السعدي ALSAADI

نسخة جديدة منقحة

1433/32

6

Ch. 3 - Part 2

* Quartiles.
* Five number summary.
* Skew.
* Outliers.
* Box plot.

STAT.110

جمال السعدي

رياضيات - إحصاء

CH.3 Part 2

جمال السعدي

$$\text{Percentile} = \frac{(\text{number of values below x}) + 0.5}{\text{Total number of values}} \times 100\%$$

ALSAADI

Example 1:

A teacher gives a 20 – point test to 10 students. The scores are:

18, 15, 12, 6, 8, 2, 3, 5, 20, 10

Find: the percentile rank of score of 12.

Solution

Arrange the data from low to high

$$2 , 3 , 5 , 6 , 8 , 10 , \underset{\times}{\boxed{12}} , 15 , 18 , 20$$

$$\text{Percentile} = \frac{6 + 0.5}{10} \times 100\% = 65\%$$

∴ A student whose score was 12 better than 65% of the class.

* Arrange the value corresponding to a given percentile

ترتيب القيمة المقابلة لنسبة مئوية معينة $C = \frac{n \cdot p}{100}$

Where: n is total number of values

: P is percentile.

- إذا كانت C عدد عشري نأخذ العدد الصحيح التالي له ثم توجد القيمة المناظرة للنسبة P.
- إذا كانت C عدد صحيح تأخذ القيمة المناظرة للعدد C والقيمة التالية وتوجد وسطها الحسابي فيكون هو القيمة المناظرة للنسبة P.

From example 1:

1. Find the value corresponding to percentile 25%

Arrange the data.....

2 , 3 , 5 , 6 , 8 , 10 , 12 , 15 , 18 , 20

ترتيب القيمة $C = \frac{n \times p}{100} = \frac{10 \times 25}{100} = 2.5$

- If: C is not whole number

Round it up to the next whole number

يقرب للعدد الصحيح التالي له أي إلى 3

$\therefore C = 3$ (Third)

∴ The value 5 corresponding to 25%

2. Find the value corresponding to percentile 60 %

2 , 3 , 5 , 6 , 8 , [10] , [12] , 15 , 18 , 20

ترتيب القيمة $C = \frac{n \times p}{100} = \frac{10 \times 60}{100} = 6$

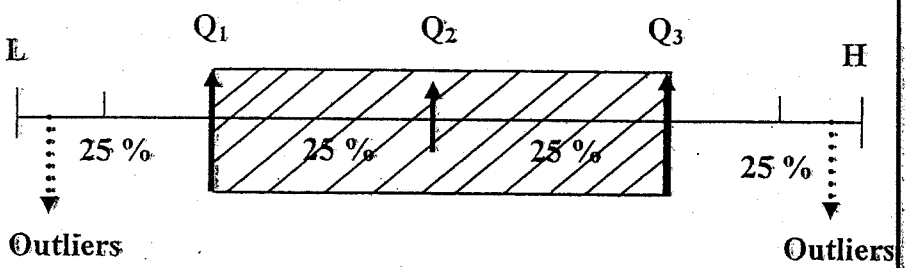
- If: C is whole number

Use the value halfway between C and C + 1

$\frac{10 + 12}{2} = 11$ ∴ corresponding to 60%

∴ Anyone scoring 11 is better than 60% of the class.

Quartiles



- Quartiles divide the distribution into four groups separated by Q_1 , Q_2, Q_3.
  - Q_2 is the median
  - Inter quartile range: IQR = $Q_3 - Q_1$

- Skew : S.K = $\dfrac{\overline{X} - D}{S}$ OR $S.K = \dfrac{3(\overline{X} - median)}{S}$

Mean, Mode, Standard Deviation

القانون الأكثر استخداماً

Where $\overline{X} = \dfrac{\sum x}{n}$ and $S = \sqrt{\dfrac{\sum x^2 - \dfrac{(\sum x)^2}{n}}{n-1}}$

- Mid Q = $\dfrac{Q_1 + Q_3}{2}$

Note

Five number summary for the data set are:

- **low value**
- **Q_1**
- **Q_2**
- **Q_3**
- **High value**

ALSAADI

Example:

For the values

33 , 27 , 51 , 30 , 31 , 23 , 38 , 40 , 42 , 43 , 29

Find: Q_1 , Q_2 , Q_3 , IQR , mid Q and Skew.

Solution

* Arrange the data from low to high: * ترتيب البيانات تصاعدياً

23 , 27 , 29 , 30 , 31 , $\boxed{33}$, 38 , 40 , 42 , 43 , 51 (Q_2)

* Q_2 (Median) = 33

* Q1 = 29 * Q_3 = 42

* IQR = $Q_3 - Q_1 = 42 - 29 = 13$

$$* \text{ Mid Q} = \frac{Q1 + Q3}{2} = \frac{29 + 42}{2} = 35.5$$

* To find skew : we find $\bar{X} = \frac{\sum x}{n} = \frac{387}{11} = 35.181$

$$\text{and } S = \sqrt{\frac{\sum x^2 - \frac{(\sum x^2)}{n}}{n-1}} = \sqrt{\frac{14307 - \frac{(387)^2}{11}}{11-1}} = 8.32$$

$$\therefore \text{ S.K} = \frac{3(\bar{x} - \text{med.})}{S} = \frac{3(35.1818 - 33)}{8.32} = 0.79 > 0$$

موجب

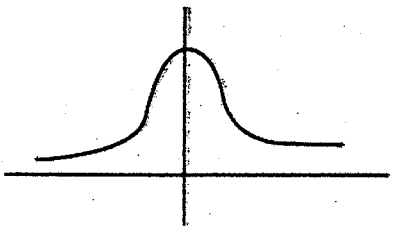
$\therefore$ Skew to the right.

Note

Skew :	$S.K = \frac{3(\bar{x} - \text{median})}{S}$

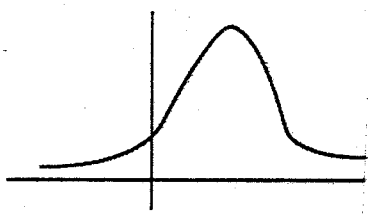
1. Symmetric : if S.K = 0

   Mean = median = mode



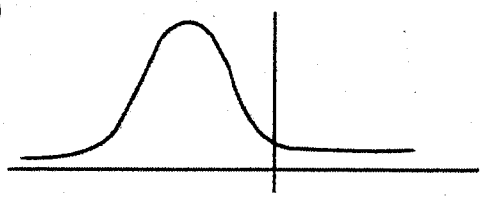
2. Skew to **the right**: if S.K > 0

   Mode < median < mean



3. Skew to **the left**: if S.K <0

   Mean < median < mode



Example:

- If the mean = 4 and mode = 2 ⇒ Mode < Mean

  then the distribution is : right skew.

- If the mode = 8 and median = 5 ⇒ Median < Mode

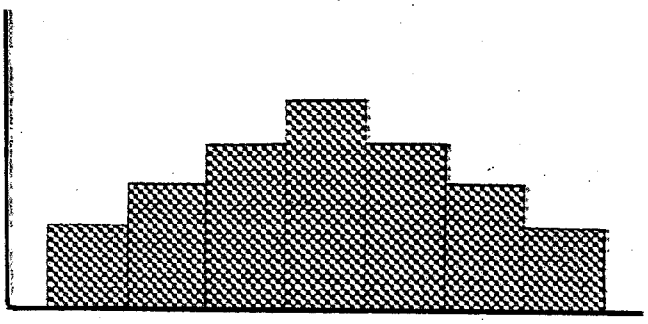
  then the distribution is : left skew.

- If the mean, median and mode are equal * الثلاثة متساويين *

  the distribution is: Symmetric.

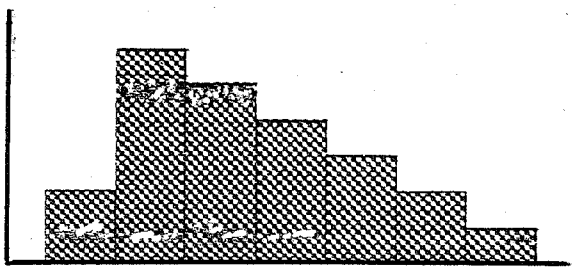
- In a *symmetrical distribution*, the data values are evenly distributed on both sides of the mean.

mean=median=mode



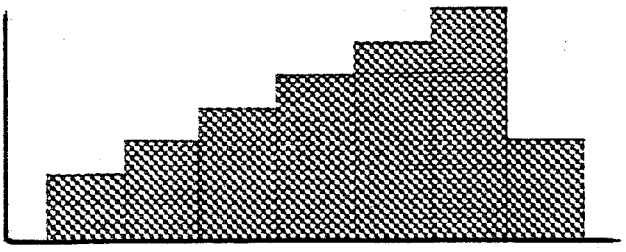
- In a *positively skewed* or *right skewed distribution*, the majority of the data values fall to the left of the mean and cluster at the lower end of the distribution.

mode<median<mean

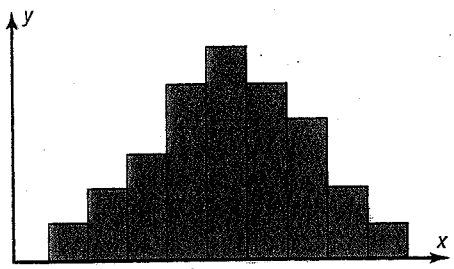


- In a *negatively skewed* or *left skewed distribution*, the majority of the data values fall to the right of the mean and cluster at the upper end of the distribution

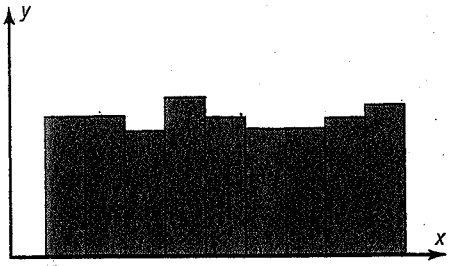
mean<median<mode



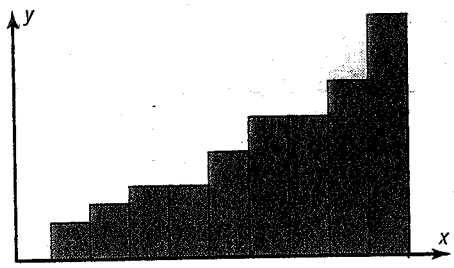
Distribution Shapes



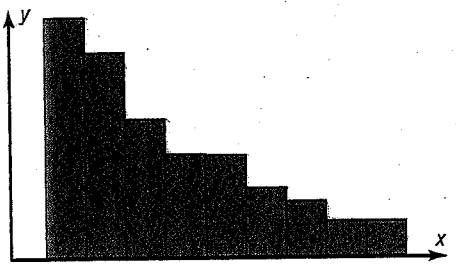
(a) Bell-shaped



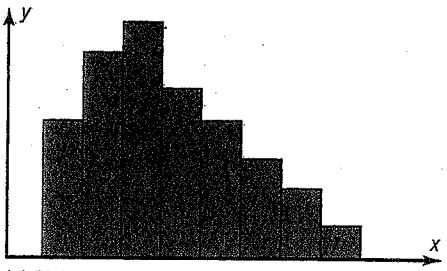
(b) Uniform



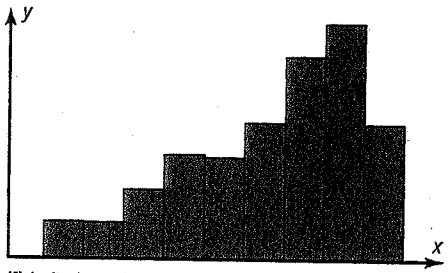
(c) J-shaped



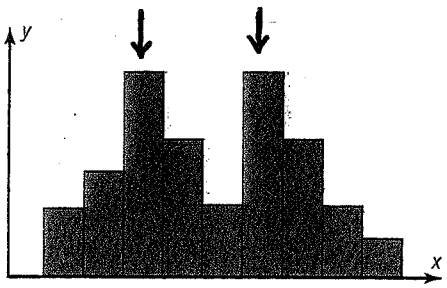
(d) Reverse J-shaped



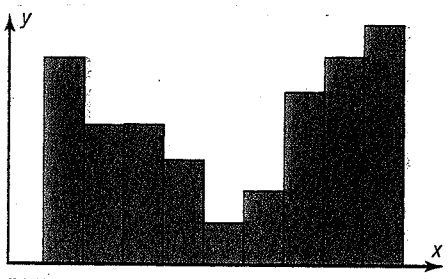
(e) Right-skewed



(f) Left-skewed



(g) Bimodal



(h) U-shaped

Exercises:

Find: Q_1 , Q_2 , Q_3 , IQR , mid Q and Skew.

For the value.

22 , 50 , 15 , 18 , 6 , 13 , 12 , 5

If: the number of data set is 6

and the mean is 8

* Find the sum of the data values.

For the following data set:

79 , 83 , 86 , 90 , 96 , 100

Find the value corresponding to the 60^{th} percentile

Find the five number summary for the following data

10 , 6 , 12 , 2 , 16 , 20 , 14

If the variance is 36.

Find the standard deviation.

Example:

Check the following data set for outliers.

5 , 6 , 12 , 13 , 15 , 18 , 22 , 50

Solution

* Q_2 (median) $= \dfrac{13 + 15}{2} = 14$

* $Q_1 = \dfrac{6 + 12}{2} = 9$

* $Q_3 = \dfrac{18 + 22}{2} = 20$

* IQR = Q3 – Q1 = 20 – 9 = 11

طريقة تحديد Outliers

١ نوجد Q_3 , Q_1 , Q_2

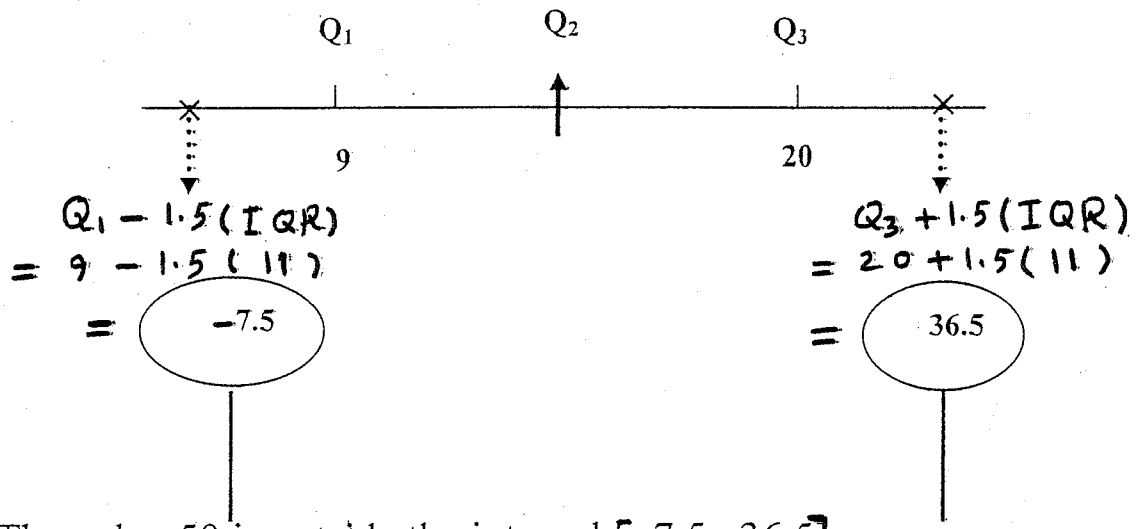
٢ نوجد IQR = $Q_3 - Q_1$

٣ نوجد $Q_1 - 1.5$ (IQR) , $Q_3 + 1.5$ (IQR)

٤ القيمة الأقل من $Q_1 - 1.5$ (IQR)

القيمة الأكبر من $Q_3 + 1.5$ (IQR)

يطلق عليها Outliers



The value 50 is outside the interval [−7.5 , 36.5]

∴ 50 it can be considered an outlier.

Pox Plot

Consists of:

- **Five – number summary of the data set:**

(1) The lowest value of the data set.

(2) The first quartile Q_1.

(3) The median Q_2.

(4) The third quartile Q_3.

(5) The highest value of the data set.

Example:

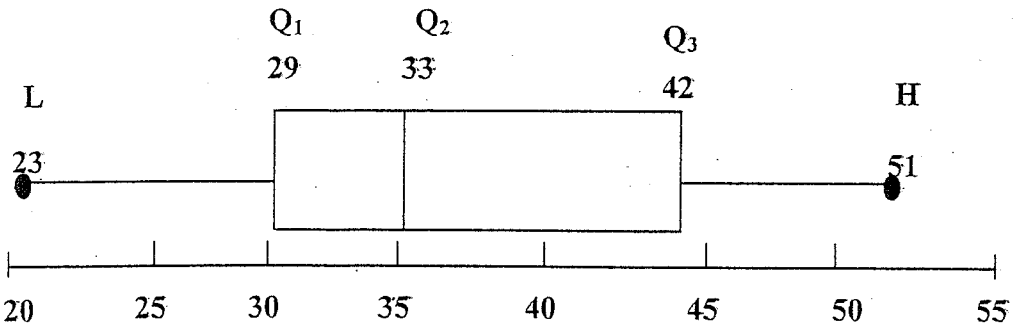
Construct a box Plot for the data.

33 , 38 , 43 , 30 , 29 , 40 , 51 , 27 , 23 , 31

Solution:

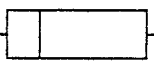
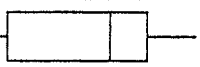
Arrange the data from low to high:

23 , 27 , (29), 30, 31 , [33], 38 , 40 , (42), 43 , 51



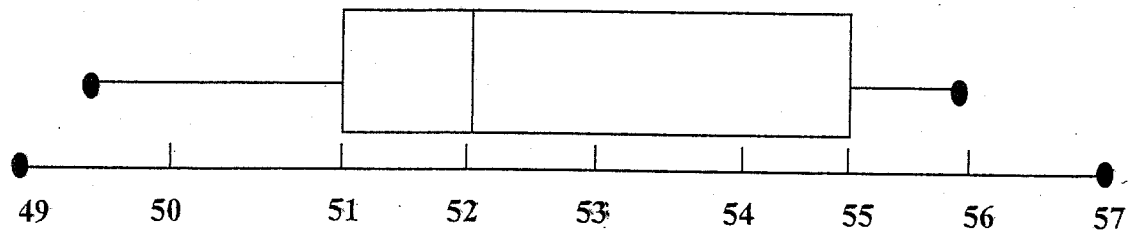
Note

- The distribution is right skew.

- The distribution is symmetric  إذا كان الوسيط Q_2 في منتصف الـ Box
- The distribution is right skew  إذا كانت المساحة الأكبر داخل الـ Box جهة اليمين
- The distribution is left skew  إذا كانت المساحة الأكبر داخل الـ Box جهة اليسار

Example:

Use the box plot to find:



1. Q_1
2. Q_3
3. MD
4. IQR
5. Min. value
6. Max .value
7. This distribution is positive or negative skew.

Solution

1. $Q_1 = 51$
2. $Q_3 = 55$
3. Median: $Q_2 = 52$
4. $IQR = Q_3 - Q_1 = 55 - 51 = 4$
5. Min. Value = 49.
6. Max. Value = 57.
7. The distribution is : right skew