الجمهورية الجزائرية الديمقراطية الشعبية
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
وزارة التكوين والتعليم المهنيين
Ministère de la Formation et de l'Enseignement Professionnels

Centre National de l'Enseignement Professionnel à Distance (CNEPD)

المركز الوطني للتعليم المهني عن بعد (م.و.ت.م.ب)

درس في الإحصاء

السلسلة رقم 03

التوزيعات الإحصائية لمتغير واحد

خطة الدرس :

I- مميزات النزعة المركزية لسلسلة إحصائية

1- المنوال أو الفئة المنوالية
2- الوسيط
3- المتوسط الحسابي
4- المتوسط الهندسي
5- المتوسط التوافقي
6- مقارنة المتوسطات الثلاثة

II- مميزات التشتت (الانتشار)

1- لماذا مميزات التشتت؟
2- المدى
3- المجال الربيعي البيني
4- العشريات و العشريات البينية
5- الانحراف المتوسط
6- التباين و الانحراف المعياري
7- معامل التباين(التغير)

WWW.DZPEN.COM | قلم الجزائر

I- مميزات النزعة المركزية لسلسلة إحصائية :

1- المنوال أو الفئة المنوالية :

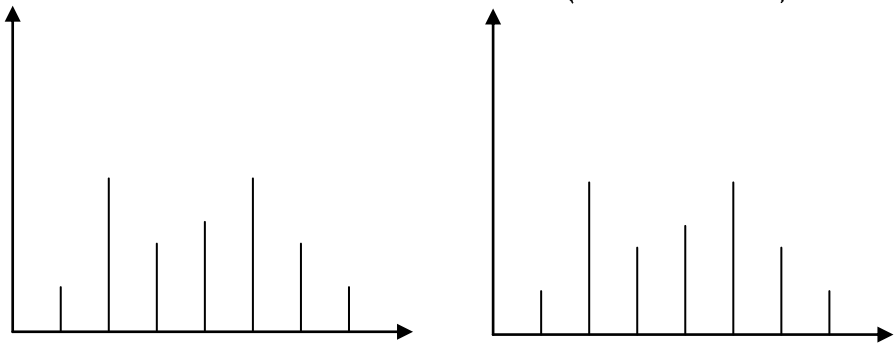
هي قيمة المتغير أو الفئة التي يوافقها أكبر تكرار.

مثال : فيمايلي دراسة عن عمر عمال مؤسسة.

العمر (فئات)	التكرار n_i
]25-20]	9
]30-25]	27
]35-30]	36
]40-35]	45
]45-40]	18
]50-45]	9
]55-50]	3
[60-55]	3

الفئة المنوالية هي الفئة]35-40[التي تضم أكبر عدد من العمال و يمكن أن نعتبر مركز الفئة هو المنوال أي 37 سنة و نصف.

ملاحظة: قد يكون لسلسلة إحصائية منوال مضاعف (منوالان) أو منـــوال متعدد (أكثر من منوالين).



فائدة مفهوم المنوال :

لإظهار فائدة المنوال، سنأخذ بائع الأحذية كمثال لذلك، إذ من المهم جدا له أن يعرف المقياس الأكثر بيعا و هو ما يمثله المنوال

2- الوسيط :

وسيط سلسلة إحصائية هو القيمة "M" للمتغير التي يكون فيها عدد الملاحظات الأصغر يساوي عدد الملاحظات الأكبر فيها، أو بحيث تكون نصف الملاحظات أصغر من "M" .

يتم تحديد الوسيط على مرحلتين:

- ترتيب الملاحظات حسب القيم المتصاعدة.
- ثم البحث عن قيمة مركزية في هذه السلسلة المرتبة

1.2- حالة متغير منقطع:

مثال:

تحصل مترشح في إحدى الإمتحانات على العلامات التالية علما بأن كل الاختبارات تحسب على 20:

العلامات	14	08	11	12	10	07
المعاملات	1	1	1	3	1	2

لنرتب هذه العلامات تصاعديا مع الإشارة إلى أنه يجب كتابة العلامة 7 مرتين (معامل2)، و العلامة 12 ثلاث مرات (معامل 3).

7 ، 7، 8 ، 10 ، 11 ، 12 ، 12، 12، 14

إن الوسيط في هذه الحالة هو 11 (هناك 4 علامات أقل منه و هناك 4 علامات أكبر منه).

لو كانت هذه السلسلة كالتالي:

6، 7 ، 7، 8 ، 10 ، 11 ، 12 ، 12، 12، 14

في هذه الحالة ، نحدد مجالا وسيطا هو المجال[10،11] أو نحدد وسيطا هو مركز المجال الوسيط أي:

$$\frac{11+10}{2}=10.5$$

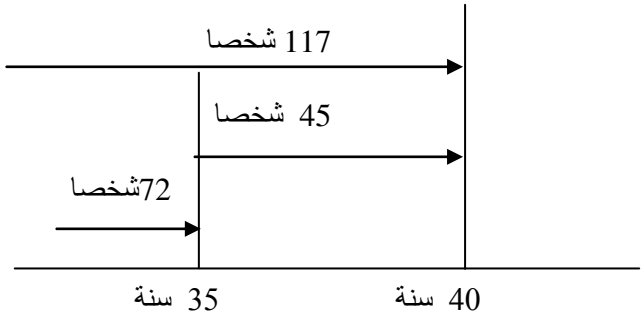
2.2- حالة متغير مستمر :

لنسترجع مثال المؤسسة التي تشغل 150 عاملا، و لنستخدم ترتيبهم حسب العمر و عمود التكرارات المتراكمة الصاعد.

العمر	25-20	30-25	35-30	40-35	45-40	50-45	55-50	60-55
التكرار المتراكم الصاعد	9	36	72	117	135	144	147	150

إن العمر الوسيط هو عمر العامل الذي يحتمل الرتبة 75 (150 ÷ 2) و لهذا نستطيع التأكيد على أن العمر الوسيط الذي نبحث عنه يكون محصورا فيما بين 35 و 40 سنة.

و للحصول على نتيجة أكثر دقة، نفرض أن أعمار الأشخاص الذين عددهم هو 45، و الذين لهم عمر يقل عن 40 سنة و يبلغ 35سنة عل الأقل، تتوزع بانتظام على المجال [35-40[مثلما يوضحه الشكل التالي:



و يتم بعدها تحديد قيمة الوسيط عن طريق الإستكمال بأجزاء نسبية أو الإستكمال الخطي ضمن المجال [35−40[، أي:

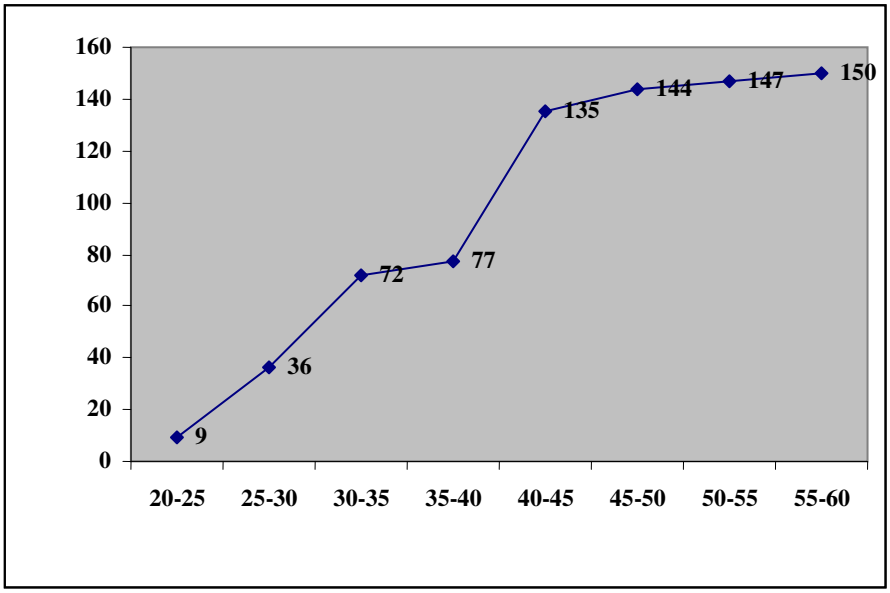
$$\text{الوسيط} = 35 + \frac{72-75}{72-117}(35-40) = 35 + \frac{1}{3}$$

و هكذا يكون العمر الوسيط هو 35 سنة و4 أشهر.

3.2- عرض الوسيط بيانيا :

يكفي رسم المضلع المتراكم الصاعد ثم نمد من النقطة التي تقسم التكرار إلى نصفين خطا مستقيما موازيا لمحور السينات إلى أن يتقاطع مع المضلع. بعدها نعين فاصلة هذا التقاطع . و هي قيمة الوسيط الذي نبحث عنه.

المثال السابق:



3- المتوسط الحسابي:

رغم أنه يمكن تحليل الظاهرة المدروسة تحليلا أوليا إنطلاقا من الجداول الإحصائية و الرسوم البيانية فقط. لكنه من المفيد إظهار مميزات سلسلة إحصائية بواسطة عدد يمثل مجموعة الظاهرة.

إن مفهوم المتوسط شائع، و المتوسط الحسابي هو القيمة المركزية. فهو مثلا يسمح بوصف المستوى العام لأحد الطلاب و لكنه يحجب في نفس الوقت مستواه الحقيقي في كل مادة.

في هذا الصدد يقول "بول فاليري PAUL VALERY" ما هو سهل هو خطا، و ما هو معقد لا يفيد".

كما أن المتوسط الحسابي لا يشكل القيمة المركزية الوحيدة إذ يمكن إستعمال المتوسط الهندسي، المتوسط التوافقي و المنوال أو الوسيط و هو يساوي حاصل قسمة مجموع قيم المؤشر الملحوظة على مجموع التكرار.

1.3- المتوسط الحسابي البسيط :

ليكن عدد N من الملاحظات المختلفة:

$x_1 , x_2 , \ldots\ldots\ldots x_n$ التي لم تتكرر إلا مرة واحدة.

نرمز للمتوسط الحسابي بـ $\bar{x}$ (يقرأ : x barre) و هو :

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{N} \quad \text{أو} \quad \bar{x} = \frac{x_1 + x_2 + \ldots\ldots + x_n}{N}$$

مثال:

تحصل طالب في الإمتحان على العلامات التالية:

10 في الرياضيات، **14** في الإقتصاد، **12** في الإحصاء، **08** في اللغات و بهذا يبلغ معدله:

$$\frac{10+14+12+8}{4} = 11$$

إن هذا المعدل هو المتوسط الحسابي البسيط لأنه كان لكل إختيار نفس الترجيح أو نفس " الثقل".

2.3- المتوسط الحسابي الترجيحي :

إذا ظهرت بعض قيم المتغير عدة مرات فإنه يجب حسابها بعدد مرات ظهورها، و هذا يعني ترجيح (ضرب) كل قيمة المتغير بالتكرار الذي يوافقها.

ليكن P عدد الملاحظات المختلفة : $x_1 , x_2 \ldots . x_p$

و أن عدد ظهور x_1 هو n_1 مرة...، و عدد ظهور x_i هو n_i مرة،...، و عدد ظهور x_p هو n_p مرة :

عندئذ يكون : $\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \ldots + n_i x_i + n_p x_p}{n_1 + n_2 + \ldots + n_i + \ldots n_p}$

حيث: (التكرار الإجمالي) $n_1 + n_2 + \ldots . + n_i + \ldots . + n_p = N$

يمكننا كتابة:

المتوسط الحسابي الترجيحي

$$\bar{x} = \frac{\sum_{i=1}^{p} n_i x_i}{\sum_{i=1}^{p} n_i} = \left[\frac{\sum_{i=1}^{P} n_i x_i}{N} \right] \quad \text{أي} \quad : \quad \sum_{i=1}^{p} n_i = N$$

مثال1 : تباع إحدى السلع في محلات مختلفة بالأسعار التالية :

سعر بـ دج (مؤشر)	عدد المحلات (التكرار)
10	20
12	40
14	30
15	10
المجموع	**100**

xi	ni	ni × xi
10	20	200
12	40	480
14	30	420
15	10	150
مجموع	$\Sigma n_i = 100$	$\Sigma n_i x_i = 1250$

لهذه السلعة متوسط هو :

$$\bar{x} = \frac{\sum n_i x_i}{\sum n_i} = \frac{1250}{100} =$$

12.05 دج

في حالة توفر التكرارات النسبية يصبح حساب المتوسط الحسابي سهلا :

$$\bar{x} = \frac{x_1 n_1 + .. + x_i\, n_i + ... + x_p\, n_p}{n_1 + ... + n_i + ... + n_p}$$

$$\bar{x} = \frac{n_1}{N} x_1 + \frac{n_2}{N} x_2 + ... + \frac{n_i}{N} x_i + ... + \frac{n_p}{N} x_p$$

و حيث : $\frac{n_1}{N} = f_1 : \frac{n_i}{N} = f_i$: $\frac{n_p}{N} = f_p$

$$\bar{x} = f_1 x_1 + f_2 x_2 + ...+ f_i x_i ... + f_p x_p$$

$$\sum_{i=1}^{p} f_i \; x_i$$

مثال2 :ليكن توزيع العائلات التالية حسب عدد أطفالها:

عدد الأطفال x_i	التكرار النسبي f_i	$f_i \times x_i$
0	0,11	0
1	0,25	0,25
2	0,32	0,64
3	0,21	0,63
4	0,07	0,28
5	0,04	0,20
6 و أكثر	0	0
مجموع	**1**	**2**

WWW.DZPEN.COM | قلم الجزائر

لهذه العائلات عدد من الأطفال يبلغ في المتوسط:

$$\bar{x} = \sum_{i=1}^{6} f_i x_i = 2$$

إن كانت السلسلة ذات متغير مستمر فإن x_i هي مراكز الفئات.

مثال : أجر ساعة عمل بالدينار الجزائري في إحدى المؤسسات هو كالتالي:

الفئات	تكرار n_i	مركز الفئة x_i	$n_i \times x_i$
من 10 إلي أقل من 12	10	11	110
من 12 إلي أقل من 14	30	13	390
من 14 إلي أقل من 16	50	15	750
من 16 إلي أقل من 20	20	18	360
من 20 إلي أقل من 24	10	22	220
	$\Sigma n_i = 120$		$\Sigma n_i\ x_i = 1830$

يبلغ الأجر المتوسط لساعة العمل في هذه المؤسسة :

$$\bar{x} = \frac{\sum x_i n_i}{\sum n_i} = \frac{1830}{120}$$

$\bar{x}$ = **15,25 دج**

3.3- خواص المتوسط الحسابي :

الخاصية الأولى :

إذا كانت جميع الملاحظات متساوية فإن متوسط هذه الملاحظات يساوي أي واحدة منها.

$$x_1 = x_2 = \ldots . x_i = \ldots xp = a$$

$$\Rightarrow x = \frac{\sum n_i x_i}{\sum n_i} = \frac{\sum n_i a}{\sum n_i} = a$$

$\bar{x} = a$

WWW.DZPEN.COM | قلم الجزائر

متوسط العدد الثابت يساوي العدد نفسه :

الخاصية الثانية:

متوسط المجموع يساوي مجموع المتوسطات $\boxed{\overline{x+y} = \overline{x} + \overline{y}}$

الخاصية الثالثة :

المجموع الجبري لانحرافات كل حدود السلسلة عن المتوسط هو مجموع منعدم.

انحرافات كل ملاحظة عن المتوسط هي :

$$(x_1 - \overline{x}) ; (x_2 - \overline{x}),\ldots.(x_i - \overline{x}),\ldots , (x_p - \overline{x}).$$

أما مجموع الانحرافات فهو :

$$\sum_{i=1}^{p} n_i (x_i - \overline{x}) = \sum n_i x_i - \sum n_i \overline{x} \qquad \mathbf{S} =$$

$$\mathbf{S = \Sigma n_i x_i - N\overline{x}} \quad \textbf{أو :} \qquad \overline{x} = \frac{\sum n_i x_i}{N} \Rightarrow N\ \overline{x} = \sum n_i x_i$$

و بتعويض $N\overline{x}$ بقيمته في S(مجموع) نحصل على :

$$\mathbf{S = Nx^{-} - Nx^{-} = 0}$$

الخاصية الرابعة :

إذا أضفنا أو طرحنا نفس العدد إلى كل حد من حدود السلسلة نحصل على متوسط جديد يساوي المتوسط القديم بعد أن نضيف إليه أو نطرح منه ذلك العدد.

فإذا أضفنا العدد الثابت "a" إلى كل حد من حدود السلسلة نحصل على متوسط جديد هو :

$$\frac{\sum n_i (x_i + a)}{\sum n_i} = \frac{\sum n_i x_i + \sum n_i a}{\sum n_i}$$

$$= \frac{\sum n_i x_i}{\sum n_i} + \frac{\sum n_i a}{\sum n_i} = \frac{\sum n_i x_i}{\sum n_i} + \frac{N.a}{N} = \boxed{\bar{x} + a}$$

و بطريقة أخرى :

$$(x + a) = \bar{x} + \bar{a} = \bar{x} + a$$

الخاصية الخامسة :

إذ ضربنا كل حدود السلسلة في نفس العدد فإن متوسط السلسلة الجديدة يساوي حاصل ضرب المتوسط القديم في هذا العدد.

$$\boxed{\overline{ax + b} = a\bar{x} + b} \qquad \boxed{\overline{(bx)} = b \times \bar{x}}$$

الخاصية السادسة :

يكون مجموع مربعات فروقات كل حدود السلسلة الإحصائية عن عدد ما الأدنى عندما يكون هذا العدد مساويا للمتوسط الحسابي لهذه السلسلة.

لتكن « P » ملاحظة $x_1 , x_2 , ..., x_p$ التي لها التكرارات : $n_1 , n_2 , ..., n_p$ حيث $\sum_{i=1}^{p} n_i = N$ لهذه السلسلة متوسط هو :

$$\bar{x} = \frac{\sum n_i x_i}{N}$$

فروقات حدودها عن الثابت « a » هي : $(x_1-a), \ (x_2-a)\ldots(x_p-a)$.
يجب حساب كل فرق منها بحسب عدد المرات التي لوحظ فيها .

و من ثم فإن مجموع مربعات هذه الفروقات هو:

$$S_1 = n_1(x_1-a)^2 + n_2(x_2-a)^2 + \ldots + n_p\,(x_p - a)^2$$

$$S_1 = \Sigma\, n_i\,(x_i-a)^2$$

و بتحليل المربع ينتج :

$$S_1 = \Sigma\, n_i\,(x_i^2 - 2ax_i + a^2) = \Sigma\, n_i\, x_i^2 - 2a\,\Sigma\, n_i\, x_i + a^2\,\Sigma\, n_i$$

$$S_1 = \Sigma\, n_i\, x_i^2 - 2aN\,\bar{x} + a^2\, N$$

و لنفرض S_2 مجموع مربعات الإنحرافات عن المتوسط:

$$S_2 = \Sigma\, n_i\,(x_i - \bar{x})^2 = \Sigma\, n_i\, x_i^2 - 2\,\bar{x}\ N\,\bar{x} + N\,\bar{x}^2$$

$$S_2 = \Sigma\, n_i\, x_i^2 - N\ \bar{x}^2$$

و الفرق $S_1 - S_2$ هو دائما مقدار موجب أو منعدم.

$$S_1 - S_2 = N\,\bar{x}^2 - 2aN\,\bar{x} + Na^2 = N(\bar{x}^2 - 2a\,\bar{x} + a^2) = N(\bar{x} - a)^2$$

ولهذا يتضح أن يكون دائما مقدار موجب أو منعدم.

و أنه أكبر دائما من S_2 إلا إذا كان $\bar{x} = a$ عندئذ يصبح S_1 الأدنى.

قد لا يسمح المتوسط الحسابي بالترجمة الصحيحة للظاهرة المدروسة حينئذ يصبح من الضروري استخدام أنواع أخرى من المتوسطات.

4- المتوسط الهندسي :

المتوسط الهندسي M_g لعدد من الملاحظات الموجبة N يساوي الجدر النوني لحاصل جداء هذه الملاحظات.

لتكن N ملاحظة : $x_1 , x_2 ,\ldots,x_n$ و $x_i \geq 0$

عند ئذ يكون :

$$\mathbf{Mg} = \sqrt[N]{x_1 \quad x \quad x_2 \; x\ldots x \; x_i \quad x\ldots x \quad x_n}$$

مثال : المتوسط الهندسي للعددين 4 و 9 هو :

$$\mathbf{Mg} = \sqrt[2]{4\times 9} = \sqrt[2]{36} = \mathbf{6}$$

بينما متوسطها الحسابي هو :

$$\bar{x} = \frac{4+9}{2} = 6{,}5$$

إن كانت السلسلة مرجحة فإن حساب المتوسط الهندسي يكون بهذه الصيغة :

$$\mathbf{Mg} = \sqrt[p]{x_1^{\,n_1} \times x_2^{\,n_2} \cdot x\ldots x \quad x_i^{\,n_i} \quad x\ldots x \quad x_p^{\,n_p}}$$

مثال :

n_i	x_i
2	2
1	16
3	**مجموع**

WWW.DZPEN.COM | قلم الجزائر

المتوسط الهندسي :

$$\mathbf{Mg} = \sqrt[3]{2^2 \times 16^1} = \sqrt[3]{64} = \mathbf{4}$$

المتوسط الحسابي: $\boxed{\bar{x} = \mathbf{6,66}}$

5- <u>المتوسط التوافقي</u> :

المتوسط التوافقي H لسلسلة من الملاحظات يساوي مقلوب المتوسط الحسابي لمقلوب هذه الملاحظات.

لتكن N ملاحظة : $x_1 , x_2 , \ldots, x_n$:

$$\boxed{H = \frac{N}{\sum_{i=1}^{N} \frac{1}{x_i}}}$$

إذا كانت السلسلة مرجحة :

$x_1 , x_2, \ldots . x_i, \ldots x_p$

$n_1, n_2, \ldots . . n_i \ldots n_p$

$$\boxed{H = \frac{\sum_{i=1}^{p} n_i}{\sum_{i=1}^{p} n_i \frac{1}{X_i}}}$$

حسب المثال السابق يبلغ المتوسط التوافقي :

$$\mathbf{H} = \frac{2+1}{\left(2 \times \frac{1}{2}\right) + \left(1 \times \frac{1}{16}\right)} = \frac{3}{1 + \frac{1}{16}} = \frac{3}{\frac{16+1}{16}} = \frac{3}{\frac{17}{16}} = \frac{48}{17} = 2,82$$

6- مقارنة المتوسطات الثلاثة :

إذا كانت الملاحظات غير متساوية، يكون المتوسط التوافقي أقل من المتوسط الهندسي الذي يقل هو الآخر عن المتوسط الحسابي.

$$H < Mg < \bar{x}$$

وأيا كانت القيمة المركزية التي نعتمدها، فإنها تحجب جزاء من الظاهرة التي ندرسها خاصة إنتشار الملاحظات حول القيمة المركزية.

II- مميزات التشتت (الإنتشار) :

1- لماذا ميزة التشتت؟ :

لنفرض أنه يتوجب علينا مقارنة مؤسستين مختلفتين من حيث أجر الساعة الذي تمنحه كل واحدة لعمالها الخمسة.

ولذلك على أساس قياسات الأجور المرتبة تصاعديا كالتالي :

المؤسسة الأولى	2,75	2,85	3,10	3,25	3,40
المؤسسة الثانية	2,35	2,40	3,10	3,20	4,30

عند مقارنة ميزات النزعة المركزية لهذين السلستين يتبين :

- أن لهذين السلستين نفس الوسيط و هو يساوي **3,10**، ولهما نفس المتوسط الحسابي و هو يساوي **3,07** .
- رغم أن لهذين السلسلتين نفس الوسيط و نفس المتوسط الحسابي إلا أنهما مختلفتين حتما : فالأجور التي تمنحها المؤسسة الثانية هي أكثر نشرا وانتشارا حول القيم المركزية من تلك التي تمنحها المؤسسة الأولى.

و لهذا فإننا نحتاج إلى تعريف وحساب مميزات التشتت تسمح لنا بقياس هذا التشتت وبالتالي مقارنة السلاسل مثل السلسلتين السابقتين.

2- <u>المدى</u> :

تنبسط الأجور في المؤسسة الأولى من 2,75 إلى 3,40، أي تمتد في مجال قدره 3,40 – 2,75 = 0,65. عندئذ نقول **<u>0,65</u>** هو مدى هذه السلسلة فالمدى هو الفرق بين أكبر قيمة و اصغر قيمة في السلسلة. و مثال ذلك مدى الأجور التي تمنحها المؤسسة الثانية الذي يبلغ : **4,30 – 2,35 = 1,59**.

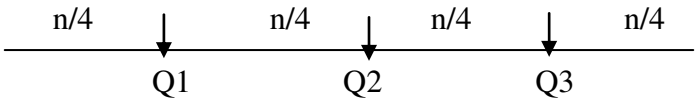
إنه مجرد التمعن في مدى كل سلسلة يؤكد لنا أن السلسلة الثانية هي أكثر نشرا و إنتشارا من الأولى.

3- <u>المجال الربيعي البيني</u> :

الربيع الأول لتوزيع إحصائي هو القيمة التي يأخذها المتغير بحيث يكون 25% من القيم التي يأخذها متغير أقل منها، و 75% أكبر منها و نرمز لهذه القيمة بالرمز Q_1.

الربيع الثالث لتوزيع إحصائي الذي نرمز له بالرمز Q_3 هو قيمة المتغير التي تكون أكبر من 75% من القيم التي يأخذها المتغير، و تكون أصغر من 25% منها.

و من هذين التعريفين ندرك أن الربيع الثاني ينطبق على الوسيط و من هذا نستخلص أن الربيعات هي القيم الثلاثة التي يأخذها المتغير و التي تقسم الملاحظات المرتبة حسب القيم المتصاعدة، إلى أربع مجموعات لكل منها نفس التكرار.



المجال الربيعي البيني هو الفرق بين الربيع الثالث و الربيع الأول.

المجال الربيعي البيني = $Q_1 - Q_3$

مثال : لنسترجع توزيع عمال إحدى المؤسسات حسب العمر و الذين يبلغ عددهم 150.

العمر (فئات)	التكرارات n_i	تكرار متراكم صاعد	تكرار متراكم نازل
]20 ، 25]	9	9	150
]25 ، 30]	27	36	141
]30 ، 35]	36	72	114
]35 ، 40]	45	117	78
]40 ، 45]	18	135	33
]45 ، 50]	9	144	15
]50 ، 55]	3	147	6
[55 ، 60]	3	150	3
	150		

الربيع الأول : $150\times\frac{25}{100}=37,5$ و بتغيير آخر هو قيمة المتغير التي تطابق الملاحظة التي يكون ترتيبها 37,5 ، و هو يتواجد ضمن الفئة $]30,35]$.

و بفرض وجود إنتظام داخل الفئات يكون :

$$Q_1 = 30 + (35 - 30)\ \frac{37,5-36}{72-36} \cong 30,2$$

الربيع الثالث : $\left(\frac{75}{100}\text{x}150=112,5\right)$، فهو موجود في الفئة $]35,40]$ و يساوي :

$$Q_3 = 35 + (40-35)\frac{112,5-72}{117-72} = 39,5$$

المجال الربيعي البيني : 9.3 $= 30.2 - 39.5 = Q - Q_3$

WWW.DZPEN.COM | قلم الجزائر

و هكذا نجد أن مجال هذه السلسلة الذي يبلغ طوله 60 – 40 = 20 يضم 100% من الملاحظات المجراة. بينما المجال الربيعي البيني الذي طوله 9,3 يتضمن 50% من الملاحظات المجراة.

ومن هذا يتضح أن المجال الربيعي البيني هو وسيلة لتقدير الإنتشار حول القيم المركزية. ففي المثال السابق نجد أن نصف التكرارات يرتكز في مجال يقل طوله عن ربع المجال الكلي.

4- العشريات و العشريات البينية:

- العشر الأول الذي نرمز له بالرمز D_1 هو قيمة المتغير التي يكون أكبر من 10% من القيم الملحوظة، و تكون أقل من 90%.
- العشر الثاني D_2 هو قيمة المتغير التي تكون أكبر من 20% من القيم الملحوظة، و تكون اقل من 80% منها.
- العشر الخامس هو الوسيط بطبيعة الحال.

وعليه فإن الفرق $D_8 - D_1$ يحدد مجالا يتضمن 80% من الملاحظات الواقعة.

5- الانحراف المتوسط :

رأينا أن ميزات التشتت تهدف إلى قياس انتشار القيم الملحوظة حول القيم المركزية.
فإن كانت القيمة المركزية المختارة هي المتوسط الحسابي، فإنه من المنطقي حساب انحراف كل قيمة من قيم المتغير عن المتوسط الحسابي، ثم جمع هذه الانحرافات و حساب متوسطها الحسابي.

وبذلك نحصل على الانحراف المتوسط. و بتعبير آخر، الانحراف المتوسط هو المتوسط الحسابي لانحرافات المتغير عن المتوسط الحسابي الذي يخص السلسلة.

$$e = \frac{\sum |x_i - \bar{x}| n_i}{\sum n_i}$$

نحسب المثال السابق :

الفئات	مراكز الفئات	التكرار n_i	$\|xi - \bar{x}\|$	$\|x_i - \bar{x}\| n_i$
]25 ، 20]	22,5	9	13	117
]30 ، 25]	27,5	27	8	216
]35 ، 30]	32,5	36	3	108
]40 ، 35]	37,5	45	2	90
]45 ، 40]	42,5	18	7	126
]50 ، 45]	47,5	9	12	108
]55 ، 50]	52,5	3	17	51
[60 ، 55]	57,5	3	22	66
		150		**882**

$$\bar{x} = 35,5 \qquad e = \frac{882}{150} = 5,88$$

وهكذا نجد أعمار عمال هذه المؤسسة لا تبتعد في المتوسط إلا بنحو 6 سنوات (5,88) عن المتوسط الحسابي لهذه الأعمار.

6- التباين و الانحراف المعياري :

1.6 - تعاريف :

أ- التباين هو المتوسط الحسابي لمربع انحرافات قيم المتغير عن المتوسط الحسابي لهذه القيم. ونرمز له بالرمز σ^2.

(يقرأ سيجما مربع) ، أو بالرمز V (x)

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2 n_i}{\sum n_i}$$

ب-الإنحراف المعياري الذي نرمز له بالرمز σ هو الجذر التربيعي الموجب للتباين.

$$\sigma = \sqrt{v(x)}$$

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

نحسب المثال السابق (عن الانحراف المتوسط) :

$(x_i - \bar{x})^2 n_i$	$(x_i - \bar{x})^2$	$\lvert x_i - \bar{x} \rvert$	n_i
1521	169	13	9
1728	64	8	27
324	9	3	36
180	4	2	45
882	49	7	18
1296	144	12	9
867	289	17	3
1452	484	22	3
8250			

$$\sigma^2 = \frac{8250}{150} = 56 \quad \Rightarrow \sigma = \sqrt{55} = 7.4$$

WWW.DZPEN.COM | قلم الجزائر

2.6- خواص التباين و الانحراف المعياري :

- التباين موجب دائما لأنه متوسط لمربعات، أي لأعداد موجبة، وكذلك الحال بالنسبة للإنحراف المعياري، فهو الجذر التربيعي الموجب للتباين.
- الإنحراف المعياري هو دائما نفس طبيعة المتغير الملحوظ و يقاس كذلك بنفس وحدة قياس المتغير.
- كلما كانت السلسلة منتشرة، كلما كانت هذه المقاييس كبيرة و بقدر ما تكون السلسلة متمركزة، تكون هذه المؤشرات صغيرة.

3.6- صيغة أخرى للتباين :

كثيرا ما يضم المتوسط $\bar{x}$ كسورا أكثر مما تضمه كل وحدة من الملاحظات x_i وكذلك الحال بالنسبة للفروقات $(x_i - \bar{x})$، ولهذا كثيرا ما يصعب حساب $(x_1 - \bar{x})$ وحساب $(x_1 - \bar{x})^2$.

و من هنا تظهر أهمية التعبير عن التباين بصيغة أخرى هي :

تباين سلسلة إحصائية يساوي متوسط مربعات الملاحظات بعد أن تنقص من هذا المتوسط مربع متوسط هذه الملاحظات.

البرهان :

ليكن :

$$v(x) = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i}{\sum_{i=1}^{k} n_i} = \frac{1}{N}\sum_{i=1}^{K} n_i (x_i - \bar{x})^2$$

بنشر كل مربع ينتج:

$$\begin{cases} (x_i - \bar{x})^2 = x_i^2 - 2x_i\,\bar{x} + \bar{x}^2 \\ (x_2 - \bar{x})^2 = x_2^2 - 2x_2\,\bar{x} + \bar{x}^2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ (x_k - \bar{x})^2 = x_k^2 - 2x_k\,\bar{x} + \bar{x}^2 \end{cases}$$

و إذا ضربنا هذه المساوات على التوالي بالتكرارات $n_1, n_2, \ldots, n_p$ ثم نجمعها و نحصل على :

$$\sum_{i=1}^{k} n_i (x_i - \bar{x})^2 = \sum_{i=1}^{k} n_i (x_i)^2 - 2\,\bar{x} \sum_{i=1}^{k} n_i x_i + N(\bar{x})^2$$

و بما أن : $N\,\bar{x} = \sum_{i=1}^{k} n_i x_i$

إذن نصل إلى أن :

$$V(x) = \frac{1}{N} \sum_{i=1}^{K} n_i (x_i)^2 - (\bar{x})^2$$

لنطبق هذا على المثال السابق :

مركز الفئة	تكرار n_i	x_i^2	$n_i\,x_i^2$
22,5	9	506,25	4.556,25
27,5	27	756,25	20.418,75
32,5	36	1.056,25	38.025,0
37,5	45	1.406,25	63.281,25
42,5	18	1.806,25	32.512.25
47,5	9	2.256,25	20.306,25
52,5	3	2.756,25	8.268,75
57,5	3	3.306,25	9.918,75
			197.287,50

$$\sigma^2 = \frac{197287{,}50}{150} - (35{,}5)^2 \quad \Rightarrow \quad \sigma^2 = 55$$

4.6- طريقة عملية لحساب التباين :

تستمد هذه الطريقة من إجراء تغيرات على قياسات المتغير x، و هي تغيير نقطة الأصل ووحدة القياس حيث ترتبط القيم الجديدة x_i بالقيم القديمة x_i حسب العلاقة التالية :

$$x_i = x_0 + a\ x'_i \quad \Rightarrow \quad \bar{x} = x_0 + a\ \bar{x}'$$

و منه :

$$x_i - \bar{x} = a\,(x'_i - \bar{x}')$$

إذن

$$V(x) = \frac{1}{N}\sum_{i=1}^{K} n_i\,(x_i - \bar{x}) \quad \text{ou} \quad V = \frac{a^2}{N}\sum_{i=1}^{K} n_i\left(x'_i - \bar{x}'\right)^2$$

لتكن $V'(x)$ تباين سلسلة x'_i أي :

$$V'(x) = \frac{1}{N}\sum_{i=1}^{K} n_i\,(x'_i - \bar{x}')^2\ ;\ \sigma' = \sqrt{v'(x)}$$

و أخيرا نجد أن :

$$\sigma = |a|\ \sigma' \qquad \text{و} \qquad V(x) = a^2\, v'(x)$$

لتطبق هذا على المثال السابق بحيث ندخل التغييرات التالية :

$\bar{x} = 35{,}5$، نأخذ $x_0 = 37{,}5$ و $a = 5$

$\left(\frac{x_i - 37,5}{5}\right)^2 n_i$	$\left(\frac{xi - 37,5}{5}\right)^2$	$\frac{xi - 37,5}{5}$	تكرار n_i	مركز الفئة
81	9	3 -	9	22,5
108	4	2 -	27	27,5
36	1	1-	36	32,5
0	0	0	45	37,5
18	1	1	18	42,5
36	4	2	9	47,5
27	9	3	3	52,5
48	16	4	3	57,5
354			**150**	

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2 n_i}{\sum n_i} = \frac{\sum (x_i - x_0)^2 n_i}{\sum n_i} - (\bar{x} - x_0)^2$$

$$\sigma^2 = a^2 \frac{\sum x_i'^2 n_i}{\sum n_i} - (\bar{x} - x_0)^2 \quad \text{avec} \quad x'_i = \frac{x_i - x_0}{a}$$

$$\sigma^2 = 5^2 \times \frac{354}{150} - (35,5 - 37,5)^2 = 5^2 \times \frac{354}{150} - 4 = 55$$

7- معامل التباين (التغير) :

يقاس المتوسط $\bar{x}$ تماما مثل الانحراف المعياري بنفس الوحدات التي يقاس بها المتغير x. و لتجاوز إشكالية الإختلاف في وحدات القياس الممكنة، يستخدم معامل التباين في التغيرات الموجبة فقط و نعرفه بأنه حاصل قسمة الإنحراف المعياري على المتوسط:

$$Cv = \sigma / \bar{x}$$

و هو مقدار نسبي لا يرتبط بوحدات القياس المستخدمة، ولهذا فهو يسمح مثلا بمقارنة توزيعات الأجور في بلدان مختلفة أو حسب التأهيلات المختلفة.