1- نبذة تاريخية عن علم الإحصاء

تطور علم الإحصاء ضمن مراحل متعددة مع التطور التكنولوجي والرياضي وزاد الاهتمام بهذا النوع من العلوم نظراً لما يوفره من اساليب وطرق علمية ورياضية لمعالجة البيانات وتحليلها، كما يعد علم الإحصاء من العلوم الأساسية التي تدخل في العديد من العلوم التطبيقية والاجتماعية لمعالجة العديد من الظواهر، أن كلمة (الإحصاء) في الماضي كانت تهدف إلى العد والحصر وسمي الإحصاء بعلم العد (The science of counting) كما أن لفظة إحصاء باللغة الانكليزية (Statistics) كانت تستعمل في بلاد أوربا للدلالة على أعمال وحسابات الدولة في شؤون الحرب والضرائب وعدد السكان والمواليد والوفيات والإنتاج الزراعي ...الخ، لذا فان مفهوم علم الإحصاء يختلف عند العامة فمنهم من يصفه بأنه العلم الذي يهتم بعملية جمع البيانات وحفظها، والبعض الأخر اتجه نحو وصفه بأنه العلم الذي يهتم بعملية جمع ووصف البيانات وتنظيمها ومن ثم تحليلها وإعادة تشكيلها مما يسهل قرأتها ومن ثم تهيئتها لغرض اتخاذ القرار المناسب، أما الآن فإن الإحصاء قد تطور كثيراً خاصة في القرن العشرين حيث اكتسب هذا العلم أهميةً كبيرة مع تطور النظرية الاحتمالية بحيث أصبح علماً مستقلاً وانتشر استخدامه من قبل العلماء في مجالات متعددة كـ(الطب، الهندسة، الزراعة، الاقتصاد، الاجتماع، الجغرافيا، الفلك... الخ)، لذا أصبح له أهميته كوسيلة وأداة في البحث العلمي لجميع العلوم.

2- الهدف من دراسة علم الإحصاء

أن الهدف أو الأهداف الأساسية من دراسة علم الإحصاء

- التعريف بطبيعة علم الإحصاء وأهميته في مجالات تطبيقه .
- التعريف بأساليب وطرق جمع البيانات وسحب العينة وكيفية عرضها لغرض الوصول الى نتائج دقيقة بأقصر الطرق واقل الكلف .
- تخطيط المشاريع سواء كانت مشاريع فردية او مشاريع حكومية بما يسهم في تطوير العمل لتحقيق أفضل النتائج .

3- تعريف علم الإحصاء، مجالاته وأهميته

اولاً : <u>تعريف علم الإحصاء</u>

عرف علم الإحصاء بتعاريف عديدة اختلفت و تباينت من حيث المضمون والشمول باختلاف مراحل تطور هذا العلم والفوائد المتوخاة منه نذكر منها :

- هو العلم الذي يعتمد الطرق العلمية في عملية جمع وتلخيص وتحليل البيانات للوصول إلى نتائج دقيقة ومقبولة تسهم في اتخاذ قرارات سليمة في ضوء هذا التحليل .
- هو الطريقة العلمية التي تختص بجمع البيانات والحقائق بشكل يسهل عملية تحليلها وتفسيرها ومن ثم استخلاص النتائج واتخاذ القرار في ضوء التحليل .
- بأنه العلم الذي يعمل على استخدام الأسلوب العلمي في طرق جمع البيانات وتبويبها وتلخيصها وعرضها و تحليلها بهدف الوصول إلى استنتاجات وقرارات مناسبة.
- هو احد فروع علم الرياضيات الذي يشمل النظريات والطرق الموجهة نحو جمع البيانات ووصفها، وتفسيرها ومن ثم صنع القرارات .

ثانياً: مجالات وأهمية علم الإحصاء وعلاقته بالعلوم الأخرى:

يعتبر علم الإحصاء من العلوم التطبيقية المهمة ووسيلة بحث علمي تسهم في دعم الدراسات والبحوث العلمية في جميع المجالات من خلال استعمال قوانينه وقواعده وأساليب جمع البيانات والمعلومات ومن ثم تحليلها بهدف الوصول إلى النتائج وأخيراً استخدام تلك النتائج في صنع القرار، ان من ابرز المجالات التي يدعمها علم الإحصاء هي :

1. التخطيط : أي وضع الخطط المستقبلية عن طريق التنبؤ بالنتائج، من هذا التعريف يتضح أن التخطيط يقوم على عملية التفكير و التقدير للمستقبل و النظر إلى البعد الزمني و التنبؤ بالمتغيرات و وضع الخطط لما يخفيه المستقبل و التأقلم مع الظروف المتغيرة.
2. الاقتصاد والاجتماع : يبرز دور الإحصاء في بحوث السكان متمثلاً بالتعداد السكاني اذ ان التخطيط السليم للتنمية الاقتصادية والاجتماعية لا ينفصلان ولا يمكن أن يتم إلا من خلال دراسات إحصائية للسكان .
3. الزراعية : يبرز دور الإحصاء بشكل مهم في المجالات الزراعية والتي تبدأ بملاحظة وجمع بيانات عن الطبيعة في الحقل أو المزرعة ثم يلي ذلك الدراسات العملية للتجارب الحقلية بعد ترتيب الملاحظات (المشاهدات) ومن تم تحليلها .
4. الصناعية
5. الطب والصحة العامة .
6. علم النفس .

4- **الطريقة الإحصائية بالبحث العلمي** :

استخدام الأسلوب الإحصائي في البحث العلمي يعني توفير البيانات والمعلومات عن الظاهرة المطلوب دراستها في ذلك البحث وهذا يعني ان أمكانية تطبيق الطريقة الإحصائية مرهوناً بإمكانية التعبير عن هذه الظاهرة أو تلك تعبيرا كميا (رقمياً)، وتمتاز الطريقة الإحصائية بكونها تهيئ أسلوبا موضوعيا محايدا للبحث له قواعده وأصول التي يجب أن يلتزم بها الباحث حتى يتجنب التحيز الشخصي والوقوع في بعض الأخطاء . كما يساعد استخدام الطريقة الإحصائية إلى وصول الباحث إلى النتائج الدقيقة بأقصر طريق وأقل كلفة ، وعليه يمكن ان نحدد المراحل الرئيسية للطريقة الإحصائية :

- ✓ تحديد المشكلة
- ✓ تحديد مجتمع الدراسة
- ✓ جمع البيانات والمعلومات
- ✓ تصنيف البيانات وتبويبها
- ✓ عرض البيانات
- ✓ حساب المؤشرات وتقدير المعالم
- ✓ التفسير والتنبؤ

5- **ينقسم علم الإحصاء الى نوعين هما**

أولاً : الإحصاء الوصفي :

ويتضمن الطرق الإحصائية المستخدمة في جمع البيانات والمعلومات عن ظاهرة معينة أو مجموع من الظواهر وكيفية تنظيم وتصنيف وتبويب البيانات مع أمكانية عرضها في جداول ورسوم بيانية وحساب بعض المؤشرات الإحصائية .

ثانياً الإحصاء الاستدلالي (ألاستنتاجي) :

يستند الإحصاء الاستدلالي على مجموعة من النظريات الإحصائية أهمها نظرية الاحتمالات ونظرية العينات التي تمثلان حلقة الوصل بين الإحصاء الوصفي والاستنتاجي، ونسعى من خلال هذا النوع من الأساليب الإحصائية إلى تقدير معالم وخصائص المجتمعات من خلال ما متوفر من معلومات وعينات مختارة فضلا عن اختبار الفرضيات الإحصائية للمجتمع على أساس البيانات المتاحة .

6- طبيعة البيانات الإحصائية

ان جمع بيانات حول ظاهرة معينة نرمز لتلك الظاهرة بالرمز y او x أما مفردات (المشاهدة) لتلك الظاهرة فنرمز لها بالرمز y_i او x_i، فعند دراسة أطوال الطلبة في جامعة معينة فأننا نرمز لصفة الطول (y) بينما طول أي طالب (المشاهدة) لتلك الصفة نرمز لها بالرمز (y_i)، وعليه يمكن تعريف المتغير على انه الظاهرة التي تظهر الاختلافات بين مفردات (مشاهدات) تلك الظاهرة، وهناك نوعين من المتغيرات هما:

1- المتغيرات الوصفية أو النوعية : وهي الظواهر أو الصفات التي لا يمكن قياسها مباشرةً بالأرقام العددية مثل صفة اللون مثل (لون العين) أو الحالة الاجتماعية (أعزب، متزوج، مطلق، أرمل) أو مستوى المعيشة (غني، متوسط الدخل، فقير) .

2- المتغيرات الكمية :هي الظواهر التي يمكن قياسها مباشرةً بأرقام عددية مثل صفة الطول، الوزن، العمر، كمية المحصول المنتج ...الخ، وهناك نوعن من المتغيرات الكمية، هما :

(المتغيرات المنفصلة أو المتقطعة) و(المتغيرات المتصلة أو المستمرة)

أولا": المتغيرات المنفصلة أو المتقطعة :

هي الظاهرة التي تكون مفرداتها (المشاهدات) تأخذ قيما عددية محددة صحيحة ولا تحتوي علي قيم كسرية مثل عدد موظفي جامعة البصرة ، عدد المصانع في كل مدينة من مدن دولة ما ، الاشخاص، عدد ثمار في شجرة ما .... الخ.

ثانيا": المتغيرات المتصلة أو المستمرة

وهي الظاهرة التي تكون مفرداتها(المشاهدات) تأخذ قيما تكون عددا صحيحا وكسرا من وحدة القياس مثل : بيانات الدخل اليومي (الدولار)لعينة من الاشخاص، الطول(مقاسا بالسنتمترات) ، الوزن(مقاسا بالكيلوغرام) ، العمر(مقاسا بالسنوات)درجة حرارة الجو(مقاسة بالدرجات.) وغيرها هي البيانات الكمية التي.

7- المجتمع والعينة

✓ المجتمع الإحصائي: عبارة عن جميع القيم أو المفردات التي يمكن أن يأخذها المتغير(الظاهرة)، ويكون المجتمع الإحصائي على نوعين هماً مجتمعاً محدوداً أي هو المجتمع الذي يمكن ملاحظة

جميع مفرداته والسيطرة عليها، أو مجتمعاً غير محدود أي هو المجتمع الذي يكون من الصعب ملاحظة و حصر جميع مفرداته .

- ✓ العينة الإحصائية : هي الجزء المختار(المسحوب) من مجتمع إحصائي بشرط أن تكون ممثلة لجميع خصائص ذلك المجتمع، ويتم اختيار العينة بأسلوب او طريقة يطلق عليها (بالمعاينة الإحصائية)، ويعتبر أسلوب العينة هو الأسلوب الأخر في اختيار البيانات عندما يكون من الصعب الحصول على بيانات دقيقة من المجتمع الإحصائي وبالتحديد المجتمع غير المحدود.

8- أسلوب تصميم البحوث

هناك اعتبارات كثيرة يتوقف عليها تصميم البحث العلمي التي يجب ان يأخذها الباحث بنظر الاعتبار وهي مسألة الحصول على البيانات والمعلومات بأقصر وقت واقل جهد واقل كلفة ولكي يتم تصميم البحث بأسلوب علمي صحيح يجب مراعاة ما يلي :

1. تحديد الغرض من البحث العلمي :

بمعنى آخر تحديد الهدف الأساسي الذي يقوم عليه البحث العلمي والذي يجب ان يكون واضحاً، دقيقاً ومحدداً لجميع أوجه الاستفادة التي سيحققها البحث والاستفادة من نتائجه .

2. أمكانية التنفيذ العلمي للبحث :

ان من ضروريات تنفيذ البحث العلمي هو تحديد المتطلبات الضرورية التي تلتزمها عملية تنفيذ البحث كالموارد المالية، والإمكانات البشرية المتاحة لتحقيق بعض فقرات البحث، والتأكد من مدى توافر البيانات والمعلومات الدقيقة عن مشكلة البحث .

3. تحديد إطار البحث :

من الضروري والمهم تحديد إطار البحث من قبل الباحث من حيث تحديد نوع وطبيعة مجال البحث، أو المجتمع الإحصائي أي تحديد الوحدات او مفردات ذلك المجتمع ومعرفة صفته او صفات المشتركة، مع تحديد نوع ذلك المجتمع أي هل هو (مجتمع محدود) او (غير محدود).

4. تحديد الأسلوب المناسب لجمع البيانات الإحصائية وهما أسلوبين (أسلوب الحصر الشامل) و(أسلوب الحصر غير الشامل (أسلوب العينات)) .

9- مصادر جمع البيانات

هناك مصدرين لجمع البيانات هما :

1- المصادر التاريخية : وهي البيانات والمعلومات المسجلة لدى أجهزة ومؤسسات ودوائر الدولة وكذلك شركات القطاع الخاص والمحفوظة لسنوات سابقة نتيجة إجراء عمليات المسح الميداني و الإحصاءات من قبل الجهات الحكومية .

2- المصادر الميدانية : وهي المعلومات والبيانات التي يتم تجميعها ميدانياً من قبل الباحث ومن مصادرها الأصلية معتمداً أما أسلوب المراسلة، المقابلة الشخصية، أو توزيع استمارات الاستبيان.

10- اساليب جمع البيانات

ان عملية جمع البيانات والمعلومات تعتبر خطوة أساسية من خطوات البناء الإحصائي للبحث العلمي، لذا تتوفر اساليب يمكن من خلالها جمع البيانات والمعلومات ولكل من هذه الأساليب مزايا وعيوب، وهذه الأساليب هي :

- أسلوب التسجيل الشامل (الحصر الشامل) :

وهو الأسلوب الذي يلجأ اليه الباحث في حالة المجتمعات الإحصائية المحدودة التي يمكن ملاحظة جميع مفرداته على سبيل المثال (التعداد العام للسكان) ويعتبر هذا الأسلوب هو الأفضل في عملية جمع البيانات لأنه يعطي صورة حقيقة وكاملة حول مفردات الظاهرة المراد دراستها.

أما عيوب هذا الأسلوب

- ✓ يتطلب إمكانيات مادية وبشرية كبيرة.
- ✓ يتطلب وقت وجهد كبير .

- أسلوب الحصر غير الشامل (أسلوب العينة):

وهو الأسلوب الذي يتم اللجوء اليه في حالة المجتمعات الإحصائية غير المحدودة التي لا يمكن حصر وملاحظة جميع مفرداته ويسمى هذا الأسلوب بالعينة بشرط أن تحمل هذه العينة

كل خصائص المجتمع الإحصائي، ويمتاز هذا الأسلوب بأنه لا يحتاج إلى إمكانات مادية وبشرية ووقت وجهد كما هو الحال في الحصر الشامل .

11- المعاينة الإحصائية

تُعد المعاينة الإحصائية إحدى اساليب وطرق اختيار المفردات من المجتمع والتي تؤلف العينة الإحصائية .

<u>مفاهيم اساسية:</u>

- ✓ **العينة** : هي جزء من المجتمع يتم اختيارها بطريقة علمية محددة، ويشترط بالعين المختارة ان تكون ممثلة للمجتمع تمثيلاً جيداً وتحمل كل خواصه بما في ذلك الاختلافات في وحداته .
- ✓ **حجم العينة** : جميع وحدات العينة المؤلفة لها ويرمز لها بالرمز (n) .
- ✓ **وحدة المعاينة**: هي الجزء الصغير الذي تجمع منه البيانات، وهي كل وحدة من وحدات المجتمع وقد تكون متشابه من حيث الحجم أو مختلفة، ويجب تعريف وحدة المعاينة تعريفاً واضحاً عند تنفيذ البحوث الميدانية بحيث يشمل جميع مفردات الظاهرة وعدم تداخلها مع المفردات الأخرى .
- ✓ **المعاينة**: هي عملية اختيار جزء من مفردات المجتمع الإحصائي للاستدلال على خواص المجتمع بأكمله عن طريق تعميم نتائج العينة، وتقوم المعاينة على علم فن التحكم وقياس دقة المعلومات الإحصائية عن طريق استخدام بعض النظريات الرياضية .

(11-1) أنواع المعاينة الإحصائية

تنقسم المعاينة الإحصائية إلى قسمين رئيسين هما :

أولاً : المعاينة العشوائية (الاحتمالية)

ثانياً: المعاينة غير العشوائية (غير الاحتمالية)

أولاً : المعاينة العشوائية (الاحتمالية)

وهي الطريقة التي يتم من خلالها اختيار مفردات العينة بصورة عشوائية بدون تدخل الباحث في اختيارها بحيث تكون لكل مفردة من مفردات العينة نفس الفرصة في الظهور أو الاختيار . ومن أنواع المعاينة العشوائية:

1– المعاينة العشوائية البسيطة :

هي طريقة اختيار عينة مكونة من (n) وحدة من بين وحدات المجتمع (N) بحيث يكون لكل عينة من العينات الممكن اختيارها فرصة متساوية (احتمال متساوٍ) في الظهور، أي أن احتمال سحب أي وحدة يكون متساوياً عند اختيار كل وحدة من وحدات العينة، ولتحديد عدد العينات العشوائية البسيطة الممكن سحبها من مجتمع إحصائي هناك طريقتين هما :

- <u>السحب بدون إرجاع</u> : ويقصد بها أن المفردة المسحوبة لا يتم اختيارها مرة أخرى، وفي هذه الطريقة نضمن أن كل مفردة تمتلك نفس الاحتمال بالاختيار، إضافة إلى ذلك أن تمتلك كل مفردة نفس عدد مرات الظهور .

فإذا كان لدينا مجتمع إحصائي مؤلف من (N) المفردات وتم سحب عينة عشوائية بسيطة بحجم (n)، فإن عدد العينات (m) الممكن سحبها من ذلك المجتمع تحسب على وفق الصيغة التالية :

$$\mathbf{m} = \binom{N}{n} \Rightarrow \boldsymbol{M} = \frac{N!}{n!(N-n)!} \qquad \ldots(1)$$

<u>مثال</u> : مجتمع إحصائي مؤلف من (5) مفردات سحبت منه عينة عشوائية بسيطة بحجم (2)، فما هو عدد العينات الممكن سحبها .

$$\mathbf{m} = \binom{5}{2} \Rightarrow \boldsymbol{m} = \frac{5!}{2!(5-2)!} \Rightarrow \boldsymbol{m} = \frac{5!}{2!3!} \Rightarrow \boldsymbol{m} = \mathbf{10}$$

<u>المعاينة العشوائية الطبقية Stratified Random Sampling :</u>

ان دقة تقدير معالم المجتمع تتوقف على حجم العينة كما تتوقف على عدم تجانس المجتمع، ويمكن وضع بعض القيود على المعاينة العشوائية البسيطة للتقليل من تأثير مشكلة عدم التجانس، وابسط هذه القيود هي تقسيم المجتمع إلى عدد من الطبقات، وهذا النوع من العينات يدعى بالمعاينة العشوائية الطبقية لذا يمكن تعريفها (وهي إحدى أنواع المعاينة العشوائية الاحتمالية التي تستخدم

في المجتمعات الإحصائية الكبيرة وغير المتجانسة حيث يتم تقسيم المجتمع الذي حجمه N إلى عدد من الطبقات(h) المتجانسة التي أحجامها $(N_1, N_2, N_3, \ldots, N_h)$ ، ومن ثم سحب عينة عشوائية بسيطة من كل طبقة)) .

وعليه :

$$N = N_1 + N_2 + \ldots + N_h$$

$$n = n_1 + n_2 + \ldots + n_h$$

حيث ان

N : حجم المجتمع $N_1, N_2, .. N_h$: حجم الطبقات

h: عدد الطبقات

n : حجم العينة العشوائية $n_1, n_2, \ldots, n_h$: حجم العينة العشوائية المسحوبة من كل طبقة

ويمكن الحصول على حجم العينة العشوائية الطبقية من العلاقة التالية:

$$\frac{n_h}{n} = \frac{N_h}{N} \quad \ldots(2)$$

مثال(1) : في إحدى كليات الإدارة الاقتصاد تم إجراء استطلاع للرأي بمستوى الخدمات التنقية والتكنولوجية المطبقة في الكلية تطلب إجراء هذا النوع من الاستطلاع استخدام أسلوب العينات الاحتمالية، فإذا علمت ان حجم العينة هو (100) طالب وطالبة وان عدد الطلبة في الأقسام العلمية كان كما يلي :

الأقسام العلمية	الإدارة	المحاسبة	الاقتصاد	الإحصاء	العلوم المالية	نظم المعلومات
عدد الطلبة	500	600	300	100	400	100

حدد الحجم العينة المناسب الممكن اختياره من كل قسم

الحل :

$$N = 500+600+300+100+400+100 = 2000$$

ويتم حساب حجم العينة لكل قسم من خلال العلاقة

$$n_h = n * \frac{N_h}{N}$$

- حجم العينة لقسم الادارة

$$n_1 = 100 * \frac{500}{2000} = 25$$

- حجم العينة لقسم المحاسبة

$$n_2 = 100 * \frac{600}{2000} = 30$$

- حجم العينة لقسم الاقتصاد

$$n_3 = 100 * \frac{300}{2000} = 15$$

- حجم العينة لقسم الاحصاء

$$n_4 = 100 * \frac{100}{2000} = 5$$

- حجم العينة لقسم العلوم المالية

$$n_5 = 100 * \frac{400}{2000} = 20$$

- حجم العينة لقسم نظم المعلومات

$$n_6 = 100 * \frac{100}{2000} = 5$$

ولو جمعنا احجام العينات التي حصلنا عليها

= 25+30+15+5+20+5 = 100

مثال(2): مجتمع إحصائي مقسم إلى (3) طبقات تم الحصول على البيانات التالية :

حجم الطبقة N_h	*A*	*750*	*C*
الجزء المختار n_h	*a*	70	50

<u>حدد قيم حجوم الطبقات والأجزاء المختارة المفقودة في الجدول إذا علمت إن حجم المجتمع N=12000 .</u>

مثـــال(3): مجتمع إحصائي غير متجانس تم تقسيمه إلى (4) طبقات حجم كل طبقة مرتبة على التوالي(1000، 3500، 4000، 1500) ويراد سحب عينة عشوائية بحجم (400) مفردة ، حدد حجوم الأجزاء المختارة من كل طبقة بحيث تشكل العينة تمثيلاً سليماً .

مثال(4): مجتمع إحصائي مقسم إلى(3) طبقات تم سحب عينة بحجم (300) وتم الحصول على البيانات التالية :

حجم الطبقة N_h	*800*	*Y*	*Z*
الجزء المختار n_h	*x*	*Y*	60

حدد قيم حجوم الطبقات والأجزاء المختارة المفقودة في الجدول اذا علمت ان حجم المجتمع N=5000 .

2– المعاينة العشوائية المنتظمة (الأسلوبية) Systemic Random Sampling

هي طريقة اختيار عينة منتظمة وذلك عن طريق اختيار وحدة المعاينة (M) والتي تسمى أيضاً بنسبة المعاينة، ويطلق على العينة المنتظمة بالعينة ذات الفترات المتساوية وذلك لتساوي الفترات بين مفرداتها، وتعد العينة الأسلوبية شائعة الاستعمال لسهولتها وبساطتها ولكن طريقة اختيار مفرداتها بهذه الصورة قد تعرض العينة لبعض التحيز .

إذا كان مجتمع إحصائي حجمه N سحبت منه عينة عشوائية بسيطة بحجم n فإن وحدة المعاينة M تحسب $M = \frac{N}{n}$، حيث يتم اختيار رقم عشوائي بين (1 و M) لتشكل رقم المفردة الأولى (K_1) ثم يتم إضافة M ومضاعفاتها على المفردة الأولى إلى أن يكتمل حجم العينة .

ويمكن حساب مفردات العينة المنتظمة باعتماد الصيغة التالية :

$$K_i = K_1 + (i-1)M$$

مثال(5)/ مدير مدرسة يريد أن يعرف رأي طلبة مدرسته في مستوى تدريس أحد المعلمين . فإذا كان عدد طلبة المدرسة 200 طالب، كيف يتم سحب عينة منتظمة بحجم 10 من هذا المجتمع.

الحل: من اجل سحب عينة منتظمة نتبع الخطوات التالية :

1-ترقيم عناصر المجتمع من 1 إلى N (اي من 1 الى 200).

2-تحديد حجمِ العيّنةَ n التي نحتاجها بالدراسة وهي n=10.

3-نجد قيمة الزيادة المنتظمة k اذ ان : M = N/n

4-يتم اختيارُ اول مفردة باختيار اي عدد صحيحُ بشكل عشوائي بين 1 إلى M .

5-ثمّ نأخذ العناصر الاخرى عن طريق الزيادة بانتظام بمقدار M الى ان يصبح عدد الارقام المختارة يساوي n اي 10 . وكالتالي:

بما ان حجم المجتمع (N=200) يتم ترقيم عناصر المجتمع من 1 الى 200 وحسب ترميز معين او خاصية محددة ثم نحدد حجم العينة ولتكن (n=10)، نحسب: $M = \frac{N}{n} = \frac{200}{10} = 20$ ، ثم نختار رقم من 1 الى 20 بصورة عشوائية وليكن K_1 =12 عندها نختار العناصر المرقمة التالية

$$K_2 = K_1 + (2-1)M \qquad = 12+(2-1)*20 = 32$$
$$K_i = K_1 + (i-1)M \qquad = 12+(3-1)*20 = 52$$

وهكذا الى ان نصل الى 10 ارقام :

12,32,52,72,92,112,132,152,172,192

ثم نختار العناصر في المجتمع التي تحمل هذه الارقام وعددها 10.

مثال(6) : تم اختيار عينة أسلوبية من مجتمع إحصائي حجمها (30) مفردة حيث تم الحصول على المفردات $(K_{10} = 200, K_{20} = 460)$، المطلوب جد حجم المجتمع والمفردة الأخيرة في العينة .

مثال(7) : تم سحب عينة عشوائية بسيطة بحجم (60) مفردة من مجتمع إحصائي متجانس فإذا علمت أن المفردة $K_{25} = 520$ و المفردة $K_{35} = 600$ ، المطلوب جد حجم المجتمع والمفردتين الأخيرتين في العينة، مع تحديد المفردات K_{55}, K_{48}, K_{30} .

3- المعاينة متعددة المراحل :

وهو أسلوب يهدف الى تقسيم المجتمع الى وحدات اولية ثم يتم اختيار عينة عشوائية من هذه الوحدات ومن ثم يتم تقسيم كل وحدة من الوحدات الأولية الى وحدات ثانوية ثم تؤخذ عينة عشوائية منها ثم يتم تقسيم كل وحدة من الوحدات الثانوية الى وحدات اصغر وتسحب من هذه الوحدات عينة عشوائية وهكذا نستمر الى ان نصل الى الجزء الذي سيتم جمع البيانات منه لتشكل عينة البحث .

ثانياً: المعاينة غير العشوائية (غير الاحتمالية) :

ومن اهمها:

1- المعاينة الحصصية : يتم تقسيم المجتمع إلى طبقات استناداً إلى معايير تقسيم معينة تتعلق بطبيعة الدراسة ثم يتم اختيار عدد من المفردات من كل طبقة وبشكل شخصي (غير عشوائي) بحيث أن عدد مفردات هذه العينات يمثل حجم العينة المطلوبة .
2- المعاينة العمدية : وهو الأسلوب الذي يتم فيه اختيار مفردات العينة بشكل متعمد من قبل الباحث معتقداً ان هذه العينة تمثل المجتمع خير تمثيل .

(12) تصنيف و تبويب البيانات

بعد التعرف على مصادر جمع البيانات وأساليب جمعها فأن البيانات التي تم الحصول عليها تدعى (بالبيانات الأولية أو البيانات غير المصنفة، إذ أن البيانات بشكلها الأولي تكون غير منظمة مما يصعب التعامل معها واستخدامها لأغراض التحليل الإحصائي، ولغرض تحقيق الفائدة من هذه البيانات لا بد من التعريف بعملية تصنيف وجمع البيانات والتي تتم من خلال المراحل التالية :

1- مراجعة البيانات : بعد أتمام عملية جمع البيانات من مصادرها الأصلية وفق الأسلوب المناسب يجب على الباحث مراجعة وتدقيق البيانات للتأكد من مطابقتها وتكاملها .

2- تصنيف البيانات : تتحقق هذه المرحلة بعد أتمام عملية المراجعة والتدقيق حيث تتم عملية التصنيف على أساس الظواهر التي جمعت منها البيانات ، فقد يكون التصنيف على أساس ظاهرة (العمر، الوزن، التحصيل الدراسي، المهنة، ...الخ) .

3- تبويب البيانات : تتم هذه العملية بعد أتمام عملية تصنيف البيانات ويقصد بها تفريغ البيانات المصنفة في جداول إحصائية خاصة بحيث كل جزء من هذه البيانات المصنفة يعبر عن ظاهرة معينة والهدف من تبويب البيانات هو ابراز البيانات وتوضيحها في أضيق حيز ممكن، ويكون تبويب البيانات على انواع هي :

أ– التبويب الزمني : تجميع البيانات المصنفة وترتيبها في جداول إحصائية على أساس وحدات زمنية (يوم، شهر، سنة، ...الخ) مثال ذلك جدولة أعداد الطلبة الخريجين لكلية الإدارة والاقتصاد للسنوات من (2015 –2019)

السنوات	عدد الطلبة الخريجين
2015	1500
2016	2000
2017	1600
2018	1400
2019	2000

ب– التبويب الجغرافي : تجميع البيانات المصنفة وجدولتها بالاعتماد على الوحدة الجغرافية (المكانية) او تقسيم أداري معين (قارات، بلدان، محافظات، أقضية، النواحي)، مثال تحديد أعداد النخيل في المحافظات الجنوبية والوسطى (البصرة، ذي قار، ميسان، كربلاء، النجف)

المحافظات	أعداد النخيل
البصرة	*10000*
ذي قار	*6500*
ميسان	*3000*
كربلاء	*4750*
النجف	*6570*

ت- التبويب الكمي : تجميع البيانات حسب صفة يمكن قياسها (الوزن، العمر، الطول، المساحة، الحجم، قيمة الأجرة اليومية للعاملين...الخ)

ث- التبويب النوعي (على أساس صفة معينة) :ويتم فيه تجميع البيانات المصنفة في جداول باعتماد صفة لا يمكن قياسها مثل (لون العين، الجنس، الحالة الاجتماعية ، المستوى العلمي، العنوان الوظيفي.. الخ) .

(13) عرض البيانات الإحصائية

هناك أسلوبين لعرض البيانات الإحصائية هما :

1- العرض الجدولي : يتم في هذا النوع تلخيص البيانات وعرضها في جداول إحصائية أما بسيطة أو مركبة

- الجداول البسيطة : يعبر عن الجداول الإحصائية بالجداول البسيطة إذا كانت تحتوي على صفة واحدة فقط .
- الجداول المركبة : وهي الجداول التي تحتوي على صفتين أو أكثر .

2- العرض البياني : في هذا النوع تعرض البيانات الإحصائية بصورة أشكال هندسية أو رسوم تصويرية أو خرائط إحصائية ومن طرق العرض البياني هي :

- المستطيل البياني
- الأشرطة البيانية
- الخط البياني
- الدائرة البيانية

- المستطيل البياني :

ويستعمل المستطيل البياني لتمثيل ظاهرة معينة أو عدت ظواهر من خلال المقارنة فيما بينها، وتتلخص فكرة المستطيل البياني في رسم قاعدة للمستطيل ومن ثم تقسيمها الى مستطيلات جزئية تمثل بيانات الظاهرة، ولغرض حساب مساحة أي ظاهرة نعتمد الصيغة التالية :

$$Y_i = A * \frac{x_i}{\sum x_i}$$

حيث أن :

Y_i : تمثل مساحة الظاهرة

A : طول قاعدة المستطيل الكلية

x_i: بيانات الظاهرة

$\sum x_i$: المجموع الكلي للظواهر

أو :

طول قاعدة المستطيل الجزئي =(البيانات الجزئية/ البيانات الكلية) * طول قاعدة المستطيل الكلي

مثال: بلغت كلفة إنتاج النفط الأبيض في إحدى المصافي (300) دولار للبرميل الواحد، وكما موضحة بالجدول :

مستلزمات الإنتاج	كلفة الإنتاج
أجور العاملين	130
مواد خام	40
مصاريف مباشرة	75
مصاريف غير مباشرة	55

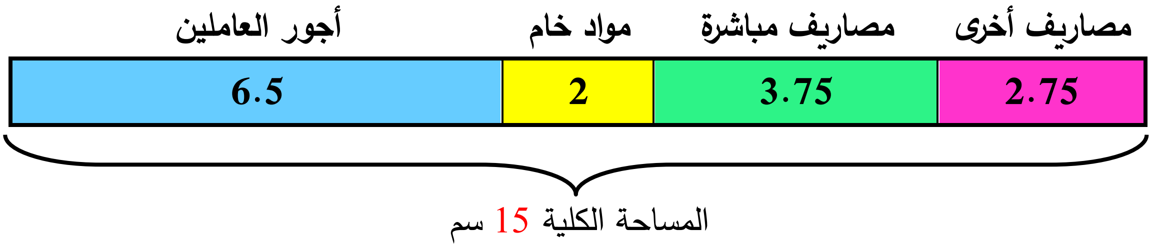
فإذا فرضنا ان طول قادة المستطيل هي (15)سم، مثل البيانات التالية باستخدام الرسم البياني

الحل

$$\text{اجور العاملين} = 15 * \frac{130}{300} = 6.5$$

وهكذا يتم حساب مساحة مستلزمات الإنتاج المتبقية

مستلزمات الإنتاج	المساحة
أجور العاملين	6.5
مواد خام	2
مصاريف مباشرة	3.75
مصاريف أخرى	2.75
مجموع المساحة الكلية	15



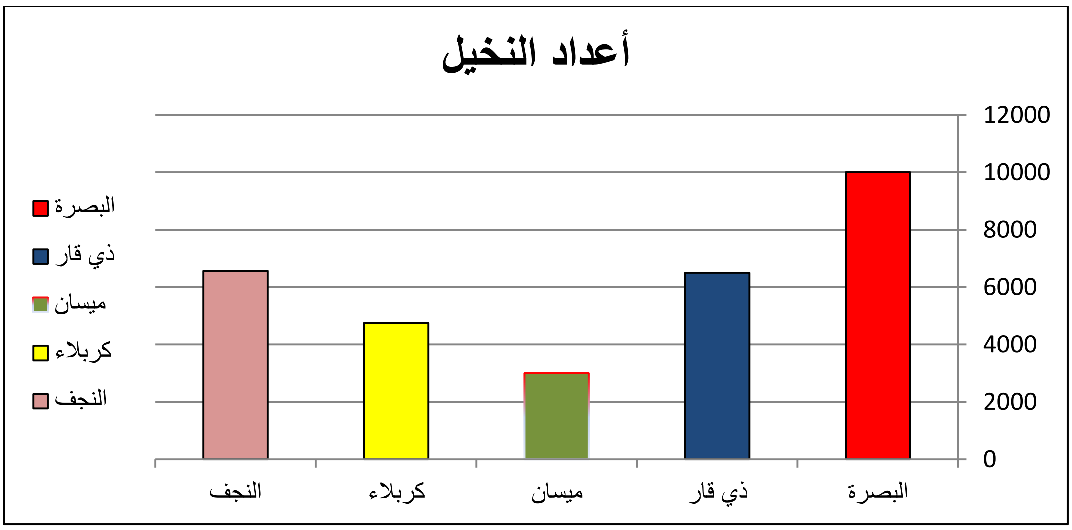
- <u>الأشرطة البيانية</u>

هي مجموعة من المستطيلات قواعدها تكون متساوية وارتفاعاتها مختلفة بحسب طبيعة البيانات ولغرض توضيح ذلك نأخذ المثال التالي :

مثال : في دراسة لتحديد أعداد النخيل في المحافظات الجنوبية والوسطى (البصرة، ذي قار، ميسان، كربلاء، النجف) تم الحصول على البيانات التالية :

المحافظات	أعداد النخيل
البصرة	10000
ذي قار	6500
ميسان	3000
كربلاء	4750
النجف	6570

<u>المطلوب:</u> استخدم الرسم بالأشرطة البيانية لتوضيح أعداد النخيل حسب المحافظات



- **الدائرة الإحصائية**

هي أحد أشكال الرسوم البيانية التي تتناسب فيها زوايا الأشكال مع القيم المراد التعبير عنها)، ويسهل فهمه بنظرة سريعة إذ تعبر أجزاءه عن قيمة الأجزاء بالنسبة للكل (أي نسبتها المئوية) بشكل بسيط، يتم استخدام المخططات الدائرية في توضيح نتائج استطلاعات الرأي والإحصائيات والبيانات المعقدة ونسب الدخل أو الإنفاق. ويتم تقسيم الدائرة البيانية بتقسيم الزاوية المركزية لها والتي تساوي (360°)، كما في الصيغة التالية :

$$Y_i = 360^\circ * \frac{x_i}{\sum x_i}$$

حيث أن :

Y_i : زاوية القطاع

x_i: بيانات الظاهرة

$\sum x_i$: المجموع الكلي للظواهر

أو : زاوية القطاع = (البيانات الجزئية/ البيانات الكلية) * 360

ويمكن توضح ذلك بالرجوع إلى بيانات المثال السابق الخاص بإعداد النخيل حسب المحافظات

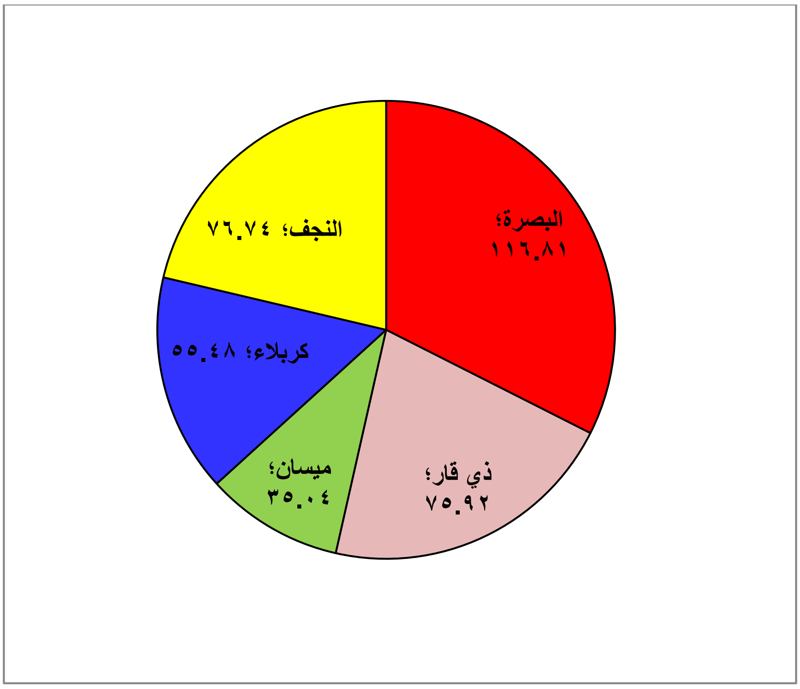
المحافظات	أعداد النخيل
البصرة	10000
ذي قار	6500
ميسان	3000
كربلاء	4750
النجف	6570
المجموع الكلي للنخيل	30820

الحل

$$\text{البصرة} = 360^\circ * \frac{10000}{30820}$$
$$= 116.81$$

وهكذا يتم حساب المساحة لجميع المحافظات:

المحافظات	أعداد النخيل
البصرة	116.81
ذي قار	75.92
ميسان	35.04
كربلاء	55.48
النجف	76.74



(14) العرض الجدولي

(14–1) التوزيع التكراري :

تلخص البيانات الإحصائية سواء كانت وصفية او كمية في جداول تسمى (بالتوزيعات التكرارية Frequency Distribution وهي عبارة عن جداول يلخص البيانات الخام (الأصلية) فيوزعها على عمودين الأول يمثل الفئات والثاني يمثل تكرارات تلك الفئات ويرمز لهذا العمود (f) .

ولغرض تكون جدول التوزيع التكراري يتطلب ذلك تكوين جدول أولي لتفريغ البيانات الأصلية وهو يتكون من ثلاث أعمدة الأول يمثل (الفئات) أو ما تسمى بصفات الظاهرة لتضمين صفة البيانات كمية كانت ام وصفية، العمود الثاني يمثل عمود العلامات ويثمل بحزمة من خمسة خطوط (𝍸) أما العمود الثالث فيمثل مجموع العلامات المتحققة في العمود الثاني .

بعض التعاريف والمصطلحات :

(1)البيانات غير المبوبة: وهي البيانات الأولية أو الأصلية التي تم جمعها من مصادرها ولكن لم يتم تبويبها .

(2)البيانات المبوبة : وهي البيانات التي رتبت ونضمت في جداول إحصائية تسمى (جداول التوزيع التكراري) .

(3)الفئات : وهي عدد من المجاميع التي تمثل صفة (ظاهرة معينة) والتي قسمت إليها قيم المتغير بحيث كل فئة تأخذ مدى معين من قيم المتغير .

أولاً : تكوين جدول توزيع تكراري لبيانات ذات صفة معينة أو بيانات رتبية

لتكوين جدول توزيع تكراري لبيانات تحمل صفة معينة (تقديرات الطلبة، لون العين، الحالة الاجتماعية، ...الخ) لا نحتاج في هذا النوع من البيانات أي قوانين رياضية وانما نحتاج فقط الى ترتيب هذه البيانات (تصاعدي او تنازلي) أو ترتيبها وفق اليه يحددها الباحث .

مثال/ إذا كانت لدينا بيانات تمثل تقديرات (60) طالب معرفة بالشكل التالي :

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

حيث تم توصيف التقديرات بالشكل التالي :

التقدير	A	B	C	D	E
وصفه	امتياز	جيد جداً	جيد	متوسط	مقبول

م// كون جدول توزيع تكراري لتقديرات الطلبة .

الفئات	العلامات	F (التكرارات)
A	𝍸 |	6
B	𝍸 |||	8
C	𝍸 𝍸 𝍸 |	16
D	𝍸 𝍸 𝍸 𝍸 ||	22
E	𝍸 |||	8
مجموع التكرارات		60

وعليه يكون جدول التوزيع التكراري

الفئات تقديرات الطلبة	التكرارات (f)
A	6
B	8
C	16
D	22
E	8

ثانياً : تكوين جدول التوزيع التكراري للبيانات الكمية

لتكوين جداول توزيع تكراريه تحمل فئاتها الصفة الكمية، يجب توضح بعض المفاهيم التالية :

الفئة : وهي عبارة عن مجاميع قسمت اليها قيم المتغير ضمن مدى معين، بحيث يكون لهذه الفئة حد ادني وحد أعلى .

المدى : وهو الفرق بين اكبر قيمة واقل قيمة بين مفردات الظاهرة، ويرمز له بالرمز R .

$$R=(X_{MAX}-X_{MIN})+1$$

$$\text{Or} \quad R=(X_{MAX}-X_{MIN})$$

مركز الفئة: وهي القيمة الواقعة في منتصف الفئة، وتحسب كما يلي :

$$X=\frac{L_c+U_c}{2}$$

L_c : الحد الأدنى للفئة

U_c : الحد الأعلى للفئة

<u>عدد الفئات</u>: هي عدد من المجاميع التي قسمت إليها قيم المتغير بحيث تحتوي هذه المجاميع جميع حالات الظاهرة ويكون عدد الفئات محصورة بين (5−20) حسب طبيعة البيانات ويجب أن يكون عدد الفئات اقل من طول الفئة ويرمز لعدد الفئات بالرمز K، ويمكن تقدير عدد الفئات بشكل تقريبي باعتماد إحدى الصيغتين :

- طريقة سترجس

$$K = 1 + 3.3log(n)$$

- طريقة يول

$$K = 2.5 * \sqrt[4]{n}$$

حيث ان : (n) تمثل عدد مفردات الظاهرة .

<u>طول الفئة :</u> هو الفرق بين فئتين متتاليتين(الحدين الاعلى او الحدين الادنى) أو الفرق بين مركزي فئتين متتاليتين، ويرمز لها بالرمز L،

كما يمكن حسابه على وفق الصيغة التالية :

$$L = \frac{R}{K}$$

الحدود الحقيقية للفئات :

- الحد الحقيقي الأدنى = مركز الفئة − $\frac{1}{2}$ (طول الفئة) .
- الحد الحقيقي الأعلى = مركز الفئة + $\frac{1}{2}$ (طول الفئة) .

<u>مثال</u> : كون جدول توزيع تكراري للأعداد التالية التي تمثل أوزان عدد من الطلبة في المعهد التقني في المنطقة الوسطى :

66	73	80	60	56	74
65	63	89	77	55	56
57	60	77	60	67	50
80	64	70	59	62	70
69	70	72	74	61	57

الحل :

نرتب البيانات ترتيباً تصاعديا أو تنازلياً او بطريقة أخرى نكتب البيانات في جدول مقسم كما في أعلاه ومن ثم نستخرج اقل وأكبر قيمة في كل صف وبعد ذلك يتم اختيار اقل القيم من عمود اقل قيمة واكبر القيم من عمود اكبر قيمة وكما يلي :

اقل قيمة	أوزان الطلبة/بالكيلو غرام						اكبر قيمة
56	66	73	80	60	56	74	80
55	65	63	89	77	55	56	89
50	57	60	77	60	67	50	77
59	80	64	70	59	62	70	80
57	69	70	72	74	61	57	74

أكبر القيم

أقل القيم

(1)حساب المدى

$R=(X_{MAX}-X_{MIN})+1$

X_{MAX}: اكبر قيمة = 89 X_{MIN} : اقل قيمة = 50

$R= (89-50)+1$

$R=40$

(2)عدد الفئات (K)

- استخدام طريقة سترجس

$$K = 1 + 3.3log(n)$$
$$K = 1 + 3.3log(30)$$
$$K = 1 + 3.3(1.48)$$
$$K = 5.88 \approx 6$$

- طريقة يول

$$K = 2.5 * \sqrt[4]{n}$$
$$K = 2.5 * \sqrt[4]{30}$$
$$K = 2.5 * (2.34)$$
$$K = 5.85 \approx 6$$

(3)طول الفئة

$$L = \frac{R}{K}$$
$$L = \frac{40}{6}$$

$L = 6.67 \approx 7$

(4)تكوين جدول العلامات

التكرارات	العلامات	الفئات أوزان الطلبة/بالكيلوغرام
4	$\vert\vert\vert\vert$	50 – 56
9	$\cancel{\vert\vert\vert\vert}\ \vert\vert\vert\vert$	57 – 63
8	$\cancel{\vert\vert\vert\vert}\ \vert\vert\vert$	64 – 70
6	$\cancel{\vert\vert\vert\vert}\ \vert$	71 – 77
2	$\vert\vert$	78 – 84
1	$\vert$	85 – 91

ملاحظة / يمكن تحديد العلامات من خلال متابعة القيمة في الجدول الأصلي وتحديد موقعها في أي فئة وتأشيرها لضمان عدم الرجوع اليها .

~~66~~	~~73~~	~~80~~	~~60~~	~~56~~	~~74~~
~~65~~	~~63~~	~~89~~	~~77~~	~~55~~	~~56~~
~~57~~	~~60~~	~~77~~	~~60~~	~~67~~	~~50~~
~~80~~	~~64~~	~~70~~	~~59~~	~~62~~	~~70~~
~~69~~	~~70~~	~~72~~	~~74~~	~~61~~	~~57~~

(5) جدول التوزيع التكراري

التكرارات	الفئات أوزان الطلبة/بالكيلوغرام
4	50 – 56
9	57 – 63
8	64 – 70
6	71 – 77
2	78 – 84
1	85 – 91
30	المجموع الكلي للتكرارات

واجب / إذا كانت لديك البيانات التالية التي تمثل أطوال عدد من نباتات الظل في احد المشاتل مقاسه بالسنتمتر وهي كما يلي :

10	7	20	60	30
35	40	25	15	10
50	31	21	55	45
39	17	22	18	52
38	14	42	51	47

ويمكن تكوين جدولين من جدول التوزيع التكراري هما :

1- جدول التوزيع التكراري النسبي **Relative Frequency Table**

2- جدول التوزيع التكراري المئوي **Percentage Frequency Table**

(1)جدول التوزيع التكراري النسبي :

ويتكون هذا الجدول من خانتين مثل جدول التوزيع التكراري هما خانة الفئات أما خانة التكرارات (f_i) تسمى بأسم (خانة التكرار النسبي) ويرمز لها بالرمز (rf_i)، ويمكن حساب التكرار النسبي لكل فئة من خلال الصيغة التالية :

$$rf_i = \frac{f_i}{\sum_{i=1}^{k} f_i}$$

حيث ان :

f_i : التكرار الفعلي للكل فئة

$\sum_{i=1}^{k} f_i$: المجموع الكلي لتكرارات الفئات ، k: تمثل عدد الفئات في جدول التوزيع التكراري

rf_i : التكرار النسبي لكل فئة

$\sum_{i=1}^{k} rf_i$: المجموع الكلي للتكرار النسبي للفئات

بحيث يجب أن يكون مجموع التكرارات النسبية مساوية إلى الواحد الصحيح، أي أن $\sum_{i=1}^{k} rf_i = 1$

<u>مثال</u> / إذا كان لديك جدول التوزيع التكراري الذي يمثل اوزان طلبة المعهد التقني/ الجامعة التقنية الوسطى والمبينة في أدناه :

الفئات أوزان الطلبة/بالكيلوغرام	التكرارات f_i
50 – 56	4
57 – 63	9
64 – 70	8
71 – 77	6
78 – 84	2
85 – 91	1

المطلوب // كون جدول توزيع تكراري نسبي

الحل /

نحسب مجموع التكرارات الكلي للفئات $\sum f_i = 30$

ثم نطبق صيغة حساب التكرار النسبي لكل فئة $rf_i = \frac{f_i}{\sum_{i=1}^{k} f_i}$

نحسب التكرار النسبي للفئة الأولى $rf_1 = \frac{4}{30} \Rightarrow rf_1 = 0.13$

الفئات أوزان الطلبة/بالكيلوغرام	التكرارات f_i	rf_i
50 – 56	4	0.13
57 – 63	9	0.30
64 – 70	8	0.27
71 – 77	6	0.20
78 – 84	2	0.07
85 – 91	1	0.03

وعليه يكون جدول التكرار النسبي

الفئات أوزان الطلبة/بالكيلوغرام	rf_i
50 – 56	0.13
57 – 63	0.30
64 – 70	0.27
71 – 77	0.20
78 – 84	0.07
85 – 91	0.03
$\sum_{i=1}^{k} rf_i$	1

(2) جدول التوزيع التكراري المئوي :

ويمكن الحصول عليه من خلال جدول التوزيع التكراري النسبي وذلك بضرب عمود التكرار النسبي لكل فئة في (100)، ويعبر عنها رياضياً

$$pf_i = rf_i * 100$$

حيث أن :

pf_i: هو التكرار المئوي لكل فئة

ومن المثال السابق تم الحصول على جدول التوزيع التكراري النسبي لاوزان الطلبة/ بالكيلوغرام وكما مبين في ادناه :

الفئات أوزان الطلبة/بالكيلوغرام	rf_i
50 – 56	0.13
57 – 63	0.30
64 – 70	0.27
71 – 77	0.20
78 – 84	0.07
85 – 91	0.03

المطلوب // حساب جدول التكرار المئوي

الحل//

نسحب التكرار المئوي للفئة الأولى (56 – 50) :

$$pf_1 = rf_1 * 100$$
$$pf_1 = 0.13 * 100$$
$$pf_1 = 13$$

الفئات أوزان الطلبة/بالكيلوغرام	التكرار النسبي rf_i	التكرار المئوي pf_i
50 – 56	0.13	13
57 – 63	0.30	30
64 – 70	0.27	27
71 – 77	0.20	20
78 – 84	0.07	7
85 – 91	0.03	3

<u>ملاحظة</u> // يجب أن يكون المجموع الكلي للتكرارات المئوية مساوية (100)

وعليه يكون جدول التوزيع التكرار المئوي :

الفئات أوزان الطلبة/بالكيلوغرام	التكرار المئوي pf_i
50 – 56	13
57 – 63	30
64 – 70	27
71 – 77	20
78 – 84	7
85 – 91	3
$\sum pf_i$	100

مثال : الجدول التالي يمثل حجم المبيعات اليومية في إحدى المحلات التجارية بالدولار خلال فترة ثلاث أشهر، ولغرض الوقوف بشكل دقيق حول حجم المبيعات، كون ما يلي :

1- جدول توزيع تكراري لتحديد حجم المبيعات .

2- جدول توزيع تكراري نسبي .

3- جدول توزيع تكراري مئوي .

26.9	77.8	8.14	9.33	46.3	71.5	46	7.9	64.5	87.8
33.7	27	71.8	33.8	59.1	28.8	80.2	81.3	43.7	41.6
7.1	64	58.3	37.9	72.2	79.9	54.3	89.7	61.3	21.5
70.1	67	15.5	2.2	28.6	31.7	4.9	26.8	99.2	0.3
10.5	28.3	99.3	14.1	47	98.1	20.3	44.2	21.7	66.3
3.2	6.2	67	39.6	39.2	62.3	21.9	72.2	67.5	64
93.7	51	40	20.2	73.9	18.2	25.5	24.5	67.3	16.4
25.4	35.9	26.5	83.1	88	37.8	19.1	80	45.4	73.8
73.9	63.8	54.2	23.2	53	52.3	63.6	46.2	24.9	55.6

الحل//

نلاحظ حجم المبيعات (القيم) هي قيم ذات مراتب عشرية وهنا لا يجوز التقريب لأنه سيؤثر على البيانات وطبيعة الظاهرة المدروسة، وعليه سنوضح خطوات تكوين جدول توزيع تكراري لهذه البيانات:

(1)حساب المدى R :

أقل القيم

أكبر القيم

اصغر قيمة في كل صف	حجم المبيعات اليومية/ بالدولار										أكبر قيمة في كل صف
7.9	26.9	77.8	8.14	9.33	46.3	71.5	46	7.9	64.5	87.8	87.8
27	33.7	27	71.8	33.8	59.1	28.8	80.2	81.3	43.7	41.6	81.3
7.1	7.1	64	58.3	37.9	72.2	79.9	54.3	89.7	61.3	21.5	89.7
0.3	70.1	67	15.5	2.2	28.6	31.7	4.9	26.8	99.2	0.3	99.2
10.5	10.5	28.3	99.3	14.1	47	98.1	20.3	44.2	21.7	66.3	99.3
3.2	3.2	6.2	67	39.6	39.2	62.3	21.9	72.2	67.5	64	72.2
16.4	93.7	51	40	20.2	73.9	18.2	25.5	24.5	67.3	16.4	93.7
19.1	25.4	35.9	26.5	83.1	88	37.8	19.1	80	45.4	73.8	88
23.2	73.9	63.8	54.2	23.2	53	52.3	63.6	46.2	24.9	55.6	73.9

$R=(X_{MAX}-X_{MIN})+1$

X_{MAX}: اكبر قيمة = 99.3 X_{MIN} : اقل قيمة = 0.3

$R= (99.3 -0.3)+1$

$R=100$

(2)عدد الفئات (K)

- استخدام طريقة سترجس

$$K = 1 + 3.3log(n)$$
$$K = 1 + 3.3log(90)$$
$$K = 1 + 3.3(1.95)$$
$$K = 7.4 \approx 7$$

- طريقة يول

$$K = 2.5 * \sqrt[4]{n}$$
$$K = 2.5 * \sqrt[4]{90}$$
$$K = 2.5 * (3.08)$$
$$K = 7.7 \approx 8$$

ونظرا لوجود اختلاف في عدد الفئات بين الطريقتين نتجه نحو افتراض إن عدد الفئات مساوية إلى (K=5) فكلما كان عدد الفئات قليل كان هو الأفضل .

(3)طول الفئة

$$L = \frac{R}{K}$$
$$L = \frac{100}{5} = 20$$

(4)تكوين جدول العلامات

التكرارات	العلامات	حدود الفئات	الفئات حجم المبيعات/بالدولار
14		$0 \le C_1 < 20$	**0 – 20**
26		$20 \le C_2 < 40$	**20 – 40**
17		$40 \le C_3 < 60$	**40 – 60**
22		$60 \le C_4 < 80$	**60 – 80**
11		$80 \le C_5 < 100$	**80 – 100**

~~26.9~~	~~77.8~~	~~8.14~~	~~9.33~~	~~46.3~~	~~71.5~~	~~46~~	~~7.9~~	~~64.5~~	~~87.8~~
~~33.7~~	~~27~~	~~71.8~~	~~33.8~~	~~59.1~~	~~28.8~~	~~80.2~~	~~81.3~~	~~43.7~~	~~41.6~~
~~7.1~~	~~64~~	~~58.3~~	~~37.9~~	~~72.2~~	~~79.9~~	~~54.3~~	~~89.7~~	~~61.3~~	~~21.5~~
~~70.1~~	~~67~~	~~15.5~~	~~2.2~~	~~28.6~~	~~31.7~~	~~4.9~~	~~26.8~~	~~99.2~~	~~0.3~~
~~10.5~~	~~28.3~~	~~99.3~~	~~14.1~~	~~47~~	~~98.1~~	~~20.3~~	~~44.2~~	~~21.7~~	~~66.3~~
~~3.2~~	~~6.2~~	~~67~~	~~39.6~~	~~39.2~~	~~62.3~~	~~21.9~~	~~72.2~~	~~67.5~~	~~64~~
~~93.7~~	~~51~~	~~40~~	~~20.2~~	~~73.9~~	~~18.2~~	~~25.5~~	~~24.5~~	~~67.3~~	~~16.4~~
~~25.4~~	~~35.9~~	~~26.5~~	~~83.1~~	~~88~~	~~37.8~~	~~19.1~~	~~80~~	~~45.4~~	~~73.8~~
~~73.9~~	~~63.8~~	~~54.2~~	~~23.2~~	~~53~~	~~52.3~~	~~63.6~~	~~46.2~~	~~24.9~~	~~55.6~~

وعليه يكون جدول التوزيع التكراري

التكرارات f_i	الفئات حجم المبيعات/بالدولار
14	**0 – 20**
26	**20 – 40**
17	**40 – 60**
22	**60 – 80**
11	**80 – 100**
90	$\sum f_i$

وسيتم حساب التكرار النسبي والمئوي بنفس الجدول كما في أدناه :

التكرار المئوي pf_i	التكرار النسبي rf_i	التكرارات f_i	الفئات حجم المبيعات/بالدولار
$0.16 * 100 = 16$	$\frac{14}{90} = 0.16$	**14**	**0 – 20**
29	**0.29**	**26**	**20 – 40**
19	**0.19**	**17**	**40 – 60**
24	**0.24**	**22**	**60 – 80**
12	**0.12**	**11**	**80 – 100**

$\sum pf_i = 100$	$\sum rf_i = 1$	$\sum f_i = 90$	المجموع الكلي

الحدود الحقيقية (الفعلية) للفئات :

عند تكوين جدول توزيع تكراري لبيانات أي ظاهرة فغالباً ما تكون بيانات هذا الجدول مقربة لأقرب وحدة قياس أو لأقرب نصف وحد قياس، وعلى هذا الأساس يمكن حساب الحدود الحقيقية للفئات والتي تكون من حدين (الحدود الحقيقة الدنيا) و(الحدود الحقيقية العليا) للفئات :

حساب الحدود الحقيقية للفئات إذا كانت مقربة لأرقام صحيحة :

الحد الحقيقي الأدنى للفئة = الحد الأدنى الفعلي للفئة - $\frac{1}{2}$

الحد الحقيقي الأعلى للفئة= الحد الأعلى الفعلي للفئة + $\frac{1}{2}$

أو

الحد الحقيقي الأدنى للفئة = مركز الفئة – ($\frac{1}{2}$)*طول الفئة

الحد الحقيقي الأعلى للفئة= مركز الفئة +($\frac{1}{2}$) * طول الفئة

التكرارات المتجمعة

تُعد التوزيعات التكرارية البسيطة إحدى الجوانب الأساسية التي تعطي تصور حول عدد المفردات في كل فئة ولكن في بعض الأحيان يتطلب الأمر معرفة عدد المفردات التي قيمها (اقل من) أو (أكثر من) من قيمة معينة في التوزيع التكراري، لذا فان التوزيع التكراري المتجمع هو التوزيع الذي يبين كمية التكرار المتجمع عند قيمة معينة من قيم المتغير العشوائي ويرمز له بالرمز F، وهناك نوعان من الجداول المتجمعة :

أولاً : جدول التوزيع المتجمع الصاعد

ثانياً: جدول التوزيع المتجمع النازل

أولاً : **جدول التوزيع المتجمع الصاعد**

يحسب التكرار المتجمع الصاعد حسب طريقة عرض جدول التوزيع التكراري البسيط، ويمكن التعبير عنه هو تجميع التكرارات من الحد الأعلى للفئة الأولى إلى أخر حد أعلى للفئة الأخير بحيث يكون تكرار الفئة الأولى هو نفسه في الجدول البسيط وتكرار الفئة الأخيرة مساويا للمجموع الكلي للتكرارات، أو هو تجميع التكرارات من الحد الأدنى للفئة الثانية إلى الحد الأعلى للفئة الأخيرة . كما موضحة في أدناه :

جدول توزيع تكراراي متجمع صاعد		جدول توزيع تكراري بسيط	
التكرارات المتجمع الصاعد Fi	الحدود العليا للفئات	التكرارات fi	الفئات
f_1	أقل من U_{C1}	f_1	$L_{C1} - U_{C1}$
$f_1 + f_2$	أقل من U_{C2}	f_2	$L_{C2} - U_{C2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$f_1 + f_2 + \cdots + f_k$	أقل من U_{Ck}	f_k	$L_{Ck} - U_{Ck}$

اما إذا كان عرض الجدول كما في أدناه :

جدول توزيع تكراراي متجمع صاعد		جدول توزيع تكراري بسيط	
التكرارات المتجمع الصاعد Fi	الحدود العليا للفئات	التكرارات fi	الفئات
f_1	أقل من L_{C2}	f_1	$L_{C1} -$
$f_1 + f_2$	أقل من L_{C3}	f_2	$L_{C2} -$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$f_1 + f_2 + \cdots + f_k$	أقل من U_{Ck}	f_k	$L_{Ck} - U_{Ck}$

مثال/ كون جدول توزيع تكراري متجمع صاعد من بيانات الجدول التالي :

التكرارات fi	الفئات
4	40 - 49
3	50 - 59
7	60 – 69
4	70 – 79
2	80 – 89
20	$\sum f_i$

الحل/

التكرار المتجمع الصاعد Fi	الحدود العليا للفئات	التكرارات fi	الفئات
4	اقل من 49	4	40 - 49
4+3=7	اقل من 59	3	50 - 59
4+3+7=14	اقل من 69	7	60 – 69
4+3+7+4=18	اقل من 79	4	70 – 79
4+3+7+4+2=20	اقل من 89	2	80 – 89
ملاحظة/ التكرار المتجمع الصاعد يبدأ بتكرار الفئة الأولى وينتهي بالمجموع الكلي للتكرارات		20	$\sum f_i$

ولو فرضنا أن الجدول بالشكل التالي:

التكرارات fi	الفئات

40 -	4
50 -	3
60 –	7
70 –	4
80 – 89	2
$\sum f_i$	20

فان جدول التوزيع التكراري البسيط بالشكل التالي، فان التكرار المتجمع الصاعد لهذا الجدول :

التكرار المتجمع الصاعد Fi	الحدود العليا للفئات	التكرارات fi	الفئات
4	اقل من 50	4	40 -
4+3=7	اقل من 60	3	50 -
4+3+7=14	اقل من 70	7	60 –
4+3+7+4=18	اقل من 80	4	70 –
4+3+7+4+2=20	اقل من 89	2	80 – 89
ملاحظة/ التكرار المتجمع الصاعد يبدأ بتكرار الفئة الأولى وينتهي بالمجموع الكلي للتكرارات		20	$\sum f_i$

ثانياً : **جدول التوزيع المتجمع النازل (الهابط)**

يحسب التكرار المتجمع النازل حسب طريقة عرض جدول التوزيع التكراري البسيط، ويمكن التعبير عنه هو تجميع التكرارات من الحد الأدنى للفئة الأولى إلى الحد الأدنى للفئة الأخيرة وبشكل تناقصي بحيث يكون تكرار الفئة الأولى هو المجموع الكلي للفئات في جدول التوزيع التكراري البسيط، بينما تكرار الفئة الأخيرة في عمود المتجمع النازل يكون مساويا إلى تكرار الفئة الأخيرة في جدول التوزيع التكراري البسيط. كما موضحة في أدناه :

جدول توزيع تكراراي متجمع صاعد		جدول توزيع تكراري بسيط	
التكرارات المتجمع النازل Fi	الحدود الدنيا للفئات	التكرارات fi	الفئات
$f_1 + f_2 + \cdots + f_k = \sum f_i$	L_{C1} فأكثر	f_1	$L_{C1} - U_{C1}$
$\sum f_i - f_1$	L_{C2} فأكثر	f_2	$L_{C2} - U_{C2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\sum f_i - (f_1 + f_2 + \cdots + f_{k-1}) = f_k$	L_{Ck} فأكثر	f_k	$L_{Ck} - U_{Ck}$

<u>مثال</u>/ كون جدول توزيع تكراري متجمع نازل(هابط) من بيانات الجدول التالي :

الفئات	التكرارات fi
40 - 49	4
50 - 59	3
60 – 69	7
70 – 79	4
80 – 89	2
$\sum f_i$	20

الحل/

الفئات	التكرارات fi	الحدود العليا للفئات	التكرار المتجمع الصاعد Fi
40 – 49	4	40 فأكثر	20
50 – 59	3	50 فأكر	20-4 = 16
60 – 69	7	60 فأكثر	20-(4+3) = 13 or 16 – 3 =13
70 – 79	4	70 فأكثر	20 –(4+3+7) =6 or 13 – 7 = 6
80 – 89	2	80 فأكثر	20-(4+3+7+4)=2 or 6 – 4 =2
$\sum f_i$	20	ملاحظة/ التكرار المتجمع النازل يبدأ بالمجموع الكلي للتكرارات وينتهي بتكرار الفئة الأخيرة	

ولو فرضنا أن الجدول بالشكل التالي

الفئات	التكرارات fi
40 -	4
50 -	3
60 –	7
70 –	4
80 – 89	2
$\sum f_i$	20

فان جدول التوزيع التكراري البسيط بالشكل التالي، فان التكرار المتجمع النازل لهذا الجدول :

الفئات	التكرارات fi	الحدود الدنيا للفئات	التكرار المتجمع الصاعد Fi
40 -	4	40 فأكثر	20
50 -	3	50 فأكر	20 – 4 =16
60 –	7	60 فأكثر	16 – 3 =13
70 –	4	70 فأكثر	13 – 7 = 6
80 – 89	2	80 فأكثر	6 – 4 = 2
$\sum f_i$	20	ملاحظة/ التكرار المتجمع النازل يبدأ بالمجموع الكلي للتكرارات وينتهي بتكرار الفئة الأخيرة	

ثالثاً: **تمثيل التكرار المتجمع الصاعد والنازل بيانياً**

يتم رسم منحنى توزيع التكرار المتجمع الصاعد والنازل من خلال جعل الفئات ممثله على المحور السيني(X) بينما التكرارات (المتجمعة) تكون ممثله بالمحور الصادي (Y)، ولتوضيح ذلك نأخذ المثال التالي

مثال/ كون جدول تكراري متجمع صاعد ونازل لبيانات جدول التوزيع التكراري تم مثله بيانياً:

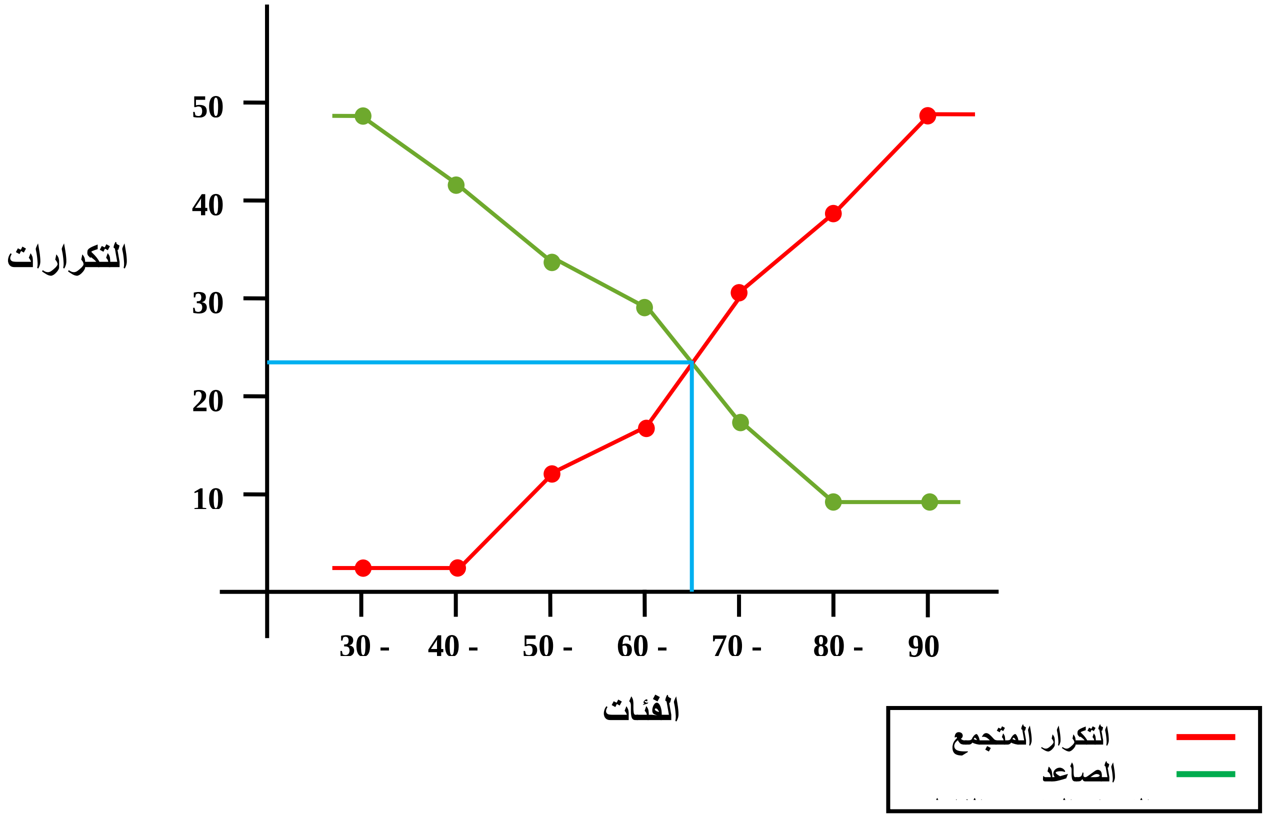
الفئات	التكرار f_i
30 -	5
40 -	10
50 -	4
60 -	12
70 -	8
80 - 90	9
$\sum f_i$	48

الحل//

الفئات	التكرار f_i	الحدود العليا للفئات	التكرار المتجمع الصاعد F_i	الحدود الدنيا للفئات	التكرار المتجمع النازل F_i
30 -	5	اقل من 40	5	30فأكثر	48
40 -	10	اقل من 50	15	40فأكثر	43
50 -	4	اقل من 60	19	50فأكثر	33
60 -	12	اقل من 70	31	60فأكثر	29
70 -	8	اقل من 80	39	70فأكثر	17
80 - 90	9	اقل من 90	48	80فأكثر	9
مجموع التكرارات	48				

وعليه يكون الرسم البياني على النحو التالي :

1- المحور (x) يمثل الفئات لجدول التوزيع التكراري .
2- المحور (y) تمثل التكرارات F_i وبما أن التكرارات المتجمعة فيها اكبر قيمة هي (48) التي تمثل مجموع التكرارات لذا تم تقسيم التكرارات على المحور (y) على هذا الأساس .
3- يتم تسقيط نقاط التكرار المتجمع الصاعد ومن ثم النازل على المحاور فيتشكل منحنى صاعد وأخر نازل وكما مبين في أدناه .



مثال (واجب) // **البيانات التالية تمثل ما تدفعه** (150) **عائلة عن أجور اشتراك الكهرباء الوطنية (الآلاف الدنانير) خلال السنة** :

أجور الكهرباء	50–100	100-150	150-200	200-250	250-300	300-350	350-400
عدد العوائل	32	35	25	20	17	13	8

المطلوب : **كون جداول التوزيع التالية**

(1) جدول التوزيع النسبي .
(2) جدول التوزيع المئوي .
(3) جدول التوزيع المتجمع الصاعد .
(4) جدول التوزيع المتجمع النازل .
(5) مثل بيانياً التوزيع المتجمع الصاعد والنازل

مثال (واجب)

وضح كلا مما يأتي، ثم أعط حول حالة :

(1) المستطيل البياني

(2) التوزيعات التكرارية الصاعدة .

(3) التبويب الزمني .

(4) الدائرة الإحصائية .

(5) التبويب الجغرافي .

مثال(واجب) // إذا كانت لديك البيانات التالية :

48	38	51	45	56	39	50	71	65	34	51	66	27	73	69	34	43	34	53	54
91	44	53	36	49	56	45	64	59	41	58	76	52	57	69	81	46	34	41	62
43	55	79	66	87	81	70	67	55	53	84	52	56	44	51	65	76	52	54	33
54	61	52	95	40	57	35	53	60	55	64	42	69	57	47	53	52	61	36	61
57	80	46	61	54	94	55	85	73	60	27	44	67	65	62	32	54	71	49	95

المطلوب// كون جدول توزيع تكراري بسيط

مقاييس النزعة المركزية

Measures of Central Tendency

قبل الدخول في موضوع مقاييس النزعة المركزية لا بد من توضيح دليل أو رمز مهم في التعبير الرياضي هو (رمز المجموع $\sum$) .

أولاً : تعريف رمز المجموع $\sum$

إذا كانت لدينا المتغير X الذي يمتلك عدد من المشاهدات $(x_i, x_1, x_2, \dots, x_n)$ فإن حاصل جمع هذه المشاهدات (المجموع الكلي لقيم الظاهرة) يعبر عنها بالشكل التالي :

$$\sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$$

وعليه يمكن الاستفادة من هذا الرمز في توصيف بعض العلاقات الرياضية التي تدخل في تحديد بعض الخصائص للمقاييس الإحصائية .

أمثله تطبقيه حول رمز المجموع

إذا كانت لدينا ظاهرتين هما x و y لكل منهما عدد من المشاهدات، حيث أن :

$$(x|x_i = x_1, x_2, \dots, x_n), (y|y_i = y_1, y_2, \dots, y_n)$$

(1) مجموع حاصل جمع متغيرين أو حاصل جمع مجموع مشاهدات كل متغير، ويمكن وضع هذه العلاقة باعتماد رمز المجموع

مجموع حاصل جمع متغيرين $= \sum_{i=1}^{n}(x_i + y_i)$

حاصل جمع مجموع كل متغير $= \sum_{i=1}^{n}(x_i) + \sum_{i=1}^{n}(y_i)$

حيث أن :

$$\sum_{i=1}^{n}(x_i + y_i) = \sum_{i=1}^{n}(x_i) + \sum_{i=1}^{n}(y_i)$$

(2) حاصل قسمة مجموع مشاهدات كل متغير $\frac{\sum_{i=1}^{n}(x_i)}{\sum_{i=1}^{n}(y_i)}$

(3) مجموع حاصل قسمة مشاهدات متغيرين $\sum_{i=1}^{n}(\frac{x_i}{y_i})$

(4) إذا كانت c كمية ثابتة فأن :

$$\sum_{i=1}^{n} c(x_i) = \mathrm{c}\sum_{i=1}^{n}(x_i)$$

ملاحظة :

$$\frac{\sum_{i=1}^{n}(x_i)}{\sum_{i=1}^{n}(y_i)} \neq \sum_{i=1}^{n}(\frac{x_i}{y_i})$$

مثال: إذا كانت لدينا بيانات ظاهرتين مختلفتين هما x و y، وكانت قيم كل ظاهرة كما يلي :

x_i	2	4	6
y_i	1	5	3

احسب كلا مما يأتي :

(1) $\sum_{i=1}^{n}(x_i + y_i)$

(2) $\sum_{i=1}^{n}(\frac{x_i}{y_i})$

(3) $\frac{\sum_{i=1}^{n}(x_i)}{\sum_{i=1}^{n}(y_i)}$

الحل :

(1) $\sum_{i=1}^{3}(x_i + y_i) = (x_1 + y_1) + (x_2 + y_2) + (x_3 + y_3)$

$= (2 + 1) + (4 + 5) + (6 + 3)$

$= 3+9+9$

$=21$

(2) $\sum_{i=1}^{n}(\frac{x_i}{y_i}) = \frac{x_1}{y_1} + \frac{x_2}{y_2} + \frac{x_3}{y_3}$

$= \frac{2}{1} + \frac{4}{5} + \frac{6}{3}$

$= 2 + 0.8 + 2$

$= 4.8$

(3) $\frac{\sum_{i=1}^{n}(x_i)}{\sum_{i=1}^{n}(y_i)} = \frac{x_1+x_2+x_3}{y_1+y_2+y_3}$

$= \frac{2+4+6}{1+5+3}$

$= \frac{12}{9}$

$= 1.33$

ويمكن تطبيق أي علاقة رياضية تحتوي على رمز المجموع اخذين بنظر الاعتبار أسبقيات العمليات الرياضية لتنفيذ أي علاقة .

ثانياً : مفهوم مقاييس النزعة المركزية

ان العملية الإحصائية تمر بعدد من المراحل بداية من عملية جمع البيانات ومن ثم وصف البيانات وصولاً إلى حساب المؤشرات الإحصائية المتعلقة بالنزعة المركزية أو بالتشتت ستختلف تبعاً لنوع البيانات الإحصائية المستخدمة سواء كانت هذه البيانات من المجتمع أو من عينة عشوائية .

تهتم مقاييس النزعة المركزية بتوفير مؤشرات كمية تمثل التوجه العام لقيم المتغير الكمي، وعليه يمكن تعريف النزعة المركزية هو تحديد القيمة النموذجية التي تميل إلى الوقوع في المركز داخل مجموعة من البيانات ولتحديد هذه القيمة النموذجية هناك عدد من المقاييس الإحصائية التي يطلق عليها بمقاييس النزعة المركزية وهذه المقاييس هي :

(1) الوسط الحسابي
(2) الوسيط
(3) المنوال
(4) الوسط الهندسي
(5) الوسط التوافقي

وكل من هذه المقاييس يتميز بعدد من المزايا والعيوب .

(1) <u>الوسط الحسابي Mean</u> : يعتبر الوسط الحسابي (Arithmetic Average or Mean)

وهو من مقاييس النزعة المركزية الأكثر شهرةً وأهمية، حيث تعتبر قيمة الوسط الحسابي هي القيمة التي تتمركز حولها جميع القيم للمتغير الكمي، يمكن حساب قيمة الوسط الحسابي حسب طبيعة البيانات .

أولاً : حساب الوسط الحسابي للبيانات غير المبوبة (القيم الأصلية)

بافتراض أن المتغير العشوائي X يمثل بيانات مجتمع محدود بحجم (N) فيمكن حساب قيمة الوسط الحسابي الذي يرمز له بالرمز (μ) ، كما في الصيغة التالية :

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N} \quad ; \qquad i = 1, 2, \dots, N$$

وهو عبارة عن القيمة التي يحصل عليها من خلال قسمة المجموع الكلي للقيم على عددها.

أما إذا افترضنا أن المتغير العشوائي x يمثل بيانات عينة مسحوبة من مجتمع وبحجم (n) فيمكن حساب الوسط الحسابي والذي يرمز له بالرمز ($\overline{x}$)، كما في الصيغة التالية :

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} \quad ; \qquad i = 1, 2, \dots, n$$

مثال/ احسب متوسط أوزان عينة من طلبة المرحلة الأولى في كلية الإدارة والاقتصاد

الأوزان xi بالكيلوغرام	55	70	60	65	77	80	66

الحل //

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad ; i = 1, 2, \dots, 7$$

$$\overline{x} = \frac{\sum_{i=1}^{7} x_i}{n}$$

$$\overline{x} = \frac{55 + 70 + 60 + 65 + 77 + 80 + 66}{7}$$

$$\overline{x} = \frac{473}{7}$$

$\overline{x} = 67.57$ كيلو غرام

ثانياً: حساب الوسط الحسابي للبيانات المبوبة

ويقصد بالبيانات المبوبة هي البيانات المنظمة بشكل جدول توزيع تكرار أي (فئات وتكرارات)، ويحسب حسب الصيغة التالية :

$$\overline{X} = \frac{x_1 f_1 + x_2 f_2 + \cdots + x_k f_k}{f_1 + f_2 + \cdots + f_k} \qquad \dots (1)$$

وتكون الصيغة العامة بالشكل التالي:

$$\overline{X} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i} \qquad \dots (2)$$

حيث أن :

x_i: يمثل مركز الفئة

f_i: التكرار الفعلي لكل فئة

$\sum_{i=1}^{k} f_i$: المجموع الكلي لتكرارات الفئات

k: تمثل عدد الفئات في جدول التوزيع التكراري

عند تقدير الوسط الحسابي للبيانات الملخصة في جدول توزيع تكراري، نلاحظ ما يلي :

1– البيانات الأصلية غير معروفة

2– عدد البيانات في كل فترة (عدد التكرارات لكل فترة) معروف .

3– يستخدم مركز الفئة كمية تقريبية لجميع البيانات في الفترة .

مثال/ من بيانات جدول التوزيع التكراري التالي احسب الوسط الحسابي

الفئات	30 –	40 –	50 –	60 –	70 –80
f_i	3	9	5	8	7

الحل//

$$\overline{X} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i}$$

لحساب الوسط الحسابي للبيانات المبوبة يتطلب حساب مراكز الفئات (x_i) :

$$\text{مركز الفئة} = \frac{\text{مجموع حدي فئتين متتاليتين}}{2}$$

$$\text{مركز الفئة الاولى} = \frac{\text{الحد الادنى للفئة الاولى} + \text{الحد الادنى للفئة الثانية}}{2}$$

$$x_1 = \frac{30 + 40}{2}$$

$$x_1 = 35$$

مركز الفئة الثانية

$$x_2 = \frac{40+50}{2}$$

$$x_2 = 45$$

وهكذا يمكن حساب مركز الفئة لجميع فئات الجدول، كما يمكن حساب مراكز الفئات بطريقة اخرى ابسط باعتماد طول الفئة وكما يلي :

نحسب قيمة مركز الفئة الأولى حسب القانون

$$\text{مركز الفئة الاولى} = \frac{\text{الحد الادنى للفئة الاولى} + \text{الحد الادنى للفئة الثانية}}{2}$$

$$x_1 = \frac{30 + 40}{2}$$

$x_1 = 35$

بعد ذلك نحدد طول الفئة بين جميع الفئات (L) وحسب المثال نلاحظ ان طول الفئة مساوي إلى (10) وهو متساوي لجميع فئات الجدول عندئذ يتم إضافة طول الفئة لمركز الفئة الأولى لنحصل على مركز الفئة الثاني

$$x_2 = 35 + 10$$

وهكذا لجميع الفئات، أي أن :

$$x_{i+1} = x_i + L \quad ; i = 1, 2, \dots, k$$

الفئات	f_i	x_i	$f_i x_i$
30 -	3	35	3*35 =105
40 -	9	35+10 =45	9*45=405
50 -	5	45+10=55	275
60 -	8	55+10=65	520
70 - 80	7	65+10=75	525
∑	32		1830

$$\bar{X} = \frac{1830}{32} \Longrightarrow \bar{X} = 57.19$$

مثال / احسب الوسط الحسابي لمتوسط أعمار الطلاب للبيانات التالية :

العمر	5 – 6	7 – 8	9 – 10	11-12	13 – 14
عدد الطلبة	2	5	8	4	1

الحل//

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

نلاحظ لدينا حسب الجدول لدينا k = 5 فئات، ولغرض حساب الوسط الحسابي لهذه البيانات نكون الجدول التالي :

الفئات (العمر) سنة	التكرارات (عدد الطلبة) f_i	مركز الفئة x_i	$f_i x_i$
5 – 6	2	$x_1 = \frac{5+6}{2} = 5.5$	11
7 – 8	5	7.5	37.5
9 – 10	8	9.5	76
11-12	4	11.5	46
13 - 14	1	13.5	13.5
$\sum$	20		184

$$\bar{X} = \frac{184}{20} \Longrightarrow \bar{X} = 9.2 \text{ سنوات}$$

تمرين:

احسب الوسط الحسابي لمتوسط الهيموغلوبين في الدم لعينة مكونة من (50) شخص تم تلخيص نتائج مستوى الهيموغلوبين في جدول توزيع تكراري، كما يلي :

مستوى الهيموغلوبين	الاشخاص
13 – 14	3
14 – 15	5
15 – 16	15
16 – 17	16
17 – 18	10
18 – 19	1

ثالثا" : الوسط الحسابي المرجح أو الموزون Weighted Mean

1– في حالة البيانات غير المبوبة:

من الناحية العملية هناك الكثير من الحالات تكون بعض المفردات اكثر اهمية من الاخرى مما يتوجب الامر اخذ ذلك بنظر الاعتبار لدى حساب الوسط الحسابي ، فمثلا عند حساب معدل درجات الطالب فأن الامر يستوجب الاخذ بنظر الاعتبار عدد الساعات الاسبوعية المخصصة لكل درس وهذا يعني ترجيح المفردات بأوزان معينة تمثل اهمية كل منها وعنده ادخال اهمية المفردات في حساب الوسط الحسابي فأن عندئذ يسمى الوسط الحسابي المرجح وبتعبير آخر لكل قيمة من المشاهدات (x_i) وزن خاص يتناسب مع اهميتها (w_i) فالوسط الحسابي لهذه القيم يحسب كما يلي:–

$$\bar{X}_w = \frac{\sum_{i=1}^{k} x_i w_i}{\sum_{i=1}^{k} w_i}$$

حيث ان $\bar{X}_w$: الوسط الحسابي الموزون

w_i : اوزان وأهمية (المفردة)

x_i : قيمة المشاهدة

مثال // اذا كانت درجات احد الطلبة في المرحلة الاولى في قسم العلوم المالية والمصرفية في الدروس المقررة في هذه المرحلة حسب الساعات الاسبوعية المحدد لكل درس ، المطلوب حساب معدل الطالب ؟

الدرجة	62	80	75	88	84	86	90
عدد الساعات	2	2	2	3	3	3	3

الحل //

$W_i\ x_i$	الأهمية w_i	الدرجة x_i
124	2	62
160	2	80
150	2	75
264	3	88
252	3	84
258	3	86
270	3	90
1478	18	المجموع

$$\bar{X}_w = \frac{\sum_{i=1}^{k} x_i w_i}{\sum_{i=1}^{k} w_i} = \frac{1478}{18} = 80.11$$

2– الوسط الحسابي الموزون في حالة البيانات المبوبة

$$\bar{X}_w = \frac{\sum_{i=1}^{k} x_i f_i w_i}{\sum_{i=1}^{k} w_i}$$

اذ ان :

x_i= مركز الفئة

f_i= التكرار

w_i= الاهمية

مثال // اوجد الوسط الحسابي الموزون للبيانات التالية التي تمثل انتاج معمل الادوية في سامراء من الادوية بالطن وعدد المكائن العاملة وعدد ساعات العمل ؟

فئات الانتاج بالطن	عدد المكائن العاملة fi	عدد ساعات العمل wi	x_i	f_iw_i	$x_if_iw_i$
2-4	4	6	3	24	72
4-6	5	5	5	25	125
6-8	6	6	7	36	252
8-10	3	4	9	12	108
10-12	2	4	11	8	88
	20			105	645

$$\bar{X}_w = \frac{\sum_{i=1}^{k} x_i f_i w_i}{\sum_{i=1}^{k} w_i} = \frac{645}{105} = 6.14$$ طن

خصائص الوسط الحسابي :-

1- مجموع انحرافات القيم عن وسطها الحسابي = صفر

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

2- مجموع مربعات انحرافات القيم عن وسطها الحسابي = اقل ما يمكن

$\sum (x_i - \bar{x})^2 =$ اقل ما يمكن

3- يأخذ الوسط الحسابي بعين الاعتبار جميع القيم في حسابه .

4- يتأثر الوسط الحسابي بالقيم الشاذة او المتطرفة لان الوسط الحسابي يأخذ بنظر الاعتبار جميع القيم .

5- هناك صعوبة في حساب الوسط الحسابي في حالة الفئات المفتوحة لأنه من الصعب تحديد مراكز الفئات وهذه المشكلة تحل بتحديد مراكز الفئات بصورة تقريبية.

(2)– الوسيط Median:–

يعرف الوسيط بأنه القيمة التي تمثل المرتبة الوسطى عندما ترتب القيم قيد الدراسة تصاعدياً او تنازلياً وهذا يعني ان نصف القيم تقل عن قيمة الوسيط والنصف الاخر يزيد عنها.

أ– ايجاد الوسيط للبيانات غير المبوبة

1– يتم ترتيب القيم تصاعدياً او تنازلياً

2– اذا كان عدد القيم فردي (n) فالوسيط يكون القيمة التي ترتيبها $\frac{n+1}{2}$ واذا كان عدد القيم زوجي (n) فالوسيط هو الوسط الحسابي للقيمتان اللتان ترتيبهما $\frac{n}{2}$ و $\frac{n}{2}+1$

مثال // اوجد الوسيط للبيانات التالية:

$X_i = 11 , 12 , 13 , 12 , 13 , 11 , 14$

الحل :– 1– ترتب البيانات ترتيب تصاعدي $11 , 11 , 12 , 12 , 13 , 13 , 14$

2 – ايجاد رتبة الوسيط

بما ان عدد القيم (n) = فردي

رتبة الوسيط $T = \frac{n+1}{2} = \frac{7+1}{2} = \frac{8}{2} = 4$

Me = 12

مثال // اوجد الوسيط للبيانات التالية:

$X_i = 11 , 12, 13 , 12 , 13 , 11 , 14 , 10$

الحل :–

1– ترتيب البيانات تصاعدياً $10 , 11 , 11, 12 , 12 , 13 , 13 , 14$

2– ايجاد رتبة الوسيط

بما ان عدد القيم زوجي = 8

فالوسيط هو الوسط الحسابي للقيمتان التي ترتيبهما $\frac{n}{2}$ ، $\frac{n}{2}+1$

$$\frac{n}{2} = \frac{8}{2} = 4$$

$$\frac{n}{2} + 1 = \frac{8}{2} + 1 = 5$$

$$\text{Me} = \frac{12+12}{2} = \frac{24}{2} = 12$$

ب – ايجاد الوسيط لبيانات مبوبة :

1 –الوسيط للبيانات المبوبة (متغير متقطع):

و يمكننا ايجاد الوسيط من الجداول التكرارية البسيطة بتحويلها الى جداول تكرارية صاعدة او نازلة .

الوسيط في حالة التكرار المتجمع الصاعد

خطوات ايجاد الوسيط لبيانات مبوبة لمتغير متقطع :

1– نجد التكرار المتجمع الصاعد

2– نجد ترتيب الوسيط والذي يساوي $T = \frac{\sum f_i}{2}$

ملاحظة : f_i تمثل هنا التكرارات الاصلية وليس التكرار المتجمع الصاعد.

3– نحدد قيمة الوسيط وهي التي تقع بين التكرارين يعني ترتيب الوسيط بين التكرارين .

4–نحدد فئة الوسيط وان مركز هذه الفئة يمثل الوسيط.

مثال/ الاتي توزيع لعينة من الاسر حسب عدد افراد الاسرة ، المطلوب حساب الوسيط لعدد افراد الاسرة.

الفئات	2-4	5-7	8-10	11-13	14-16	17-19	20-22	المجموع
التكرارات	6	9	12	20	14	11	8	80

الحل/

الفئات	التكرارات f_i	الحدود العليا للفئات	التكرار المتجمع الصاعد F_i
2–4	6	4 فاقل	6
5–7	9	7 فاقل	15
8–10	12	10 فاقل	27
11–13	20	13 فاقل	47
14–16	14	16 فاقل	61
17–19	11	19 فاقل	72
20–22	8	22 فاقل	80
المجموع	80		

ترتيب الوسيط :

$$T = \frac{\sum f_i}{2} = \frac{80}{2} = 40$$

أي الوسيط يقع بين التكرارين المتجمع الصاعد الثالث والرابع أي بين 27 و47

و عليه فان فئة الوسيط هي الفئة (11–13) لأنها الاقرب الى 47

وان الوسيط يمثل مركز هذه الفئة

وعليه فالوسيط يساوي 12 وكالاتي:

$$Me = \frac{11+13}{2} = 12$$

2- الوسيط في حالة البيانات المبوبة (متغير مستمر)

نطبق القانون التالي:

$$Me = L_k + \left[\frac{\frac{\sum fi}{2} - F_{k-1}}{F_{k+1} - F_{k-1}}\right] \times h_k$$

L_k = هي الحد الادنى لفئة الوسيط

$\frac{\sum fi}{2}$ = رتبة الوسيط في حالة مجموع التكرارات عدد زوجي

$\frac{\sum fi+1}{2}$ = رتبة الوسيط في حالة مجموع التكرارات عدد فردي

F_{K-1}= التكرار المتجمع الصاعد السابق لفئة الوسيط

F_{K+1}= التكرار المتجمع الصاعد اللاحق لفئة الوسيط

h_k= طول الفئة (طول فئة الوسيط)

مثال // اوجد الوسيط للبيانات التالية التي تبين توزيع 100 طالب من طلبة كلية ما حسب صفة الوزن.

الحل 1- : نجد التكرار المتجمع الصاعد وكالتالي:

الفئات	التكرارات f_i	الحدود العليا للفئات	التكرار المتجمع الصاعد F_i
60-62	5	اقل او يساوي62	5
63-65	15	65 = =	20
66-68	45	68 = =	65
69-71	27	71 = =	92
72-74	8	74 = =	100
المجموع	100		

2 – نجد ترتيب الوسيط من الصيغة :

$$T = \frac{\sum f_i}{2} = \frac{100}{2} = 50$$

3–نحدد فئة الوسيط والتي تقع بينالتكرارين الثاني والثالث وهي (68–66) لان 50 اقرب الى 65

اي الوسيط يقع بين التكرارين الثالث والرابع أي بين 20 و65

4–نطبق صيغة القانون

$$Me = L_k + \left[\frac{\frac{\sum fi}{2} - F_{k-1}}{F_{k+1} - F_{k-1}}\right] \times h_k$$

$$Me = 66 + \left[\frac{50-20}{92-20}\right] * 3 = 67.25$$

<u>ملاحظات عن الوسيط :–</u>

1– يستخدم الوسيط كمقياس للنزعة المركزية بدلاً عن الوسط الحسابي عندما تكون هناك قيمة شاذة في التوزيع .

2– يستخدم في حالة الفئات المفتوحة

3– الوسيط قليل الحساسية للمتغيرات التي تحدث في قيم البيانات الاصلية لانه يهتم بالقيم الواقعة في المنتصف ويهمل الاطراف على عكس الوسط الحسابي الذي يعتبر شديد الحساسية لانه يأخذ بعين الاعتبار جميع القيم في حسابه .

4– يمكن استخدامه في حالة المتغيرات الوصفية التي لا تعبر عنها في الارقام كما هو الحال في ترتيب الاشخاص وفقاً لخصائصهم .

(3)– المنوال Mode:–

هي القيمة الاكثر شيوعاً او تكراراً في التوزيع وهو ابسط مقاييس النزعة المركزية

أ– المنوال في حالة البيانات غير المبوبة

مثال 1// اوجد المنوال للبيانات التالية : $7\ ,\ 4\ ,\ 8\ ,\ 6\ ,\ 4$

$Mo = 4$

مثال 2// اوجد المنوال للبيانات التالية: $7\ ,\ 6\ ,\ 3\ ,\ 8\ ,\ 5\ ,6\ ,\ 3$

$Mo = 3$

التوزيع ثنائي المنوال $Mo = 6$

مثال3// اوجد المنوال للبيانات التالية:, $7\ ,\ 9\ ,\ 3\ ,\ 8\ ,\ 5\ ,6$

لا يوجد منوال

ب – المنوال في حالة البيانات المبوبة

$$\boxed{Mo = L + \left[\frac{d1}{d1+d2}\right] \times C}$$

L = هي الحد الادنى الحقيقي للفئة المنوالية

$d1$ = الفرق بين تكرار فئة المنوال و تكرار الفئة السابقة لها

$d2$ = الفرق بين فئة تكرار المنوال وتكرار الفئة اللاحقة لها

C = طول الفئة

ملاحظة : الفئة المنوالية هي الفئة التي تقابل اكبر التكرارات

مثال // اوجد المنوال للبيانات التالية التي تمثل توزيع طلبة كلية الصيدلة حسب صفة الوزن

الفئات	التكرار fi
62−62	5
63−65	15
66−68	45 (الفئة المنوالية)
69−71	27
72−74	8

الحل :-

$d1 = 45 - 15 = 30$

$d2 = 45 - 27 = 18$

$Mo = L + [\frac{d}{d1+d2}] \times C$

$Mo = 65.5 + [\frac{30}{30+18}] \times 3$

$Mo = 65.5 + \frac{30}{48} \times 3$

$Mo = 65.5 + 0.625 \times 3$

$Mo = 65.5 + 1.88 = 67.38$

مميزات المنوال :

1- مقياس سهل حسابه ولا يتأثر بالقيم الشاذة .

2- يمكن ايجاده للقيم الوصفية والتوزيعات التكرارية المفتوحة.

عيوب المنوال :

1- عند حساب المنوال لا تؤخذ جميع البيانات بالاعتبار .

2- قد يكون لبعض البيانات اكثر من منوال وبذلك لا يمكن تحديد قيمة وحيدة للمنوال .

العلاقة بين الوسيط والمنوال والوسط الحسابي

توجد علاقة بين الوسط الحسابي والمنوال والوسيط في حالة التوزيعات التكرارية احادية المنوال وتعطى من خلال المعادلة التالية :

الوسط الحسابي − المنوال = 3(الوسط الحسابي – الوسيط)

ثانيا": ـ مقاييس التشتت:

ان مقاييس النزعة المركزية غير كافية لوصف مجموعة من البيانات بشكل كامل اذ مقاييس التشتت تسمح بتباين مدى اختلاف البيانات فيما بينها ومدى تجانس البيانات مع بعضها و متوسطها الحسابي وتأتي أهمية التشتت من حقيقة انه ربما يتساوي المتوسطات لأكثر من مجموعة ولكن هذه المجموعات تكون مختلفة من حيث التجانس لذا فانه من الخطورة القول بأنه هذه المجموعات متشابه ولتوضيح ذلك نعطي المثال الآتي :

مثال :- لو كان لدينا ثلاثة مصانع A , B , C وسحبنا عينة عشوائية من أربعة عمال من كل مصنع وكما مبين في الجدول الآتي : والمطلوب هل أن أجور العاملين في المصانع الثلاثة متجانسة .

العمال	المصنع A	المصنع B	المصنع C
1	2	3.5	3.5
2	4	4.5	3.5
3	3	3	3.5
4	5	3	3.5
الوسط الحسابي	3.5	3.5	3.5

الحل :-

نقوم بحساب الوسط الحسابي للمصانع الثلاث :

نجد أن الوسط الحسابي للمصانع الثلاث متساوي وهذا يدل على أن أجور العمال في المصانع الثلاث تتوزع توزيعا متشابها وهذه نتيجة مضللة حيث نلاحظ من بيانات العينة للمصنع C متجانسة تماما أي أن درجة التشتت = صفر .

وبالتالي فان وسطها الحسابي يمثلها تمثيلا كاملا أما في المصنع (A , B) فنلاحظ أن المصنع (B) تكون فيه البيانات أكثر تجانسا من المصنع (A) ولذا فان تشتت البيانات في المصنع (B) اضعف من المصنع (A) .

و من أهم مقاييس التشتت :

1. المدى Rang
2. التباين Variance والانحراف المعياري Standard Deviation

المدى Rang :

1. **في حالة البيانات غير المبوبة :** يعرف المدى لمجموعة من البيانات على انه الفرق بين اكبر قيمة واقل قيمة .

$$R = X_L - X_S$$

حيث ان :

R: المدى

X_L : اكبر قيمة في البيانات ، X_S : اصغر قيمة في البيانات

مثال : فيما يلي بيانات مستخرجة عن تغيب الطلاب عن الكلية بالأيام

رقم الطالب	1	2	3	4	5	6	7	8	9
عدد الأيام	0	9	15	13	18	5	22	50	40

المطلوب // حساب المدى .

الحل :- يمكن ترتيب البيانات تصاعديا او تنازلي

50 40 22 18 15 13 9 5 0

حساب المدى $R = 50 - 0 = 50$

2- في حالة البيانات المبوبة : هو الفرق بين الحد الأعلى للفئة الأخيرة والحد الأدنى للفئة الأولى

المدى = (الحد الأعلى للفئة الاخيرة) - (الحد الأدنى للفئة الاولى)

مثال : من البيانات التالية احسب قيمة المدى .

الفئة (الوزن gm)	التكرار f_i
60 – 62	5
63 – 65	18
66 – 68	42
69 – 71	27
72 – 74	8
مجموع التكرارات	100

الحل:

المدى $R = 74 - 60 = 14$

ملاحظة : كلما كان المدى قليلا كلما كان ذلك دليلا على انخفاض التشتت .

2. التباين والانحراف المعياري :- Variance

يعتبر التباين والانحراف المعياري من أهم مقاييس التشتت التي تقيس مدى انتشار القيم عن بعضها البعض او عن قيمة معينة ، والتباين يعتمد على التشتت أو تباعد البيانات عن متوسطها ,فالتباين يكون كبيرا أذا كانت البيانات متباعدة عن متوسطها والعكس بالعكس، ويعرف التباين بأنه متوسط مربع انحرافات القيم عن الوسط الحسابي ويرمز له بالرمز (S^2) .

اما الانحراف المعياري فيعرف بانه الجذر التربيعي للتباين ويرمز له بالرمز (S) .

(أ) البيانات غير المبوبة :

يتم احتساب التباين من العلاقة التالية :

$$S^2 = \frac{\sum_i^n (X_i - \bar{X})^2}{n-1}$$

حيث ان : X_i : تمثل قيم المشاهدات (المتغير)

ويتم حساب الانحراف المعياري كما في الصيغة التالية :

$$S = \sqrt{\frac{\sum_i^n (X_i - \bar{X})^2}{n-1}}$$

مثال / اوجد التباين والانحراف المعياري للعينة التالية التي تمثل مجموعة من الأوزان مقاسه بالكيلوغرام (4 , 2.5 , 3.5 , 7 , 8).

الحل :

نجد التباين باستخدام الصيغة :

$$S^2 = \frac{\sum_i^n (X_i - \bar{X})^2}{n-1}$$

حجم العينة $n=5$

متوسط العينة هو $\bar{X}$

$$\bar{X} = \frac{\sum X_i}{n} = \frac{25}{5} = 5$$

نقوم بعمل الجدول التالي :

X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
4	-1	1
2.5	-2.5	6.25
3.5	-1.5	2.25
7	2	4
8	3	9
25	0	22.5

ويتم حساب التباين :

$$S^2 = \frac{22.5}{5-1} = 5.625$$

ويتم حساب الانحراف المعياري

$$S = \sqrt{S^2} = \sqrt{5.625} = 2.37$$

(ب) البيانات المبوبة :

يتم احتساب التباين من العلاقة التالية :

$$S^2 = \frac{\sum_i^n f_i(X_i-\bar{X})^2}{\sum f_i - 1}$$

حيث ان : X_i : تمثل مراكز الفئات

والمتوسط الحسابي يحسب من العلاقة :

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

ويتم حساب الانحراف المعياري كما في الصيغة التالية :

$$S = \sqrt{\frac{\sum_i^n f_i(X_i-\bar{X})^2}{\sum f_i - 1}}$$

مثال :- اوجد التباين والانحراف المعياري للبيانات التالية التي تمثل أوزان شرائح النحاس من الجدول الآتي :

الفئة (الوزن)	التكرار f_i	مركز الفئة X_i	$f_i X_i$	$X_i-\bar{X}$	$(X_i-\bar{X})^2$	$f_i(X_i-\bar{X})^2$
2-4	2	3	6	-3.8	14.44	28.88
4-6	1	5	5	-1.8	3.24	3.24
6-8	4	7	28	0.2	0.04	0.16
8-10	2	9	18	2.2	4.84	9.68
10-12	1	11	11	4.2	17.64	17.64
المجموع	10		68			59.6

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

نجد المتوسط

$$\bar{X} = \frac{68}{10} = 6.8$$

نجد التباين :

$$S^2 = \frac{\sum_i^n f_i(X_i-\bar{X})^2}{\sum f_i - 1}$$

$$S^2 = \frac{59.6}{10-1} = 6.62$$

ويتم حساب الانحراف المعياري كما يلي:

$$S = \sqrt{S^2} = \sqrt{6.62} = 2.57$$

Correlation Analysis: تحليل الارتباط

الارتباط : وهو من ادوات التحليل الوصفي ويهدف الى معرفة ان كانت هناك علاقة بين متغيرين مستقلين اوبين متغير مستقل (X) ومتغير تابع (معتمد) (Y) او بين مجموعة متغيرات مستقلة (X_i) ومتغير تابع (Y). ومقياس العلاقة يدعى معامل الارتباط (Correlation Coefficient) ويرمز له بالرمز (r) وقيمته تقع بين (-1,+1) ، ($-1\leq r \leq +1$) .

- ففي حالة r=1 تكون العلاقة تامة وهذا يدل على ان المتغيرات معتمدة(غير مستقلة)
- في حالة r=0 لا توجد أي علاقة وهذا يدل على ان المتغيرات مستقلة ، او ان هناك علاقة غير خطية .

- والاشارة تدل على اتجاه العلاقة فعندما تكون الاشارة (+) موجبة يقال ان الارتباط موجب والتي تعني ان كل زيادة في المتغير المستقل (X) تؤدي الى زيادة في المتغير التابع (Y) .
- ويكون الارتباط سالبا (-) اذا كانت الزيادة في قيمة (X) تؤدي الى نقصان في (Y).
- اما في حالة الزيادة في (X) ولا تؤدي الى أي تغيير في (Y) فذلك يشير الى عدم وجود أي علاقة بين المتغيرين ..

والمخطط التالي يوضح درجات قوة معامل الارتباط

قوي جدا *	قوي *	متوسط *	ضعيف *	ضعيف *	*	متوسط *	قوي *	قوي جدا *
-1	-0.9	-0.7	-0.5	0	0.5	0.7	0.9	+1

علاقة عكسية ← ← → علاقة طردية

معامل الارتباط البسيط : Simple Correlation Coefficient

ويستخدم لقياس العلاقة بين متغيرين ذات قيم رقمية ويعتبر (معامل الارتباط بيرسون) من اهم الطرق المستخدمة في حالة الارتباط البسيط .

حساب معامل الارتباط البسيط :

هو حاصل قسمة التباين المشترك للمتغيرين الى حاصل ضرب انحرافهما المعياريين وصيغته كما يأتي:

$$r = \frac{Sxy}{\sqrt{[S_{xx}\ ]*[S_{yy}]}}$$

$$S_{xy} = \sum_{i=1}^{n} XY - \frac{(\sum_{i=1}^{n} X)(\sum_{i=1}^{n} Y)}{n}$$

$$S_{xx} = \sum_{i=1}^{n} X^2 - \frac{(\sum_{i=1}^{n} X)^2}{n}$$

$$S_{yy} = \sum_{i=1}^{n} Y^2 - \frac{(\sum_{i=1}^{n} Y)^2}{n}$$

حيث ان :

r: تمثل معامل الارتباط البسيط بين المتغيرين (x,y)

S_{xy}: تمثل التباين(التغاير) المشترك بين المتغيرين (x,y)

S_{xx} : تمثل التباين للمتغير x

S_{yy} : تمثل التباين للمتغير y

مثال: البيانات التالية تمثل الدرجات النهائية لـ (8) طلاب في مادتي الاحصاء X والرياضيات Y.

X	85	65	60	45	80	65	85	95
Y	77	67	62	52	72	57	82	87

المطلوب : ايجاد معامل الارتباط بين المتغيرين X وY مع التفسير ؟

الحل : نعمل الجدول التالي:

	X	Y	XY	X^2	Y^2
	85	77	6545	7225	5929
	65	67	4355	4225	4489
	60	62	3720	3600	3844
	45	52	2340	2025	2704
	80	72	5760	6400	5184
	65	57	3705	4225	3249
	85	82	6970	7225	6724
	95	87	8265	9025	7569
المجموع	580	556	41660	43950	39692

$$S_{xy} = \sum_{i=1}^{n} XY - \frac{(\sum_{i=1}^{n} X)(\sum_{i=1}^{n} Y)}{n}$$

$$= 41660 - \frac{(580)(556)}{8} = 41660 - 40310 = 1350$$

$$S_{xx} = \sum_{i=1}^{n} X^2 - \frac{(\sum_{i=1}^{n} X)^2}{n}$$

$$= 43950 - \frac{(580)^2}{8} = 43950 - 42050 = 1900$$

$$S_{yy} = \sum_{i=1}^{n} Y^2 - \frac{(\sum_{i=1}^{n} Y)^2}{n}$$

$$= 39692 - \frac{(556)^2}{8} = 39692 - 38642 = 1050$$

$$r = \frac{Sxy}{\sqrt{[S_{xx}] * [S_{yy}]}} = \frac{1350}{\sqrt{[1900] * [1050]}}$$

$$= \frac{1350}{\sqrt{[1995000]}} = \frac{1350}{1412.44} = 0.95579$$

r = 0.96

التفسير:

هناك علاقة ارتباط قوي جدا والاشارة الموجبة تدل على هناك علاقة طردية بين المتغيرين أي انه كلما ارتفعت درجة الطالب في مادة الاحصاء فمن المتوقع ان ترتفع درجته في مادة الرياضيات ايضا .

تحليل الانحدار : Regression Analysis

تحليل الانحدار : هو وسيلة احصائية تستخدم لتحليل البيانات التي تحتوي على متغيرين فاكثر عندما يكون الهدف هو اكتشاف طبيعة هذه العلاقة، ويعد تحليل الانحدار من اكثر الطرق الاحصائية استعمالا في مختلف العلوم لأنه يصف العلاقة بين المتغيرات على هيئة معادلة .

استخدامات تحليل الانحدار :

ويستخدم تحليل الانحدار لعدة اغراض اهمها :

1- وصف البيانات

2- تقدير المعلمات لإمكانية الاستدلال على اهمية وقوة العلاقة بين المتغيرات.

3- التنبؤ من خلال تقدير الاستجابة .

4- السيطرة : اذ يمكن السيطرة على قيم المتغير المعتمد وذلك بتغيير قيم المتغيرات التوضيحية .

ويقسم الانحدار الى نوعين :

1- الانحدار البسيط 2- الانحدار المتعدد

وسوف نتناول في دراستنا فقط تحليل الانحدار البسيط

تحليل الانحدار الخطي البسيط : Simple Linear Regression Analysis

ونعني بالبسيط عندما يكون المتغير المعتمد (Y) دالة بدلالة المعلمات (B) ومتغير توضيحيا واحدا هو المتغير المستقل (X) فقط وتكون صيغة معادلة الانحدار الخطي البسيط كما ياتي:

$$Y = B_0 + B_1 X + U_i$$

حيث ان:

Y: يمثل المتغير التابع او المعتمد Dependent Variable

B_0: تمثل معلمة الحد الثابت وتمثل متوسط الاستجابة (Y) عندما (X=0).

B_1: تمثل معلمة الميل وتسمى ايضا معلمة الانحدار ، وتمثل مقدار التغير في Y عند زيادة قيمة المتغير X وحدة واحدة .

X : ويمثل المتغير التوضيحي (المستقل)

U: يمثل حد الخطأ العشوائي .

تقدير المعلمات :

هناك عدة طرائق لتقدير معلمات معادلة الانحدار ومن اهم هذه الطرائق هي طريقة المربعات الصغرى الاعتيادية **Ordinary Least Square** ، ويرمز لها بالرمز اختصارا (OLS) والتي تستند على مبدأ تصغير مجموع مربعات الخطأ.

ويتم احتساب المعلمات المقدرة حسب الصيغ التالية :

$$\hat{B}_1 = \frac{\sum_{i=1}^{n} XY - \frac{(\sum_{i=1}^{n} X)(\sum_{i=1}^{n} Y)}{n}}{\sum_{i=1}^{n} X^2 - \frac{(\sum_{i=1}^{n} X)^2}{n}}$$

$$\hat{B}_0 = \bar{Y} - \hat{B}_1 \bar{X}$$

مثال 1:

البيانات التالية تمثل دراسة انتاج محصول البطاطا بدلالة المساحة المزروعة من ذلك المحصول .

المطلوب: اوجد معادلة الانحدار المقدرة .

المزرعة	1	2	3	4	5	6	7	8	9	10
المحصول(الف دونم) (Y)	140	500	400	300	356	240	200	34	70	19
المساحة المزروعة (هكتار) (X)	50	200	110	80	120	75	89	6	11	3

مثال 2:

من بيانات الجدول الاتي:

X	1	2	3	4	5	6
Y	4	6	7	7	9	11

المطلوب: اوجد معادلة الانحدار التقديرية .

مثال 3:

من البيانات التالية اوجد معادلة الانحدار التقديرية .

n = 8

$$\sum X = 2108 \quad ; \quad \sum Y = 5264 \quad ; \quad \sum XY = 1373536 \quad ; \quad \sum X^2 = 567498$$

حل المثال(1)

	Y	X	X^2	XY
	140	50	2500	7000
	500	200	40000	100000
	400	110	12100	44000
	300	80	6400	24000
	356	120	14400	42720
	240	75	5625	18000
	200	89	7921	17800
	34	6	36	204
	70	11	121	770
	19	3	9	57
المجموع	2259	744	89112	254551

$$\hat{B}_1 = \frac{\sum_{i=1}^{n} XY - \frac{(\sum_{i=1}^{n} X)(\sum_{i=1}^{n} Y)}{n}}{\sum_{i=1}^{n} X^2 - \frac{(\sum_{i=1}^{n} X)^2}{n}}$$

$$= \frac{254551 - \frac{(744)(2259)}{10}}{89112 - \frac{(744)^2}{10}} = \frac{254551 - 168069.6}{89112 - 55353.6} = \frac{86481.4}{33758.4} = 2.56$$

$$\bar{X} = \frac{\sum X_i}{n} = \frac{744}{10} = 74.4$$

$$\bar{Y} = \frac{\sum Y_i}{n} = \frac{2259}{10} = 225.9$$

$$\hat{B}_0 = \bar{Y} - \hat{B}_1 \bar{X}$$

= 225.9 - (2.56) (74.4)

= 225.9 - 190.464 = 35.436

اذن معادلة الانحدار التقديرية هي:

$\hat{Y} = 35.436 + 2.56 X$

حل المثال 2:

نكون الجدول التالي :

	Y	X	X^2	XY
	4	1	1	4
	6	2	4	12
	7	3	9	21
	7	4	16	28
	9	5	25	45
	11	6	36	66
المجموع	44	21	91	176

$$\hat{B}_1 = \frac{\sum_{i=1}^{n} XY - \frac{(\sum_{i=1}^{n} X)(\sum_{i=1}^{n} Y)}{n}}{\sum_{i=1}^{n} X^2 - \frac{(\sum_{i=1}^{n} X)^2}{n}}$$

$$= \frac{176 - \frac{(21)(44)}{6}}{91 - \frac{(21)^2}{6}} = \frac{176 - 154}{91 - 73.5} = \frac{22}{17.5} = 1.26$$

$$\bar{X} = \frac{\sum X_i}{n} = \frac{21}{6} = 3.5$$

$$\bar{Y} = \frac{\sum Y_i}{n} = \frac{44}{6} = 7.33$$

$$\hat{B}_0 = \bar{Y} - \hat{B}_1 \bar{X}$$

$= 7.33 \; - (1.26)\,(3.5)$

$= 7.33 \; - \; 4.41 \; = \; 2.92$

اذن معادلة الانحدار التقديرية هي:

$\hat{Y} = 2.92 + 1.26\ X$

حل المثال 3:

$$\hat{B}_1 = \frac{\sum_{i=1}^{n} XY - \frac{(\sum_{i=1}^{n} X)(\sum_{i=1}^{n} Y)}{n}}{\sum_{i=1}^{n} X^2 - \frac{(\sum_{i=1}^{n} X)^2}{n}}$$

$$= \frac{1373536 - \frac{(2108)(5264)}{8}}{567498 - \frac{(2108)^2}{8}} = \frac{1373536 - 1387064}{567498 - 555458} = \frac{-13528}{12040} = -1.12$$

$$\bar{X} = \frac{\sum X_i}{n} = \frac{2108}{8} = 263.5$$

$$\bar{Y} = \frac{\sum Y_i}{n} = \frac{5264}{8} = 658$$

$$\hat{B}_0 = \bar{Y} - \hat{B}_1 \bar{X}$$

$= 658 \quad - (-1.12)(263.5)$

$= 658 \quad + 295.12 \quad = 953.12$

اذن معادلة الانحدار التقديرية هي:

$\hat{Y} = 953.12 - 1.12 \; X$