د. سلام حسين

مقاييس التشتت

Measures of Dispersion

These Measures give us an idea about the amount of dispersion in a set of observations. they give the answers in the same units as the units of the original observations.

The Measurses are Commonly used are:

1) The Range
2) the Quartile Deviation
3) The Mean Deviation
4) Standard Deviation and variance
5) Coefficient of variation
6) standard Error.

Examples:-

1) Calculate rang from the following

Size	60-62	63-65	66-68	69-70	72-74
Number	5	18	42	27	8

Sol L = Mid value of the highest class = 73
S = Mid value of the lowest class = 61
Rane = L - S
= 73 - 61 = 12

2) The information represent the prices for shirts in the market and the number of customers which bought these shirts.

prices	20-24	25-29	30-34	35-39	40-44
no. of customers	1	2	6	5	6

Find the standard deviation and variance.

classes	f_i	X_i	$X_i f_i$	$(X_i-\bar{X})^2$	$f_i(X_i-\bar{X})^2$
20-24	1	22	22	175.56	175.56
25-29	2	27	54	68.06	136.12
30-34	6	32	192	10.56	63.36
35-39	5	37	185	3.06	15.3
40-44	6	42	252	45.56	273.36
	$\Sigma = 20$		$\Sigma = 705$		$\Sigma = 663.7$

$\bar{X} = \frac{\Sigma f_i X_i}{\Sigma f_i} = \frac{705}{20} = 35.25$

$S = \sqrt{\frac{\Sigma f_i (X_i - \bar{X})^2}{\Sigma f_i - 1}} = \sqrt{\frac{663.7}{20-1}} = 5.91$, $S^2 = (5.91)^2 = 34.93$

Then the standard deviation is high also the variance

3) The information represent the pressure for five machines: 30, 20, 40, 60, 50, Find the Mean Deviation

Sol $\bar{X} = \frac{\sum x_i}{n} = \frac{200}{5} = 40$

x_i	$x_i - \bar{x}$	$\lvert x_i - \bar{x} \rvert$
30	-10	10
20	-20	20
40	0	0
60	20	20
50	10	10

$\sum \lvert x_i - \bar{x} \rvert = 60$

$$M.D = \frac{1}{n} \sum_{i=1}^{n} \lvert x_i - \bar{x} \rvert = \frac{1}{5}(60) = 12$$

4) The information represent the prices for shirts in the market and the number of customers which bought these shirts

Prices	20-24	25-29	30-34	35-39	40-44
no. of customers	1	2	6	5	6

Find the Mean Deviation.

classes	f_i	x_i	$x_i f_i$	$\lvert x_i - \bar{x} \rvert$	$f_i \lvert x_i - \bar{x} \rvert$
20-24	1	22	22	13.25	13.25
25-29	2	27	54	8.25	16.5
30-34	6	32	192	3.25	17.5
35-39	5	37	185	1.75	8.75
40-44	6	42	252	6.75	40.5
	$\sum = 20$		$\sum = 705$		$\sum = 96.5$

$$\bar{X} = \frac{\sum f_i x_i}{\sum f_i} = \frac{705}{20} = 35.25$$

$$M.D = \frac{\sum f_i |x_i - \bar{x}|}{\sum f_i} = \frac{96.5}{20} = 4.825$$

5) A student made a grade of 84 in mathematics and grade of 92 in physics, where the mean grade of all students in math. is 76 and standard deviation was 10, but the mean of all students and the standard deviation were 82, 16 respectively of the physics. In which subject the student have higher ability.

Sol * standard score is used for comparing two observations from two different samples in order to determine.

$$Z = \frac{X - \bar{X}}{S}$$, for math. $$Z = \frac{84-76}{10} = 0.8$$

for physics $Z = \frac{92-82}{16} = 0.5$, then we see that the student ability in math. is higher.

6) In two factories A and B located in the same industrial area, the average weekly wages (in dinar) and the standard deviation are as follows:-

Factory	Average	Standard deviation	no. of workers
A	34.5	5	476
B	28.5	4.5	524

- Which factory A or B pays out a large amount as weekly wages?
- Which factory A or B has greater variability in wages

(coefficient of variation [C.V])

Sol:- [This measure used to Compare the variability in one data set with that in another in situation in which a direct Comparison of standard deviation is not convenient and realistic].

Given $N_1 = 476, \bar{X}_1 = 34.5, S_1 = 5$

$N_2 = 524, \bar{X}_2 = 28.5, S_2 = 4.5$

(1) Total wages paid by A $= 34.5 \times 476 = 16.422$

" " " B $= 28.5 \times 524 = 14.934$

Therefore A payes out larger amount as weekly wages.

(2) C.V. of deviation of weekly wages of A & B are

$$C.V(A) = \frac{S_1}{\bar{X}_1} \times 100 = 14.49$$

$$C.V(B) = \frac{S_2}{\bar{X}_2} \times 100 = 15.79$$

B has greater variability in individual wages, since C.V. of factory B is greater than A.

Problems:-

① A study of 1000 companies gives the following information

Profit	0-10	10-20	20-30	30-40	40-50	50-60
No. of Companies	10	20	30	50	40	30

calculate the standard deviation of the profit earned.

Profit	No.(f_i)	X_i	$X_i f_i$	$X_i-\bar{X}$	$(X_i-\bar{X})^2$	$f_i(X_i-\bar{X})^2$	$f_i\lvert X_i-\bar{X}\rvert$
0-10	10	5	50	-30	900	9000	300
10-20	20	15	300	-20	400	8000	400
20-30	30	25	750	-10	100	3000	300
30-40	50	35	1750	0	0	0	0
40-50	40	45	1800	10	100	4000	400
50-60	30	55	1650	20	400	12000	600
	$\Sigma=180$	$\Sigma=$	$\Sigma=6300$			$\Sigma=36000$	$\Sigma=2000$

$$\bar{X}=\frac{\Sigma f_i X_i}{\Sigma f_i}=\frac{6300}{180}=35$$

1) $$S=\sqrt{\frac{\Sigma f_i(X-\bar{X})^2}{\Sigma f_i-1}}=\sqrt{\frac{36000}{179}}=\sqrt{201}=14.17$$

② Variance $=S^2=(14.17)^2=200.7 \Rightarrow 201$

③ Mean deviation $=M.D=\frac{\Sigma f_i\lvert X_i-\bar{X}\rvert}{\Sigma f_i}=\frac{2000}{180}=11.11$

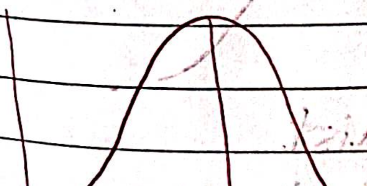
Measures of shape مقاييس الأشكال

The measures of centeral tendency and dispersion can descibe the dist. but they are not sufficient to descibe the nature of the dist., For this purpose we use other two statical measures that compare the shape to the normal curve called skewness (الالتواء) and kurtosis (التفرطح).

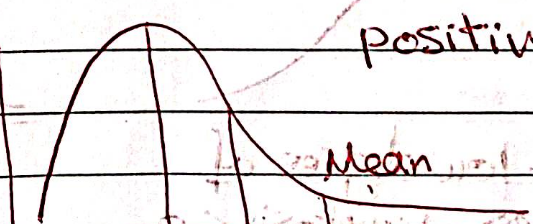
skewness : is a statistical number that tells us if a dist. is symmetric or not.
A dist. is symmetric if the right side of the dist. is similar to the left side of the dist.

1. If a dist. is symmetric, then the skewness = 0 i.e (normal dist) : median = mean = mode
2. If skewness is greater than zero, then it's called right-skewed or that the right tail is longer than the left tail.
3. If skewness is less than zero, then it's called left-skewed or that the left tail is longer than the right tail.

positive skewed



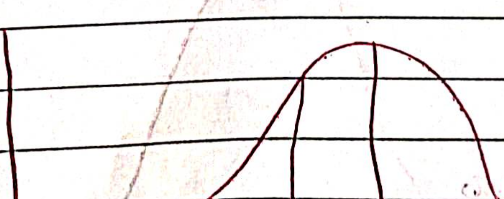
Mean



Mean

Mode Median

Normal dist.
No skew



Negative skewed

Mean Median Mode

The formula of skewness is:-

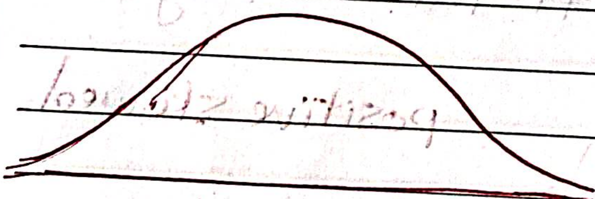
$$\text{Skewness} = \frac{\sum (X-\bar{X})^3}{(n-1) \cdot S^3}$$

Kurtosis: Kurtosis is a statistical number that tells us if a dist. is taller or shorter than a normal dist.

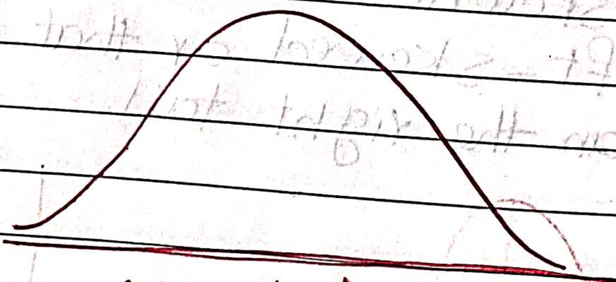
1. If a dist. is similar to the normal dist., the kurtosis = 0
2. If kurtosis is greater than (0), then it has a higher peak compared to the normal dist.
3. If kurtosis is is less than (0), then it's flatter than a normal dist.

there are three types of distribution:

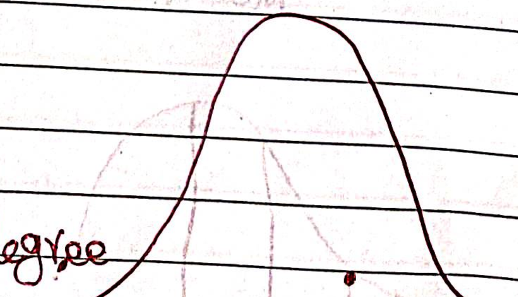
1. Leptokurtic: sharply peaked with fat tails, and less variab
2. Meso kurtic: Medium peaked.
3. Platykurtic: Flattest peak and highly dispersed.



platy, low degree of peakedness, kurtosis < 0



Normal dist
Meso, kurtosis = 0



lepto, High degree of peakedness
kurtosis > 0

The formula of kurtosis is :-

$$\text{kurtosis} = \frac{\sum (x-\bar{x})^4}{(n-1)\cdot S^4}$$

Examples:-

1) Calculate sample skewness and sample kurtosis from the following grouped data

class	frequency
2 – 4	3
4 – 6	4
6 – 8	2
8 – 10	1

Sol:

classes	Mid value (X)	f	fx	$(x-\bar{x})$	$f(x-\bar{x})^2$	$f(x-\bar{x})^3$	$f(x-\bar{x})^4$
2–4	3	3	9	2.2	14.52	−31.9	70.27
4–6	5	4	20	0.2	0.16	−0.03	0.006
6–8	7	2	14	1.8	6.48	11.66	20.98
8–10	9	1	9	3.8	14.44	54.87	208.5
		n=10	Σ=52		Σ=35.6	Σ=34.56	Σ=299.7

$$\text{Mean} = \frac{\sum f_i x_i}{\sum f_i} = \frac{52}{10} = 5.2$$

$$\text{standard deviation (S.D)} = \sqrt{\frac{\sum f_i (x_i-\bar{x})^2}{\sum f_i}} = \sqrt{\frac{35.6}{10}} = 1.88$$

$$\text{Skewness} = \frac{\sum (x-\bar{x})^3}{(n-1)\,S^3} = \frac{34.56}{9\cdot(1.88)^3} = 0.48$$

$$\text{Kurtosis} = \frac{\sum (x-\bar{x})^4}{(n-1)\,S^4} = \frac{299.79}{9\cdot(1.88)^4} = 2.12$$