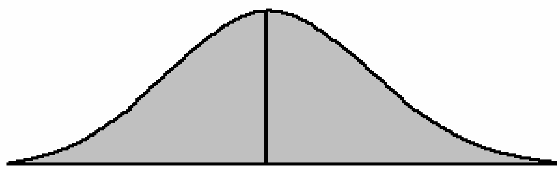
الفصـــــــل الخامس

بعض المقاييس الأخرى لوصف البيانات

1/5 مقدمـــــة

عند تمثيل بيانات الظاهرة في شكل منحني تكراري ، فإن هذا المنحني يأخذ أشكالا مختلفة ، فقد يكون هذا المنحنى متماثل بمعنى أن له قمة في المنتصف ، ولو أسقطنا عمودا من قمته على المحور الأفقي لشطره نصفين متماثلين ، مثل منحنى التوزيع الطبيعي ، كما هو مبين بالشكل التالي .



منحنى التوزيع الطبيعي (منحنى متماثل)

وعندما يكون الشكل متماثل ، فإن الوسط والوسيط والمنوال كلهم يقعون على نقطة واحدة ، ولكن في كثير من الحالات يكون هناك قيم كبيرة في البيانات تجذب إليها الوسط الحسابي ، وهذا معناه أن المنحنى التكراري سوف يكون له ذيل جهة اليمين ، مشيرا بوجود التواء جهة الـيمين ، وكـذلك العكس لو أن البيانات بها قيم صغيرة ، فإنها تجذب الوسط إليها، ويدل المنحني التكراري على وجـود التواء جهة اليسار، كما يمكن من خلال الشكل البياني معرفة ما إذا كان توزيع البيانات منبـسط، أو مدبب، وهذا من الناحية البيانية، إلا أن هناك مقاييس كثيرة لوصف البيانات تعتمد في حسابها علـى مقاييس النزعة المركزية والتشتت معا، ومنها مقاييس الالتواء، والتفرطح، وبعض المقـاييس الأخـرى سوف يتم عرضها فيما بعد.

2/5 مقاييس الالتواء Skewness

هناك طرق كثيرة لقياس الالتواء ومنها ما يلي:

1/2/5 طريقة "بيرسون Person" في قياس الالتواء

تأخذ هذه الطريقة في الاعتبار العلاقة بين الوسط والوسيط والمنوال، في حالة ما إذا كان التوزيع قريب من التماثل وليس شديد الالتواء ، وهذه العلاقة هي: (5-1)

$$\boxed{\text{المنوال} = \text{الوسط الحسابي} - 3(\text{الوسط الحسابي} - \text{الوسيط})} \qquad (٥-١)$$

ومن ثم فإن طريقة "بيرسون" في قياس الالتواء ، تتحدد بالمعادلة التالية.

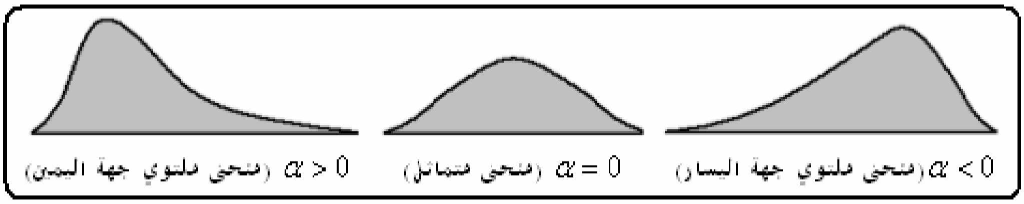
$$\boxed{\alpha = \frac{3(Mean - Median)}{Standard\ Deviation} = \frac{3(\bar{x} - Med)}{S}} \qquad (٥-٢)$$

حيث أن a (ألفا) هو معامل الالتواء "لبيرسون"، $\bar{x}$ الوسط الحسابي، Med هو الوسيط، S هو الانحراف المعياري، ويمكن من خلال الإشارة التي يأخذها هذا المعامل الحكم على شكل الالتواء، كما يلي :

- إذا كان (الوسط الحسابي = الوسيط) كان قيمة المعامل $(a = 0)$، ويدل ذلك على أن منحنى التوزيع التكراري متماثل.
- إذا كان (الوسط الحسابي > الوسيط) كان قيمة المعامل $(a > 0)$، ويدل ذلك على أن منحنى التوزيع التكراري ملتوي جهة اليمين .
- إذا كان (الوسط الحسابي < الوسيط) كان قيمة المعامل $(a < 0)$، ويدل ذلك على أن منحنى التوزيع التكراري ملتوي جهة اليسار.

شكل (5-1)

أشكال التواء البيانات



2/2/5 طريقة "المئين" في قياس الالتواء

المئين ينتج من ترتيب البيانات تصاعديا، ثم تقسيمها البيانات إلى 100 جزء، يفصل بينها قيم تسمى المئين، وعلى سبيل المثال يعرف المئين 15 ويرمز له بالرمز (v_{15}) على أنه القيمة التي يقل عنها 15% من القيم، ولحساب قيمة المئين p ، ونرمز له بالرمز (v_p)، يتبع نفس الفكرة المستخدمة في حساب الربيع كما يلي:

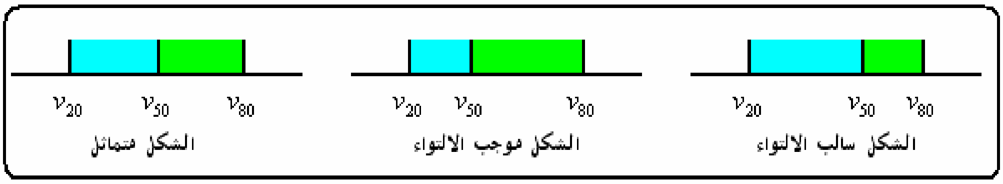
- ترتب القيم تصاعديا: $x_{(1)} < x_{(2)} < \ldots < x_{(n)}$
- رتبة المئين: $R = (n+1)\left(\frac{p}{100}\right)$.
- إذا كانت الرتبة R عدد صحيح فإن ($v_{15} = x_{(R)}$).
- أما إذا كانت الرتبة R عدد كسري فإن قيمة المئين (v_p) تحسب بالمعادلة التالية:

$$\boxed{v_p = x_{(l)} + (R - l)(x_{(u)} - x_{(l)})} \qquad (٥-٣)$$

وتعتمد فكرة المئين في قياس الالتواء على مدى قرب المئين v_p، والمئين v_{100-p}، من المئين v_{50} ، وكمثال على ذلك ، عند قياس الالتواء باستخدام المئين 20 ، والمئين 80 ، يلاحظ على الرسم التالي

حالات الالتواء :

شكل (5-2)



ومن الشكل أعلاه يلاحظ الآتي :

- إذا كان بعد المئين (v_{80}) عن المئين (v_{50}) يساوي بعد المئين (v_{20}) عن المئين (v_{50}) كـان التوزيع متماثلا .
- إذا كان بعد المئين (v_{80}) عن المئين (v_{50}) أكبر من بعد المئين (v_{20}) عن المئين (v_{50}) كـان التوزيع موجب الالتواء .
- إذا كان بعد المئين (v_{80}) عن المئين (v_{50}) أقل من بعد المئين (v_{20}) عن المئين (v_{50}) كـان التوزيع سالب الالتواء .

وبشكل عام يمكن الحكم على شكل التوزيع باستخدام معامل الالتواء المئيني، ويأخذ المعادلة التالية.

$$\boxed{\alpha_{p,100-p} = \frac{(v_{100-p} - v_{50}) - (v_{50} - v_p)}{v_{100-p} - v_p}} \qquad (٥-٤)$$

حيث أن : $v_p < v_{50} < v_{100-p}$ ويفضل استخدام هذا المعامل في حالة البيانات التي تحتوي علـى قيم شاذة ، وأيضا البيانات التي لا نعرف لها توزيع محدد، وعندما نستخدم المئين 25 $(v_{25} = Q_1)$، المئين 75 $(v_{75} = Q_3)$ نحصل على معامل الالتواء الربيعي ، وهو :

$$\boxed{v_q = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_1)}} \qquad (٥-٥)$$

مثـــال (5-1)

كانت درجات 8 طلاب في الاختبار النهائي في مقرر **122** إحص ، كالتالي.

66 85 52 78 80 91 74 58

والمطلوب : **1**- حساب معامل الالتواء بطريقة " بيرسون " .

2- حساب معامل الالتواء الربيعي .

الحـــــل

1- حساب معامل الالتواء بطريقة "بيرسون" .

في هذه الحالة يتم تطبيق المعادلة رقم (**5-2**) كما يلي:

- حساب الوسط الحسابي ، والانحراف المعياري :

الدرجة x	x^2
66	4356
85	7225
52	2704
78	6084
80	6400
91	8281
74	5476
58	3364
584	43890

$$\sum x = 584 \ , \ \sum x^2 = 43890$$

ويكون :

$$\bar{x} = \frac{\sum x}{n} = \frac{584}{8} = 73$$

$$s = \sqrt{\frac{\sum x^2 - (\sum x)^2 / n}{n-1}} = \sqrt{\frac{43890 - (584)^2 / 8}{8-1}}$$

$$= \sqrt{\frac{1258}{7}} = \sqrt{179.71428} = 13.406$$

- حساب الوسيط :

موقع الوسيط : *(n+1)/2=(8+1)/2=4.5*

52	58	66	74	78	80	85	91
1	2	3	4	5	6	7	8

2.25　　　4.5　　　6.75

$$Med = 74 + 0.5(78 - 74) = 76$$

- معامل الالتواء "بيرسون"

$$s.c = \frac{3(\bar{x} - Med)}{S} = \frac{3(73 - 76)}{13.406} = -0.67$$

إذا منحنى توزيع درجات الطلاب ملتوي جهة اليسار .

2- معامل الالتواء الربيعي .

لحساب معامل الالتواء الربيعي ، يتم تطبيق المعادلة رقم (**5-5**).

- حساب الربيع الأدنى .

موقع الرباعي : *(n+1)/4=(8+1)(1/4)=2.25*

إذا

$$Q_1 = 58 + (2.25 - 2)(66 - 58) = 60$$

- حساب الرباعي الأعلى.

موقع الرباعي : *(n+1)/(3/4)=(8+1) (3/4)=6.75*

إذا

$$Q_3 = 80 + (6.75 - 6)(85 - 80) = 83.75$$

- الوسيط (الربيع الثاني)

$$Med(Q_2) = 76$$

إذا معامل الالتواء الربيعي هو :

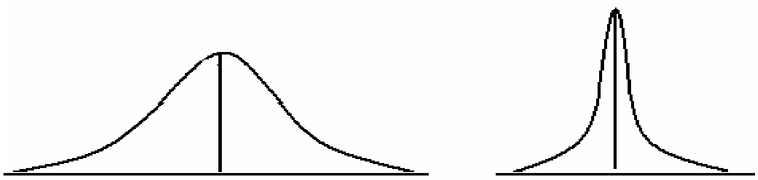
$$a_q = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_1)} = \frac{(83.75 - 76) - (76 - 60)}{(83.75 - 60)}$$

$$= \frac{-8.25}{23.75} = -0.35$$

إذا توزيع درجات الطلاب ملتوي جهة اليسار.

3/5 التفرطح Kurtosis

عند تمثيل التوزيع التكراري في شكل منحنى تكراري ، قد يكون هذا المـــنحنى منبـــسط ، أو مدبب ، فعندما يتركز عدد أكبر من القيم بالقرب من منتصف المنحنى، ويقل في طرفيه، يكون المنحنى مدببا ، وعندما يتركز عدد أكبر على طرفي المنحنى ، ويقل بالقرب من المنتصف يكون المنحنى مفرطحا ، أو منبسطا، ويظهر ذلك من الشكل التالي :



منحنى مدبب   منحنى مفرطح

ويمكن قياس التفرطح باستخدام عدد من الطرق، ومنها طريقة العزوم ، حيث يحسب معامل التفرطح (K) بتطبيق المعادلة التالية :

$$k = \frac{\frac{1}{n}\sum(x-\bar{x})^4}{s^4} \qquad (٥-٦)$$

حيث أن المقدار $\sum(x-\bar{x})^4/n$ هو العزم الرابع حول الوسط ، s هو الانحراف المعياري . ومعامل التفرطح في التوزيع الطبيعي يساوي 3 ، ومن ثم يمكن وصف منحنى التوزيع من حيث التفرطح ، والتدبب كما يلي :

- إذا كان $k=3$ كان منحنى التوزيع معتدلا .
- إذا كان $k>3$ كان منحنى التوزيع مدببا .
- إذا كان $k<3$ كان منحنى التوزيع منبسطا (مفرطحا) .

وبالتطبيق على بيانات المثال رقم (5-1) نجد أن: $\bar{x} = 73$

x	66	85	52	78	80	91	74	58	584
$(x-\bar{x})$	-7	12	-21	5	7	18	1	-15	0
$(x-\bar{x})^2$	49	144	441	25	49	324	1	225	1258
$(x-\bar{x})^4$	2401	20736	194481	625	2401	104976	1	50625	376246

ومن البيانات أعلاه نجد أن:

$$s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = \sqrt{\frac{1258}{7}} = 13.406$$

$$\frac{1}{n}\sum(x-\bar{x})^2 = \frac{1}{8}(376246) = 47030.75$$

إذا معامل التفرطح هو:

$$K = \frac{47030.75}{(13.406)^4} = \frac{47030.75}{(32299.58)} = 1.456$$

إذا شكل توزيع بيانات الدرجات مفرطح.

4/5 مقاييس أخرى لوصف البيانات

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

1/4/5 معامل الاختلاف Variation Coefficient

أحد مقاييس المستخدمة لقياس درجة التشتت، وفيه يحسب قيمة التشتت كنسبة مئوية من قيمة مقياس النزعة المركزية ، ومن ثم يفضل استخدام معامل الاختلاف عند مقارنة درجة تشتت بيانات مجموعتين أو أكثر مختلفة لها وحدات قياس مختلفة، بدلا من الانحراف المعياري ، أو الانحراف الربيعي ،

لأن معامل الاختلاف يعتمد على التغيرات النسبية في القيم عن مقياس النزعة المركزية، بينما يعتمد الانحراف المعياري أو الانحراف الربيعي على التغيرات المطلقة للقيم، فعند مقارنة درجة تشتت بيانات الأطوال بالسنتمتر، وببيانات الأوزان بالكيلوجرام، لا يمكن الاعتماد على الانحراف المعياري في هذه المقارنة، وإنما يستخدم معامل الاختلاف، ومن ثم يطلق عليه بمعامل الاختلاف النسبي، وفيما يلي بعض هذه المعاملات.

• معامل الاختلاف النسبي

ويحسب معامل الاختلاف النسبي بتطبيق المعادلة التالية:

$$\boxed{v.c = \frac{s}{\bar{x}} \times 100} \qquad (٥-٧)$$

• معامل الاختلاف الربيعي

ويحسب هذا المعامل بتطبيق المعادلة التالية:

$$\boxed{v.c_q = \frac{(Q_3 - Q_1)/2}{Med} \times 100} \qquad (٥-٨)$$

مثال (5-2)

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

المقاييس	المجموعة الأولى	المجموعة الثانية
$\bar{x} =$	173	198
$s =$	23	25

والمطلوب مقارنة درجة تشتت المجموعتين:

الحـــــــل:

- معامل الاختلاف النسبي للمجموعة الأولى:

$$v.c_1 = \frac{s}{\bar{x}} \times 100 = \frac{23}{173} \times 100 = 13.3\%$$

- معامل الاختلاف النسبي للمجموعة الثانية:

$$v.c_2 = \frac{s}{\bar{x}} \times 100 = \frac{25}{195} \times 100 = 12.8\%$$

يلاحظ أن درجة تشتت أوزان المجموعة الثانية أقل من درجة تشتت أوزان المجموعة الأولى.

2/4/5 تقدير مدى الانحراف المعياري

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

$$\frac{\text{المدى}}{6} < \text{الانحراف المعياري} < \frac{\text{المدى}}{4}$$

$$\frac{Rang}{6} < \hat{S} < \frac{Rang}{4} \qquad (٥-٩)$$

وإذا كان المدى الذي يقع فيه الانحراف المعياري صغير دل ذلك على أن تشتت البيانات صغير، أما إذا كان المدى كبير دل ذلك على وجود تشتت كبير في البيانات، وإذا وقع الانحراف المعياري خارج المدى دل ذلك على وجود قيم شاذة.

3/4/5 الدرجة المعيارية Standardized degree

تقيس الدرجة المعيارية لقيمة معينة عدد وحدات الانحراف المعياري التي تزيد بها تقل بها هذه القيمة عن الوسط الحسابي، فإذا كان $X_1, X_2, \ldots, X_n$ ، هي قيم المشاهدات، وعددها n ، وكان $\overline{X}$ هو الوسط الحسابي لهذه القيم، S هو الانحراف المعياري، فإن الدرجة المعيارية للقيمة X ، ويرمز لها بالرمز Z ، تحسب باستخدام المعادلة التالية:

$$z = \frac{x - \overline{x}}{s} \qquad (٥-١٠)$$

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

مثـــــال(5-3)

في المثال (5-2) السابق إذا تم اختيار أحد الأغنام من المجموعة الأولى بعد تطبيق البرنامج ، ووجد أن وزنه 178 كيلوجرام، وبالمثل أحد الأغنام من المجموعة الثانية، ووجد أن وزنه 180 كيلوجرام ، قارن بين هذين القيمتين من حيث أهمية كل منها في المجموعة التي تنتمي إليها.

الحـــــل

البيانات المتاحة عن كل من المجموعتين هي:

	المجموعة الأولى	المجموعة الثانية
$\overline{X} =$	173	198
$S =$	23	25

القيمة.	178	180

للمقارنة بين الوحدتين من حيث أهمية وزن كل منها في المجموعة التي تنتمي إليها، يتم حساب الدرجة المعيارية لوزن كل منها، بتطبيق المعادلة (5-10).

- الدرجة المعيارية لوزن الوحدة المسحوبة من المجموعة الأولى (178 Kg.) هي:

$$z = \frac{x - \overline{x}}{s} = \frac{178 - 173}{23} = 0.22$$

- الدرجة المعيارية لوزن الوحدة المسحوبة من المجموعة الثانية (180 Kg.) هي:

$$z = \frac{x - \overline{x}}{s} = \frac{180 - 198}{25} = -0.75$$

- نجد أن الوزن 178 كيلوجرام يزيد عن الوسط الحسابي بـ 0.22 انحراف معياري ، بينما نجد أن الوزن 180 كيلوجرام يقل عن الوسط الحسابي بـ 0.75 انحراف معياري . ومن ثم الوزن الأول أهميته النسبية أعلى من الوزن الثاني.

4/4/5 القاعدة العملية

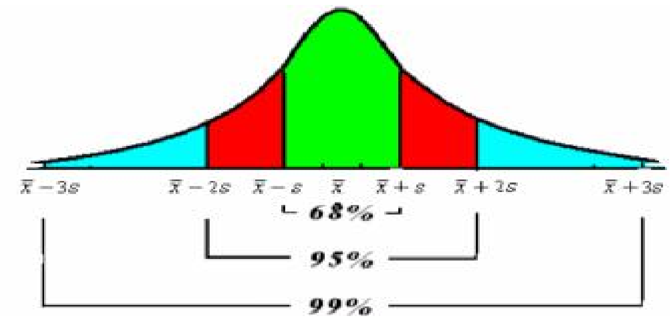
إذا كان لدينا المشاهدات التالية: $x_1, x_2, \ldots, x_n$ ، وكان $\overline{x}$ هو الوسط الحسابي لهذه المشاهدات، s هو الانحراف المعياري لها ، يكون منحنى توزيع هذه المشاهدات متماثل، إذا تحقق الآتي:

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

ويمكن بيان ذلك من الشكل التالي:

شكل (5-3)

شكل توزيع القيم طبقا للقاعدة العملية



5/4/4 القاعدة النظرية

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

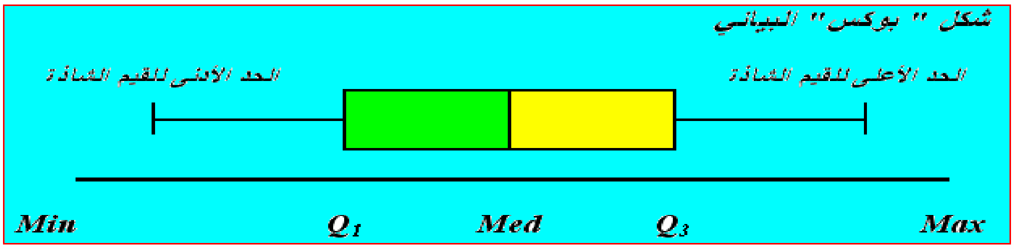
وطبقا لهذه القاعدة، فإنه على الأقل 75% من قيم المشاهدات تقع في المدى $\overline{x} \pm 2s$ ، على الأقل 89% من قيم المشاهدات تقع في المدى $\overline{x} \pm 3s$.

6/4/5 شكل "بوكس" Box Plot

شكل "بوكس" 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 شعيرة، والشكل التالي يبين رسمة "بوكس" البياني:

شكل (5-4)

رسمة بوكس البياني

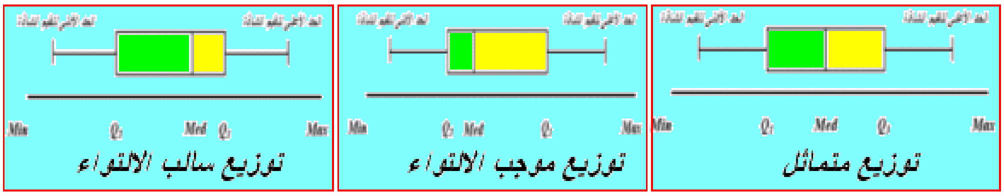


ويمكن استخدام شكل "بوكس" البياني، أعلاه في وصف البيانات من حيث الآتي:

1- من حيث التماثل: إذا كان الوسيط Med يقع في المنتصف على بعد متساوي من الرباعيين Q_3 , Q_1 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 (5-5)

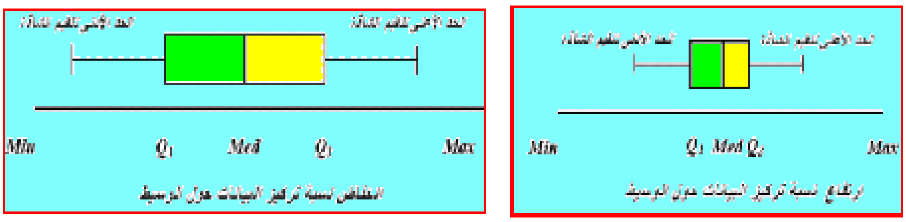
وصف شكل الالتواء باستخدام رسمة بوكس البياني



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 تركيز البيانات حول الوسيط، والشكل التالي يبين ذلك.

شكل (5-6)

وصف درجة تركز البيانات باستخدام رسمة بوكس البياني

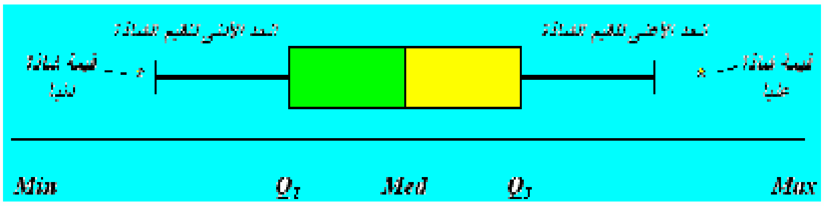


3- من حيث وجود قيم شاذة: إذا وقعت قيم بعض المشاهدات خارج الحدين الأدنى والأعلى الشاذ ،

كانت هذه القيم شاذة ، وتظهر هذه القيم على الرسم في شكل نجوم (*) ، والشكل التالي يبين طريقة عرض القيم الشاذة الدنيا والعليا على الرسم .

شكل (5-7)

تحديد القيم الشاذة باستخدام رسمة بوكس البياني



طريقة حساب حدي القيم الشاذة

لحساب الحدين الأعلى والأدنى للقيم الشاذة، يتبع الخطوات التالية:

- حساب الانحراف الرباعي: $Q = (Q_3 - Q_1)/2$
- حساب الحد الأدنى للقيم الشاذة *(Low)*، وهو: $Low = Q_1 - 3Q$
- حساب الحد الأعلى للقيم الشاذة *(Upp)* ، وهو: $UPP = Q_3 + 3Q$

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

مثـــــال(5-4)

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

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

والمطلوب:

1- رسم شكل "بوكس" البياني

2- اكتب تحليل وصفي لهذه البيانات.

الحــــــــــل

1- رسم شكل " بوكس البياني "

- ترتيب القيم تصاعديا .

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

- تحديد أقل وأعلى إنفاق استهلاكي، وحساب الرباعيات:

$Min = 2$ $\qquad$ $Max = 18$

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 الأعلى 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 الشاذ .

الانحراف الربيعي: $Q = (10 - 5.25)/2 = 2.375$

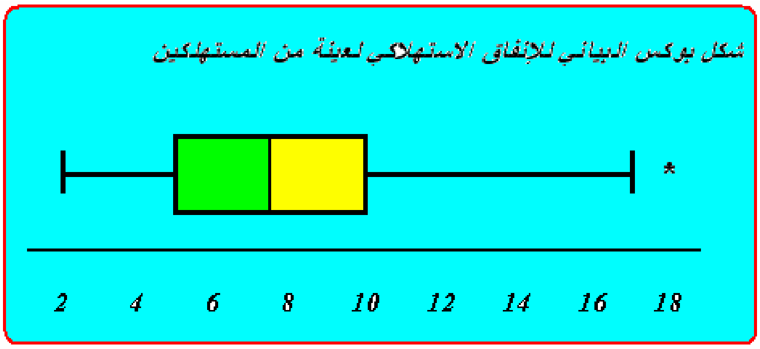
الحد الأدنى للقيم الشاذة:

$$Low = Q_1 - 3Q = 5.25 - 3(2.375) = -1.875$$

الحد الأعلى للقيم الشاذة :

$$Upp = Q_3 + 3Q = 10 + 3(2.375) = 17.125$$

- رسم شكل "بوكس"



2- تحليل وصفي من خلال الشكل أعلاه:

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

ويمكن استخدام شكل "بوكس" البياني لمقارنة مجموعتين أو أكثر .