و (C) فإن :

الجمع	P(A) + P(B)	$P(A\cup B) = P(A) + P(B) - P(A\cap B)$
	P(A) + P(B) + P(C)	$P(A\cup B\cup C) = P(A) + P(B) + P(C)$ $- P(A\cap B) - P(A\cap C) - P(B\cap C) - P(A\cap B\cap C)$
الضرب	P(A) × P(B)	$P(A\cap B) = P(A) \times P(B)$ للأحداث المستقلة

إختبار - ت (T-Test)

ويستخدم هذا الإختبار للمقارنة بين متوسطين حسابيين كل منهما لمجتمع (Population) مختلف عن الأخر . والهدف من هذا الإختبار هو للتأكد من إن الفرق بين المتوسطين الناتجين من عينتي المجتمعين المختلفين هو فرق ثابت (أي له دلالة) أم إنه فرق ناتج عن الصدفة وظروف إختبار العينة أي إنه إذا تكرر عدة مرات فإن هذا الفرق لن يظهر مرة أخرى (أي إن الوسطين الحسابيين متساوين و لا يوجد فرق بينهما).

والظروف الواجب توافرها في المجتمعين (العينتين) المراد مقارنة وسطهما الحسابيين ($\overline{X}_1, \overline{X}_2$) هي :

1. أن تكون العينتين مستقلتين عن بعضهما البعض (أي لاتوثر أحدهما على الأخرى بأي شكل من الأشكال).
2. المتغير المعتمد أو متغير الإستجابة (Response Variable) (y) يجب أن يكون متغير كمي لا وصفي.
3. يجب أن يكون المتغير (y) ذو توزيع إعتيادي (Normal Distribution) في كل عينة.

ولإجراء هذا الإختبار توجد ثلاث حالات لتفاصيل العينات ، وكما يلي:

- **الحالة الأولى :** العينتين متساويتان من حيث الحجم (Sample sizes) و التباين (Variances) ، أي $(n_1 = n_2 = n)$ ،$(S_1^2 = S_2^2)$:

$$\boxed{t = \frac{\overline{X_1} - \overline{X_2}}{S_{X_1X_2} \times \sqrt{\frac{2}{n-1}}}, \quad S_{X_1X_2} = \sqrt{\frac{S_{X_1}^2 + S_{X_2}^2}{2}}}$$

حيث:

($\overline{X_1}$) : الوسط الحسابي للعينة الأولى.

($\overline{X_2}$) : الوسط الحسابي للعينة الثانية.

($S_{X_1X_2}$) : الإنحراف المعياري الكلي (Pooled standard Deviation) (للعينتين معاً ككل).

(S_{X_1}) : الإنحراف المعياري للعينة الأولى.

(S_{X_2}) : الإنحراف المعياري للعينة الثانية.

($S_{X_1X_2} \times \sqrt{\frac{2}{n-1}}$) : الخطأ القياسي للفرق بين الوسطين الحسابيين للعينتين.

- الحالة الثانية : العينتين غير متساويتين من حيث الحجم (Sample sizes) ومتساويتان بالتباين (Variances) ، أي $(n_1 \neq n_2)$ ،$(S_1^2 = S_2^2)$:

$$t = \frac{\overline{X_1} - \overline{X_2}}{S_{X_1X_2} \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

$$S_{X_1X_2} = \sqrt{\frac{(n_1)S^2_{X_1} + (n_2)S^2_{X_2}}{n_1 + n_2 - 2}}$$

حيث : $(n_1 + n_2 - 2)$ الحجم الكلي للعينتين معاً = العدد الكلي لدرجات الحرية المستخدمة في هذا الإختبار.

● الحالة الثالثة : العينتين غير متساويتين من حيث الحجم (Sample sizes) والتباين (Variances) ، أي $(n_1 \neq n_2), (S_1^2 \neq S_2^2)$:

$$t = \frac{\overline{X_1} - \overline{X_2}}{S_{\overline{X_1} - \overline{X_2}}},$$

$$S_{\overline{X_1} - \overline{X_2}} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

حيث :
(S^2) : هو التقدير غير المتحيز للتباين .

● مستوى الدلالة الإحصائية (α) (Statistical level of significance)

يختار الباحث في العادة مستوى دلالة الفرق الذي يقبله بين المجموعتين في دراسته منذ البداية ليرفض الفرض أو يقبله إذا كانت القيمة المستخرجة أقل من تلك الموجودة عند ذلك المستوى الذي قبل به، وقيم الدلالة الإحصائية تكون في الغالب في معظم البحوث عند المستويات الأتية :

D.F	α					
	0.1	0.05	0.025	0.01	0.005	0.001
1	3.078	6.314	12.706	31.821	63.657	318.313
2	1.886	2.92	4.303	6.965	9.925	22.327
3	1.638	2.353	3.182	4.541	5.841	10.215
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.893
6	1.44	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.782
8	1.397	1.86	2.306	2.896	3.355	4.499
9	1.383	1.833	2.262	2.821	3.25	4.296
10	1.372	1.812	2.228	2.764	3.169	4.143
11	1.363	1.796	2.201	2.718	3.106	4.024
12	1.356	1.782	2.179	2.681	3.055	3.929
13	1.35	1.771	2.16	2.65	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.12	2.583	2.921	3.686
17	1.333	1.74	2.11	2.567	2.898	3.646
18	1.33	1.734	2.101	2.552	2.878	3.61
19	1.328	1.729	2.093	2.539	2.861	3.579

20	1.325	1.725	2.086	2.528	2.845	3.552
21	1.323	1.721	2.08	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.5	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.06	2.485	2.787	3.45
26	1.315	1.706	2.056	2.479	2.779	3.435
27	1.314	1.703	2.052	2.473	2.771	3.421
28	1.313	1.701	2.048	2.467	2.763	3.408
29	1.311	1.699	2.045	2.462	2.756	3.396
30	1.31	1.697	2.042	2.457	2.75	3.385
31	1.309	1.696	2.04	2.453	2.744	3.375
32	1.309	1.694	2.037	2.449	2.738	3.365
33	1.308	1.692	2.035	2.445	2.733	3.356
34	1.307	1.691	2.032	2.441	2.728	3.348
35	1.306	1.69	2.03	2.438	2.724	3.34
36	1.306	1.688	2.028	2.434	2.719	3.333
37	1.305	1.687	2.026	2.431	2.715	3.326
38	1.304	1.686	2.024	2.429	2.712	3.319
39	1.304	1.685	2.023	2.426	2.708	3.313
40	1.303	1.684	2.021	2.423	2.704	3.307
41	1.303	1.683	2.02	2.421	2.701	3.301
42	1.302	1.682	2.018	2.418	2.698	3.296
43	1.302	1.681	2.017	2.416	2.695	3.291
44	1.301	1.68	2.015	2.414	2.692	3.286
45	1.301	1.679	2.014	2.412	2.69	3.281
46	1.3	1.679	2.013	2.41	2.687	3.277
47	1.3	1.678	2.012	2.408	2.685	3.273
48	1.299	1.677	2.011	2.407	2.682	3.269
49	1.299	1.677	2.01	2.405	2.68	3.265
50	1.299	1.676	2.009	2.403	2.678	3.261
51	1.298	1.675	2.008	2.402	2.676	3.258
52	1.298	1.675	2.007	2.4	2.674	3.255
53	1.298	1.674	2.006	2.399	2.672	3.251
54	1.297	1.674	2.005	2.397	2.67	3.248
55	1.297	1.673	2.004	2.396	2.668	3.245
56	1.297	1.673	2.003	2.395	2.667	3.242
57	1.297	1.672	2.002	2.394	2.665	3.239
58	1.296	1.672	2.002	2.392	2.663	3.237
59	1.296	1.671	2.001	2.391	2.662	3.234
60	1.296	1.671	2	2.39	2.66	3.232
61	1.296	1.67	2	2.389	2.659	3.229
62	1.295	1.67	1.999	2.388	2.657	3.227
63	1.295	1.669	1.998	2.387	2.656	3.225
64	1.295	1.669	1.998	2.386	2.655	3.223
65	1.295	1.669	1.997	2.385	2.654	3.22
66	1.295	1.668	1.997	2.384	2.652	3.218
67	1.294	1.668	1.996	2.383	2.651	3.216
68	1.294	1.668	1.995	2.382	2.65	3.214
69	1.294	1.667	1.995	2.382	2.649	3.213

70	1.294	1.667	1.994	2.381	2.648	3.211
71	1.294	1.667	1.994	2.38	2.647	3.209
72	1.293	1.666	1.993	2.379	2.646	3.207
73	1.293	1.666	1.993	2.379	2.645	3.206
74	1.293	1.666	1.993	2.378	2.644	3.204
75	1.293	1.665	1.992	2.377	2.643	3.202
76	1.293	1.665	1.992	2.376	2.642	3.201
77	1.293	1.665	1.991	2.376	2.641	3.199
78	1.292	1.665	1.991	2.375	2.64	3.198
79	1.292	1.664	1.99	2.374	2.64	3.197
80	1.292	1.664	1.99	2.374	2.639	3.195
81	1.292	1.664	1.99	2.373	2.638	3.194
82	1.292	1.664	1.989	2.373	2.637	3.193
83	1.292	1.663	1.989	2.372	2.636	3.191
84	1.292	1.663	1.989	2.372	2.636	3.19
85	1.292	1.663	1.988	2.371	2.635	3.189
86	1.291	1.663	1.988	2.37	2.634	3.188
87	1.291	1.663	1.988	2.37	2.634	3.187
88	1.291	1.662	1.987	2.369	2.633	3.185
89	1.291	1.662	1.987	2.369	2.632	3.184
90	1.291	1.662	1.987	2.368	2.632	3.183
91	1.291	1.662	1.986	2.368	2.631	3.182
92	1.291	1.662	1.986	2.368	2.63	3.181
93	1.291	1.661	1.986	2.367	2.63	3.18
94	1.291	1.661	1.986	2.367	2.629	3.179
95	1.291	1.661	1.985	2.366	2.629	3.178
96	1.29	1.661	1.985	2.366	2.628	3.177
97	1.29	1.661	1.985	2.365	2.627	3.176
98	1.29	1.661	1.984	2.365	2.627	3.175
99	1.29	1.66	1.984	2.365	2.626	3.175
100	1.29	1.66	1.984	2.364	2.626	3.174
infinity	1.282	1.645	1.96	2.326	2.576	3.09

مثال:

كانت درجات مجموعتين من الطلاب (ذكور و إناث) عدد كل منهما ستة في مادة دراسية معينة كما يلي:

Male X1				**Female X2**			
	X_1	$X_1 - \overline{X_1}$	$\left(X_1 - \overline{X}_1\right)^2$		X_2	$X_2 - \overline{X_2}$	$\left(X_2 - \overline{X}_2\right)^2$
1	**5**	**0**	**0**	**1**	**3**	**- 3**	**9**
2	**10**	**5**	**25**	**2**	**12**	**6**	**36**
3	**8**	**3**	**9**	**3**	**15**	**9**	**81**
4	**4**	**-1**	**1**	**4**	**4**	**-2**	**4**
5	**2**	**-3**	**9**	**5**	**1**	**-5**	**25**
6	**1**	**-4**	**16**	**6**	**1**	**-5**	**25**

sum	30	0	60	sum	36	0	180`

$$S^2_{X_1} = \frac{\sum\left(X_1 - \overline{X_1}\right)^2}{n} = \frac{60}{6} = 10$$

$$S^2_{X_2} = \frac{\sum\left(X_2 - \overline{X_2}\right)^2}{n} = \frac{180}{6} = 30$$

$$S_{X_1X_2} = \sqrt{\frac{10+30}{2}} = \sqrt{\frac{40}{2}}$$

$$t = \frac{\overline{X_1} - \overline{X_2}}{S_{X_1X_2} \times \sqrt{\frac{2}{n-1}}} = \frac{5-6}{\sqrt{\frac{40}{2}} \times \sqrt{\frac{2}{5}}} = \frac{1}{\sqrt{8}}$$

$$\therefore t = 0.35$$

وعند تحري دلالة القيمة العددية ل (t) يتم إجراء ما يلي:

1. إن درجة الحرية المعتمدة في الحسابات هنا هي (n-1= 6 – 1=5) .
2. عند الرجوع إلى الجدول الخاص بدرجات الحرية نجد ما يلي:

D.f	0.1	0.05	0.025	0.01	0.005	0.001
5	1.476	2.015	2.571	3.365	4.032	5.893

حيث إن القيم العددية المدرجة عند كل قيمة من قيم (α) تمثل قيمة (t) التي يجب إعتمادها للتحليل .

3. مقارنة قيمة (t) المحسوبة مع قيم (t) الموجودة في الجدول عند أي قيمة من قيم (α) حيث نجد إنها أقل (غير ذا قيمة مقارنة لما في الجدول) من أي مما في الجدول مما يدل على إن الفرق بين قيمتي الوسطين الحسابيين لكلا العينتين هو غير مهم ولا يؤخذ بنظر الإعتبار.
4. أما إذا كانت القيمة المحسوبة ل(t) أكبر مما في الجدول عند أي قيمة من النسب الثلاث (0.05 ,0.01, 0.001) كان الفرق بين الوسطين الحسابيين ذو قيمة عند قيمة (α) التي هي أقل من المحسوبة.

مثال

أجري إختبار على (عينتين) مجموعتين (ذكور) و (إناث) مختلفتي في الحجم وكانت نتائج الإختبار كما يلي:

Male X1				Female X2			
	X_1	$X_1 - \overline{X_1}$	$\left(X_1 - \overline{X}_1\right)^2$		X_2	$X_2 - \overline{X_2}$	$\left(X_2 - \overline{X}_2\right)^2$
1	5	0	0	1	15	1	1
2	10	5	25	2	19	5	25
3	8	3	9	3	16	2	4
4	4	-1	1	4	10	-4	16
5	2	-3	9	5	10	-4	16
6	1	-4	16				
sum	30	0	60	sum	70	0	62

$$S^2_{X_1} = \frac{\sum\left(X_1 - \overline{X_1}\right)^2}{n_1} = \frac{60}{6} = 10$$

$$S^2_{X_2} = \frac{\sum\left(X_2 - \overline{X_2}\right)^2}{n_2} = \frac{62}{5} = 12.4$$

$$S_{X_1X_2} = \sqrt{\frac{(n_1)S^2_{X_1} + (n_2)S^2_{X_2}}{n_1 + n_2 - 2}} = \sqrt{\frac{(6)10 + (5)12.4}{6+5-2}} = \sqrt{\frac{60+62}{9}} = 3.68$$

$$t = \frac{\overline{X_1} - \overline{X_2}}{S_{X_1X_2} \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{14 - 5}{3.68 \times \sqrt{\frac{1}{6} + \frac{1}{5}}} = \frac{9}{2.23} = 4.04$$

درجة الحرية المستخدمة هي (5+6-2=9) ، ومن خلال الجدول نجد

D.f	**0.1**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**
9	1.383	1.833	2.262	2.821	3.25	4.296

ومن خلال مقارنة قيمة (t) المحسوبة مع ما موجود في الجدول نجد إنها أكبر من نظيرتها عند المستوى (0.005) أي لها دلالة إحصائية.
وعليه فإن الفرق بين قيمتي الوسطين الحسابيين هو فرق واقعي ولم يكن من جراء الصدفة.

تحليل التباين (ANOVA) (Analysis of Variance)

● تحليل التباين

هو مجموعة من النماذج الإحصائية (statistical models) مع إجراءات مرافقة لهذه النماذج تمكن من مقارنة الأوساط الحسابية لمجتمعات إحصائية مختلفة عن طريق تقسيم التباين variance الكلي الملاحظ بينهم إلى أجزاء مختلفة.

أول طرق تحليل التباين تم وضعها من قبل الإحصائي رونالد فيشر في العشرينات والثلاثينات من القرن العشرين لذلك تعرف أحيانا بتحليل فيشر للتباين.

قد سبق دراسة اختبارات الفروض لتساوي متوسطي مجتمعين ولكن هناك دراسات لتساوي متوسطات ثلاث مجتمعات أو أكثر، وهل يمكن التعميم لأكثر من مجتمعين، نعم قد يكون ذلك ولكن وجود عقبات رئيسية تجعلنا نبحث عن طريقة أخرى ومن هذه عقبات هي:

1. الجهد المبذول في المقارنة بين كل مجتمعين وخاصة إذا كثر عدد المجموعات الثنائية .
2. إذا كان لدينا العديد من المستويات، فالمقارنة الثنائية بينهم تفقد الكثير من المعلومات المتوفرة لدينا عن المجتمع محل الدراسة وهو ما ينقص من دقة حساب معالم المجتمع.

أسلوب آخر لاختبار تساوي الأوساط الحسابية يعرف بتحليل التباين (Analysis of variance) أو ANOVA هو الذي قدمه العالم فيشر (Ronald A.Fisher) كأسلوب لتحليل البيانات للتجارب المختلفة وهو عبارة عن مجموعة من الطرق الإحصائية المساعدة لاختبارات الفروض أبسطها (one way ANOVA) وهو طريقة لاختبار دلالة الفرق بين الأوساط الحسابية لعدة عينات بمقارنة واحدة، ويعرف أيضاً بطريقة تؤدي لتقسيم الاختلافات الكلية لمجموعة من المشاهدات التجريبية لعدة أجزاء للتعرف على مصدر الاختلاف بينها ولذا فالهدف هنا فحص تباين المجتمع لمعرفة مدى تساوى متوسطات المجتمع ولكن لا بد من تحقيق ثلاثة أمور قبل استخدامه وهي:

1) العينات عشوائية ومستقلة.

2) مجتمعات هذه العينات كلاً لها توزيع طبيعي.

3) تساوي تباين المجتمعات التي أخذت منها العينات العشوائية المستقلة.

ولتوضيح ما سبق بمقارنة متوسطات ثلاث مجتمعات باستخدام ثلاث عينات (تحقق فيها الشروط الثلاثة السابقة) موضحة بالجدول الآتي:

Obsrvs. No.	Smpl.1(X_1)	Smple.2(X_2)	Smpl.3(X_3)
1	33	27	40
2	32	28	41
3	33.5	26.5	40.5
4	31.5	26.5	38.5
Sum	130	108	160
Mean ($\overline{X}$)	32.5	27	40
Standard deviation (S)	0.91	0.71	1.08

السؤال: هل في البيانات ما يكفي لوجود فرق بين الأوساط الحسابية ؟

الجواب: نعم (بمجرد النظر) فالتشتت (التباين) ظاهر 40، 27، 32.5 (الأوساط الحسابية) بمقارنته بالتشتت بين العينات (وحداتها 40 ، 41 ، 40.5، 38.5) فيبدو معدوماً.

إذا أخذنا البيانات الآتية:

Obsrvs. No.	Smpl.1(X_1)	Smple.2(X_2)	Smpl.3(X_3)
1	10	50	40
2	60	20	15
3	27.5	11	65
Sum	97.5	81	120
Mean ($\overline{X}$)	32.5	27	40
Standard deviation (S)	25.4	20.4	25

فالبيانات هنا لها نفس الأوساط الحسابية في البيانات السابقة ولكن التشتت (داخل لعينات) كبيراً بما هو عليه في الأوساط الحسابية.

فالدليل على وجود الفرق بين متوسطات الجدول الأول واضح ولا يظهر ذلك بوضوح في بيانات الجدول الثاني بالرغم من تساوي الأوساط الحسابية في الحالتين ولذا يتبين لنا القصد من تحليل التباين والذي يعني الفرق بين الأوساط الحسابية والذي يقاس بالتشتت داخل البيانات.

اختبار تساوي أكثر من متوسطين

إن كان لدينا عدد (K) من العينات (أو المجتمعات) وضمن كل عينة من هذه العينات عدد (n) من القراءات (المشاهدات) ، يمكن وضع الفرضيات الأتية :

1. فرضية العدم (H_o)(Null Hypothesis) (وتنص على تساوي الأوساط الحسابية للعينات):

$$\boxed{H_o = \overline{X_1} = \overline{X_2} = \overline{X_3} = ... = \overline{X_k}}$$

2. فرضية عدم تساوي الأوساط الحسابية (H_1) (وتنص على عدم تساوي الأوساط الحسابية الحسابية للعينات أو على الأقل عدم تساوي إثنان منها):

$$\boxed{\begin{array}{l} H_1 = \overline{X_1} \neq \overline{X_2} \neq \overline{X_3} \neq .... \neq \overline{X_k} \\ H_1 = \overline{X_1} >,< \overline{X_k} \end{array}}$$

وبفرض أن العينات مأخوذة من مجتمعات طبيعية ولها نفس التباين، فلإختبار يرتكز على مقارنة التباين داخل العينات وبينها بحساب التباين المشترك بطريقتين فالأولى لا تعتمد على صحة أو عدم صحة الفرض الصفري بينما تتأثر الطريقة الثانية بالفرض الصفري فإن تبين وجود خلاف معنوي بين الطريقتين (في الحساب) أخذنا بعدم صحة الفرض الصفري لأن عدم تساوي الأوساط الحسابية قد أثر على الحساب الثاني فتسبب في تجاوزه الحساب الأول فرفض (H_o) ونفصل ذلك بحسابين للتباين (S^2) كالآتي:

للتبسيط لنأخذ عينات من المجتمعات محل الدراسة لها نفس الحجم وحيث أن التباين في المجتمعات متماثل فنقدر التباين (S^2) بمتوسط التباينات في العينات أي أن:

$$\boxed{S^2 = \sum_{i=1}^{K} \frac{S_i^2}{K}}$$

ويرمز لهذا الحساب بالرمز (S_w^2) لكونه يمثل التباين داخل المجموعات (Group Within) أي:

$$\boxed{S_w^2 = \sum_{i=1}^{k} \frac{S_i^2}{k}}$$

وفي حالة تساوي حجم العينات . ويلاحظ عدم اعتماد هذا الحساب على صحة أو عدم صحة (H_o) لأن كل تباين (S_i^2) محسوب بطريقة مستقلة عن الآخرين وبافتراض صحة (H_o) فيعني أن العينات مأخوذة من مجتمع واحد، ونعلم تباين الأوساط الحسابية مأخوذة من مجتمع تباينه (S^2) ويساوي $(\frac{S^2}{n})$ وحسابه يتم بالعلاقة الأتية :

$$\boxed{S_{\overline{X}}^2 = \frac{\sum_{i=1}^{k} \left(\overline{X_i} - \overline{X}\right)^2}{K-1}}$$

حيث يمثل ($\overline{X}$) الوسط الكلي للمتوسطات وعليه فيمكن حساب (S^2) بضرب ($S^2_{\overline{X}}$) بالحجم المشترك للعينات (n) ، ويرمز للناتج ب (S^2_B) أي:

$$\boxed{S^2_B = n.S^2_{\overline{X}}}$$

و(S^2_B) يمثل التباين بين المجموعات (Between Group) وهذان الحسابان للتباين المشترك (S^2_W) و (S^2_B) أحدهم لا يعتمد على صحة أو عدم صحة (H_o) في حين الآخر يوجب صحة (H_o) أي أن جميع العينات المأخوذة يجب أن تكون من نفس المجتمع فتطابق الحسابين يعني صحة (H_o) وإلا تعارضت البيانات مع (H_o) ويجب أن نعلم أن اختلاف حجم العينات يجعل قيمة الحساب الأول (S^2_W) كالآتي:

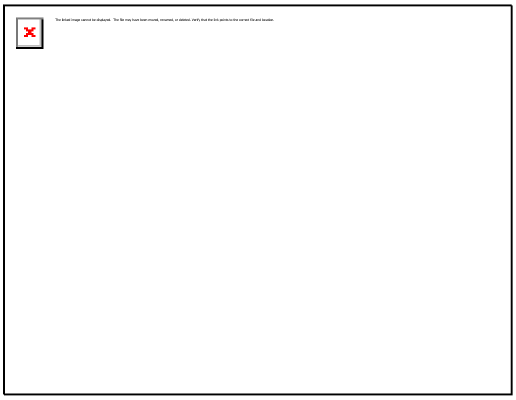
$$\boxed{S^2_W = \frac{(n_1-1)S_1^2 + (n_1-1)S_2^2 + \ldots + (n_k-1)S_k^2}{n_1+n_2+\ldots+n_k-K}}$$

وهو امتداد لحساب المجتمع واستخدم للاستدلال الإحصائي لمتوسطين حال تساوي تباين المجتمعين ويكون الحساب الثاني كالآتي:

$$\boxed{S_B^2 = \frac{\sum_{i=1}^{k} n_i.\left(\overline{X_i} - \overline{X}\right)^2}{K-1}}$$

ويمكن استخدم الصيغ التالية:

النسبة ($\frac{S^2_B}{S^2_W}$) تعرف بتوزيع (F) (F-distribution) وهو توزيع ملتوي لجهة اليمين يعتمد على درجتي حرية (البسط ، المقام) وهما (k-1) للبسط و (N-k) للمقام حيث (N) مجموع إحجام العينات ، فإذا كان لدينا اختبار لقياس دلالة الفرق بين الحسابين (F) نوجد (F_α) حيث (α) مستوى الدلالة (معامل الدلالة الإحصائية) المستخدم للفرضية (H_o) التي ترفض إذا كان ($F_\alpha < or > F$) وإلا نؤكد بوجود الاختلاف بين الأوساط الحسابية .



منحنى توزيع F حسب درجات الحرية

جدول تحليل التباين (ANOVA table)

إذا أردنا إجراء اختبار الفروض بين متوسطات عددها (k) من العينات العشوائية المستقلة وبفرض (n_1) عدد مفردات العينة الأولى ، (n_2) عدد مفردات العينة الثانية ، ...، (n_k) عدد مفردات العينة k وأن يمكن وضع الجدول التالي لبيانات العينات في تحليل التباين:

Observations	Samples				
	Smpl(1) **X_1**	**Smpl(2)** **X_2**	**Smpl(3)** **X_3**	**Smpl (…)** **$X_{…}$**	**Smpl(k)** **X_k**
1	X_{11}	X_{21}	X_{31}	$X_{…1}$	X_{k1}
2	X_{12}	X_{22}	X_{32}	$X_{…2}$	X_{k2}
:	$X_{1:}$	$X_{2:}$	$X_{3:}$	$X_{…:}$	$X_{k:}$
N	X_{1n1}	X_{2n2}	$X_{…n:}$	$X_{…n:}$	X_{knk}
sum	$T_1=X_{11}+X_{12}+…+X_{1n1}$	T_2	T_3	T_n	T_k
Mean	$\bar{X}_1 = T_1/n_1$	$\bar{X}_2$	$\bar{X}_3$		$\bar{X}_K = T_K/n_K$
S.D	S_1	S_2	S_3		S_K
Total No. of observations (N) = $n_1+n_2+…+n_K$					
Total No. of data (T) = $T_1+T_2+…+T_K$					

من الجدول يتبين لنا:

1. الانحراف بين قيمة المشاهدة والوسط الحسابي العام وهو الانحراف الكلي ويرمز له ($X_{ji} - \bar{X}$) حيث ($i=1,2,3....k$) (للعينات) و ($j=1,2,3,...n$) (للمشاهدات).
2. الاختلاف بين الوسط الحسابي بكل عينة ($\bar{X}_k$) والوسط الحسابي العام ($\bar{X}$) .
3) الاختلاف بين قيمة كل مشاهدة داخل العينة والوسط الحسابي لنفس العينة ($X_{ji} - \bar{X}_i$) وهو الانحراف داخل العينات ويرجع هذا الاختلاف لأسباب عشوائية بحتة.

بناء على ما سبق يمكن النظر للجدول التالي (جدول تحليل التباين) الذي يبين الخطوات اللازمة لحساب F (قيمة إحصائية الاختبار) حيث K عدد مستويات المتغير المستقل:

Source of var. مصدر التباين	Sum. Of squares (SS) مجموع المربعات	Degree of freedom (DF) درجة الحرية	Mean Squares (MS) متوسط مجموع المربعات	Calculated (F) قيمة(F)المحسوبة	Tabulated (F)=(F_α) قيمة(F) الجدولية
Between Samples بين العينات	SS_B	K-1	S_B^2	$\frac{S_B^2}{S_W^2}$	$F_{\alpha(k-1),(N-k)}$
within Sample(error) ضمن العينة الواحدة	SS_W	N-K	S_W^2		
Total	$SS_T = SS_B + SS_W$	N-1			

حيث الكميات الموجودة في الجدول يمكن حسابها من خلال المعادلات الأتية:

$$SS_B = \sum_{j=1}^{K}\left(\frac{T_j^2}{n_j}\right) - \frac{T^2}{N},$$

$$SS_W = \sum_{j=1}^{K}\left(\sum_{i=1}^{n_k} X_{ji}^2\right) - \sum_{j=1}^{K}\left(\frac{T_j^2}{n_j}\right),$$

$$S_B^2 = \frac{SS_B}{K-1},$$

$$S_W^2 = \frac{SS_W}{N-K}$$

وعلية ، فمن أجل إجراء إتبار تحليل التباين ، يجب إجراء التالي:

1. من خلال المعطيات يجب تحديد قيم كل من (عدد النماذج k ، عدد المشاهدات أو القراءات في كل عينة n_j ، عدد المشاهدات الكلي $N=n_1+n_2+\ldots+n_k$).
2. نحسب مجموع قيم المشاهدات في العينة الواحدة ، وليكن (T_j)، ثم نحسب مربعه (T_j^2).
3. نحسب المجموع الكلي للمشاهدات في كل العينات و ليكن ($T=T_1+T_2+\ldots+T_k$)، ثم نحسب مربعه (T^2).
4. نحسب مربع كل قيمة من قيم المشاهدات في كل عينة (X_{ji})، حيث ($j=1,2,..,k$) يشير إلى رقم العينة ، ($i=1,2,\ldots,n_k$) يشير إلى رقم المشاهدة في العينة ، ويكون المربع هو (X_{ji}^2).
5. نحسب مجموع مربعات القيم في العينة الواحدة وليكن ($\sum_{i=1}^{n} X_{ji}^2$).
6. نحسب المجموع الكلي (لجميع العينات) لمحموع المربعات ،و ليكن ($\sum_{j=1}^{k}\left(\sum_{i=1}^{n} X_{ji}^2\right)$).
7. نحسب قيمة مصدر التباين بين العينات (SS_B): $\boxed{SS_B = \sum_{j=1}^{K}\frac{T_j^2}{n_j} - \frac{T^2}{N}}$
8. ومنها نحسب قيمة متوسط مجموع المربعات (S_B^2): $\boxed{S_B^2 = \frac{SS_B}{K-1}}$
9. نحسب قيمة مصدر التباين ضمن العينة الواحدة (SS_W): $\boxed{SS_W = \sum_{j=1}^{k}\left(\sum_{i=1}^{n} X_{ji}^2\right) - \sum_{j=1}^{K}\frac{T_j^2}{n_j}}$
10. ومنها نحسب قيمة متوسط مجموع المربعات (S_W^2) : $\boxed{S_W^2 = \frac{SS_W}{N-K}}$
11. وأخيراً يتم حساب قيمة (F) ومقارنتها مقيمة (F_α) اجدولية .

مثال:

في دراسة لتأثير وجود الطلاب في الصفوف على تحصيلهم في مادة الإحصاء، قام أستاذ الإحصاء بأخذ عينات عشوائية ومستقلة من ثلاثة صفوف (يقوم بتدريسها) كل منها مكون من خمسة طلاب وقام الأستاذ برصد درجاتهم والجدول التالي يبينها.

بمستوى دلالة (α= 0.05) اختبر ما إذا كان متوسط النتائج في اختبارات الأداء يختلف في تحصيل الطلاب.

Class 1	Class 2	Class 3
66	96	58
65	87	62
88	66	77
92	55	90
60	78	80

الحل:

الاختبار : $(H_o = \overline{X}_1 = \overline{X}_2 = \overline{X}_3)$

الإختبار : على الأقل متوسطان فقط غير متساويين.

نستكمل الجدول كالآتي:

No.	**Class1**		**Class 2**		**Class 3**	
	X_1	X_1^2	X_2	X_2^2	X_3	X_3^2
1	**66**	**4356**	**96**	**9216**	**58**	**3364**
2	**65**	**4225**	**87**	**7569**	**62**	**3844**
3	**88**	**7744**	**66**	**4356**	**77**	**5929**
4	**92**	**8464**	**55**	**3025**	**90**	**8100**
5	**60**	**3600**	**78**	**6084**	**80**	**6400**
Sum (T)	**371**	**28389**	**382**	**30250**	**367**	**27637**

$$\mathrm{T} = \mathrm{T}_1 + \mathrm{T}_2 + \mathrm{T}_3 = 371 + 382 + 367 = 1120$$

$$T^2 = 1254400$$

$$N = n_1 + n_2 + n_3 = 5 + 5 + 5 = 15$$

$$k = 3$$

$$SS_B = \sum_{k=1}^{K}\left(\frac{T_k^2}{n_k}\right) - \frac{T^2}{N},$$

$$SS_W = \sum_{k=1}^{K}\left(\sum_{i=1}^{n_k} X_{ki}^2\right) - \sum_{k=1}^{K}\left(\frac{T_k^2}{n_k}\right) = \sum_{1}^{n} X_1^2 + \sum_{1}^{n} X_1^2 + \sum_{1}^{n} X_1^2 - \sum_{k=1}^{K}\left(\frac{T_k^2}{n_k}\right),$$

$$S_B^2 = \frac{SS_B}{K-1}, S_W^2 = \frac{SS_W}{N-K}$$

$$\therefore SS_B = \frac{137641}{5} + \frac{145924}{5} + \frac{134689}{5} - \frac{1254400}{15} = \frac{418254}{5} - \frac{1254400}{15} = 24.1333,$$

$$S_B^2 = \frac{24.1333}{3-1} = 12.1,$$

$$SS_W = 28389 + 30250 + 27637 - \frac{418254}{5} = 86276 - 83650.5 = 2625.5,$$

$$S_W^2 = \frac{2625.5}{15-3} = 218.8,$$

$$\therefore F = \frac{S_B^2}{S_W^2} = \frac{12.1}{218.8} = 0.055$$

ومن الجداول نجد قيمة $\left(F_{\alpha(k-1,N-k)}\right)$ لتكون $\left(F_{\alpha_{(2,12)}} = 3.89\right)$ وهي أكبر من قيمة (F) المحسوبة وهي (F=0.055) وعليه يجب قبول فرضية العدم (H_o) بعدم وجود إختلاف بين الأوساط عند قيمة (α=0.05).

جدول النتائج

Source of var.	Sum. Of squares (SS)	Degree of freedom (DF)	Mean Squares (MS)	Calculated (F)	Tabulated (F)=(F_α)
Between Samples	SS_B=24.1333	K-1= 3-1=2	S_B^2=12.1	$\frac{S_B^2}{S_W^2}$=0.055	$F_{\alpha_{(k-1),(N-k)}}$=3.89
within Sample(error)	SS_W=2625.5	N-K= 15-3=12	S_W^2=218.8		
Total	$SS_T = SS_B + SS_W$=2649.6333	N-1= 15-1=14			

No.	A		B		C		D		E	
	X_A	X^2_A	X_B	X^2_B	X_C	X_C^2	X_D	X_D^2	X_E	X_E^2
1	68	4624	72	5184	60	3600	48	2304	64	4096
2	72	5184	53	2809	82	6724	61	3721	65	4225
3	77	5929	63	3969	64	4096	57	3249	70	4900
4	42	1764	53	2809	75	5625	64	4096	68	4624
5	53	2809	48	2304	72	5184	50	2500	53	2809
T	312	20310	289	17075	353	25229	280	15870	320	20654

$$N = n_A + n_B + n_C + n_D + n_E = 5+5+5+5+5 = 25$$

$$K = 5$$

$$T = T_A + T_B + T_C + T_D + T_E = 312 + 289 + 353 + 280 + 320 = 1554$$

$$\mathrm{T}^2 = 2414916$$

$$SS_B = \sum_1^K \frac{T_K^2}{n_K} - \frac{T^2}{N} = \frac{97344}{5} + \frac{83521}{5} + \frac{124609}{5} + \frac{78400}{5} + \frac{102400}{5} - \frac{2414916}{25}$$

$$= \frac{486274}{5} - \frac{2414916}{25} = 97254.8 - 96596.64 = 658.16$$

$$\mathrm{S}_\mathrm{B}^2 = \frac{SS_B}{K-1} = \frac{658.16}{4} = 164.54$$

$$\mathrm{SS_W} = \sum_{k=1}^{K}\left(\sum_{i=1}^{n_k} X_{ki}^2\right) - \sum_{k=1}^{K}\left(\frac{T_k^2}{n_k}\right) = 20310 + 17075 + 25229 + 15870 + 20654 - 79254.8$$

$$= 99138 - 97254.8 = 1883.2$$

$$\mathrm{S}_\mathrm{W}^2 = \frac{SS_W}{N-K} = \frac{1883.2}{20} = 94.16$$

$$F = \frac{S_B^2}{S_W^2} = \frac{164.54}{94.16} = 1.75$$

$$F_{\alpha(4,20)} = 0.175$$

ANOVA TEST
Relationship between quality and temperature

An industrial plant was maintained at different temperatures on four successive days, and on each day three samples were taken from the process and analyzed for quality; a score was awarded in arbitrary units, and the resulting scores are given in the table below

Sample	100° C		120° C		140° C		160° C	
	X_1	X_1^2	X_2	X_2^2	X_3	X_3^2	X_4	X_4^2
1	41	1681	54	2916	50	2500	38	1444
2	44	1936	56	3136	52	2704	36	1296
3	48	2304	53	2809	48	2304	41	1681
T(sum)	133	5921	163	8861	150	7508	115	4421
T^2	17689		26569		22500		13225	

Is there evidence to suggest that average score depends on temperature? If so, determine between which temperatures there are significant differences at the *5%* level.

Solution

$$k = 4$$

$$n_1 = n_2 = n_3 = n_4 = 3$$

$$N = 12$$

$$T = T_1 + T_2 + T_3 + T_4 = 133 + 163 + 150 + 115 = 561$$

$$\Rightarrow T^2 = 314721,$$

$$\sum_{j=1}^{K}\left(\sum_{i=1}^{n} X_{ji}^2\right) = 5921 + 8861 + 7508 + 4421 = 26711$$

$$SS_B = \sum_{k=1}^{K}\frac{T_k^2}{n_K} - \frac{T^2}{N} = \left(\frac{17689}{3} + \frac{26569}{3} + \frac{22500}{3} + \frac{13225}{3}\right) - \frac{314721}{12}$$

$$= \frac{79983}{3} - \frac{314721}{12} = 26661 - 26226.75 = 434.25$$

$$S_B^2 = \frac{SS_B}{k-1} = \frac{434.25}{3} = 144.75$$

$$SS_W = \sum_{j=1}^{K}\left(\sum_{i=1}^{n} X_{ji}^2\right) - \sum_{k=1}^{K}\frac{T_k^2}{n_K} = 26711 - 26661 = 50$$

$$S_W^2 = \frac{SS_W}{N-K} = \frac{50}{8} = 6.25$$

$$\therefore F = \frac{S_B^2}{S_W^2} = \frac{144.75}{6.25} = 23.16$$

The calculated F-ratio, 23.16, is now compared with the tabulated upper percentage points of the F-distribution on 3 and 8 degrees of freedom. At the 1% significance level the tabulated value is (7.59).
The result is thus highly significant, giving very strong evidence that temperature affects score.