Chapter 3(part 3)

RELATIONSHIPS OF THE MEAN, MEDIAN AND MODE

اوق واشمل مقاييس dispersion

***Variance and standard deviation** better S.D هو

*the standard deviation is the most used measure dispersion

*the value of the standard deviation tells how closely the values of a data set are clustered around The mean.

Variance" raw data"

Population

$$\sigma^2 = \frac{\sum x2 - \frac{(\sum x)2}{n}}{n}$$

sample

$$s^2 = \frac{\sum x2 - \frac{(\sum x)2}{n}}{n-1}$$

Standard deviation "raw data"

Population

$$\sigma = \sqrt{\frac{\sum x2 - \frac{(\sum x)2}{n}}{n}} = \sqrt{\sigma^2}$$

sample

$$s = \sqrt{\frac{\sum x2 - \frac{(\sum x)2}{n}}{n-1}} = \sqrt{s^2}$$

Example

refer to data in table of the 2002 total payroll (in millions of dollars) of five MLB teams , find the variance and standard deviation of these date

Solution

X	X^2
62	3844
93	8649
126	15876
75	5625
34	1156
$\sum x = 390$	$\sum x^2 = 35150$

$$S^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{35150 - \frac{390^2}{5}}{5-1} = \frac{35150-30420}{4} = 1182.50$$

$$S = \sqrt{1182.50} = 34.387498$$

Note two observation   نتعامل مع + بس

1) the values of the variance and the standard deviation are never negative

2) the measurement units of variance are always the square of the measurement units of the original data

Grouped Data

Variance

Population

$$\sigma^2 = \frac{\sum m^2 f - \frac{(\sum mf)^2}{n}}{n}$$

sample

$$s^2 = \frac{\sum m^2 f - \frac{(\sum mf)^2}{n}}{n-1}$$

Standard deviation

Population

$$\sigma = \sqrt{\frac{\sum m^2 f - \frac{(\sum mf)^2}{n}}{n}}$$

Sample

$$s = \sqrt{\frac{\sum m^2 f - \frac{(\sum mf)^2}{n}}{n-1}}$$

Example

The next gives the frequency distribution of the daily commuting times in minutes from home to work for all employees of a company. Calculate the variance and standard deviation.

Daily commuting time (minutes)	Number of employees
0 to less than 10	4
10 to less than 20	9
20 to less than 30	6
30 to less than 40	4
40 to less than 50	2

Solution

Daily commuting time (minutes)	F	M	Mf	M^2f
0 to less than 10	4	5	20	100
10 to less than 20	9	15	135	2025
20 to less than 30	6	25	150	3750
30 to less than 40	4	35	140	4900
40 to less than 50	2	45	90	4050
	N = 25		$\sum mf = 535$	$\sum m^2 f = 14825$

$$\sigma^2 = \frac{\sum m^2 f - \frac{(\sum mf)^2}{n}}{n} = \frac{14825 - \frac{(535)^2}{25}}{25} = \frac{3376}{25} = 135.04$$

CamScanner

Hence, the standard deviation is

$\sigma = \sqrt{\sigma^2} = \sqrt{135.04} = 11.62$ minutes

Example

The next table gives the frequency distribution of the number of order received each day during the part 50 days at the office of a mail-order company calculate the variance and standard deviation

Number of orders	F
10 – 12	4
13 – 15	12
16 – 18	20
19 – 21	14

Solution

Number of orders	F	M	Mf	M²f
10 – 12	4	11	44	484
13 – 15	12	14	168	2352
16 – 18	20	17	340	5780
19 – 21	14	20	280	5600
	N = 50		$\sum mf = 832$	$\sum m^2 f = 14216$

$$s^2 = \frac{\sum m^2 f - \frac{(\sum mf)^2}{n}}{n-1} = \frac{14.216 - \frac{(832)^2}{50}}{50-1} = 7.8520$$

Hence, the standard deviation is

$s = \sqrt{s^2} = \sqrt{7.8520} = 2.75$ orders

μ = bell-shaped = symmetrical

Empirical Rule

طريقة حساب الاحتمالات

For a bell-shaped distribution approximately

1) 68% of the observations lie within one standard deviation of the mean

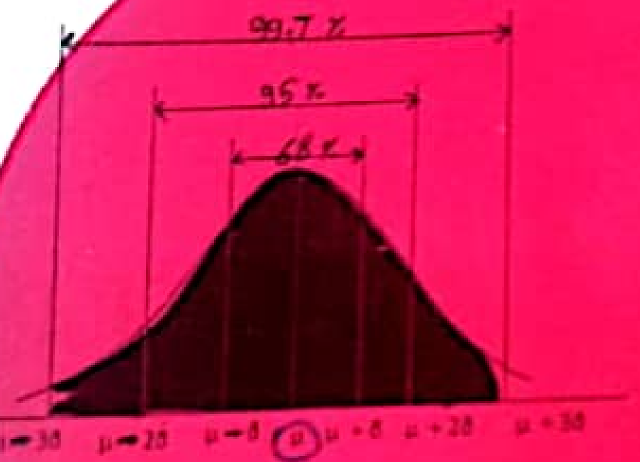
2) 95% of the observations lie within two standard deviations of the mean

3) 99.7% of the observations lie within three standard deviations of the mean

Illustration of the empirical rule

μ = Zero

$K = \frac{X - \mu}{\sigma}$



Example

The age distribution of a sample of 5000 persons is bell-shaped with a mean of 40 years and a standard deviation of 12 years determine the approximate percentage of people who are 16 to 64 years old

Solution

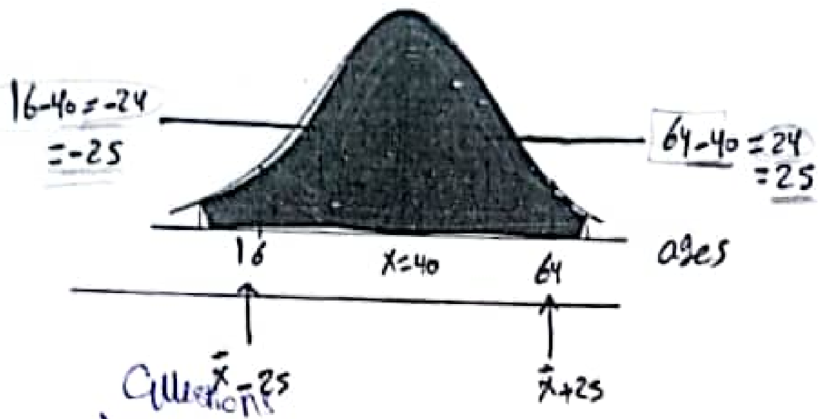
From the given information for this distribution

X = 40 and s = 12 years

k = (16 – 40) / 12=- 2

k = (64 – 40) / 12=2

each of the two points, 16 and 64, is 24 units away from the mean because the area within two standard deviation of the mean is approximately 95% for a bell-shaped curve , So approximately 95% of the people in the sample are 16 to 64 years old...



Questions
1- Find C.V
2- What is more consistency

The coefficient of variation → معامل الاختلاف

Used to compare two or more distribution

CV (coefficient variation = $(\frac{\sigma}{\mu} \times 100)$ القصار هو الافضل

Example كل ما الاختلاف بيزيد الـ error بيزيد

In a small business firm two types' are employed typist A and typist B typist A types out on an average 30 pages per days with a standard deviation of 6 typist B on an average types out 45 pages with standard deviation of 10 which typist show greater consistency in his output ? اختلاف قليل ← تناسق عالي

ثبات عالي

Solution

Coefficient of variation for A = $(\frac{\sigma}{\mu} \times 100)$

$A = \frac{6}{30} \times 100 = 20\%$ more Variation CV.A

01005681563

Coefficient of variation for B = $(\frac{\sigma}{\mu} \times 100)$

$\boxed{B = \frac{10}{45} \times 100 = 22.2\%}$ more consistency in CV.B

$CV_B > CV_A$

Typist A is greater consistency than typist B standing hire

موقف الطالب اي

* **standardized variable ,standard scores** الوضع النسبي

الوضع النسبي للباقي اي

Population $z = \frac{x-\mu}{\sigma}$ الوضع النسبي

Sample $z = \frac{x-\bar{x}}{s}$

كلما زاد الرقم كلما يعني وضع

Example

A student has scored $\boxed{68}$ marks in statistics for which the average marks were $\boxed{60}$ and the standard deviation was $\boxed{10}$ in the paper on marketing. he scored $\boxed{74}$ marks for which the average marks were 68 and the standard deviation was 15 .in which paper ,statistics or marketing was his relative standing higher ?

Solution

The Standardized variable Z measures the deviation of x from the mean x in terms of standard deviation s.

For statistics Z = (68 – 60) / 10 = 0.8 for $\boxed{\text{marketing,}}$

X = (74 – 68) / 15 = 0.4

Since the standard score is 0.8 in statistics as compared to 0.4 in marketing his relative standing was $\boxed{\text{higher in statistics.}}$

Test of skeweness

العلامة
1- right (+)
2- left (-)
3- symmetric (zero)

skeweness is Present when:

1) mean, median and mode don't equal

2) data plotted graph not normal (not symmetry)

3) sum of the positive deviation from the median not equal to the sum of the negative deviations

4) quartiles are not equidistant from the median

5) frequencies are not equal distribution at point mode

متوزعه بصورة غير متساوية

أداة القياس

Measures of skeweness :

SKP → person
skeweness

(واحدة)
1) karl person skeweness

* skeweness = mean - mode

لو لقيت ال mode هستخدم ده

* coefficient of skeweness = $\dfrac{\text{mean} - \text{mode}}{\text{standard deviation}}$ → Population only

Mode = 3 median - 2 mean

$$Skp = \frac{mean-(3median-2mean)}{\sigma}$$

open-ended مينفعش استخدمه في حالة -

$$Skp = \frac{mean-3median+2mean}{\sigma}$$

$$Skp = \frac{3mean-3median}{\sigma}$$

$$Skp = \frac{3(mean-median)}{\sigma}$$

لو لقيت ال Median هستخدم ده

Ex 1: Given the following data , calculate the karl person's coefficient of skeweness :

$\sum x = 452$ $\qquad$ $\sum x^2 = 24270$

Mode = 43.7 N = 10

Solution

$$\sigma = \sqrt{\frac{\sum x2-\left(\frac{452}{n}\right)2}{n}}$$

$$\sigma = \sqrt{\frac{24270-\left(\frac{452}{10}\right)2}{10}} = 19.59$$

** $\bar{x} = \frac{\sum x}{n} = \frac{452}{10} = 45.2$

$$Skp = \frac{mean-mode}{\sigma} = \frac{45.2-43.7}{19.59} = 0.08$$

So positive skeweness so the extent of skeweness is marginal

القيمة موجبة لذلك له قيمة تذكر

Ex 2: from the following data ,calculate the measure of skeweness using the mean ,median and standard deviation:

X	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70	70 - 80
F	18	30	40	55	38	20	16

Solution : Mean >>

	F	M
10 – 20	18	15
20 – 30	80	25
30 – 40	40	35
40 – 50	55	45
50 – 60	38	55
60 – 70	20	65
70 – 80	16	75
	N = 217	

$$Mean = \frac{\sum mf}{n} = \frac{(18\times15)+(30\times25)+(40\times35)+(55\times45)+(38\times55)+(20\times65)+(16\times75)}{217}$$

Mean = 43.71

1 median : >>

X	F	Comulative
10 – 20	18	18
20 – 30	30	48
30 – 40	40	88
40 – 50	55	143
50 – 60	38	181
60 – 70	20	201
70 – 80	16	217

$M = \frac{n+1}{2} = \frac{217+1}{2} = 109$

$\text{Median} = L_1 + \frac{L2-L1}{f}(m - c)$

$= 40 + \frac{50-40}{0.5}(109 - 88) = \boxed{43.82}$

Standard deviation

	F	M	Mf	M²f
10 – 20	18	15	270	4050
20 – 30	30	25	750	18750
30 – 40	40	35	1400	49000
40 – 50	55	45	2475	111375
50 – 60	38	55	2090	114950
60 – 70	20	65	1300	84500
70 – 80	16	75	1200	90000
			9485	472625

$$\sigma = \sqrt{\frac{\Sigma m2f - \frac{(\Sigma mf)^2}{n}}{n}}$$

$$\sigma = \sqrt{\frac{472625 - \frac{9485^2}{217}}{217}} = 16.4$$

coefficient of skeweness $= \frac{3(mean-median)}{\sigma}$

$$= \frac{3(43.71-43.82)}{16.4} = -0.02$$

Negative skeweness so the extent of skeweness is extremely negligible

القيمة سالبه لذلك له قيمة لا تذكر

SKP

+ Marginal

- negligible

2) bowley's measure

Bowley developed a measure a measure of skeweness, which is based on quartile values

Skeweness = skp $= \frac{Q3+Q1-2M}{Q3-Q1}$

1) symmetric

25% Q_1 25% Q_2 25% Q_3 25%

$Q_3 - Q_2 = Q_2 - Q_1$

skewed

$Q_3 - Q_2 \neq Q_2 - Q_1$

Skewed

$Q_3 - Q_2 \neq Q_2 - Q_1$

1) Right

Q_1 ____ Q_2 ____ Q_3

Skewed right +

Mean > median > mode

2) left

Q_1 ____ Q_2 ____ Q_3

skewed left –

mode > median > mean

Note coefficient of quartile deviation $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

Ex:

For a distribution Boweley's coefficient of skewness is = 0.56, $Q_1 = 16.4$ and median=24.2

Solution

$$Skp = \frac{Q_3 + 16.4 - (2 \times 24.2)}{Q_3 - 16.4}$$

$$-0.56 = \frac{Q_3 + 16.4 - 48.4}{Q_3 - 16.4}$$

Or $-0.56\,(Q3 - 16.4) = Q_3 - 32$

Or $-0.56 + 9.184 = Q_3 - 32$

Or $-0.56Q_3 - Q_3 = -32 - 9.184$

$-1.56Q_3 = -41.184$

$$Q_3 = \frac{-41.184}{1.56} = 26.4$$

coefficient of quartile deviation $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

$= \frac{26.4-16.4}{26.4+16.4} = \frac{10}{42.8} = 0.234$ approx

Ex:

Calculate an appropriate measure o skewness from the following data:

Values in Rs	Frequency
Less than 50	40
50 – 100	80
100 – 150	130
150 – 200	60
200 and above	30

Solution

Open ended > so we use bowely's measure

Value in Rs	Frequency	Cumulative frequency
Less than 50	40	40
50 – 100	80	120
100 – 150	130	250
150 – 200	60	310
200 and above	30	340

$Q_1 = L_1 + \frac{L_2 - L_1}{f}(m - c)$

Now $m = (\frac{n+1}{4})$ item $= \frac{341}{4} = 85.25$ which lies in 50 – 100 class

$Q_1 = 50 + \frac{100-50}{80}(85.25 - 40) = 78.25$

$M = (\frac{n+1}{4})$ item $= \frac{341}{4} = 170.25$ which lies in 100 – 150 class

$$M = 100 + \frac{150-100}{130}(170.5 - 120) = 119.4$$

$$Q_3 = L_1 + \frac{L_2 - L_1}{f}(m - c)$$

$$M = 3(341) / 4 = 255.75$$

$$Q_3 = 150 + \frac{200-150}{60}(255.75 - 250) = 154.79$$

Bowley's coefficient of skeweness is :

$$\frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1} = \frac{154.79+78.28-(2\times 119.4)}{154.79-78.28} = \frac{-5.73}{76.51} = -0.075 \text{ approx}$$

This shows that there is a negative skewness , which has a bery negligible magnitude