الفصل الرابع: مقاييس التشتت

1/4 المدى 66
2/4 نصف المدى الربيعي أو الانحراف الربيعي 67
3/4 الانحراف المتوسط 68
4/4 التباين 69
5/4 الانحراف المعياري 71
6/4 معامل الاختلاف 73
7/4 خصائص واعتبارات مقاييس التشتت 74
8/4 كشف تشتت النقاط البعيدة بمخطط العلبة Box Plot 75
1/8/4 التوزيع الطبيعي 75
2/8/4 مخطط العلبة البياني(علبة الشوارب) 76
3/8/4 طريقة حساب حدود المجالات مخطط العلبة 78

مسألة 05 (مقاييس التشتت) 81
مسألة 06 (مقاييس التشتت) 82

قائمة المراجع 83

الفصل الرابع: مقاييس التشتت

الدراسة الوصفية تبنى أيضا على قياس التشتت أو التبعثر للإحصائيات وتوزعها في مجال تغيرها وحول الوسط لهذا التوزيع، كما تقيّم مدى فعالية مقاييس النزعة المركزية، فمثلا كلما تجمعت القيم حول المتوسط الحسابي كلما كان مقياس التشتت صغيرا كان هذا مؤشرا على أن مقياس النزعة المركزية يمثل بياناته أفضل تمثيل. وهي تساعد أيضا على مقارنة التوزيعات مع بعضها البعض. مقاييس التشتت تصنف إلى صنفين: الاول يوضح مدى وكيفية اختلاف المقاييس فيما بينها؛ الثاني، يبين مدى تناثر أو تجمع القيم حول المتوسطات.

أهم مقاييس التشتت:

- المدى: وهو أبسط مقاييس التشتت، ويعبر عن الفرق بين اكبر مشاهدة أو اصغر مشاهدة.
- الانحراف الربيعي: يعرف على أنه نصف المدى الربيعي. والمدى الربيعي هو الفرق بين الربيع الثالث والربيع الأول، وذلك بعد ترتيب البيانات تصاعدياً أو تنازلياً .
- الانحراف المتوسط : وهو الوسط الحسابي لمجموع انحرافات القيم عن الوسط الحسابي بالقيمة المطلقة .
- التباين: مجموع مربع انحرافات القيم عن الوسط الحسابي مقسوما على عددها، إذ يقوم على أساس متوسط انحراف القيم عن الوسط الحسابي، كما أنه يأخذ بذلك جميع القيم بعين الاعتبار دون أن يستبعد أي قيمة ،كما و تسمح قيمته في معرفة القدرة التمثيلية للوسط الحسابي.

سبق أن درسنا عرض البيانات الإحصائية ووصفها، ثم مقاييس النزعة المركزية التي تصف هذه البيانات عددياً، وكذلك تمت دراسة مقاييس المكانة التي تستعمل في وصف مكانة الفرد النسبية، فهي تصف ترتيبه بالنسبة للمجتمع وتحدد موقعه النسبي بالمقارنة مع جميع الأفراد في مجال البحث، إلا أن هذه المقاييس غير كافية لتحديد صفات التوزيعات التكرارية والبيانات الإحصائية، فربما يكون لدينا ظاهرتان متساويتان في مقاييس النزعة المركزية كالوسيط والوسط الحسابي إلا أنهما مختلفتان، مثال (01-04):

إذا كانت درجات الحرارة في مدينتين كالقدس وغزة خلال أشهر العام 2013 هي:

JERUSALEM	18	18	17	17	15	20	24	30	28	30	28	25
GAZA	18	14	20	21	23	27	30	30	28	25	16	18

هاتان المدينتان لهما نفس الوسط الحسابي ($\overline{X} = 22{,}5$) وقيمة الوسيط متقاربة وهذا يعني أن معدل درجات الحرارة في المدينتين متساوٍ، إلا أنه بالنظر إلى البيانات في كل من المدينتين نجد اختلافاً بينهما، وهذا يدل أن الوسط الحسابي لا يكفي لوصف البيانات أو الحكم على تشابهها. إذن لابد من استعمال مقاييس أخرى تبين لنا مدى اختلاف البيانات فيما بينها ومدى التفاوت والتغاير بين مفرداتها.

التشتت أو التغاير: يقصد به الدرجة التي تتجه بها البيانات الرقمية للانتشار حول قيمة وسطى تسمى تشتت أو تغاير البيانات، أو بمعنى آخر، التشتت يقصد به درجة التفاوت أو الاختلاف بين قيم مجموعة فإذا كانت هذه

القيم متقاربة من بعضها البعض يكون التشتت صغيراً وإذا كانت متباعدة عن بعضها البعض أي متباينة يكون التشتت كبير.

1/4 المدى العام (RANGE)

يرمز له بـ E يعرف بمجال التغير وهو الفرق بين اكبر قيمة وادنى قيمة.

المدى = اكبر قيمة – اصغر قيمة ؛ $E = X_{max} - X_{min}$

مثال (02-04): اوجد المدى العام لمعطيات المثال (01-14)

$E_1=30-15=15$ مدينة القدس؛ تغيرات درجة الحرارة على مدار أشهر السنة يتراوح بين 15 درجة حرارية.

$E_2=30-14=16$ مدينة غزة؛ تغيرات درجة الحرارة على مدار أشهر السنة يتراوح بين 16 درجة حرارية.

2/4 المدى الربيعي أو الانحراف الربيعي (INTERQUARTILE AMPLITUDE)

يرمز له بـ IQ هو عبارة عن الفرق بين الربيع الثالث والربيع الأول. نحسب قيمته بالصيغة الآتية:

$$IQ=Q_3-Q_1$$

حيث Q_1 هو الربيع الأول و Q_3 هو الربيع الثالث.

وهو يحتوي على 50% الوسطية من المجتمع الاحصائي المدروس.

مثال (03-04): ليكن الاحصائيات التالية لمردودية إنتاج القمح لمجموعات فلاحية بولاية عين الدفلى.

- اوجد قيمة المدى الربيعي (IQ)

Xi (قنطار/هكتار)	ni	ecc
[100 – 110[	27	27
[110 – 120[	24	51
[120 – 130[	11	62
[130 – 140[	4	66
[140 – 150[	4	70
[150 – 160[	2	72

الحل

1. نقوم بإيجاد التكرار المتجمع التصاعدي.
2. نحدد رتبة الربيع الأول والثالث $RQ_1=72/4=18$; $RQ_3=3(72)/4=54$
3. نحسب قيمة الربيعي

$$Q_1 = 100 + \frac{18-0}{27}(10) = 100 + 06{,}66 = 106{,}66$$

$$Q_3 = 120 + \frac{54-51}{11}(10) = 120 + 02{,}72 = 122{,}72$$

4. نحسب المدى الربيعي:

$$IQ=Q_3-Q_1=122{,}72-106{,}66=16{,}06$$

التعليق: 50% الوسطية لمردودية إنتاج القمح للمجموعات الفلاحية بولاية عين الدفلى يتراوح بين 123 و107 أي بفارق 16 قنطار/هكتار مما يدل على قلة التشتت.

3/4 الانحراف المتوسط :

هو عبارة عن مجموع انحرافات البيانات عن وسطها الحسابي ولكن هذا المجموع يساوي صفراً دائماً لأن مجموع الانحرافات الموجبة عن الوسط الحسابي يساوي مجموع الانحرافات السالبة، ولذلك لابد من حذف الإشارة السالبة لنحصل على مقياس ذي معنى. وللتخلص من الإشارة السالبة نأخذ القيمة المطلقة. ونرمز له EM

$$EM = \frac{\sum_{i=1}^{k} n_i |x_i - \bar{X}|}{\sum_{i=1}^{k} n_i}$$

مثال (04-04): ليكن احصائيات مثال (04-03)

احسب الانحراف المتوسط؟

الحل

xi (ق/هـ)	ci	ni	ni*ci	$x_i - \bar{X}$	$\lvert x_i - \bar{X}\rvert$	$n_i\lvert x_i - \bar{X}\rvert$
100-110	105	27	2835	-11,66	11,66	314,82
110-120	115	24	2760	-1,66	1,66	39,84
120-130	125	11	1375	8,34	8,34	91,74
130-140	135	4	540	18,34	18,34	73,36
140-150	145	4	580	28,34	28,34	113,36
150-160	155	2	310	38,34	38,34	76,68
$\sum$	/	72	8400	/	106,68	709,8

1. نحسب الوسط الحسابي.

$$\bar{X} = \frac{8400}{72} = 116{,}66$$

2. نحسب الفرق بين بين كل واحدة من القيم والوسط الحسابي وبالقيمة المطلقة.
3. نجمع كل الفروق السابقة بدون النظر للإشارة (بالقيم المطلقة) مضروبة في التكرار.
4. نحسب الانحراف المتوسط

$$EM = \frac{\sum_{i=1}^{k} n_i |C_i - \bar{X}|}{\sum_{i=1}^{k} n_i} = \frac{709{,}8}{72} = 9{,}85$$

ملاحظة: نستطيع حساب الانحراف المتوسط بالنسبة للوسيط وذلك باستبدال المتوسط الحسابي بالوسيط

$$EM_{Me} = \frac{\sum_{i=1}^{k} n_i |x_i - Me|}{\sum_{i=1}^{k} n_i}$$

4/4 التباين (THE VARIANCE):

هو عبارة عن متوسط مربعات انحرافات القيم عن متوسطها الحسابي. ونرمز له بالرمز $V(x)$

أ. في حالة سلسلة بسيطة

تعطى بالعلاقة التالية: $V(x) = \frac{\sum_{i=1}^{n} (x_i - \bar{X})^2}{n}$

مثال (05-04): احسب التباين للبيانات التالية:

5 8 1 2

الحل

Xi	Xi-$\bar{X}$	$(X-\bar{X})^2$
2	-2	4
1	-3	9
8	4	16
5	1	1
16	/	30

1. نحسب الوسط الحسابي.

$$\bar{X} = \frac{\sum x_i}{n} = \frac{16}{4} = 4$$

2.- نحسب الفرق بين بين كل واحدة من القيم والمتوسط الحسابي.

3.- نحسب مربعات هذه الانحرافات، ثم نجمعها.

4.- نعوض في الصيغة السابقة

$$V(x) = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n} = \frac{30}{4} = 7{,}5$$

ب. في حالة بيانات مبوبة

تعطى بالعلاقة التالية:

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

يمكن ان ننشر العلاقة السابقة كما يلي:

$$V(x) = \frac{\sum_{i=1}^{k} n_i (x_i^2 - 2x_i\bar{X} + \bar{X}^2)}{\sum_{i=1}^{k} n_i}$$

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

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

$$= \frac{\sum_{i=1}^{k} n_i x_i^2}{\sum_{i=1}^{k} n_i} - \frac{2\bar{X}\sum_{i=1}^{k} n_i\,x_i}{\sum_{i=1}^{k} n_i} + \frac{\bar{X}^2\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} n_i}$$

$$= \frac{\sum_{i=1}^{k} n_i x_i^2}{\sum_{i=1}^{k} n_i} - 2\bar{X}\,\bar{X} + \bar{X}^2$$

$$= \frac{\sum_{i=1}^{k} n_i x_i^2}{\sum_{i=1}^{k} n_i} - 2\bar{X}^2 + \bar{X}^2$$

$$= \frac{\sum_{i=1}^{k} n_i x_i^2}{\sum_{i=1}^{k} n_i} - \bar{X}^2$$

مثال (04-06): البيانات التالية لعدد الاطفال في 69 الاسرة في قرية من قرى ولاية عين الدفلى.

xi	ni	ni.xi	ni.(xi)^2
1	12	12	12
2	8	16	32
3	14	42	126
4	9	36	144
5	11	55	275
6	15	90	540
$\sum$	69	251	1129

احسب التباين؟

الحل:

1. نحسب المتوسط الحسابي.

$$\bar{X} = \frac{\sum_{i=1}^{k} n_i x_i}{\sum_{i=1}^{k} n_i} = \frac{251}{69} = 3.64$$

(متوسط عدد الاطفال في الاسرة هو 3.64 طفل)

2. نضرب كل تكرار في مربع قيم المتغير المقابلة له، ثم نجمعهم.
3. نعوض في الصيغة المنشورة السابقة

$$V(x) = \frac{\sum_{i=1}^{k} n_i x_i^2}{\sum_{i=1}^{k} n_i} - \bar{X}^2 = \frac{1129}{69} - (3{,}64)^2 = 3{,}13$$

التعليق: تباين عدد الاطفال لأسر هذه القرية هو 3,13 (طفل)^2.

ملاحظة: - يمكن استعمال العلاقة الأولى إلا أن الأخيرة هي الأبسط.

- وحدة القياس مربعة.
- في حالة متغير مستمر نستبدل Xi بـ Ci.

5/4 الانحراف المعياري (STANDARD DEVIATION):

الانحراف المعياري هو عبارة عن الجذر التربيعي للتباين، نظير اعادة أس وحدة القياس الى الحالة العادية، وهذه هي الغاية من ايجاد هذا المقياس، وعليه يعتبر أهم مقاييس التشتت وأكثرها استخداما ويدخل في حساب كثير من المقاييس الإحصائية الأخرى. وله عدة رموز منها: $SD_x;\ \sigma_x$

أ. حالة البيانات البسيطة

$$SD_x = \sqrt{V(x)} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n}}$$

مثال (04-07): باستخدام معطيات مثال (04-05)

أحسب الانحراف المعياري

الحل

$$SD_x = \sqrt{V(x)} = \sqrt{7{,}5} = 2{,}73$$

ب. حالة البيانات المبوبة

$$SD_x = \sqrt{V(x)} = \sqrt{\frac{\sum_{i=1}^{k} n_i(x_i - \bar{X})^2}{\sum_{i=1}^{k} n_i}}$$

$$SD_x = \sqrt{V(x)} = \sqrt{\frac{\sum_{i=1}^{k} n_i x_i^2}{\sum_{i=1}^{k} n_i} - \bar{X}^2}$$

مثال (08-04): باستخدام معطيات مثال (06-04)

أحسب الانحراف المعياري

الحل

$$SD_x = \sqrt{V(x)} = \sqrt{3,13} = 1,77$$

التعليق: الانحراف المعياري لعدد الاطفال الأسر لهذه القرية هو 1,77 طفل.

مثال (09-04): ليكن توزيع الأجر الاسبوعي لمجموعة عمال كما يلي:

أحسب الانحراف المعياري

xi(1000دج)	ni	ci	ni.ci	ci-$\bar{X}$	(ci-$\bar{X}$)^2	ni (ci-$\bar{X}$)^2
8-10	11	9	99	-3,28	10,76	118,34
10-12	15	11	165	-1,28	1,64	24,58
12-14	10	13	130	0,72	0,52	5,18
14-16	9	15	135	2,72	7,40	66,59
16-18	5	17	85	4,72	22,28	111,39
	50	/	614			326,08

الحل

1. نحسب المتوسط الحسابي بعد حساب مراكز الفئات

$$\bar{X} = \frac{614}{50} = 12,28$$

متوسط الأجر الاسبوعي هو 12280 دج

2. نحسب التباين

$$V(x) = \frac{\sum_{i=1}^{k} n_i(x_i - \bar{X})^2}{\sum_{i=1}^{k} n_i} = \frac{326,1}{50} = 6,52$$

تباين الأجر الاسبوعي هو 6520 دج

3. نحسب الانحراف المعياري

$$SD_x = \sqrt{V(x)} = \sqrt{6,52} = 2,55$$

الانحراف المعياري للأجر الاسبوعي هو 2550 دج

6/4 معامل الاختلاف (CV)

هو النسبة المئوية للتغير الفعلي للتشتت نحصل عليه من قسمة الانحراف المعياري على المتوسط، إذا كان التشتت المطلق هو الانحراف المعياري SDx والمتوسط هو الوسط الحسابي $\overline{X}$ فإن التشتت النسبي يسمى بمعامل الاختلاف (أو معامل التشتت) ويعرف كالآتي:

$$CV = \frac{SD_x}{\overline{X}} \times 100$$

ولكن تغير أو تشتت **1** متر عند قياس مسافة 1000 متر يختلف في تأثيره عن نفس تغير 1 متر في مسافة 20 متر. ومقياس لهذا التأثير تحصل عليه بالتشتت النسبي.

تعبير عنه كنسبة يفيد عند مقارنة توزيعات ذات وحدات مختلفة (الوزن والطول) أو توزيعات ذات وحدات متساوية لكن وسطها الحسابي مختلف. أحد عيوب معامل الاختلاف هو أنه يصبح عديم الفائدة عندما تكون $\overline{X}$ قريبة من الصفر.

مثال (10-04): ليكن ينتج مصنع لإطارات السيارات نموذجين من الإطارات A و B النموذج A متوسط عمره 10000 KM وانحرافه المعياري 1200 KM . النموذج B متوسط عمره 11000 KM وانحرافه المعياري KM 1800 .

نود معرفة أي النموذجين أحسن؟

الحل:

نستخدم معامل التشتت وبعد ذلك نقارن:

النموذج A :

$$CV_A = \frac{SD_x}{\overline{X}} \times 100 = \frac{1200}{10000} \times 100 = 0{,}12 \times 100 = 12\%$$

النموذج B :

$$CV_B = \frac{SD_x}{\overline{X}} \times 100 = \frac{1800}{11000} \times 100 = 0{,}1636 \times 100 = 16{,}4\%$$

التعليق: معامل التشتت للنموذج A اقل من النموذج B، وبالتالي النموذج A أحسن من النموذج B.

7/4 خصائص واعتبارات مقاييس التشتت

1.– المدى العام والمدى الربيعي والانحراف المتوسط والانحراف المعياري قيمتهم تكون بنفس الوحدات المستخدمة في السؤال. التباين يكون بنفس الوحدات لكن قيمته مربعة. معامل الاختلاف (CV) يشير إلى ما هي نسبة الانحراف المعياري الموجودة في الوسط الحسابي.

2.– عند حساب الانحراف المتوسط والتباين والانحراف المعياري نأخذ في الحسبان كل البيانات. بينما المدى الربيعي يهتم فقط بقيمة الربيع الثالث (Q_3) والأول (Q_1) ، ولحساب المدى العام نركز فقط على أطراف القيم (أكبر وأصغر قيمة).

3.– الانحراف المتوسط والمدى الربيعي دائماً قيمتهم اصغر من الانحراف المعياري (لنفس البيانات أو التوزيع). إذا كان التوزيع طبيعي (سيتم شرحة فيما بعد) فالعلاقة بين هذه المقاييس الثلاثة ثابتة.

4.– عندما يكون الوسط الحسابي ($\bar{X}$) هو الأكثر ملائمة في مقاييس النزعة المركزية، فالانحراف المعياري كمقياس تشتت هو الملائم. وعندما يكون الوسيط هو الملائم، فالمدى الربيعي كمقياس تشتت هو الملائم.

$$\bar{X} \rightarrow SDx$$
$$Me \rightarrow IQ$$

5.– لا يمكن أن نحسب الانحراف الربيعي إذا كانت الفئة الربيع الثالث (Q_3) أو الفئة الربيع الأول (Q_1) مفتوحة.

6.– إذا كل الفئات تحتوي على نفس الطول يمكننا حساب التباين أو الانحراف المعياري بالطريقة المباشرة والطريقة المختصرة، لكن إذا وجدنا إحدى الفئات بطول مختلف يمكن حساب التباين (Sx) أو الانحراف المعياري (Sx) فقط بواسطة الطريقة المباشرة.

8/4 كشف تشتت النقاط البعيدة بمخطط العلبة Box Plot

كشف تشتت النقاط الطرفية هو ذو أهمية بمكان نظرا لتأثيره المباشر على مقاييس النزعة المركزية والتشتت (اهمها المتوسط الحسابي). كل الظواهر احصائية يكون مجال تغيراتها غالبا مركز في المنطقة الوسطى، خاصة اذا كان حجم العينة كبير بشكل كافي (على الاقل 30)، وهي بذلك تتبع التوزيع الطبيعي.

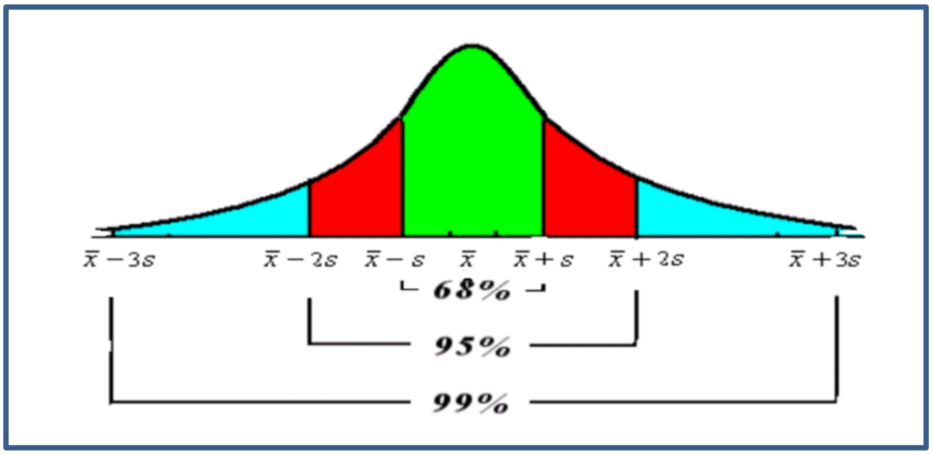
1/8/4 التوزيع الطبيعي

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

- 68% تقريبا من قيم هذه المشاهدات تتراوح بين $\bar{x} \pm SDx$.
- 95% تقريبا من قيم هذه المشاهدات تتراوح بين $\bar{x} \pm 2.SDx$.
- 99% تقريبا من قيم هذه المشاهدات تتراوح بين $\bar{x} \pm 3.SDx$.

ويمكن بيان ذلك من الشكل التالي:

شكل (5-1): شكل التوزيع الطبيعي المتناظر (المتماثل)



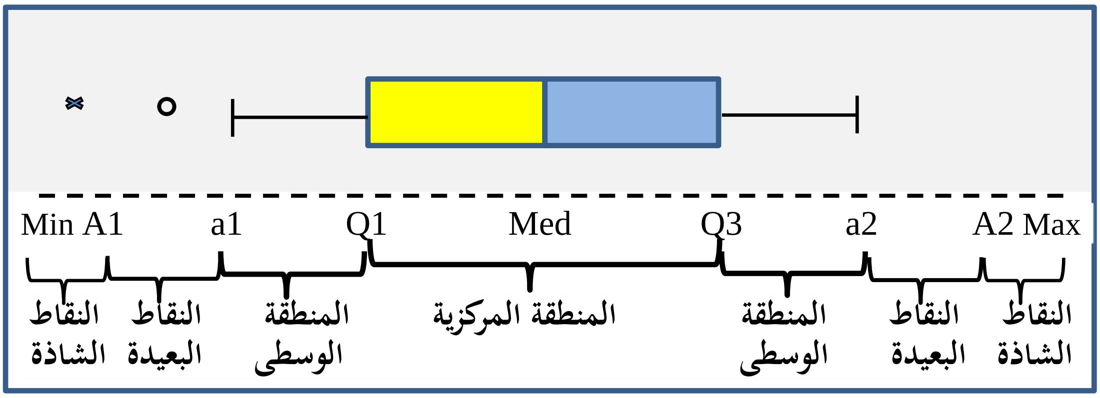
تسمى هذه القاعدة بقاعدة "تشيبشيف" ، وفكرة هذه القاعدة: في أي توزيع من التوزيعات النظرية ، فإنه على الأقل $\left(1-1/k^2\right)$ % من قيم المشاهدات تقع في المدى $\bar{x}\pm ks$ ، $k>1$.

وطبقا لهذه القاعدة، فإنه على الأقل 75% من قيم المشاهدات تقع في المدى $\bar{x}\pm 2s$، على الأقل 89% من قيم المشاهدات تقع في المدى $\bar{x}\pm 3s$.

<u>2/8/4 مخطط العلبة البياني (علبة الشوارب)</u>

أحسن تمثيل بياني لكشف تشتت النقاط البعيدة والشاذة هو مخطط العلبة البياني (علبة الشوارب) شكل مخطط العلبة البياني هو صندوق يشبه المستطيل، بداية حافته اليسرى هو الربيع الأول Q_1 ونهاية حافته اليمنى هو الربيع الثالث Q_3، ويقسم الربيع الثاني (الوسيط) Med المستطيل إلى جزأين، ويخرج من كل حافة من حافتيه خط مستقيم، والشكل التالي يبينه:

شكل (5-2): شكل مخطط العلبة البياني

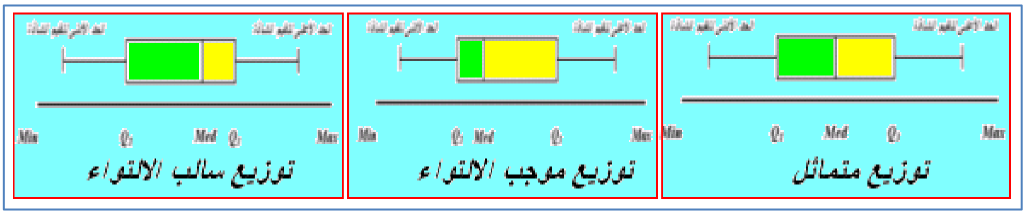


ويمكن استخدام مخطط العلبة البياني، أعلاه في وصف البيانات من حيث الآتي:

1- من حيث التماثل: إذا كان الوسيط Med يقع في المنتصف على بعد متساوي من الرباعيين Q_1 وQ_3

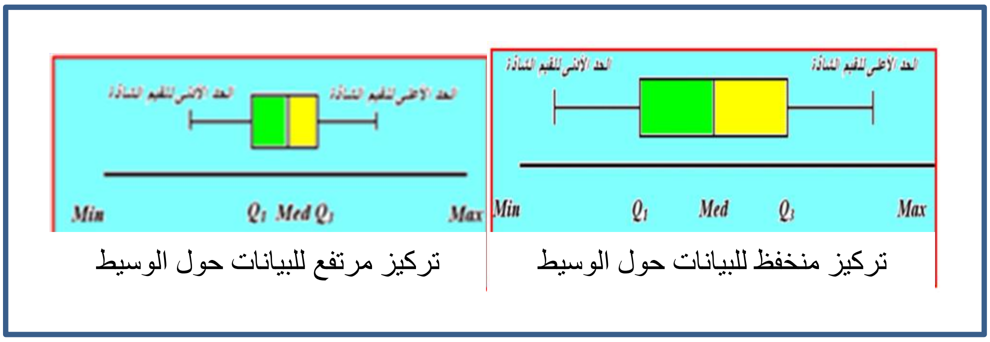
كان التوزيع متماثلا ، وإذا كان الوسيط Med أقرب إلى الرباعي الأول Q_1 من الرباعي الثالث Q_3 كان التوزيع موجب الالتواء ، وأما إذا كان الوسيط Med أقرب إلى الرباعي الثالث Q_3 من الرباعي الأول Q_1 كان التوزيع سالب الالتواء . ويظهر ذالك كما في الشكل التالي :

شكل (5-3): شكل الالتواء باستخدام مخطط العلبة



2- من حيث تركز البيانات: إذا كان الصندوق Box ضيق دل ذلك على تركز نسبة كبيرة من البيانات حول الوسيط، وإذا كان الصندوق واسع دل ذلك على انخفاض نسبة تركز البيانات حول الوسيط، والشكل التالي يبين ذلك.

شكل (5-4): درجة تركز البيانات



3- من حيث وجود القيم البعيدة والشاذة: إذا وقعت قيم بعض المشاهدات خارج الحدين الأدنى والأعلى للشارب (الخط المستقيم الخارج من العلبة) ، كانت هذه القيم "بعيدة" بالدرجة الاولى ثم "شاذة" بالدرجة الثانية، وتظهر القيم البعيدة على الرسم في شكل دائرة (o) ، وتظهر القيم الشاذة على الرسم في شكل نجوم (*) والشكل التالي يبين طريقة عرض القيم الشاذة الدنيا والعليا على الرسم .

3/8/4 طريقة حساب حدود المجالات لمخطط العلبة

لحساب حدين المنطقة الوسطى الأعلى والأدنى للخط المستقيم الخارج من العلبة (الشارب)، حيث يتوقف عند آخر نقطة تقع في المنطقة الوسطى المحصورة بين الربيعيين وكل من a1 وa2 اللذان يحسبان كما يلي:

a1 = Q1-1,5 IQ ;

a2 = Q3+1,5 IQ ;

IQ= Q3-A1

لحساب حدين منطقة النقاط البعيدة حيث تمثل النقاط في شكل دائرة صغيرة المحصورة بين كل من a1 ،a2 و A1 ، A2 اللذان يحسبان كما يلي:

A1 = a1-1,5 IQ ; A2 = a2+1,5 IQ

IQ= Q3-A1

وإذا وقعت قيم خارج الحدين تعتبر هذه القيم من القيم الشاذة.

مثال(4-11):

فيما يلي الإنفاق الاستهلاكي بالألف خلال الشهر لعينة حجمها 12 أسرة:

6 10 18 3 9 10 5 6 11 8 2 7

والمطلوب:

1- رسم مخطط العلبة البياني

2- اكتب تحليل وصفي لهذه البيانات.

الحل

1- رسم شكل مخطط العلبة

- ترتيب القيم تصاعديا .

2 3 5 6 6 7 8 9 10 10 11 18

- تحديد أقل وأعلى إنفاق استهلاكي، وحساب الرباعيات:

Min = 2   *Max = 18*

الرباعي الأدنى Q_1 :

موقع الرباعي $(n+1)(1/4)=(13/4)=3.25$

إذا قيمة Q_1 هي: $Q_1 = 5 + 0.25(6-5) = 5.25$

الوسيط Med:

موقع الوسيط $(n+1)(1/2)=(13/2)=6.5$

إذا قيمة Med هي: $Med = 7 + 0.5(8-7) = 7.5$

الرباعي الثالث Q_3:

موقع الرباعي $(n+1)(3/4)=(13)(3/4)=9.75$

إذا قيمة Q_3 هي: $Q_3 = 10 + 0.75(10-10) = 10$

- حساب الحدين الأعلى والأدنى للنقاط البعيدة

المدى الربيعي: $IQ = (10-5.25) = 4.75$

الحد الأدنى للقيم البعيدة:

$$a1 = Q_1 - 1{,}5\,IQ = 5.25 - 1{,}5(4.75) = -1.875$$

الحد الأدنى للقيم البعيدة :

$$a2 = Q_3 + 3Q = 10 + 1{,}5(4.75) = 17.125$$

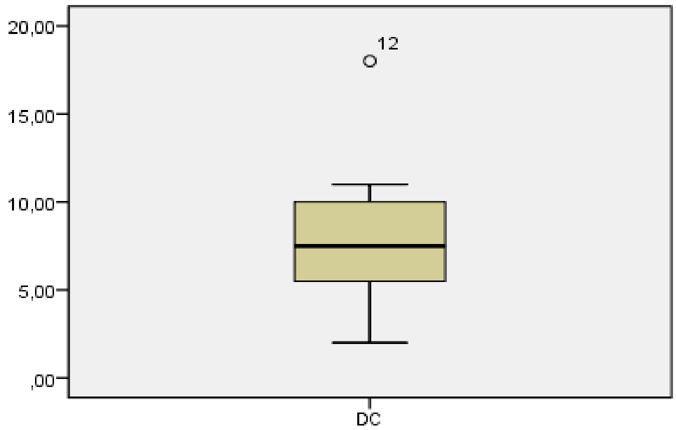
الحد الأعلى للقيم البعيدة:

$$A1 = a_1 - 1{,}5\,IQ = -1.875 - 1{,}5(2.375) = -9$$

الحد الأعلى للقيم البعيدة :

$$A2 = a_2 + 1{,}5Q = 17.125 + 1.5(4.75) = 24.25$$

- مخطط العلبة (علبة الشوارب)



يدل الرقم 12 الظاهر في الشكل السابق على رقم المشاهدة وهي المشاهدة الاخيرة قيمتها 18.

2- من خلال الشكل أعلاه:

- درجة التماثل : التوزيع قريب من التماثل لوقوع الوسيط في المنتصف .
- تركز البيانات: حوالي 60% من القيم تتركز حول الوسيط.
- القيم الشاذة: توجد قيمة بعيدة عليا هي القيمة 18.

يمكن استخدام مخطط العلبة البياني لمقارنة مجموعتين أو أكثر .

مسألة 05 (مقاييس التشتت)

الجزء الأول:

تبين البيانات التالية، توزيع الأجور الشهرية لـ 80 عاملا في مؤسسة ما :

فئة الأجور	1100-1200	1200-1300	1300-1400	1400-1500	1500-1600	1600-1700	1700-1800
عدد العمال	2	2	10	11	12	10	4

المطلوب:

1- أحسب كل من المدى العام و الانحراف المعياري؟

2- أحسب المدى الربيعي والانحراف المتوسط؟

3- أحسب معامل الاختلاف؟

4- نفرض أن لدينا توزيع تكراري أخر لمؤسسة ثانية حيث أن الانحراف المعياري لأجورها هو ضعف الانحراف المعياري لأجور المؤسسة الأولى و بنفس الوسط الحسابي. ما هو الفرق بين توزيع الأجور عمال المؤسستين؟

الجزء الثاني:

أجريت عملية مراقبة الوزن لعينة من علب عصير الفواكه وذلك قصد مراقبة الوزن المتفق عليه مع المصالح التجارية، فكانت النتائج كالتالي:

الوزن (غ)	أقل من 242	244-242	246-244	248-246	250-248	252-250	254-252	256-254	أكثر من 256
عدد العلب	15	20	23	27	30	18	12	3	2

المطلوب:

1- هل يمكن حساب المتوسط الحسابي؟

2- ما هو مقياس التشتت المناسب لهذا الجدول المفتوح؟

3- ليكن طول الفئة الأولى والأخيرة يساوي 2، أوجد كل من متوسط وزن العلبة ومقياس التشتت المناسب؟

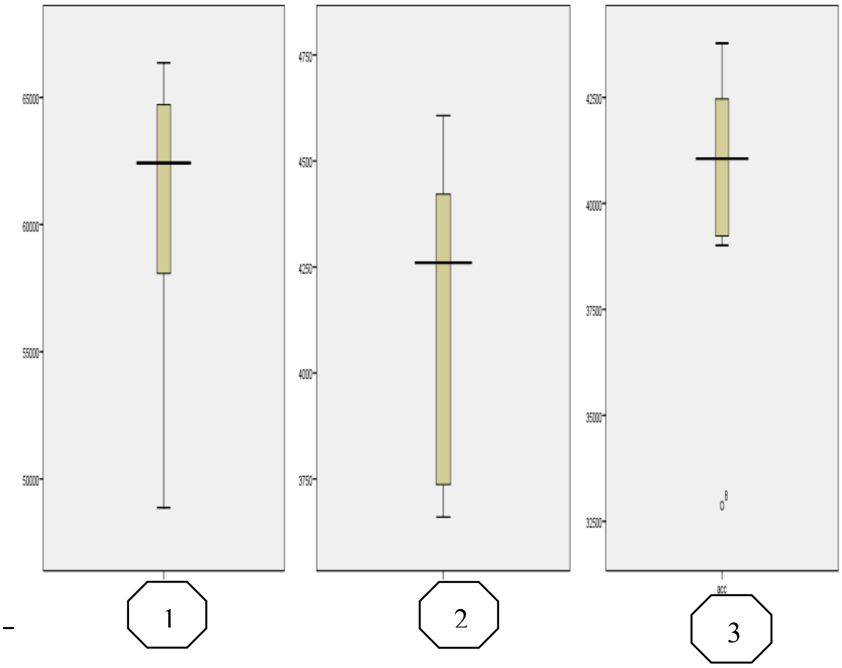
4-تحت قيد السؤال السابق احسب الانحراف المتوسط للوسيط؟

مسألة 06 (مقاييس التشتت)

المعطيات التالية تخص حوادث المرور الطرقات في الجزائر

السنوات	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
عدد القتلى	4343	4356	3711	4120	4177	4422	4607	3660	4598	3737
عدد الجرحى	63699	64714	58082	60120	61139	64708	64979	52435	66361	48875
عدد الحوادث	43227	43777	39233	40885	39010	40481	41224	32873	41467	42467

لدراسة ظاهرة حوادث المرور أجرينا تمثيل مخطط العلبة للسلاسل الثلاثة كما هي أسفله باستخدام برنامج الحلول الاحصائية للمنتجات والخدمات(SPSS version 18)



1- أحسب المتوسط الحسابي والتباين؟

2- حدد أي متغير تمثل مخططات العلب الثلاثة مع إعادة رسمها؟

3- في المخطط رقم 3؛ هل المشاهدة رقم 08 تمثل نقطة بعيدة أم متطرفة (شاذة)؟