Khullakitab.com (./)

(https://www.(https://www.

Grade 12 Mathematics Note

Home » (http://www.khullakitab.com/) Grade 12 » (grade-12/subjects/notes) Mathematics » (mathematics/chapters/grade-12/13/notes) Statistics

⟵ **Back to subjects (grade-12/subjects/notes)**

Chapter List

Chapters

Go Top

Previous (differential-equations/notes/mathematics/grade-12/41/notes) Next (probability/notes/mathematics/grade-12/43/notes)

Statistics

Statistics

Range

The difference between the highest and lowest value of any set of data is called range of the data.

R = L- S

Coefficient of range = $\frac{L-S}{L+S}$

For example, let us consider the following data set:

23, 65,82,59,55,25,85,56

Largest value (L) =85

Smallest value (S) =23

R = L- S85 -23 = 62

Mean

A mean is defined as the average of the numbers. The data may be individual, discrete and continuous .The method of calculating the means depend upon the nature of data. It is denoted by $\overline{x}$

For individual data:$\overline{x} = \frac{x_1+x_2+x_3+ x_4\ldots\ldots x_n}{N} = \frac{\Sigma x}{N}$

For discrete data: $\overline{x} = \frac{f_1x_1+f_2x_2+f_3x_3+ x_4\ldots\ldots f_nx_n}{f_1+ f_2+ f_3\ldots\ldots\ldots\ldots f_n} = \frac{\Sigma fx}{N}$

For continuous data: $\frac{f_1x_1+f_2x_2+f_3x_3+ x_4\ldots\ldots f_nx_n}{f_1+ f_2+ f_3\ldots\ldots\ldots\ldots f_n} = \frac{\Sigma fx}{N}$

For weighted mean = $\frac{\Sigma Wx}{W}$

Go Top

Combined mean

If x_1 and $x_2$be the given data, $\overline{x}_1$and $\overline{x}_2$are the arithmetic mean of two component series.

Combined mean($\overline{X}_{12}$)= $\frac{n_1\bar{x}_1+ n_2\bar{x}_2}{n_1+n_2}$

Partition values

Individual series

Q_1 = values of $\left(\frac{N+1}{4}\right)^{th}$ item

M_d = Q_2 =$2\left(\frac{N+1}{4}\right)^{th}$ item

Q_3= values of $3\left(\frac{N+1}{4}\right)^{th}$ item

D_3= 3^{th} values of $3\left(\frac{N+1}{10}\right)^{th}$ item

P_{20}= 20^{th} percentile of $20\left(\frac{N+1}{100}\right)^{th}$ item

Discrete series

Q_1 = values of $\left(\frac{N+1}{4}\right)^{th}$ item

M_d = Q_2 =$2\left(\frac{N+1}{4}\right)^{th}$ item

Q_3= values of $3\left(\frac{N+1}{4}\right)^{th}$ item

D_5= 5^{th} values of $5\left(\frac{N+1}{10}\right)^{th}$ item

P_{20}=20^{th} percentile of $20\left(\frac{N+1}{100}\right)^{th}$ item

Continuous series

Q_1 = L + $\frac{\frac{N}{4}-c.f}{f}$ *i

M_d = Q_2= L +$\frac{\frac{N}{2}-c.f}{f}$ *i

Q_3= L + $\frac{\frac{3N}{4}-c.f}{f}$ *i

D_8=$L + \frac{\frac{8N}{10}-c.f}{f} * i$

P_{60}=$L + \frac{\frac{60N}{100}-c.f}{f} * i$

L = lower limit

c.f = cumulative frequency

f = corresponding value of frequency

h= class size

Mode

The mode is the value of the set of data that occurs maximum number of times.

For individual and discrete series

M_0=l +$\frac{\Delta_1}{\Delta_1+\Delta_2}$*h

Δ_1=f_1- f_0

Δ_2=f_1- f_2

f_1 = maximum frequency

f_0 = frequency of preceding modal class

f_2 = frequency of following modal class

h = class interval

The measures of the data that shows the range or spread of the data from the central value is called measure of the **variability and dispersion**

Range, mean deviation, quartile deviation and standard deviation are the four measures of dispersion.

Quartile deviation

Quartile Deviation (Q_D) means the semi variation between the upper quartiles (Q_3) and lower quartiles (Q_1) in a distribution. Q_3 - Q_1 is referred as the interquartile range.

Quartile range = $\frac{Q_3 - Q_1}{2}$

Coeffof Q.D= $\frac{Q_3 - Q_1}{Q_3 + Q_1}$

Mean deviation

The mean deviation is the mean of the absolute deviations of a set of data about the mean. For a sample size N, the mean deviation is defined by

M.D = $\frac{1}{N}\sum_{i=1}^{N}|x_i - \bar{x}|$

Where $\bar{x}$ is the mean of the distribution

The mean deviation for a discrete and continuous distribution defined by

M.D = $\frac{1}{N}\sum_{i=1}^{N} f\,|x_i - \bar{x}\,|$

Coefficient of mean deviation = $\frac{M.D}{\bar{X}}$

Merits of Mean deviation

It has least sampling fluctuations as compared to Range, Percentile Range and Quartile Deviation.

This calculation has its base upon measurement than an estimate.

It is rigidly defined one of the main focus point of any measure used for statistical Analysis.

It we calculate it from median it is less affected by extreme terms.

As it is based on the deviations about an average, it gives us better measure for comparison.

Demerits of Mean deviation

Algebraic sign are ignored.

It doesnot give the satisfactory result when the deviation are taken from mode .

Standard deviation

The Standard Deviation is a measure of how spreads out numbers are. It is denoted by σ. it is the square root of the Variance.

Go Top

For individual series

σ = $\sqrt{\frac{\Sigma d^2}{N}}$ where d = $|x_i - \bar{x}|$ where I = 1…n

For discrete and continuous series

σ = $\sqrt{\frac{\Sigma fd^2}{N}}$ where d = $\left| {{{\rm{x}}_{\rm{i}}} - {\rm{x\: }}

No translation available for Unknown or missing character (0xFFFD) from Extra Math style (0x6E from "˘MT Extra˘").

Please add a translation for "char/0xFFFD" in the file LaTeX.tdl, or try using AMS-LaTeX.} \right|$where I = 1 \ldots n for discrete and d =$\left| {{\rm{m}} - {\rm{x\: }}

No translation available for Unknown or missing character (0xFFFD) from Extra Math style (0x6E from "˘MT Extra˘").

Please add a translation for "char/0xFFFD" in the file LaTeX.tdl, or try using AMS-LaTeX.} \right|$m = mid value

Coefficient of variance= $\frac{\sigma}{x}$

If assumed mean is taken

$\sigma = \sqrt{\left(\frac{\Sigma d^2}{N}\right) - \left(\frac{\sum^{d}}{N}\right)^2}$ $\sigma = \sqrt{\left(\frac{\Sigma fd^2}{N}\right) - \left(\frac{\sum^{fd}}{N}\right)^2}$

$\sigma = h * \sqrt{\left(\frac{\Sigma fd'^2}{N}\right) - \left(\frac{\sum^{fd'}}{N}\right)^2}$

d=x-a,d'= $\frac{x-a}{h}$ a= assumed meanh= common factor

Properties of standard deviation

Standard deviation is used to measure spread or dispersion around the mean of a data set.

Standard deviation is zero or negative

For the data with approximately the same mean, the greater the spread, the greater the standard deviation

If all values of a data set are the same, the standard deviation is zero.

Merits of Standard Deviation:

It is rigidly defined.

It is based on all the observations of the series and hence it is representative.

It is amenable to further algebraic treatment.

It is least affected by fluctuations of sampling.

Demerits:

It is more affected by extreme items.

It cannot be exactly calculated for a distribution with open-ended classes.

It is relatively difficult to calculate and understand.

Example 1

Compare the following figures in respect of their dispersion by quartile measurements

Go Top

Height(in cms)	156	158	161	162	164	164	165

Weight in(kgs)	51	53	56	57	50	65	70

Soln:

Let us arrange the given series in ascending order according to their height.

Height(in cms)	156	158	161	162	164	164	165
Weight in(kgs)	51	53	56	57	50	65	70

Here, n = 7.

For height, Q_1= Values of $\left(\frac{n+1}{4}\right)^{th}$ item = Value of $\left(\frac{7+1}{4}\right)^{th}$ item.

= Value of 2^{nd} item = 158.

Q_3 = Value of $3\left(\frac{n+1}{4}\right)^{th}$ item = Value of 6^{th} item = 164.

So, coefficients of Q.D. = $\frac{Q_3-Q_1}{Q_3+Q_1}$ = $\frac{164-158}{164+158}$ = $\frac{6}{332}$ = 0.019.

For weight.

Q_1 = Value of $\left(\frac{n+1}{4}\right)^{th}$ item = Value of 2^{nd} item = 53.

Q_3 = Value of 3 $\left(\frac{n+1}{4}\right)^{th}$ item = Value of 6^{th} item = 65.

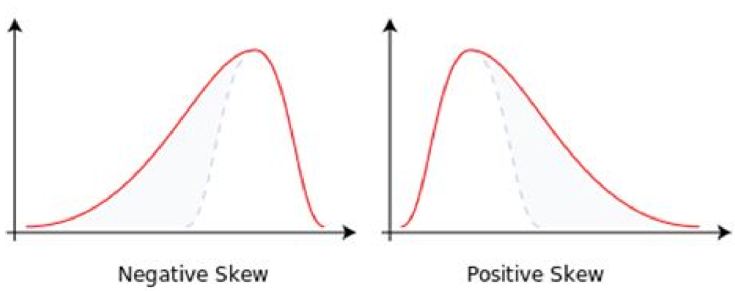
So, coefficient of Q.D. = $\frac{Q_3-Q_1}{Q_3+Q_1}$ = $\frac{65-53}{65+53}$ = $\frac{12}{118}$ = 0.0102.

Here, the coefficient of Q.D. for height < coefficient of Q.D. for weight,

So, weight is more variable than height.

Skewness

skewness is a measure of the departure of the variable about its mean. The skewness value can be positive or negative, or undefined.



Skewness is said to be positive if the curve is elongated more towards the right . For positive skewness

Mean> median > Mode

Skewness is said to be negative if the curve is elongated more towards the left . For negative skewness

Mean <median <Mode

For symmetrical curve, value of the skewness is zero

Mean= median= mode

Measure of the skewness

First measure of skewness = mean – modeor mean – median

Second measure of skewness =($Q_3 - Q_1 - 2M_d$)

Karl Pearson coefficient of Skewness S_k = $\frac{3(\text{mean} - \text{median})}{\text{Standard Deviation.}}$= $\frac{{3\left({{\rm{X\: }}

No translation available for Unknown or missing character (0xFFFD) from Extra Math style (0x6E from "˘MT Extra˘").

Please add a translation for "char/0xFFFD" in the file LaTeX.tdl, or try using AMS-LaTeX.{\rm{\: }}--{\rm{Me}}} \right)}}{{{\rm{s}}.{\rm{D}}}}{\rm{\: }} =\frac{{3\left({{\rm{X\: }}

No translation available for Unknown or missing character (0xFFFD) from Extra Math style (0x6E from "˘MT Extra˘").

Please add a translation for "char/0xFFFD" in the file LaTeX.tdl, or try using AMS-LaTeX.{\rm{\: }}--{{\rm{M}}_0}} \right)}}{{{\rm{s}}.{\rm{D}}}}$

Example 2

Coefficient of skewness based on mean , mode and the standard deviation from the following frequency distributiondata .

Wages(x)	Frequency(f)
100	2
110	6
120	10
130	8
140	4
	N = 30

Soln:

Calculation of Coefficient of skewness,

Wages(x)	Frequency(f)	fx	x^2	fx^2
100	2	200	10,000	20,000
110	6	660	12,100	72,600
120	10	1,200	14,400	144,000
130	8	1,040	16,900	135,200
140	4	560	19,600	78,400
	N = 30	$\sum^{fx}$ = 3,660		$\sum^{f} x^2$ = 450,200

Now, $\bar{x} = \frac{\sum^{fx}}{N} = \frac{3660}{30}$ = 122.

Or, $\sigma = \sqrt{\frac{\sum^{f} x^2}{N} - \left(\frac{\sum^{fx}}{N}\right)^2} = \sqrt{\frac{450200}{30} - (122)^2} = \sqrt{15006.67 - 14884} = \sqrt{122 - 67}$ = 11.07.

Since, the maximum frequency of given data is 10. So, M_o = 120.

So, $S_k(P) = \frac{\bar{x}-M_O}{\sigma} = \frac{122-120}{11.07} = 0.18$.

Correlation

Correlation is the method to find the relation of two variables in which change in valueof one variable isaccompanied by change in value of other variable .

Types of correlation

Positive and negative

Linear and non –linear

Simple, multiple and partial

Method of studying co-relation

Scatter diagram

It is the graphical representation of the data to display the relation between two variables

Karl Pearson's correlation coefficient

Pearson's correlation coefficient is the covariance of the two variables divided by the product of their standard deviations. The form of the definition involves a "product moment", that is, the mean (the first moment about the origin) of the product of the mean-adjusted random variables

$$r=\frac{\mathrm{cov}(x,y)}{\sqrt{\mathrm{var}(X)}\sqrt{\mathrm{var}\ (y)}}$$

$$\mathrm{Cov}\ (X,Y) = \frac{1}{n}\sum^{(X-\bar{X})}\left(Y-\bar{Y}\right)$$

$$r = \frac{\sum^{(X-\bar{x})}(Y-\bar{Y})}{\sqrt{\mathrm{var}(X)}\sqrt{\mathrm{var}\ (y)}}$$

or it can be written as

$$r = \frac{\sum^{XY} --n\,\bar{\bar{X}}\,\bar{Y}}{\sqrt{\sum^{x^2} -n\bar{x}^2}\sqrt{Y^2-n\bar{Y}^2}}$$

Regression

Regression analysis is a statistical process for estimating the relationships among variables. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables (or 'predictors'). Furthermore, regression analysis helps one understand how the typical value of the dependent variable (or 'criterion variable') changes when any one of the independent variables is varied, while the other independent variables are held fixed. Most commonly, regression analysis estimates the conditional expectation of the dependent variable given the independent variables – that is, the average value of the dependent variable when the independent variables are fixed. Less commonly, the focus is on a quantile, or other location parameter of the conditional distribution of the dependent variable given the independent variables. In all cases, the estimation target is a function of the independent variables called the regression function. In regression analysis, it is also of interest to characterize the variation of the dependent variable around the regression function which can be described by a probability distribution.

Lines of Regression

Linear regression is an approach for modeling the relationship between a scalar dependent variable y and one or more explanatory variables (or independent variables) denoted X. The case of one explanatory variable is called simple linear regression.

Go Top

Regression equation y on x

Y= a+ bx ………….i

$\sum y=na+b\sum x$

$\frac{\sum y}{n}$ = na+b $\frac{\sum x}{n}$

$\overline{y}$= a+ b$\overline{x}$………..ii

(y- $\overline{y}$) = b (x - $\overline{x}$) subtracting (i-ii)

Properties of regression coefficient

The correlation coefficient is the geometric mean of two regression coefficients

The value of the coefficient of correlation cannot exceed unity

The sign of both the regression coefficients will be same, i.e. they will be either positive or negative.

The coefficient of correlation will have the same sign as that of the regression coefficients.

The average value of the two regression coefficients will be greater than the value of the correlation.

The regression coefficients are independent of the change of origin, but not of the scale.

Example 3

Calculate karl's pearson's coefficient between two variables

X	6	2	10	4	8
Y	9	11	?	8	7

Soln:

Let the missed number of Y series be a then,

Or, $\overline{Y}$ = $\frac{9+11+a+8+7}{5}$

Or, 8 * 5 = 15 + a.

So, a

= 5

Calculation of correlation Co – efficient.

X	Y	x = X – $\overline{X}$	y = Y – $\overline{Y}$	x^2	y^2	xy
6	9	0	1	10	1	0
2	11	-4	3	16	9	-12
10	5	4	-3	16	9	-12
4	8	-2	0	4	0	0
8	7	2	-1	4	1	-2
				$\sum x^2$ = 40	$\sum y^2$ = 20	$\sum xy$ = -26

Now, r = $\frac{\sum xy}{\sqrt{\sum x^2}\sqrt{\sum y^2}} = -\frac{26}{\sqrt{40}\sqrt{20}} = -\frac{26}{28.28}$ = -0.92.

Previous (differential-equations/notes/mathematics/grade-12/41/notes) Next (probability/notes/mathematics/grade-12/43/notes) Go Top

Home (http://www.khullakitab.com/) About (about)

Notes (?type=notes#middle) FAQ (khullakitab-faq)

Solutions (?type=solutions#middle)

Pastpaper (?type=pastpapers#middle)

Practice Papers (?type=#middle)

BLOG

FORUM (forum)

Go Top