جامعة جيجل

كلية العلوم الاجتماعية والإنسانية

قسم التعليم الأساسي للعلوم الاجتماعية

السنة الأولى

محاضرات في الإحصاء الوصفي

مدخل للإحصاء الوصفي
مقاييس النزعة المركزية
مقاييس التشتت

أستاذ المقياس:

د. غديري

السنة الجامعية: 2022 / 2023

مدخل إلى الإحصاء الوصفي

- المصطلحات الإحصائية:

- 1 - المجتمع الإحصائي (Population): هو مجموعة العناصر أو الأفراد التي ينصب عليها الاهتمام في دراسة معينة أو مجموعة المشاهدات والقياسات الخاصة بمجموعة من الوحدات الإحصائية والتي تخص ظاهرة من الظواهر القابلة للقياس مثل: مجتمع من الطلبة، مجتمع من الأسر، مجتمع من المؤسسات وغير ذلك ، ويمكن تقسيم المجتمع الإحصائي إلى نوعان:

- **مجتمع الهدف (Target Population):** هو عبارة عن مجموع الأشخاص أو العناصر المستهدفين بالدراسة، والذين نريد تعميم نتائج الدراسة عليهم، مثل طلبة الجامعات، موظفي القطاع العام في القطاع الصحي في دولة ما...إلخ.

- **مجتمع الدراسة (Study Population):** هو مجموعة الأفراد الذين أتيح لنا الحصول منهم على البيانات أو المعلومات المتعلقة بالظاهرة أو المشكلة التي يراد بحثها ودراستها.

وغالبا ما يكون المجتمع الإحصائي كبيرا وبالتالي دراسة جميع مفرداته قد يكون أمر غير متيسر وعليه نلجأ لدراسة جزء من مفرداته يطلق عليه تسمية عينة.

- 2 - العينة (Sample): هي مجموعة جزئية من مفردات المجتمع الإحصائي، وقد جرت العادة على اختيار مفردا ته بحيث يكون لكل مفردة من مفردات المجتمع نفس الفرصة بأن تكون ضمن مفردات العينة، وذلك حتى تمثل المجتمع أحسن تمثيل، ويختلف حجم العينة حسب أهمية الدراسة وحسب الإمكانيات المادية والبشرية المتاحة للقيام بهده الدراسة، ويعتبر أسلوب المعاينة من الأساليب المتبعة في أغلب الدراسات الميدانية وهذا لاستحالة جمع المعلومات الإحصائية من الوحدات التي تشكل المجتمع المدروس أو ما يسمى الحصر الشامل.

3 - الوحدة الإحصائية (Statistical Unit): هي العنصر أو الجزء الذي تجرى عليه الدراسة الإحصائية أو المعاينة، ويشترط في الوحدة أن تكون خاضعة لتعريف دقيق وواضح.

- 4 - الظاهرة الإحصائية (Statistical Phenomenon): هي صفة لعناصر تختلف من عنصرلآخر في الشكل أو النوع أو الكمية ويطلق على الصفة تحت الدراسة اسم متغير (Variable) مثلا: القامة ،السن، الوزن، الإنتاج، الادخار، الاستثمار، الاستهلاك....

- 5 - المتغيرة (Variables): هي الصفات أو السمات التي يتصف بها أفراد عينة ما، حيث تتغير هذه الصفات من عنصر لآخر مثل: أطوال الأطفال تحت سن معين في مدينة ما، سعر مادة معينة في سوق ما وغيرها.

- 6 - المعلمة (Parametre): هو المقياس أو الثابت الذي يصف بعض خصائص المجتمع، ونحصل عليه من خلال تحليل البيانات لهذا المجتمع، وهذه المقاييس أو الثوابت نحصل عليها من خلال اعتماد عملية المسح الشامل في العادة، مثل: متوسط دخل الفرد في بلد معين يعتبر معلمة وذلك لأنه يعكس المستوى المعيشي لأفراد ذلك البلد، ويرمز للمعلمة بأحد الحروف اللاتينية مثلا متوسط المجتمع الإحصائي وتباينه.

-المتغيرة الإحصائية:

المتغيرة الإحصائية هي تلك الصفة أو الكمية القابلة للتغير من فرد لآخر أو من مشاهدة لأخرى والتي تسمح بتفريق هؤلاء عن أولئك وتصنفهم ويطلق على القيمة التي تعطى لها اسم قيمة المتغير الإحصائي، ويمكن تصنيف المتغيرة الإحصائية إلى قسمين:

-1- المتغيرة النوعية "الكيفية" (Qualitative Variables):

هي تلك المتغيرات أو الظواهر التي لا يمكن قياسها عدديا بل قياس تكرارها فقط، أو هي عبارة عن صفات أو أنواع ليست عددية وتنقسم بدورها إلى:

- **بيانات نوعية قابلة للترتيب**: يمكن ترتيبها حسب رتبة معينة تصاعديا أو تنازليا مثل: مستويات النمو الاقتصادي، المستوى التعليمي، الرتب العسكرية، تقديرات النجاح وغيرها.

- **بيانات نوعية غير قابلة للترتيب**: مثل: الجنسية، أنواع الأمراض، الحالة العائلية وغيرها.

2- المتغيرة الكمية "العددية" (Quantitative Variables):

هي تلك المتغيرات التي يمكن التعبير عنها عدديا بأرقام حقيقية وقياسها رقميا وهي أكثر المتغيرات انتشارا واستعمالا لأن لغة الإحصاء هي لغة الأرقام مثال على ذلك: الإنتاج، الاستهلاك، الوزن، الطول، الاستثمار وغيرها وتنقسم بدورها إلى قسمين:

- **المتغيرة الكمية المنفصلة "المتقطعة"**

هي تلك المتغيرات التي تأخذ قيما صحيحة لا يمكن تجزئتها مثل: عدد الأطفال في الأسرة، عدد الطلاب في مراحل التعليم المختلفة، عدد الغرف بالبيت، عدد قطع الغيار المنتجة...إلخ.

- **المتغيرة الكمية المتصلة "المستمرة"**

هي تلك المتغيرات التي تأخذ كل القيم الممكنة ال الدراسة، ونظرا للعدد غير المنتهي لهذه القيم نقسم مجال الدراسة إلى مجالات جزئية تسمى فئات.

وبصفة عامة يعتبر المتغير مستمرا إذا كان مرتبط بالزمن (السرعة، السن...)، أو الكتلة (الوزن، الكثافة...)، النقود (الدخل، الأجر، السعر ...) أو الفضاء (الطول، المساحة...)، ونرمز للمتغيرة بالرمز X_i وللقيمة التي تقابلها بالرمز x_i .

العرض الجدولي للبيانات الإحصائية

-العرض الجدولي للبيانات الإحصائية:

1-تعريف العرض الجدولي: العرض الجدولي للبيانات يقصد به وضع البيانات الأولية الخاصة بالظاهرة بعد جمعها في جداول احصائية تتكون في الأساس من عمودين

(سطرين)، يبين العمود (السطر) الأول قيم الظاهرة أو المتغير المدروس، وتكون هذه القيم على شكل صفات أو قيم نقطية أو مجالات (فئات)، أما العمود (السطر) الثاني فيحتوي على تكرارات هذه الصفات أو القيم أو ا لمشاهدات.

2 - أنواع الجداول الإحصائية:

تختلف الجداول الإحصائية باختلاف نوع البيانات من ناحية والغرض من الدراسة من ناحية أخرى ومن أهمها:

- جداول التوزيع التكراري البسيطة:

يمثل طريقة تنظيم البيانات الخام للظاهرة (المتغير) وتبويبها في جداول تضم صفات أو قيم الظاهرة والتكرارات المناظرة لها لغرض دراستها وتحليلها، ويستخدم هذا النوع من الجداول لوصف وتلخيص البيانات التي تتعلق بظاهرة واحدة فقط سواء كانت كيفية أو كمية، ولتوضح شكل جداول التوزيع التكراري البسيطة نقوم بعرض الجدول التالي:

الجدول رقم: الشكل العام لجدول التوزيع التكراري البسيط

في حالة متغير كيفي نضع في هذه الخانات صفة المتغير.
في حالة متغير كمي منفصل نكتب قيم المتغير ويجب أن ترتب تصاعديا أو تنازليا.
في حالة متغير كمي متصل تكتب قيم المتغير على

المتغير X_i	التكرار المطلق n_i
x_1	n_1
x_2	n_2
.	.
.	.
.	.
x_k	n_k
المجموع $\sum$	$N = \sum n_i$

نضع في هذه الخانات عدد المفردات المقابلة لكل صفة أو قيمة أو فئة للمتغير الإحصائي

تختلف طريقة العرض الجدولي حسب نوع المتغير لذلك نميز بين الحالات التالية:

بيانات المتغيرات الكمية المنفصلة (المتقطعة):

هي المتغيرات التي تأخذ بيانا او أرقام عددية صحيحة فقط مثل عدد طلبة الجامعة أو عدد العمال وغيرها ،ولغرض تبويب بيانات المتغيرات المنفصلة يتم تصنيفها إلى مجموعات متشا بهة، ثم وضعها في جداول مكونة عمودان، يخصص العمود الأول لقيم الظاهرة (المتغير) بعد ترتيبها والعمود الثاني للتكرارات، والمثال أدناه يعطي توضيحا لذلك:

مثال (2-02): البيانات التالية تمثل عدد الأفراد في عينة مكونة من 30 أسرة:

2	3	2	4	5	2	4	4	5	2	2	4	3	4	5
3	5	4	3	4	5	4	5	3	4	5	3	4	5	3

المطلوب: عرض البيانات في جدول توزيع تكراري؟

الحل:

الجدول: توزيع الأسر حسب عدد الأفراد (متغير كمي منفصل)

حجم الأسرة (X_i)		عدد الأسر (التكرارات) (n_i)
2		5
3		7
4		10
5		8
المجموع		30

بيانات المتغيرات الكمية المتصلة (المستمرة):

هي أكثر المتغيرات استخداما ويمكن أن تأخذ مفردا ا أرقام صحيحة وكسرية، فعند دراسة متغير كمي مستمر يضم مجال الدراسة ما لا نهاية من القيم، ولتعذر وضع كل هذه القيم نقسم هذا القيم إلى مجالات جزئية تسمى فئات، حيث يحدد عدد الفئات حسب حجم العينة

وحسب توزيع الوحدات الإحصائية على مجال الدراسة ،ولتكوين جدول التوزيع التكراري لمتغيرة كمية متصلة نتبع الخطوات التالية:

1- تحديد المدى (Range): المدى هو ا ال الذي تنتشر فيه البيانات وهو الفرق بين أكبر قيمة في البيانات وأصغر قيمة لها.

المدى = أكبر قيمة – أصغر قيمة

$$R = X_{max} - X_{min}$$

تحديد عدد الفئات: يتم تحديد عدد الفئات المطلوبة لتشكيل جدول التوزيع التكراري باستخدام بعض المعادلات الرياضية ومن هذه المعادلات

3-حديد طول الفئة: يتم تحديد طول الفئة بالعلاقة التالية:

طول الفئة = المدى ÷ عدد الفئات

$$L=R \div K$$

4- تحديد حدود الفئات:

في هذه المرحلة يتم تحديد بداية و نهاية كل فئة، على أن تكون بداية الفئة الأولى أصغر من أو تساوي أصغر قيمة في البيانات ونهاية الفئة الأخيرة أكبر من أكبر قيمة في البيانات.

5-تحديد عدد القيم أو المشاهدات:

يتم تحديد عدد القيم أو المشاهدات التي تقع في كل فئة على أن تكون لكل قيمة فئة واحدة فقط تنتمي إليها ،وهذا ما يسمى بالتكرار (n_i) والتأكد من أن مجموع التكرارات يساوي عدد القيم.

6-تحديد مراكز الفئات:

عند تكوين جداول التوزيع التكراري بفئات تضيع القيم الإحصائية الأصلية للمفردات وتصبح لا نعرف عنها شيئا سوى انها تنتمي إلى فئة معينة محدودة بحدين معلومين، ولتخطي هذه المشكلة نقوم باستخراج ما يسمى بمركز الفئة والذي نقصد به منتصف الفئة والذي نحصل عليه بالصيغة التالية:

$$\text{مركز الفئة} = \frac{\text{الحد الأدنى للفئة + الحد الأعلى للفئة}}{2}$$

$$C_i = \frac{L_i + L_{i+1}}{2}$$

حيث: L_i: الحد الأدنى للفئة i و L_{i+1} : الحد الأعلى للفئة i.

- الجدولة:

عملية الجدولة هي إفراغ البيانات في جدول التوزيع التكراري، مع مراعاة أن يكون لكل قيمة فئة واحدة وواحدة فقط والتأكد من أن مجموع التكرارات يساوي عدد القيم.

بعد إعداد جدول التوزيع التكراري يكون من المناسب في أغلب الأحيان عرض البيانات في شكل توزيع تكراري نسبي للتعبير عن الأهمية النسبية لتكرار كل فئة بالنسبة لإجمالي التكرارات، ويحسب التكرار النسبي بالصيغة التالية:

$$\text{التكرار النسبي} = \frac{\text{تكرار الصفة (المتغير أو الفئة)}}{\text{إجمالي التكرارات}}$$

$$f_i = \frac{n_i}{\sum n_i}$$

مع العلم أن مجموع التكرارات تساوي عدد القيم أو المشاهدات $\sum n_i = N$ و $\sum f_i = 1$

ويمكن تحويل التكرار النسبي إلى تكرار نسبي مئوي وهذا بضربه في 100، وتعطى علاقة التكرار النسبي المئوي بالصيغة التالية:

$$f_i\% = f_i \times 100$$

مع العلم أن: $\sum f_i\% = 100\%$

يفيد التكرار النسبي في تقليص الشكل البياني عندما يكون عدد القيم كبيرا، بينما يفيد التكرار النسبي المئوي في إظهار الشكل البياني عندما يكون عدد القيم صغيرا.

كما يمكن أن نحتاج إلى معلومات إضافية عن البيانات، فمثلا: قد نحتاج إلى معرفة المفردات التي تقل قيمتها أو تزيد عن حد معيّن، وهذه المعلومات نحصل عليها من خلال إيجاد التكرارات المتجمعة الصاعدة والنازلة:

❖ التكرار المتجمع الصاعد:

يمثل التكرار المتجمع الصاعد لفئة معينة مجموع الأفراد الذين تقل قيمتهم الإحصائية عن الحد الأعلى للفئة المقابلة.

❖ التكرار المتجمع النازل):

يمثل التكرار المتجمع النازل لفئة معينة مجموع الأفراد الذين تزيد قيمتهم الإحصائية عن الحد الأدنى للفئة المقابلة.

باستخدام معطيات المثال السابق نقوم بإيجاد التكرارات النسبية والنسبية المئوية وكذا التكرارات المتجمعة الصاعدة والنازلة:

الجدول: يوضح التكرار النسبي والنسبي المئوي والتكرار المتجمع الصاعد والنازل

فئات الأجر X_i	التكرار n_i	التكرار النسبي f_i	التكرار النسبي المئوي $f_i\%$	التكرار المتجمع الصاعد $n_i \uparrow$	التكرار المتجمع النازل $n_i \downarrow$
$]23-14]$	4	0,095	9,5	4	42
$]32-23]$	5	0,12	12	9=5+4	38=4−42
$]41-32]$	7	0,17	17	16=7+9	33=5−38
$]50-41]$	7	0,17	17	23=7+16	26=7−33
$]59-50]$	11	0,26	26	34=11+23	19=7−26
$]68-59]$	8	0,19	19	42=8+34	8=11−19
$\sum$	42	1	100	–	–

العرض البياني للبيانات الإحصائية

II-العرض البياني للبيانات الإحصائية:

بالإمكان وصف وتلخيص البيانات باستخدام الرسومات البيانية والأشكال الهندسية، إذ تمكن هذه الأخيرة من القيام بتحليل سريع للظاهرة المدروسة، وتستخدم أ نواع مختلفة للعرض البياني حسب نوع المتغير المدروس.

II-1 - العرض البياني في حالة متغير كمي منفصل:

I-2 - العرض البياني في حالة متغير كمي متصل (مستمر):

إن العروض البيانية للمتغير الكمي المتصل هي أكثر العروض البيانية استعمالا ومن أهمها:

II-2 -1- المدرج التكراري:

وهو عبارة عن مستطيلات (أعمدة) متلاصقة تمثل تكرارات أو قيم كل فئة من الفئات، حيث أن طول كل منها يتناسب مع التكرار المقابل، وقاعدة كل منها تساوي طول الفئة المقابلة، حيث توضع الفئات على محور السينات، بينما توضع التكرارات على محور العينات .

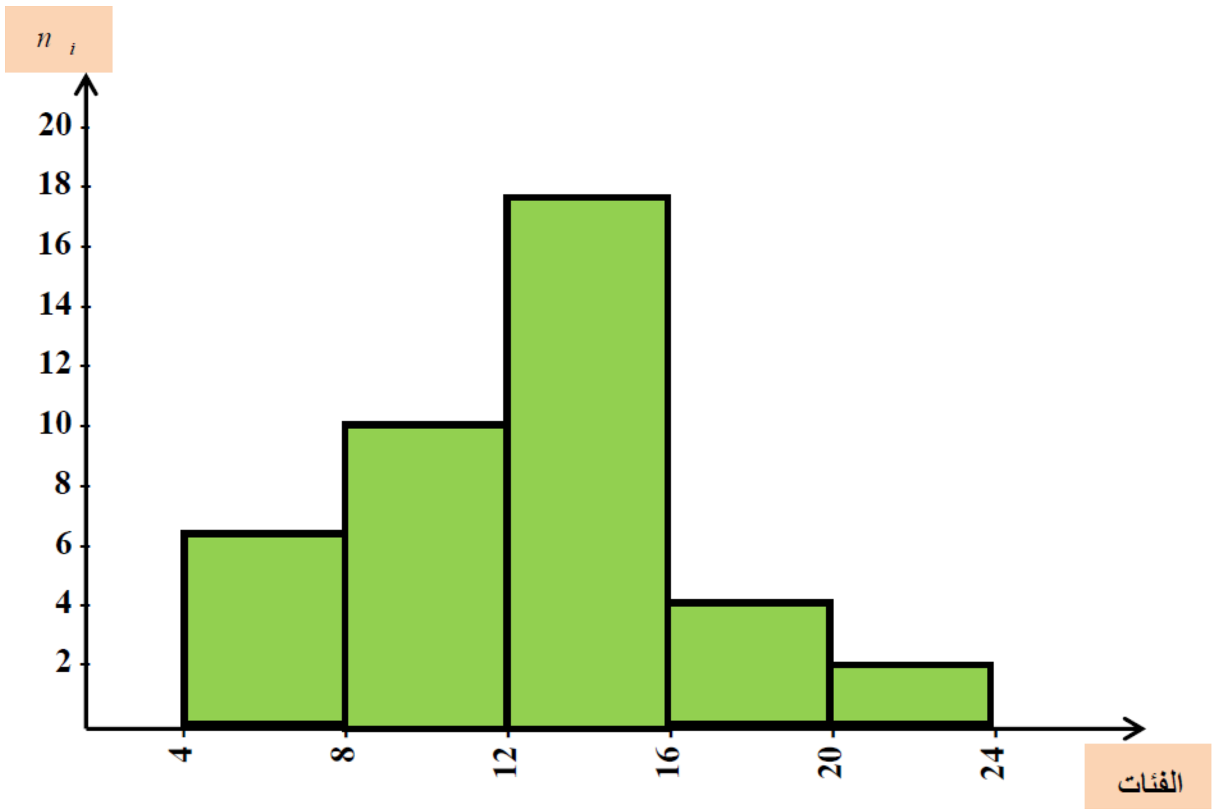
- **المدرج التكراري في حالة فئات متساوية الطول:** عندما تكون الفئات متساوية الطول فإن قاعدة المقارنة ثابتة ومتساوية وعليه نقوم برسم المدرج التكراري مباشرة، ولتوضيح ذلك نأخذ المثال التالي:

مثال: الجدول التالي يبيّن توزيع النفقات اليومية (الوحدة: 10 دج) لعينة من 40 طالب.

فئات النفقات X_i	$]8-4]$	$]12-8]$	$]16-12]$	$]20-16]$	$]24-20]$	$\sum$
التكرار n_i	**6**	**10**	**18**	**4**	**2**	**40**

المطلوب: تمثيل هذه البيانات باستخدام المدرج التكراري؟

المدرج التكراري لتوزيع النفقات

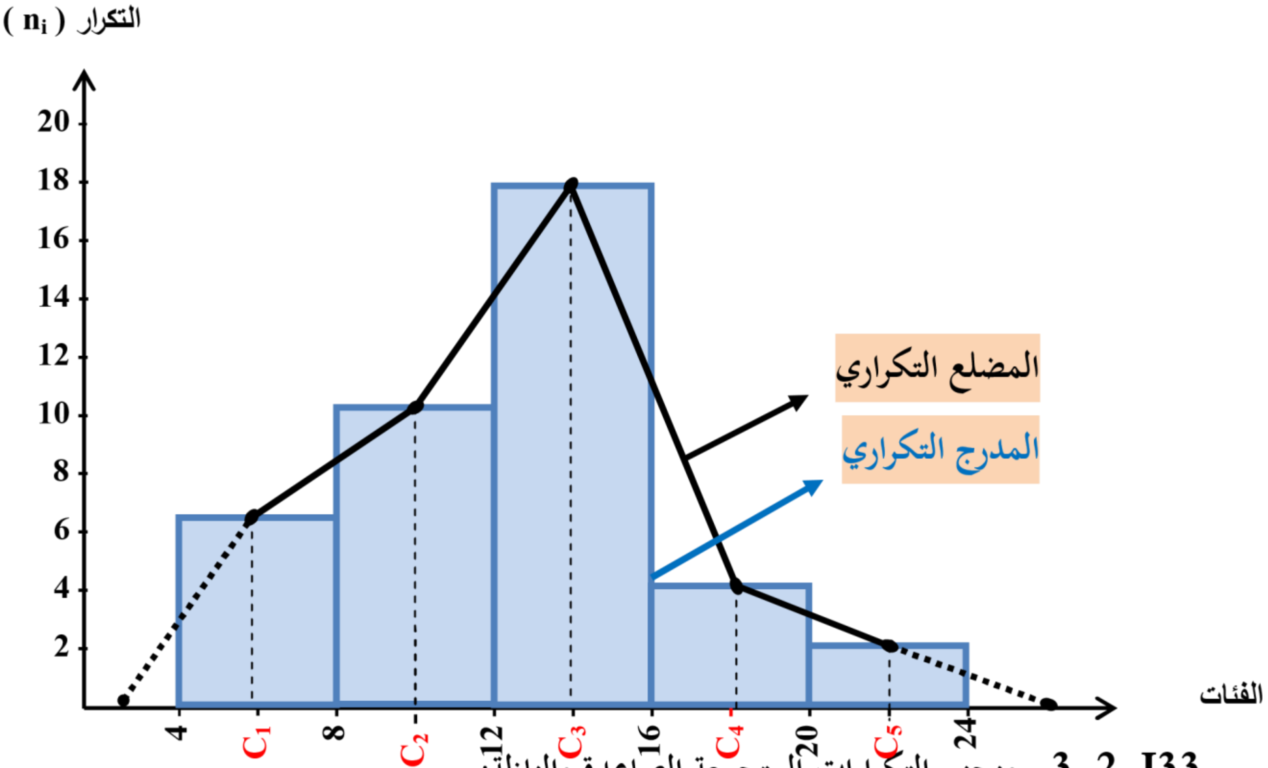


2- المضلع التكراري:

هو مجموعة من القطع المستقيمة المتصلة والمنكسرة تتحدد بنقاط إحداثيا تها: مراكز الفئات والتكرارات المقابلة لها.

ولتوضيح رسم المضلع التكراري نأخذ المثال ، ونقوم برسم المدرج والمضلع التكراري

الشكل: المدرج والمضلع التكراري



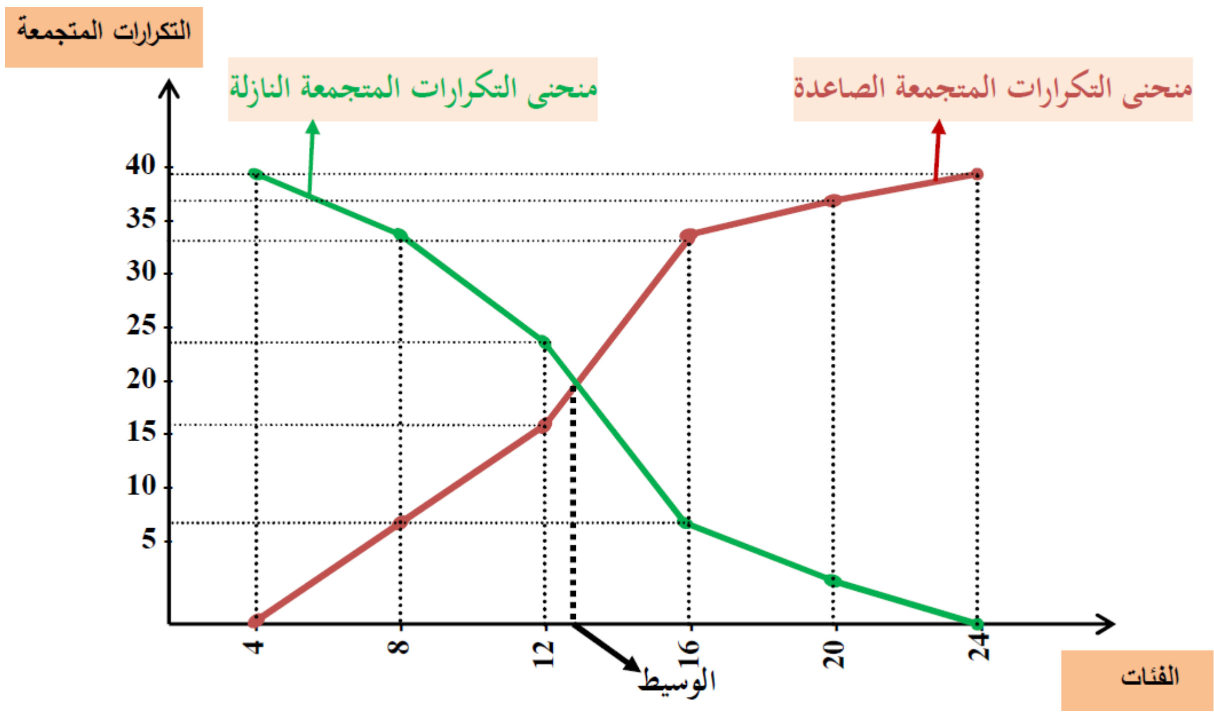
I33-2-3- منحنى التكرارات المتجمعة الصاعدة والنازلة:

يتم رسم منحنى التكرار المتجمع الصاعد عن طريق إيصال مجموعة النقاط ذات الإحداثيات التالية: الحدود العليا للفئات والتكرارات المتجمعة الصاعدة المقابل لها، أما منحنى التكرار المتجمع النازل فيتم رسمه بإيصال مجموعة النقاط ذات الإحداثيات: الحدود الدنيا للفئات والتكرارات المتجمعة النازلة المقابلة لها، وتسمى نقطة التقاطع بين المنحنيين بالوسيط ، ولغرض رسم منحنى التكرار المتجمع الصاعد والنازل نأخذ المثال

الجدول: التكرا رات المتجمع الصاعدة والنازلة

فئات النفقات X_i	التكرار n_i	$n_i\uparrow$	$n_i\downarrow$
$]8-4]$	6	6	40
$]12-8]$	10	16	34
$]16-12]$	18	34	24
$]20-16]$	4	38	6
$]24-20]$	2	40	2
$\sum$	40	–	–

منحنى التكرا رات المتجمعة الصاعدة والنازلة



مقاييس النزعة المركزية

I-مفهوم النزعة المركزية:

مقاييس النزعة المركزية هي تلك المقاييس التي تبحث في تقدير قيمة تتمركز حولها أغلبية القيم وهذه القيمة المتوسطة أو المتمركزة هي رقم واحد يعبر عن أو يمثل جميع بيانات تلك المجموعة.

إن مقاييس النزعة المركزية لا تحل محل البيانات التفصيلية ولكنها تعطي فكرة واضحة عن الظاهرة قيد الدراسة.

II-قياس النزعة المركزية:

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

II-1 - المتوسط الحسابي (Arithmetic Mean):

يعتبر المتوسط الحسابي من أشهر وأكثر متوسطات النزعة المركزية استخداما وشيوعا في الإحصاء وهو مركز التوازن لأي ظاهرة، ويمكن تعريفه بأنه: " عبارة عن حاصل قسمة مجموع قيم البيانات **i** على عددها **n** في حالة العينة، وعلى **N** في حالة ا لمجتمع" ، ويرمز له بالرمز x.

II-1 -1- طرق حساب المتوسط الحسابي:

بالنظر إلى أهمية المتوسط الحسابي في التحليل الإحصائي، فهناك عدة طرق لحسابه، حيث يمكن حسابهمن البيانات الأولية (غير المبوبة) أو البيانات المبوبة.

II-1 -1-1- حساب المتوسط الحسابي من البيانات الأولية:

الطريقة المباشرة:

يستعمل هذا المتوسط في حالة البيانات غير المبوبة، أي عندما يكون لقياسات المتغير المدروس نفس المستوى من الأهمية، فإذا كانت لدينا: $x_1 , x_2 ,......, x_n$ تمثل قيم ظاهرة ما، فإن المتوسط الحسابي لهذه القيم هو مجموع هذه القيم مقسوما على عددها، وتعطى علاقة المتوسط الحسابي بالصيغة التالية:

$$\overline{X} = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

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

حيث **n** : تمثل عدد القيم

II-1 -1-2- حساب المتوسط الحسابي من البيانات المبوبة:

الطريقة المباشرة:

في بعض الحالات، تكون القيم المراد حساب المتوسط الحسابي لها ليس لها نفس الأهمية، بل أهميات نسبية تختلف باختلاف عامل الترجيح الخاص ا، في مثل هذه الحالات يتطلب الأمر استخدام صيغة أخرى تسمى بصيغة المتوسط الحسابي المرجح، وهنا نميز حالتين:

✓ **حالة متغير كمي منفصل:**

إذا كانت: $x_1 , x_2 ,\ldots\ldots, x_k$ تمثل مفردات (قيم) ظاهرة معينة X، وكانت $n_1 ,n_2 ,\ldots\ldots, n_k$ تمثلالتكرارات المقابلة لها، في هذه الحالة تعطى علاقة المتوسط الحسابي بالصيغة التالية:

$$\overline{X} = \frac{x_1 n_1 + x_2 n_2 + \ldots\ldots + x_k n_k}{n_1 + n_2 + \ldots\ldots + n_k} = \frac{\sum_{i=1}^{k} x_i n_i}{\sum_{i=1}^{k} n_i}$$

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

حيث x_i : تمثل قيمة المفردة i ($i=1,2,\ldots\ldots,k$)

n_i : تمثل تكرار المفردة i fi : التكرار النسبي

k : عدد المفردات المختلفة عن بعضها البعض

مثال: البيانات الإحصائية التالية تمثل توزيع عائلات حي معيّن حسب عدد الغرف.

الجدول: توزيع العائلات حسب عدد الغرف

عدد الغرف X_i	1	2	3	4	$\sum$
عدد العائلات n_i	5	10	20	30	65

المطلوب: إيجاد متوسط الغرف المملوكة من طرف العائلات ؟

الحل : لدينا:

$$\overline{X} = \frac{\sum_{i=1}^{k} x_i n_i}{\sum_{i=1}^{k} n_i} \quad \text{أو} \quad \overline{X} = \sum_{i=1}^{k} x_i f_i$$

نقوم بإضافة عمود ثالث إلى الجدول تكون نتيجته حاصل ضرب قيمة المتغير

الإحصائي في التكرار المقابل لهذه القيمة

$(x_i \times n_i)$.

الجدول: يوضح حساب المتوسط الحسابي

عدد الغرف X_i	التكرارات n_i	$x_i n_i$
1	5	5
2	10	20
3	20	60
4	30	120
$\sum$	65	205

ومنه متوسط الغرف المملوكة من طرف العائلات هو:

$$\overline{X} = \frac{\sum_{i=1}^{k} x_i n_i}{\sum_{i=1}^{k} n_i} = \frac{\sum_{i=1}^{4} x_i n_i}{\sum_{i=1}^{4} n_i} = \frac{205}{65} = 3,15 \approx 3$$

✓ **حالة متغير كمي متصل:**

في هذه الحالة يكون جدول التوزيع التكراري على شكل فئات عددها k ، فإذا كانت: c_1 , c_2 ,……, c_k تمثل مراكز الفئات، وكانت n_1 ,n_2 ,….., n_k تمثل التكرارات المقابلة لتلك الفئات، فإننا نعتبر التكرارات تتجمع حول مراكز الفئات وعليه نستخدم الصيغة التالية لحساب المتوسط الحسابي:

$$\overline{X} = \frac{c_1 n_1 + c_2 n_2 + \ldots\ldots + c_k n_k}{n_1 + n_2 + \ldots\ldots + n_k} = \frac{\sum_{i=1}^{k} c_i n_i}{\sum_{i=1}^{k} n_i}$$

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

حيث c_i : مركز الفئة **i** (**i=1,2,…..,k**)

n_i : تكرار الفئة **i**

k : عدد الفئات

fi : التكرار النسبي

مثال:

الجدول الإحصائي التالي يبيّن توزيع عمال مؤسسة ما حسب أجورهم الشهرية (الوحدة: 10^3 دج).

الجدول: توزيع العمال حسب أجورهم الشهرية

الأجر X_i	$]20-10]$	$]30-20]$	$]40-30]$	$]50-40]$	$]60-50]$	$]70-60]$
عدد العمال n_i	**18**	**30**	**25**	**17**	**12**	**8**

المطلوب: إيجاد متوسط الأجر الشهري للعمال ؟

الحل:

حساب متوسط الأجر الشهري للعمال: لحساب متوسط الأجر الشهري للعمال نتبع الخطوات التالية:

1- إيجاد مجموع التكرارات n_i.

2- حساب مراكز الفئات c_i .

3- ضرب مركز كل فئة في التكرار المقابل لها ($c_i \times n_i$)، وحساب ا لمجموع.

4- حساب المتوسط الحسابي بتطبيق العلاقة:

$$\overline{X} = \frac{\sum_{i=1}^{k} c_i n_i}{\sum_{i=1}^{k} n_i}$$

أو

$$\overline{X} = \sum_{i=1} Ci \ fi$$

الجدول :يوضح حساب المتوسط الحسابي

$c_i \times n_i$	مركز الفئة c_i	التكرار n_i	فئات الأجر X_i
270	15	18	$]20-10]$
750	25	30	$]30-20]$
875	35	25	$]40-30]$
765	45	17	$]50-40]$
660	55	12	$]60-50]$
520	65	8	$]70-60]$
3840	–	110	$\sum$

ومنه متوسط أجر عمال هذه المؤسسة هو

$$\overline{X} = \frac{\sum_{i=1}^{k} c_i n_i}{\sum_{i=1}^{k} n_i} = \frac{\sum_{i=1}^{6} c_i n_i}{\sum_{i=1}^{6} n_i} = \frac{3840}{110} = 34,9 \times 10^3 \ DA$$

II-2 - الوسيط (Mediane):

الوسيط لمجموعة من البيانات هو القيمة التي تقسم تلك البيانات المرتبة ترتيبا تصاعديا أو تنازليا إلى قسمين متسا وبين، حيث يكون عدد القيم الأكبر منه مساويا لعدد القيم الأصغر منه، ويرمز للوسيط بالرمز M_e.

II-2-1- حساب الوسيط:

II-2-1-1- الوسيط للبيانات الأولية (غير المبوبة):

لحساب الوسيط من البيانات الأولية يجب أولا ترتيب البيانات (القيم) تصاعديا أو تنازليا وهنا نميز حالتين:

الوسيط للبيانات غير المبوبة المفردة:

إذا كان عدد القيم (n) فردي فإن الوسيط هو القيمة التي ترتيبها

$$\frac{n+1}{2}$$

وبالتالي قيمة الوسيط هي المفردة (القيمة) التي ترتيبها يقابل رتبة الوسيط

$$\left(M_e = X_{\frac{n+1}{2}}\right)$$

مثال: أوجد الوسيط للبيانات التالية:

9	**15**	**12**	**10**	**3**	**6**	**8**

الحل : ترتيب البيانات ترتيبا تصاعديا

X_7	X_6	X_5	X_4	X_3	X_2	X_1
15	**12**	**10**	9	**8**	**6**	**3**

$M_e = X_4 = 9$

عدد القيم n=7 عدد فردي هي وعليه فإن رتبته

$$\frac{n+1}{2}=\frac{7+1}{2}=4$$

وبالتالي قيمة الوسيط هي القيمة التي ترتيبها يقابل رتبة الوسيط

$$M_e = X_{\frac{n+1}{2}} = X_{\frac{7+1}{2}} = X_4 = 9$$

الوسيط للبيانات غير المبوبة الزوجية:

إذا كان عدد القيم (n) زوجي فرتبة الوسيط هي:

$$(\frac{n}{2},\frac{n}{2}+1)$$

، وبالتالي قيمة الوسيط تساوي المتوسط الحسابي للمفردتين المقابلتين لرتبة الوسيط أي:

$$\cdot(M_e = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2})$$

مثال: أوجد الوسيط للبيانات التالية:

7	13	1	3	10	8	6	16

الحل:

X_8	X_7	X_6	X_5	X_4	X_3	X_2	X_1
1	**3**	**6**	7	8	**10**	**13**	**16**

$$M_e = \frac{7+8}{2} = 7,5$$

عدد القيم (n=8 عدد زوجي)، وعليه فإن رتيبة الوسيط هي

$$\left(\left(\frac{8}{2}, \frac{8}{2}+1\right) = (4,5)\right)$$

، وبالتالي قيمة الوسيط تساوي المتوسط الحسابي للمفردتين المقابلتين لرتبة الوسيط أي:

$$M_e = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2} = \frac{X_{\frac{8}{2}} + X_{\frac{8}{2}+1}}{2} = \frac{X_4 + X_5}{2} = \frac{7+8}{2} = 7,5$$

II-2-1-2- الوسيط للبيانات المبوبة:

لحساب الوسيط من البيانات المبوبة فإننا نميز بين حالتين:

✓ **حالة متغير كمي منفصل:**

لحساب الوسيط لبيانات مبوبة في جدول توزيع تكراري (حالة متغير كمي منفصل)، نتبع الخطوات التالية:

1- نقوم بحساب التكرار المتجمع الصاعد؛

2– نحدد رتبة الوسيط ($\frac{N}{2}$) حيث $N = \sum_{i=1}^{k} n_i$؛

3- نبحث في العمود (السطر) الخاص بالتكرار المتجمع الصاعد عن القيمة التي تكرارها المتجمع الصاعد يساوي رتبة الوسيط

$$\left(\frac{N}{2}\right)$$

أو أعلى منها مباشرة، لتكون القيمة التي تقابلها هي قيمة الوسيط.

مثال: البيانات الإحصائية التالية تمثل توزيع عائلات حي معيّن حسب عدد الغرف.

عدد الغرف X_i	1	2	3	4	$\sum$
عدد العائلات n_i	5	10	20	30	65

المطلوب: إيجاد الوسيط لهذه البيانات ؟

لحساب الوسيط لهذه البيانات نتبع الخطوات التالية:

نقوم بحساب التكرار المتجمع الصاعد؛

الجدول يوضح كيفية حساب الوسيط

عدد الغرف X_i	عدد العائلات n_i	$n_i \uparrow$
1	5	5
2	10	15
3	20	35
4	30	65
$\sum$	65	-

الوسيط

رتبة الوسيط $\frac{N}{2} = 32,5$

2- نحدد رتبة الوسيط $\frac{N}{2} = \frac{65}{2} = 32,5$ حيث $N = \sum_{i=1}^{k} n_i = 65$؛

3- نبحث في العمود (السطر) الخاص بالتكرار المتجمع الصاعد عن القيمة التي تكرارها المتجمع الصاعد يساوي رتبة الوسيط أو أعلى منها مباشرة 35 أكبر مباشرة من 32.5 وعليه تكون قيمة الوسيط القيمة التي تقابل القيمة 35.

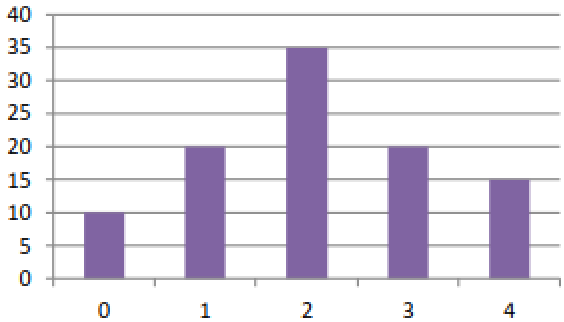
II-3 - المنوال: هو قيمة المتغير الاحصائي الذي له أكبر تكرار ونرمز له بـ Mo.

- **حالة متغير كمي منقطع:** تكون قيمة المنوال دقيقة ومحددة.

مثال: أنظر المثال السابق الذي يمثل توزيع 100 أسرة حسب عدد الأطفال:

عدد الأطفال	0	1	2	3	4	المجموع
عدد الأسر	10	20	35	20	15	100

يتضح من الجدول أن المنوال هو 2 لأنه يتوافق مع أكبر تكرار وهو 35، أما بيانيا فهو القيمة الموافقة لأطول عمود:



- **حالة متغير كمي مستمر:** في هذه الحالة أكبر تكرار يقابل فئة $[a_{mo}, b_{mo}[$ وتسمى الفئة المنوالية، ولحساب القيمة الدقيقة للمنوال نستعمل القاعدة التالية:

$$Mo = a_{mo} + C_{mo}\left(\frac{\alpha_1}{\alpha_1+\alpha_2}\right)$$

حيث:

a_{mo} : الحد الأدنى للفئة المنوالية

C_{mo} : طول الفئة المنوالية

α_1 : تكرار الفئة المنوالية – تكرار الفئة السابقة لها.

α_2 : تكرار الفئة المنوالية – تكرار الفئة الموالية لها.

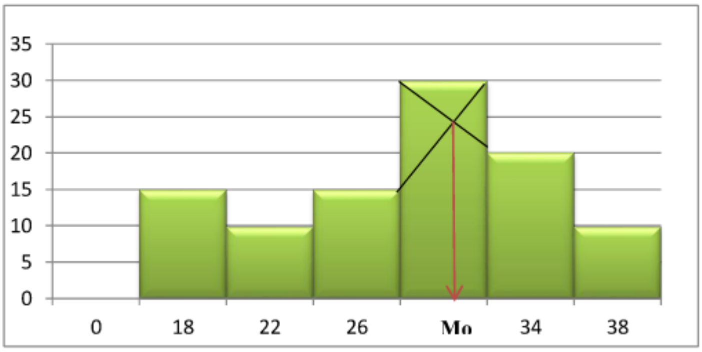
مثال: التوزيع التالي يمثل أعمار 100 طالب:

العمر	[16, 20[	[20, 24[	[24, 28[	[28, 32[	[32, 36[	[36, 40[	المجموع
عدد التلاميذ	15	10	15	30	20	10	100

من الجدول نلاحظ أن الفئة التي لها أكبر تكرار هي [28, 32[الفئة المنوالية ومنه فالمنوال يساوي:

$$Mo = 28 + 4 * 15/(15+10) = 30,4$$

كما يمكن تحديد قيمة المنوال بيانيا في حالة الفئات وذلك من خلال تمثيل المدرج التكراري، حيث نقوم برسم مستقيم يصل بين الحد الأعلى للفئة ما قبل المنوالية والحد الأعلى للفئة المنوالية ومستقيم آخر يصل بين الحد الأدنى للفئة المنوالية والحد الأدنى للفئة ما بعد المنوالية ونقوم بإسقاط نقطة التقاطع على محور الفواصل فنتحصل على المنوال.



❖ **الحالات الممكنة لقيم المنوال:**

1. توزيع تكراري وحيد المنوال: أي التوزيع الذي له قيمة واحدة فقط توافق أكبر تكرار.
2. توزيع تكراري ثنائي المنوال: أي التوزيع الذي له قيمتين لهما أكبر تكرار.
3. توزيع تكراري متعدد المنوال: هو التوزيع الذي له أكثر من منوالين، في هذه الحالة لا يكون للمنوال أي مدلول إحصائي، ولا يستعمل كمقياس إلا في حالات استثنائية كالانتخابات.
4. توزيع تكراري عديم المنوال: في هذه الحالة كل قيم المتغير الإحصائي لها نفس المستوى أو الأهمية أو التكرار أو بدون تكرار.

❖ **العلاقة بين الوسط الحسابي والوسيط والمنوال**

تعتمد العلاقة بين الوسط، الوسيط والمنوال على شكل التوزيع التكراري، فمن خلال رسم المنحنى الذي يمثل التوزيع نجد أن هناك ثلاثة أشكال هي:

المنحنى معتدل التوزيع: عندما يكون؛

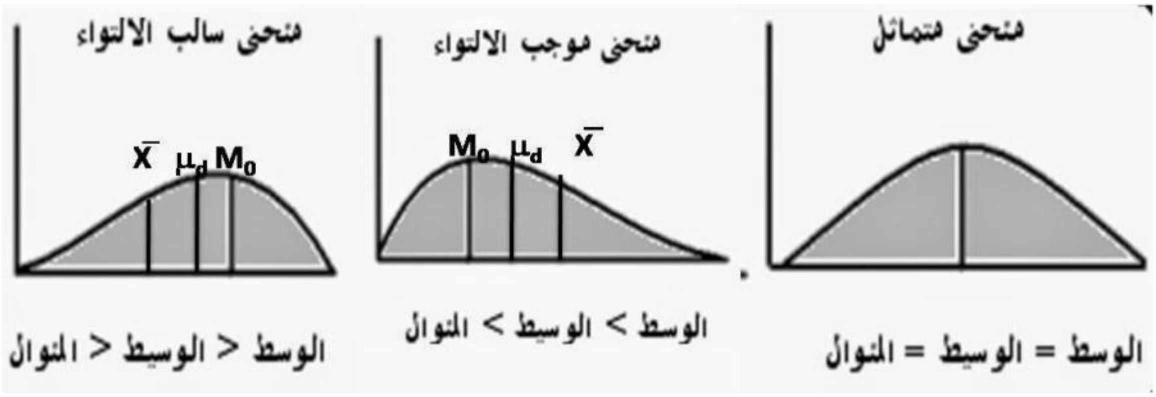
الوسط = الوسيط = المنوال

المنحنى موجب الالتواء: عندما يكون؛

الوسط > الوسيط > المنوال

المنحنى سالب الالتواء: عندما يكون؛

الوسط < الوسيط < المنوال



وبشكل عام فالعلاقة بين مقاييس النزعة المركزية تكتب من الشكل:

$$Mo = 3Me - 2\overline{X}$$

مثال: إذا علمت أن قيمة الوسط $\overline{X} = 5$ وقيمة الوسيط $Me = 10$

المطلوب: أحسب قيمة المنوال، ثم حدد نوع التواء التوزيع.

$$Mo = 3Me - 2\overline{X} = 3*10 - 2*5 = 20$$

نلاحظ أن: $\overline{X} < Me < Mo$ ؛ ومنه فالتوزيع ملتوي التواء سالب.

مقاييس التشتت

مقايس التشتت

توجد عدة طرق إحصائية لقياس التشتت تختلف فيما بينها من حيث الدقة والسهولة في العمل نذكر منها :

1- المدى

المدى يمثل الفرق بين أعلى قيمة في مجموعة بيانات وأدناها، وهو مقياس يبين مدى تباعد القيم في سلسلة عددية ما. إذا كان المدى عددًا كبيرًا فإن القيم في السلسلة متباعدة عن بعضها ومشتتة، وإذا كانت قيمته صغيرة فإن القيم متقاربة و يحسب بالطريقة التالية :

أ – المدى في حالة بيانات أولية (غير مبوبة):

يحسب بالعلاقة التالية:

$$R = X_{Max} - X_{Min}$$

مثال 1 : إليك السلسة الإحسائية التالية :

8,9,9,9,12,14,14,16,17,18,20,20,31,25,62

$$\boldsymbol{R = 54}$$

ب- المدى في حالة بيانات مبوبة :

في حالة توزيع تكراري لمتغير متصل يمثل المدى:

المدى = الحد الأعلى للفئة الأخيرة – الحد الأدنى للفئة الأولي

2- المدى الربيعي والانحراف الربيعي:

1.2- الربيعيات :

* الربيعي الأول لسلسلة إحصائية مرتبة ترتيبا تصاعديا هو أصغر قيمة للطابع الإحصائي حيث يكون 25% علي الأقل من الحدود لها قيمة أصغر أو يساوي Q_1 .

* الربيعي الثالث لسلسلة إحصائية مرتبة ترتيبا تصاعديا هو أصغر قيمة للطابع الإحصائي حيث يكون 75% علي الأقل من الحدود لها قيمة أصغر أو يساوي Q_3 .

أ – الربيعي الأول والثالث في حالة بيانات أولية (غير مبوبة):

* إذا كان N يقبل القسمة علي 4 يحسب الربيعي الأول والثالث بالطريقة التالية:

$$\begin{Bmatrix} Q_1 = x\left(\frac{N}{4}\right), & Q_1 \text{ رتبة } \frac{N}{4} \\ Q_3 = x\left(\frac{3N}{4}\right), & Q_3 \text{ رتبة } \frac{3N}{4} \end{Bmatrix}$$

مثال 2 : إليك السلسلة التالية :

2,2,2,3,3,3,4,4,5,5,5,6   N=12

حساب Q_1 و Q_3 :

$$Q_1 = x(3) = 2, \qquad Q_1 \text{ رتبة } \frac{N}{4} = 3$$

$$Q_3 = x(9) = 5, \qquad Q_3 \text{ رتبة } \frac{3N}{4} = 9$$

* إذا كان N لا يقبل القسمة علي 4 في هذه الحالة نحسب الربيعي الأول انطلاقا من متوسط القمتين $x\left(\frac{N}{4}\right)$ والتي تليها و نحسب الربيعي الثالث انطلاقا من متوسط القمتين $x\left(\frac{3N}{4}\right)$ والتي تليها.

مثال 3 : إليك السلسلة التالية :

15, 17, 18, 19, 20, 20, 21, 22, 25   N=9

$$Q_1 = \frac{x(2) + x(3)}{2} = \frac{17 + 18}{2} = 17.5$$

$$Q_3 = \frac{x(6) + x(7)}{2} = \frac{21 + 22}{2} = 21.5$$

ب – الربيعي الأول والثالث في حالة بيانات مبوبة:

*** Q_1 و Q_3 في حالة كمي متقطع :**

بنفس الطريقة التي تم حساب بها الوسيط، حيث نحسب التكرار المتجمع الصاعد حيث:

- Q_1 هو القيمة X_i التي تقابل ت.م.ص الذي يحقق $\frac{N}{4} \leq \uparrow N_i$.

- Q_3 هو القيمة X_i التي تقابل ت.م.ص الذي يحقق $\frac{3N}{4} \leq \uparrow N_i$.

*** Q_1 في حالة كمي متصل:**

لحساب الربيعي الأول نتبع الخطوات التالية :

- نقوم بحساب التكرار المتجمع الصاعد

- نحدد رتبة الربيعي الأول $\frac{N}{4}$.

- نحدد فئة الربيعي الأول وهي الفئة المقابلة لتكرار المتجمع الصاعد الذي يحقق الشرط التالي: $\frac{N}{4} \leq \uparrow N_i$.

- نطبق القانون التالي :

$$Q_1 = L + \frac{\frac{N}{4} - N_0}{n_{Q_1}} \times K$$

حيث: K : هو طول فئة الربيعي الأول.

N_0 : هو ت.م.ص للفئة ماقبل فئة الربيعي الأول.

n_{Q_1} : تكرار فئة الربيعي الأول.

L : الحد الأدنى لفئة الربيعي الأول.

*** Q_3 في حالة كمي متصل:**

لحساب الربيعي الثالث نتبع الخطوات التالية :

- نقوم بحساب التكرار المتجمع الصاعد.

- نحدد رتبة الربيعي الثالث $\frac{3N}{4}$.

- نحدد فئة الربيعي الثالث وهي الفئة المقابلة لتكرار المتجمع الصاعد الذي يحقق الشرط التالي: $\frac{3N}{4} \leq \uparrow N_i$.

- نطبق القانون التالي :

$$Q_3 = L + \frac{\frac{3N}{4} - N_0}{n_{Q_3}} \times K$$

حيث : K : هو طول فئة الربيعي الثالث.

N_0 : هو ت.م.ص للفئة ما قبل فئة الربيعي الثالث.

n_{Q_3} : تكرار فئة الربيعي الثالث.

L : الحد الأدنى لفئة الربيعي الثالث.

مثال 4:

الجدول التالي يبين توزيع عمال مؤسسة حسب أجورهم الشهرية (الوحدة 10^3 دج)

الأجر X_i	$[10-20]$	$[20-30]$	$[30-40]$	$[40-50]$	$[50-60]$	$[60-70]$	$\sum$
عدد العمال N_i	18	30	25	17	12	8	110
$N_i \uparrow$	18	48	73	90	102	110	//

المطلوب :

1- أحسب Q_1 و Q_3 .

الحل :

لدينا :

Q_1 رتبة $\frac{N}{4} = \frac{110}{4} = 27{,}5$ 48> 27,5 ومنه فئة الربيعي الأول هي [20,30]

$$Q_1 = 20 + \frac{27.5 - 18}{30} \times 10$$

$$Q_1 = 23{,}16 \times 10^{3} \ DA$$

بنفس الطريقة نجد :

$$Q_3 = 45{,}58 \times 10^{3} \ DA$$

2.2- المدى الربيعي:

هو الفرق بين الربيعي الثالث والربيعي الأول ويرمز له بالرمز I_Q يستعمل في المقارنة بين توزعين إحصائيين أو أكثر حيث :

$$I_Q = Q_3 - Q_1$$

من المثال 4 نجد: $I_Q = 22.22 \times 10^3$

3.2- انحراف الربيعي:

هو نصف المدي الربيعي .

$$E_Q = \frac{I_Q}{2} = \frac{Q_3 - Q_1}{2}$$

من المثال 4 السابق نجد :

$$E_Q = \frac{I_Q}{2} = 11,11 \times 10^3$$

4.2- معامل الاختلاف الربيعي:

يرمز له بالرمز *C.Q.V* ويعطي بالعلاقة التالية:

$$C.Q.V = \frac{I_Q}{Q_2} \times 100$$

التمرين 1 :

الجدول التالي يبين الإنفاق الشهري لعينة من الأسر في مدينة جيجل (الوحدة 10^3 دج)

الأجر X_i	$[20-25]$	$[25-30]$	$[30-35]$	$[35-40]$	$[40-45]$	$[45-50]$	$\sum$
عدد العمال n_i	2	15	30	24	19	10	100
$N_i \uparrow$	2	17	47	71	90	100	//

المطلوب :

1- أحسب المدى العام.

2- أحسب الانحراف الربيعي، أحسب معامل الاختلاف الربيعي، ماذا تستنتج ؟

الحل :

1- المدى العام: **R**=50-20=30

2- حساب الانحراف الربيعي:
- حساب الربعي الاول :

لدينا :

Q_1 رتبة $\frac{N}{4} = \frac{100}{4} = 25$ 25<47 ومنه فئة الربيعي الأول هي [30,35]

$$Q_1 = 30 + \frac{25-17}{30} \times 5$$

$$Q_1 = 31.33 \times 10^{3} \quad DA$$

بنفس الطريقة نجد :

$$Q_3 = 41.05 \times 10^{3} \quad DA$$

إذن المدى الربيعي هو $\boldsymbol{I_Q = 9.72}$

الانحراف الربيعي هو:

$$E_Q = \frac{I_Q}{2} = \frac{Q_3 - Q_1}{2} = 4.86$$

حساب معامل الاختلاف الربيعي:
أولا نحسب الوسيط :

$$Q_2 = 35 + \frac{50-47}{24} \times 5 = 35.62 \times 10^{3} \quad DA$$

$$\boldsymbol{C.Q.V} = \frac{I_Q}{Q_2} \times 100 = \frac{9.72}{35.62} \times 100 = 27.28\%$$

استنتاج : نقول أن تشتت هذه البيانات هو تشت صغير نتيجة لكون قيمة معامل الاختلاف الربيعي يساوي 27,28%.

3- الانحراف المتوسط:

الانحراف المتوسط أو متوسط الانحرافات هو ناتج مجموع القيمة المطلقة (الناتج الموجب) لانحرافات القيم عن وسطها الحسابي مقسوما على عددها، ويرمز له بالرمز $E_{\bar{X}}$.

1.3- الانحراف المتوسط لبيانات غير مبوبة :

إذا كانت $X_1, X_2, \dots \dots \dots X_k$ تمثل قيم ظاهرة معينة X، فإن الانحراف المتوسط يحسب بالعلاقة التالية :

$$E_{\bar{X}} = \frac{1}{N}\sum_{i=1}^{k}|X_i - \bar{X}|$$

مثال: أوجد الانحراف المتوسط للبيانات التالية : N=5   2,3,5,6,9

الحل : $\bar{X} = \frac{1}{N}\sum_{i=1}^{5} X_i = 5$،

$$E_{\bar{X}} = \frac{1}{N}\sum_{i=1}^{5}|X_i - \bar{X}| = 2$$

2.3- الانحراف المتوسط لبيانات مبوبة :

أ- في حالة متغير منفصل:

إذا كانت $X_1, X_2, \dots \dots \dots X_k$ تمثل قيم ظاهرة معينة X، وكانت كانت $n_1, n_2, \dots \dots \dots n_k$ تمثل التكرارات المقابلة لها فإن الانحراف المتوسط يحسب بالعلاقة التالية:

$$E_{\bar{X}} = \frac{1}{N}\sum_{i=1}^{k} n_i|X_i - \bar{X}|$$

حيث: X_i : تمثل قيم المتغير.

n_i: تمثل التكرار المقابل لقيم المتغير.

ب- في حالة متغير متصل:

إذا كانت $C_1, C_2, \dots \dots \dots C_k$ تمثل مراكز الفئات لظاهرة معينة X، وكانت كانت $n_1, n_2, \dots \dots \dots n_k$ تمثل التكرارات المقابلة لها فإن الانحراف المتوسط يحسب بالعلاقة التالية:

$$E_{\bar{X}} = \frac{1}{N}\sum_{i=1}^{k} n_i|c_i - \bar{X}|$$

مثال : إليك السلسلة التالية :

X_i	1	2	3	4	5	$\sum$
n_i	4	3	2	1	2	12

المطلوب: أوجد المتوسط الحسابي والانحراف المتوسط.

الحل:

$$\bar{X} = \frac{1}{12}\sum_{i=1}^{5} n_i \times X_i = \frac{30}{12} = 2.5$$

X_i	1	2	3	4	5	$\sum$
$\lvert X_i - \bar{X} \rvert$	1.5	0.5	0.5	1.5	2.5	6.5

$$E_{\bar{X}} = \frac{1}{N}\sum_{i=1}^{5} n_i \lvert X_i - \bar{X} \rvert = 3$$

مثال:

احسب الانحراف المتوسط من جدول التوزيع التكراري الآتي والذي يبين درجات 30 طالب في امتحان ما.

الفئات	التكرار fi	مراكز الفئات	$f_i \times C_i$	$\lvert C_i - \bar{X} \rvert$	$f_i \times \lvert C_i - \bar{X} \rvert$
$[12 - 14]$	3	13	39	5,7	17,1
$[15 - 17]$	8	16	128	2,7	21,6
$[18 - 20]$	10	19	190	0,3	3
$[21 - 23]$	7	22	154	3,3	23,1
$[24 - 26]$	2	25	50	6,3	12,6
$\sum$	30	//	561	//	77,4

الوسط الحسابي : $\bar{X} = 18,7$

الانحراف المتوسط :

$$E_{\bar{X}} = 2,58$$

التمرين 1: أدرس تشتت البيانات باستخدام الانحراف المتوسط.

X_i	6	9	10	12	14	15	36	$\sum$
n_i	1	1	1	1	1	2	1	8

التمرين 2 :

الجدول التالي يمثل توزيع 120 شركة حسب ما بيعاتها الشهرية (الوحدة مليون دج).

الفئات	$[2-8]$	$[8-14]$	$[14-20]$	$[20-26]$	$[26-32]$
التكرارات	35	15	20	15	35

المطلوب :

1- أحسب المتوسط الحسابي .

2- أحسب الانحراف المتوسط والانحراف الربيعي.

حل التمرين 2 :

X_i	n_i	C_i	N↓	N↑	$n_i \times C_i$	$\lvert C_i - \bar{X} \rvert$	$n_i \lvert C_i - \bar{X} \rvert$
$[2-8]$	35	5	120	35	175	12	420
$[8-14]$	15	11	85	50	165	6	90
$[14-20]$	20	17	70	70	340	0	0
$[20-26]$	15	23	50	85	345	6	90
$[26-32]$	35	29	35	120	1015	12	420
$\sum$	120	//	//	//	2040		1020

1- حساب المتوسط الحسابي:

$$\bar{X} = \frac{1}{120}\sum_{i=1}^{5} n_i \times C_i = \frac{2040}{120} = 17$$

2- حساب الانحراف المتوسط:

$$E_{\bar{X}} = \frac{1}{N}\sum_{i=1}^{5} n_i \lvert C_i - \bar{X} \rvert = \frac{1020}{120} = 8.5$$

- حساب الانحراف الربيعي:

$$E_Q = \frac{I_Q}{2} = \frac{Q_3 - Q_1}{2}$$

الربيعي الاول هو $Q_1 = 7.14$ والربيعي الثالث هو $Q_3 = 26.86$

إذن الانحراف الربيعي هو: $E_Q = \mathbf{9.86}$

تمارين مقترحة

التمرين 1:

الجدول الآتي يبين العمر الذي أصيب فيه 100 شخص بمرض السكري لأول مرة؟

العمر بالسنة	[0-10[	[10-20[	[20-30[	[30-40[	[40-50[	[50-60[	[60-70[	[70-80[	[80-90[
العدد	1	3	10	14	18	34	12	6	2

المطلوب :

1 - أحسب الربيعي الأول والثاني والثالث.

1- أدراس تشتت البيانات باستعمال الانحراف المتوسط.

التمرين 2:

الجدول التكراري التالي يبين توزيع الأوزان لعينة من الطلاب.

الوزن	[60-64 [	[64-68[	[68-72[	[72-76[	[76-80[	[80-84[	[84-88[
عدد الطلبة	5	12	20	26	20	12	5

المطلوب :

1 - أحسب الربعي الأول والثاني والثالث.

2- دراسة تشتت البيانات باستعمال الانحراف المتوسط.

4. التباين والانحراف المعياري: (حالة المسح بالعينة)

1.4- التباين : هو مقياس لاختلاف البيانات وتشتتها، وهو متوسط مربعات انحرافات القيم عن وسطها الحسابي، ويرمز له بالرمز δ^2 أو $V(X)$ في حالة بيانات المجتمع و S^2 في حالة عينة .

أ- حساب التباين لبيانات أولية (غير مبوبة):

إذا كانت $X_1, X_2, \dots \dots \dots .. X_k$ تمثل قيم ظاهرة معينة X، وكان $\bar{X}$ المتوسط الحسابي فإن التباين العينة يحسب بالطريقة التالية :

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{k}(X_i - \bar{X})^2$$

مثال :

احسب كلاً من التباين والانحراف المعياري للقيم التالية:

12, 15, 11, 17,18, 20, 19

X_i	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$
12	12-16=-4	16
15	15-16=-1	1
11	11-16=-5	25
17	17-16=1	1
18	18-16=2	4
20	20-16=4	16
19	19-16=3	9
$\sum X_i = 112$	$\bar{X} = 16$	$\sum(X_i - \bar{X})^2 = 72$

حساب التباين:

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{k}(X_i - \bar{X})^2 = \frac{1}{6}\sum_{i=1}^{7}(X_i - \bar{X})^2 = \frac{72}{6} = 12$$

حساب الانحراف المعياري:

$$\mathrm{S} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{k} n_i(X_i - \bar{X})^2} = 3.46$$

ب- التباين لبيانات مبوبة:

* في حالة متغير متقطع:

إذا كانت $X_1, X_2, \dots \dots \dots .. X_k$ تمثل قيم ظاهرة معينة X، وكانت كانت $n_1, n_2, \dots \dots \dots .. n_k$ تمثل التكرارات المقابلة لها وكان $\bar{X}$ المتوسط الحسابي فإن التباين يحسب بالطريقة التالية :

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{k} n_i\left(X_i - \overline{X}\right)^2$$

*** في حالة متغير متصل:**

إذا كانت $C_1, C_2, \dots \dots \dots . . C_k$ تمثل مراكز فئات ظاهرة معينة X، وكانت كانت $n_1, n_2, \dots \dots \dots . . n_k$ تمثل التكرارات المقابلة لها وكان $\overline{X}$ المتوسط الحسابي فإن التباين يحسب بالطريقة التالية :

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{k} n_i\left(C_i - \overline{X}\right)^2$$

2.4- الانحراف المعياري:

الانحراف المعياري هو الجدر التربيعي لتباين ويعتبر من أهم مقاييس التشتت وأكثرها استخداما في النظريات والقوانين الاحصائية ويرمز له بالرمز δ في حالة بيانات المجتمع و S في حالة عينة .

$$\mathrm{S} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{k} n_i\left(X_i - \overline{X}\right)^2}$$

تمرين :

نعتبر السلسلة الاحصائية التي تمثل أوزان طلاب قسم يحتوي علي 40 طالب (كلغ).

القامات	$[50-55]$	$[55-60]$	$[60-65]$	$[65-70]$	$[70-75]$
التكرارات	3	10	12	10	5

المطلوب :

1 - أحسب المتوسط الحسابي .

2- أحسب التباين والانحراف المعياري.

الحل : نشكل الجدول التالي :

X_i	n_i	C_i	$n_i \times C_i$	$(C_i - \bar{X})^2$	$n_i(C_i - \bar{X})^2$
$[50 - 55]$	3	52.5	157.5	110.25	330.75
$[55 - 60]$	10	57.5	575	30.25	302.5
$[60 - 65]$	12	62.5	750	0.25	3
$[65 - 70]$	10	67.5	675	20.25	202.5
$[70 - 75]$	5	72.5	362.5	90.25	451,25
$\sum$	40	//	2520	//	1290

1- حساب المتوسط الحسابي :

$$\bar{X} = \frac{1}{40}\sum_{i=1}^{5} n_i \times C_i = \frac{2520}{40} = 63$$

2- حساب التباين:

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{k} n_i(C_i - \bar{X})^2 = \frac{1290}{39} = 33.07$$

- الانحراف المعياري:

$$S = \sqrt{\frac{1}{N-1}\sum_{i=1}^{k} n_i(X_i - \bar{X})^2}$$

$$S = 5.75$$

3.2. التباين والانحراف المعياري (حالة المسح أو أخذ جميع مفردات المجتمع الاحصائي)

1. التباين: هو الوسط الحسابي لمربعات الفروق بين قيم المتغير الاحصائي والوسط الحسابي $\bar{X}$ ويرمز له بالرمز $V(X)$ أو σ^2 وتكتب علاقته بالشكل التالي:

- حالة البيانات البسيطة (غير المبوبة):

$$V(X)=\frac{\sum_{i=1}^{k}(X_i-\bar{X})^2}{n}=\frac{1}{n}\sum_{i=1}^{k}(X_i-\bar{X})^2$$

- حالة متغير كمي منقطع:

$$V(X)=\frac{\sum_{i=1}^{k}n_i(X_i-\bar{X})^2}{n}=\frac{1}{n}\sum_{i=1}^{k}n_i(X_i-\bar{X})^2=\sum_{i=1}^{k}f_i(X_i-\bar{X})^2$$

حالة متغير كمي مستمر: نفس العلاقة السابقة فقط هنا X_i تعبر عن مراكز الفئات :

$$V(X)=\frac{\sum_{i=1}^{k}n_i(X_i-\bar{X})^2}{n}=\frac{1}{n}\sum_{i=1}^{k}n_i(X_i-\bar{X})^2=\sum_{i=1}^{k}f_i(X_i-\bar{X})^2$$

2. الانحراف المعياري: هو مقياس يحدد مدى تباعد أو تقارب المشاهدات عن وسطها الحسابي، وتكتب علاقته بالشكل التالي:

- حالة البيانات البسيطة (غير المبوبة):

$$\sigma(X)=\sqrt{\frac{\sum_{i=1}^{k}(X_i-\bar{X})^2}{n}}=\sqrt{\frac{1}{n}\sum_{i=1}^{k}(X_i-\bar{X})^2}$$

- حالة متغير كمي منقطع:

$$\sigma(X)=\sqrt{\frac{\sum_{i=1}^{k}n_i(X_i-\bar{X})^2}{n}}=\sqrt{\frac{1}{n}\sum_{i=1}^{k}n_i(X_i-\bar{X})^2}=\sqrt{\sum_{i=1}^{k}f_i(X_i-\bar{X})^2}$$

- حالة متغير كمي مستمر:

$$\sigma(X)=\sqrt{\frac{\sum_{i=1}^{k}n_i(X_i-\bar{X})^2}{n}}=\sqrt{\frac{1}{n}\sum_{i=1}^{k}n_i(X_i-\bar{X})^2}=\sqrt{\sum_{i=1}^{k}f_i(X_i-\bar{X})^2}$$

ومنه فإن: $\sigma(X)=\sqrt{V(X)}$

مثال: توزيع 100 أسرة حسب عدد الأطفال:

عدد الأطفال	0	1	2	3	4	المجموع
عدد الأسر	10	20	40	20	15	100

المطلوب: أحسب التباين والانحراف المعياري.

ولتسهيل الحسابات نعتمد على الجدول التالي:

X_i	n_i	n_iX_i	$n_iX_i^2$	$X_i-\bar{X}$	$(X_i-\bar{X})^2$	$n_i(X_i-\bar{X})^2$	f_i	$f_iX_i^2$	$f_i(X_i-\bar{X})^2$
0	10	0	0	-2	4	40	0,1	0	0,4
1	20	20	20	-1	1	20	0,2	0,2	0,2
2	40	80	160	0	0	0	0,4	1,6	0
3	20	60	180	1	1	20	0,2	1,8	0,2
4	10	40	160	2	4	40	0,1	1,6	0,4
المجموع	100	200	500	—	—	120	1	5,2	1,2

- الوسط الحسابي:

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{k}n_iX_i=\frac{200}{100}=2$$

- التباين:

$$V(X)=\frac{1}{n}\sum_{i=1}^{k}n_i(X_i-\bar{X})^2=\frac{1}{100}\sum_{i=1}^{5}n_i(X_i-2)^2=\frac{120}{100}=1{,}2$$

$$V(X)=\frac{1}{n}\sum_{i=1}^{k}n_iX_i^2-\bar{X}^2=\frac{1}{100}\sum_{i=1}^{5}n_iX_i^2-2^2=\frac{520}{100}-4=1{,}2$$

$$V(X)=\sum_{i=1}^{k}f_i(X_i-\bar{X})^2=\sum_{i=1}^{5}f_i(X_i-2)^2=1{,}2$$

$$V(X)=\sum_{i=1}^{k}f_iX_i^2-\bar{X}^2=\sum_{i=1}^{5}f_iX_i^2-2^2=5{,}2-4=1{,}2$$

- الانحراف المعياري:

$$\sigma(X)=\sqrt{V(X)}=\sqrt{1{,}2}=1{,}1$$

3. خصائص التباين والانحراف المعياري

للعلم أن: X متغير، a ثابت و b ثابت، فإن:

- $V(a)=0$ ، $V(aX+b)=V(aX)+V(b)=a^2V(X)$ ، $V(aX)=a^2V(X)$
- $\sigma(a)=0$ ، $\sigma(a+X)=\sigma(a)+\sigma(X)=\sigma(X)$ ، $\sigma(aX)=a\sigma(X)$
- الانحراف المعياري للمقدار الثابت معدوم، أي أنه إذا كان لدينا القراءات التالية:

$X_1=X_2=X_3=..........X_k=a$ حيث أن a مقدار ثابت فإن الانحراف المعياري $\sigma(X)=0$.

- إذا أضيف (أو طرح) مقدار ثابت من قيم المتغير فإن الانحراف المعياري للقيم الجديدة يساوي الانحراف المعياري للقيم الأصلية، فمثلا لتكن القيم الأصلية: $X_1, X_2, X_3 .... X_n$ ، تمت إضافة مقدار ثابت a إلى كل قيم المتغير فإن الانحراف المعياري للقيم الجديدة: $Y_1 = X_1 + a, Y_2 = X_2 + a, ... Y_k = X_k + a$ هو: $\sigma(Y) = \sigma(X)$.
- إذا ضربت (أو قسمت) كل قيمة من قيم المتغير الإحصائي في (على) مقدار ثابت a فإن الانحراف المعياري للقيم الجديدة يساوي الانحراف المعياري للقيم الأصلية في (على) المقدار الثابت a، فمثلا إذا كان لدينا قيم المتغير الإحصائي: $X_1, X_2, X_3 .... X_n$ ، وتم ضرب القيم في مقدار ثابت a فإن الانحراف المعياري للقيم الجديدة: $Y_1 = aX_1, Y_2 = aX_2, ... Y_k = aX_k$ هو: $\sigma(Y) = a\sigma(X)$.
- إذا أضفنا (أو طرحنا) قيمة ثابتة b من قيم المتغير الإحصائي وفي نفس الوقت يضرب (أو يقسم) في (على) قيمة ثابتة a، حيث نتحصل على التوليفة الخطية التالية: $Y = aX_i + b$، فإن الانحراف المعياري للقيم الجديدة هو: $\sigma(Y) = a\sigma(X)$

ملاحظة: يستعمل الانحراف المعياري للمقارنة بين توزيعين إحصائيين أو أكثر، والتوزيع الإحصائي الأقل تشتتا هو الأفضل أي الذي له أصغر انحراف معياري.

4. معامل الاختلاف: هو أحد المقاييس المستخدمة لقياس درجة التشتت، يستعمل في المقارنة بين توزيعات إحصائية غير المتجانسة (وحدات القياس مختلفة)، وهو نسبة بين الانحراف المعياري والمتوسط الحسابي، يرمز له بـ CV ويحسب بالعلاقة التالية:

$$CV = \frac{\sigma(X)}{\bar{X}} * 100$$

مثال: الجدول التالي يمثل مقاييس خاصة بأجور عمال مؤسستين الأولى في الجزائر (الدينار الجزائري) والثانية في تونس (الدينار التونسي):

المؤسسة التونسية (100دت)	المؤسسة الجزائرية (100دج)	المقياس
173	198	$\bar{X}$
23	25	$\sigma(X)$

المطلوب: قارن درجة تشتت الأجور في المؤسستين.

بالاعتماد على المقياسين السابقين نقول أن الأجور في المؤسسة الجزائرية أكثر تشتتا لأن انحرافها المعياري أكبر من الانحراف المعياري للأجور في المؤسسة التونسية، لكن في الواقع نجد أن قيمة الدينار الجزائري تختلف عن قيمة الدينار التونسي، فلا يمكننا المقارنة بينهما باستعمال الانحراف المعياري، وعليه نستعمل معامل الاختلاف الذي هو عبارة عن نسبة مئوية ولا يحتوي على وحدة قياس حيث:

المؤسسة الجزائرية: $CV = \frac{25}{189} * 100 = 12,62\%$

المؤسسة التونسية: $CV = \frac{23}{173} * 100 = 13,29\%$

من خلال نتائج معامل الاختلاف نلاحظ أن الأجور في المؤسسة الجزائرية أقل تشتتا من الأجور في المؤسسة التونسية.