2 Nov., 2020

UNIT - 3
Statistical Analysis

Introduction to Descriptive Analysis

Descriptive Statistics - Used to describing data set.
Data - Raw data (as it collected)
Data Array - It is arrangement of data set either in ascending order or descending order.

After arranging data will get to know
- what is lowest value of data set
- what is highest value of data set.

Range - Highest Value - Lowest Value

Measures Of Central Tendency

Measures of central tendency are those measures which either lie at the centre or have a tendency to move towards the centre.

Median - It is a positional or locational measure.

Don't calculate median for raw data, always calculate median with arranged data.
Median is the value of the middle most item

In case of odd no. of values there will be only one median.
In case of even no. of readings there are two median, in that case median will be calculated by taking average value of both median.

Mode - Most occuring score is mode and the highest value in data case
In group data it will be data in which largest frequency is available.

Mean - Mean is very simple to be calculated and most commonly used measure of central tendency

Demerit - Mean is affected by extreme values

Concept of Population & Sample

Relevant population that we want to study in our research problem.

Sample is a miniature part of population

Sampling - It is a process of drawing a sample from population
Population is concerned with description statistics

Population Size N

Sample Size n

Population Mean μ

Sample Mean $\overline{X}$

Population Standard Deviation Σ

Sample Standard Deviation s

N	n
μ	$\overline{X}$
Σ	s

Range - It is a measure of variability

Max. value - Min Value

If range is high variability is high & risk is also high

If range is less variability is less and risk is also less.

3rd Nov, 2020

Mean Absolute Deviation - Absolute value mean ignoring negative sign.

7 Step procedure for solving statistics questions:

1. Understanding fundamental concepts
2. Select any 1 formula & memorised it.
3. Write down the formula
4. Read the question carefully & identify what information given in the question
5. Draw the table. Mention column no.
6. Pick up information from table & put it in formula
7. Calculations (Find Results)

Mean Absolute Deviation indicates the degree to which individual values in the data set are distributed around the mean.

$$\text{Mean Absolute Deviation} - \frac{\Sigma |\bar{x}i - \mu|}{N}$$

Population	Sample
N	n
μ	$\bar{x}$
σ	S

Ex - There are 10 students in a class their marks in English are given below:

Students	Marks
1	41
2	32
3	27
4	25
5	20
6	29
7	33
8	28
9	42
10	18

Calculate Mean Absolute Deviation

1 S. NO.	2 Students Code	3 Marks in English	4 $xi-\mu$	5 $\lvert xi-\mu \rvert$
1	01	41	$41-29.5=11.5$	11.5
2	02	32	$32-29.5=2.5$	2.5
3	03	27	$27-29.5=-2.5$	2.5
4	04	25	$25-29.5=-4.5$	4.5
5	05	20	$20-29.5=-9.5$	9.5
6	06	29	$29-29.5=-0.5$	0.5
7	07	33	$33-29.5=3.5$	3.5
8	08	28	$28-29.5=-1.5$	1.5
9	09	42	$42-29.5=-12.5$	12.5
10	010	18	$18-29.5=-11.5$	11.5
		295		60

$$MAD = \frac{|xi-\mu|}{N}$$

$$= \frac{60}{10} = 6$$

Variance use of square
Variance = Standard Deviation

$$V = (SD)^2$$

$$SD = \sqrt{V}$$

Standard deviation is taken up as the positive under root of the variance.

Standard Deviation - Standard deviation is the measure of dispersion of a set of data from its mean. It measures the absolute variability of a distribution; the higher the dispersion of variability, the greater is the standard deviation & greater will be the magnitude of the deviation of the value from their mean.

$$\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{N}}$$

→ Concept was introduced by Karl Pearson in 1893.

→ Important and widely used measure of dispersion

→ It is free from those defects which affected other methods & satisfies most of the properties of a good measure of dispersion.

→ Also known as root-mean square deviation as it is the square root of means of squared deviations from arithmetic mean.

Significance Of Standard Deviation

→ Used to measure risks involved in an investment instrument.

→ For investors it is a mathematical basis for decisions to be made regarding their investment in financial market.

→ Used in deals involving stocks, mutual funds, ETF's & others.

→ Standard Deviation is also known as volatility because it provides a sense of how dispersed the data in sample is from the mean.

Calculation Of Standard Deviation

* Ungrouped Data
  1. Actual Mean
  2. Assumed Mean

* Grouped Data
  1. Actual Data
  2. Assumed Data

Computation Of Standard Deviation - Ungrouped & Mean

- When deviation is taken from the actual mean, the following is applied:

$$\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{N}}$$

- When deviation is taken from the assumed mean, the following is applied.

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

$$D = X - A$$

Ex - Find the standard deviation from the weekly wages of labour working in the factory

Workers	A	B	C	D	E	F	G	H	I	J
Weekly Wages	1320	1310	1315	1322	1326	1340	1325	1321	1320	1331

Sol. : Actual Mean Method

Workers	Weekly Wages	$(x-\bar{x})$	$(x-\bar{x})^2$
A	1320	-3	9
B	1310	-13	169
C	1315	-8	64
D	1322	-1	1
E	1326	3	9
F	1340	17	289
G	1325	2	4

H	1321	-2	4
I	1320	-3	9
J	1331	8	64
N=10	$\Sigma X = 13230$	$\Sigma(x) = 0$	$\Sigma(x)^2$ 622

$$X = \frac{\Sigma x}{N} = \frac{13230}{10} = 1323$$

$$\sigma = \sqrt{\Sigma (x - x)^2}$$

$$\sigma \sqrt{\frac{622}{10}} = 7.89$$

Assumed Mean Method

Workers	Weekly Wages	(X-A) A=1310	d^2
A	1320	10	100
B	1310	0	0
C	1315	5	25
D	1322	12	144
E	1326	16	256
F	1340	30	900
G	1325	15	225
H	1321	11	121
I	1320	10	100
J	1331	21	441
N=10	$\Sigma X = 13230$	130	2312

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

$$\sigma = \sqrt{\frac{2312}{10} - \sqrt{\left(\frac{130}{10}\right)^2}}$$

$$= \sqrt{231.2 - 169}$$

$$= \sqrt{62.2} = 7.89$$

NOTE : Assumed mean should be preferred when actual mean is not a whole number, because it simplifies calculation.

Computation of Standard Deviation - Grouped Mean

→ When deviation is taken from the actual mean, the following is applied :

$$\sigma = \sqrt{\frac{\Sigma f x^2}{N} - \left(\frac{\Sigma f x}{N}\right)^2}$$

→ When deviation is taken from the assumed mean, the following is applied

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

Ex- Calculate the standard deviation from the following distribution of marks by using all the methods.

Marks	No. of Students
1-3	40
3-5	30
5-7	20
7-9	10

Solution : Actual Mean Method

Marks	f	x	fx	fx²
1-3	40	2	80	160
3-5	30	4	120	480
5-7	20	6	120	720
7-9	10	8	80	640
	100		400	2000

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

$$= \sqrt{\frac{2000}{100} - \left(\frac{400}{100}\right)^2}$$

$$= \sqrt{20 - 16}$$

$$= \sqrt{4} = 2 \text{ marks}$$

Taking assumed mean as 2

Marks	f	x	D = (x − 2)	fd	fd²
1-3	40	2	0	0	0
3-5	30	4	2	60	120
5-7	20	6	4	80	320
7-9	10	8	6	60	160
	100			200	800

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

$$\sigma = \sqrt{\frac{800}{100} - \left(\frac{200}{100}\right)^2}$$

$$\sigma = \sqrt{8 - 4}$$

$$\sigma = \sqrt{4}$$

$= 2$ Marks.

Correcting Incorrect Value of Standard Deviation

→ While calculating the Mean & standard deviation some values may enter wrong during tabulation

→ For ex - sometimes 15 may be misread as 51 or 159 may be misread as 59, & we have may have used these wrong values in the calculations.

→ Two ways to correct the error i.e.

- Either to make all calculation afresh, which is very tedious task

- Correcting the value by adjustment & readjustment of right hand side figures, is easy & less time consuming as compare to the former method.

Merits of Standard Deviation

- SD is well defined
- Its value is always definite
- It is based on all the observation of data
- It gives more accurate idea of how data is distributed.
- It is less affected by the fluctuation of sampling than most other measure of dispersion.
- The squaring of deviation makes them positive & the difficulty about algebric signs which was experienced in case of mean deviation is not found here.

Uses Of Standard Deviation

- Despite the drawbacks mentioned above the standard deviation is the best measure of dispersion & should be used wherever possible

- SD gives greater weight to extreme values

Demerits Of Standard Deviation

- Not easy to calculate

- It gives more weight to extreme items & less to those which are near the mean.

18th Nov 2020

CO-RELATION

Coefficient of Variance

Population

$$\frac{\sigma}{\mu} \qquad \frac{SD}{Mean}$$

Can also be calculated for sample

$$\frac{s}{\bar{x}}$$

Ex - An investor is considering 2 investment proposals. First to invest in mutual fund and second is to buy gold.
For the 1st average return of 100 ₹ with SD of 10
For 2nd ~~return~~ average return is ₹ 80
SD is 5
Which investment proposal should be selected & why

$$CV = \frac{\sigma}{\mu}$$

Mutual fund $= \frac{10}{100} = 0.1 \Rightarrow 10\%$

Gold $= \frac{5}{80} = 0.0625 \Rightarrow 6.25\%$

Lower C.V value will be selected for investment

Co-relation - Co-relation describes degree of relationship between two or more variables

When there are two variables refers to simple co-relation

more than two mixed / multiple

3 variables partial co-relation

Partial corelation is when, we assumed at least one of the variable to be constant & then we calculate coefficient of variable for remaining variable.

Coefficient of corelation also measures direction of correlation

If x & y are corelated & value of r is +1 then these 2 variables are perfectly positive co-related.

-1 perfectly negatively co-related

If value of $r = 0$ then these 2 variables are not related to each other

Range of coefficient of corelation from -1 to +1

Coefficient Of Co-relation Method

(rp) Karl Pearson's Method (When values are given)

Spearman Method (rs) (When ranks are given)

When we get rank data we used spearman rank corelation method

$$rp = \frac{N\Sigma xy - \Sigma x \Sigma y}{\sqrt{N\Sigma x^2 - (\Sigma x)^2}\ \sqrt{N\Sigma y^2 - (\Sigma y)^2}}$$

S.NO	ROll NO.	x	y	xy	x^2	y^2

1st Dec. 2020

Q. There are 5 students and their marks in subject A & B are given

① S.NO	② ROLL NO.	③ X	④ Y	⑤ XY	⑥ X^2	⑦ Y^2
1	A	13	7	91	169	49
2	B	15	3	45	225	9
3	C	10	2	20	100	4
4	D	12	8	96	144	64
5	E	17	4	68	289	16
		67	24	320	927	142

Sol.

$$\frac{N\Sigma XY - \Sigma X \Sigma Y}{\sqrt{N\Sigma X^2 - (\Sigma X)^2}\sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$$

$$\frac{5 \times 320 - 67 \times 24}{\sqrt{5 \times 927 - (67)^2}\sqrt{5 \times 142 - (24)^2}}$$

$$\frac{1600 - 1608}{\sqrt{19564}}$$

$$= 0.5672$$

Types Of Correlation

- Positive & Negative Correlation
- Simple, Partial & Multiple Correlation
- Linear & Non Linear Correlation

Methods Of Studying the Correlation

1. Scatter Diagram - Simplest device for studying correlation between two variables is a special type of dot chart.

2. Karl Pearson's Coefficient Of Correlation - Karl Pearson's Coefficient of correlation denoted by 'r' the coefficient of correlation 'r' measures the degree of linear relationship between two variables say x & y.

3. Spearman's Rank Coefficient Of Correlation

Karl Pearson's Coefficient Of Correlation

Ex Find the Pearson's correlation

Firm	1	2	3	4	5	6	7
Price	11	13	15	17	18	19	20
Demand	30	29	24	24	21	18	15

Sol.

1 S.NO	2 Firm	3 Price (x)	4 Demand (y)	5 xy (3×4)	6 x^2 $(3)^2$	7 y^2 $(4)^2$
1	1	11	30	330	121	900
2	2	13	29	377	169	841
3	3	15	24	360	225	576
4	4	17	24	408	289	576
5	5	18	21	378	324	441
6	6	19	18	342	361	324
7	7	20	15	300	400	225
		Σ 113	Σ 161	Σ 2495	Σ 1889	Σ 3883

$$r_p = \frac{N\Sigma XY - \Sigma X \Sigma Y}{\sqrt{(N\Sigma X^2 - (\Sigma X)^2)\,(N\Sigma Y^2 - (\Sigma y)^2)}}$$

$$= \frac{17465 - 18193}{\sqrt{(13223 - 12769)(27181 - 25921)}}$$

$$= \frac{-728}{572040}$$

$$= -0.0012$$

Q. There are 5 students their marks in subjects A & B are given r_p.

Sol.

S.NO.	ROll NO.	X	Y	xy	x^2	y^2
1	1	13	7	91	169	49
2	2	15	3	45	225	9
3	3	10	2	20	100	4
4	4	12	8	96	144	64
5	5	17	4	68	289	16
		67	24	320	927	142

$$r_p = \frac{N\Sigma XY - \Sigma X \Sigma Y}{\sqrt{(N\Sigma x^2 - (\Sigma X^2)(N\Sigma y^2 - (\Sigma y^2))}}$$

$$\cancel{r_p = \frac{17465 - 18193}{\sqrt{(13223 - 12769)(27181 - 25921)}}}$$

$$= \frac{5 \times 320 \; 67 \times 24 -}{\sqrt{(5 \times 927 - (67)^2)(5 \times 142 - (24)^2)}}$$

$$= \frac{1600 - 1608}{\sqrt{19564}}$$

$$= \frac{-8}{139.87}$$

$$= \underline{\underline{-0.057}}$$

3. Dec 2020

Difference between parametric & non-parametric

Parametric	Non-Parametric
Parametric are those test in which parametric related information are available to us	Non-Parametric are those test in which parametric related information are not available to us instead of rank, frequency etc. are given
Eg- Karl Pearson	Eg- Spearsman

Note - from parametric we can obtain non-parametric information but the reverse is not true.

Rank Correlation Coefficient (R)
Problems where ranks are not given

→ If the ranks are not given, then we need to assign ranks to the data series. The highest value in the series can be assigned rank 1. We need to follow the same scheme of ranking for the other series.

→ Then calculate the rank correlation coefficient in similar way as we do when the ranks are given.

Ex – An examination of ten applicants for a clerical post was taken by a firm. From the marks obtained by the applicants in the accountancy and statistics paper

Sol.

Applicants	Marks in Statistics	R_1	Marks in Accountancy	R_2	$(R_1-R_2)^2$ D^2
A	92	10	86	9	1
B	89	9	83	7	4
C	87	8	91	10	4
D	86	7	77	5	4
E	83	6	68	4	4
F	77	5	85	8	9
G	71	4	52	2	4
H	63	3	82	6	9
I	53	2	37	1	1
J	50	1	57	3	4

$$R = 1 - \frac{6\Sigma D^2}{N(N^2-1)}$$

$$R = 1 - \frac{6 \times 44}{990}$$

$$= 1 - 0.267$$

$$= 0.733$$

Merits Of Rank Correlation Coefficient :

→ Simpler to understand
→ Great advantage to the qualitative data.
→ Only method can be used on the given ranks & not on actual data.
→ Ascertains rough degree of correlation

Demerits
→ Cannot be used in correlation of grouped frequency distribution.
→ Methods cannot be applied where N is exceeding 30 unless we are given the ranks not the actual value of the variables.

4th Dec 2020

REGRESSION

It was given by Francis Galton in 1877

Types Of Regression

1. Simple & Multiple Regression
2. Linear & Curvi Linear Regression
3. Partial & Total Regression
4. Stepwise Regression

Regression Lines
- It is line for best fit
- useful for forecasting

Difference between Correlation & Regression

Basis	Correlation	Regression
Purpose	To predict which are the most dependable forecasts	To predict or estimate the unknown variable with the help of known variable
Scope	Limited Application	Wider applications
Usage	To represent linear relationship between two variables	To fit a best line & estimate one variable on the basis of another variable

Nature of variables	Both the variables are mutually dependent	One variable is dependent & other variable is independent
Responding Nature	Independent of the change of origin or change of scale.	Independent of the change of origin but dependent on the change of scale.
Measures	Measures the degree to which two variables move together	Describes the fundamental level the nature of any linear relationship between two variables.

* We calculate value of a & b by the least squares method.

Involves following steps :-

1. Draw a regression line based on past data.
2. Plot the data points in two dimensions
3. Measure the perpendicular distance of these (data) point from the regression line.
4. Find the square of these distance (data points)
5. That line is called the line of best fit for which these squares are least.

Methods of least square involves two normal equations (Y on X)

1. $\Sigma Y = n \times a + b\Sigma X$
2. $\Sigma XY = a\Sigma X + b\Sigma X^2$

$$b = \frac{\Sigma XY - n\bar{X}\bar{Y}}{\Sigma X^2 - n\bar{X}^2}$$

$$a = \bar{Y} - b\bar{X}$$

Q.

Truck (X)	Age X	Rep. Exp (Y)	XY	X^2
101	5	7	35	25
102	3	7	21	9
103	3	6	18	9
104	1	4	4	1

$\Sigma X = 12$ $\Sigma Y = 24$ $\Sigma XY = 78$ $\Sigma X^2 = 44$

$$\bar{X} = \frac{\Sigma X}{n} = \frac{12}{4} = 3$$

$$\bar{Y} = \frac{\Sigma X}{n} = \frac{24}{4} = 6$$

$$b = \frac{\Sigma XY - n\bar{X}\bar{Y}}{\Sigma X^2 - n\bar{X}^2}$$

$$= \frac{78 - 4(3)(6)}{44 - (4)(3)^2}$$

$$\frac{78-72}{44-36}$$

$$= \frac{6}{8}$$

$= .75 \rightarrow$ slope of line

$a = \bar{Y} - b\bar{X}$

$a = 6 - (.75) \times (3)$

$a = 6 - 2.25$

$a = 3.75 \rightarrow$ the Y intercept

~~$= 1 - \frac{666}{336}$~~

~~$= 0.982$~~

$= a + bX$

$3.75 + 0.75$

$= 3.75 + 0.75(4)$

$= 3.75 + 3$

$= 6.76$

Mathemetically speaking regression X on Y is also possible

X on Y

$X = a + bY$

Two Normal Equation would be

1. $\Sigma X = na + b\Sigma Y$
2. $\Sigma XY = a\Sigma Y + b\Sigma Y^2$

5th Dec. 2020

Time Series Analysis

Only difference between regression & time series

In time series X variable is time itself we analise the effect of time on dependent variable

Similarities - Two variables are there dependence & independence relationship both are forecasting tool.

Components Of Time Series
Basically there are 4 components

1. Secular / Longterm Trend
2. Seasonal variations - Variations within the year
3. Cyclical variations - Business Cycle ↑ Peak ↓ Troughs curve forms
4. Irregular variations - Residual

Models -
Additive Model - Analysis based on (TSCI)
Multiplicative Model - (T × C × S × I)
Combination Model - Multiplication & Addition (TC + SI)

Steps in Analysing Time Series

There are 4 steps

1. If seasonal component is present in time series then first of all we analyse it deseasonal the data within the year.

2. Now, secular trend will be analysed in which trend line will be developed which will be used for forecasting.

3. Once the trend line is obtained then we will analyse the cyclical component.

4. After analysing the components whatever variation is remaining is irregular component which cannot be mithametically analysed.

Analysis Of Trend
Trend lines will be analyse by using method of least squares.
To avoid lengthy calculations we will be using time coding method.
There are two variations
1. Odd No. of time ~~series~~ periods
2. Even no. of time periods.

Q For time series following data are given

S.No.	x	y	deviations $2(X-\bar{X})$	xy	x^2
1	2002	80	$-3.5 \times 2 = -7$	-560	49
2	2003	90	$-2.5 \times 2 = -5$	-450	25
3	2004	92	$-1.5 \times 2 = -3$	-276	9
4	2005	83	$-0.5 \times 2 = -1$	-83	1
5	2006	94	$0.5 \times 2 = 1$	94	1
6	2007	99	$1.5 \times 2 = 3$	297	9
7	2008	92	$2.5 \times 2 = 5$	460	25
8	2009	95	$3.5 \times 2 = 7$	665	49
		725	$\Sigma x = 0$	147	168

- When there are two even number of time periods then mean of middlemost 2 values will be the actual mean.

$$\bar{X} = \frac{\Sigma X}{N} = \frac{2005 + 2006}{2} = 2005.5$$

⇒ Steps for even number of years question:

1) When even number of time periods are there we calculate mean.
2) Mean will be given by mean of two middlemost values.
3) Calculation deviation of each time period from this mean
4) $x = 2 \times (X - \bar{X})$

$$a = \bar{Y} = \frac{\Sigma Y}{N} = \frac{725}{8} = 30.625$$

$$b = \frac{\Sigma xy}{\Sigma x^2} = \frac{147}{168} = 0.875$$

Trend line equation →

$$Y = a + bx$$
$$Y = 30.625 + 0.875x$$

when $x = 2012$ then $y = ?$

$$\Rightarrow x = 2(X - \bar{X})$$
$$x = 2(2012 - 2005.5)$$
$$x = 2(6.5) = 13$$

$$\Rightarrow y = a + bx$$
$$y = 30.625 + 0.875(13)$$
$$y = 102$$

10/12/2020

INDEX NUMBERS

An index number is a statistical value that measures the change in a variable with respect too time.

1. Price index - compares price levels from one time period to another.
2. Quantity index - how much quantity changes once a period of time.

- Bowley's Index - It is a given by calculating arithmetic mean of Laspeyer's index & Paasche Index

$$= \frac{\text{Laspeyres Index} + \text{Paasche Index}}{2}$$

Q. There are 4 commodities A, B, C & D. Their prices & quantities in the year 2015 & current year 2019 are given. Calculate laspeyres index & paasche index.

Commodity	Base Year 2015		Current Year 2019	
	Po	Qo (million to)	Pn	Qn
A	2	8	4	6
B	5	10	6	5
C	4	14	5	10
D	2	19	2	13

$$\text{Laspeyres Index} = \frac{\Sigma P_n q_0}{\Sigma P_0 q_0} \times 100$$

$$\text{Paasche Index} = \frac{\Sigma P_n q_n}{\Sigma P_0 q_n} \times 100$$

Commodity	Po	Qo	Pn	Qn	PoQo	PnQo	PnQn	PoQn
A	2	8	4	6	16	32	24	12
B	5	10	6	5	50	60	30	25
C	4	14	5	10	56	70	50	40
D	2	19	2	13	38	38	26	26
					160	200	130	103

$$\text{Laspeyres index} = \frac{200}{160} \times 100 = 125\%.$$

∴ Price has increased by 25%.

$$\text{Paasche index} = \frac{130}{103} \times 100 = 126.21\%.$$

∴ Price has increased by 26.21%.

$$\text{Bowley's Index} = \frac{125 + 126.21}{2} = 125.6$$

$$\text{Fisher's index} = \sqrt{125 \times 126.21}$$

$$= 125.60$$

15th Dec 2020

UNIT-IV
PROBABILITY

Terminology

1. Trial: Tossing a coin, single toss of a coin.
2. Experiment: Series of a trials
3. Event :- Outcome of a trial

Mutually Exclusive Events: are those events when one of the event occurs, it includes the possibility of occurrence of another one event.

Non-mutually exclusive events - Events when overlap from mutually exclusive events

Collectively Exhaustive Events - Events whose sum total of problem value is equal to one

sum of prob. of event $= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 1$

Complementary Events -

prob(A)

prob(not A) = 1 - prob(A)

4. Sample Space - Set of all possible outcome resulting from series of trial or experiment

Fundamental Concepts :-

Two fundamental concepts of prob. are :

→ The value of prob. of any event lies b/w 0 to 1

$$0 \leq P(\text{event}) \leq 1$$

(Impossible) (Certain)

Range of prob. value = 0 to 1

1.→ Value of prob. of an events is equal to 0, then event will never occur.

→ If prob. value is 1, the event will always experiment to occur.

2. The sum of prob. values for all the possible outcome of an activity must be equal to one.

Approaches of Probability

- Classical / Priority
- Relative frequency of occurrence
- Subjective

16th Dec 2020

1. Probability Addition Rule (either / or)

Two conditions are there under this rule

(i) Events are mutually exclusive
(Tail come, head will not come vice - versa)

(ii) Events are not mutually exclusive

⇒ probability of either A or B given by
prob. of A or B = P(A) + P(B)

Ex - In a single roll of dice, what is the prob. that an even no. will come out.

$= \frac{3}{6} = \frac{1}{2}$ $\left[\frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} \right]$

→ N.M.E where something overlap
How + will changes with it

pack of 52 playing cards prob. of ace of diamond $= \frac{1}{52}$

prob. of ace $\frac{4}{52} = \frac{1}{13}$

prob. of diamond card $= \frac{13}{52}$ $\frac{1}{4}$

- Two cards are drawn of either ace or king

$= \frac{4}{52} + \frac{4}{52} = \frac{8}{52}$ $\frac{2}{13}$

→ Addition Rule (when events are not mutually exclusive)

Prob. of either A or B happening

$$P(A) \text{ or } P(B) = P(A) + P(B) - P(A \,\&\, B)$$

Ex- From a well shuffuled pack of 52 cards, prob of card drawn either an Ace or a diamond

(overlap 2 diamond)

$$\frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52} \quad \frac{4}{13}$$

18th Dec 2020

→ Statistically independents events : are those events when the occurrence of one event does not affect the prob of other event.

→ Statistically dependent events : are those events when the occurrence of one event affects the prob. of occurrence of the event.

Ex. 52 cards
1 card is excluded
king been drawn $= \frac{4}{51}$
Ace been drawn $= \frac{3}{51}$
get affected

NOTE : Without replacement will give statistical dependence.
With replacement will give statistically independence.

Multiplication and Joint Probability Rule

(i) When the events are SIE
(ii) When the events are SDE.

Ex (i) Tossed a coin H is obtained first
H_1 & T_2 T is obtained second

Find out prob. of H in first draw & tail in second draw

$P(H_1 \& T_2) = P(H_1) \times P(T_2)$

$\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

H is obtained in third

$P(H_1 \& T_2 \& H_1) = P(H_1) \times P(T_2) \times (PH_1)$

$= \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$

1) Conditional Probability

Both are diff

$P\left(\frac{B}{A}\right)$ · Prob of B given A has already occured

$P\left(\frac{A}{B}\right)$ = Prob of A given B has already occured

Q. Tossed a coin H is obtained in first draw

A. What is the prob of T in second draw given that head has already tossed in first draw

$P\left(\frac{T_2}{H_1}\right) = P(T_2)$

When the two events are statistically independent then the prob of second event is not at all affected by the prob of first event

$P\left(\frac{B}{A}\right) = P(B)$

$P\left(\frac{A}{B}\right) = P(A)$

$\boxed{P\ \frac{B}{A} \neq P\left(\frac{A}{B}\right)}$

NOTE :· When 2 event are SI, then conditional prob is meaningless & it is equal to the prob. of 1 single which is on the left side or numerator place

- Probability of single event is called marginal probability

Ex - A bag contains 10 balls, 3 of them are black & 7 are red (Blindfolded)

→ What is prob of black ball is drawn
$= \frac{3}{10}$

→ 1 ball is drawn, what is prob. of red ball is drawn $= \frac{7}{10}$

- What is the prob. either A or B drawn
P(A) or P(B) $= \frac{3}{10} + \frac{7}{10} = \frac{10}{10} = 1$

- What is the prob. of red ball in second draw & black ball in first draw
$= \frac{3}{10} \times \frac{7}{10} = \frac{21}{100} = 0.21$

$P(A \& B) = P(B \& A)$ when statistically independence

- What is the prob of red balls in 2nd draw given that black ball is drawn in the 1st draw if red not replaced back $= P\left(\frac{R_2}{B_1}\right) = \frac{7}{9}$

$= \frac{6}{}$

19th Dec 2020

· Conditional prob. under SD = $P\left(\frac{B}{A}\right) = \frac{P(B \cap A)}{P(A)}$

$$P\left(\frac{A}{B}\right) = \frac{P(A \cap B)}{P(B)}$$

What is the prob of black ball being drawn in 2nd draw, given that red ball in 1st draw & it was not replaced back

prob. $\left(\frac{B_2}{R_1}\right) = \frac{3}{9}$

If black ball in place of red $= \frac{2}{9}$

• What is the prob. of red balls in the 2nd draw given that red ball was drawn in 1st draw

$$P\left(\frac{R_2}{R_1}\right) = \frac{6}{9}$$

$$P\left(\frac{R_2}{B_1}\right) = \frac{7}{9}$$

Joint prob. under SD

Marginal prob. under SD is given by the sum total of all the joint prob. in which that event appears.

Ex => 3 events A, B, C

Joint prob. AB, BC, AC

$A = AB + AC$

$B = \cancel{BA}AB + BC$

$C = BC + AC$

Marginal prob. of A = Joint prob. of AB + JP of AC

Ex - A Box contains 10 balls which have foll. distribution on the basis of colour & pattern

(a) 3 are coloured & lettered
b) 1 is coloured & numbered
c) 2 are white & lettered
d) 4 are white & numbered

$C = 4$, $W = 6$, $L = 5$, $N = 5$

1 What is the prob. of coloured ball being drawn
$= \frac{4}{10}$

2. White $= \frac{6}{10}$

3. Lettered $= \frac{5}{10}$

4. Numbered $= \frac{5}{10}$

5. Coloured & Lettered ball $= \frac{3}{10}$

6. Coloured & Numbered $= \frac{1}{10}$

7. White & lettered $= \frac{2}{10}$

8. White & numbered $= \frac{4}{10}$

· What is the prob. of a coloured ball given that it is numbered.

$$P\frac{(C)}{N} = \frac{P(C \cap N)}{N} = \frac{1}{10}$$

· What is the prob. of white given that it is lettered $= P\frac{(W)}{L} = P(W \cap L) = \frac{2}{10}$

$$\frac{P(L)}{N} = \frac{P(L \cap N)}{N} = P\left(\frac{L}{N}\right) = P(L)$$

$$= \frac{5}{10} = \frac{1}{2}$$

19th Dec 2020, 2nd Class

Probability Distribution

It is the distribution of the prob. values

Types Of Prob Distribution
1. Probability distribution of discrete variable
2. Probability distribution of continuous variable

All countables are discrete variables (broken in between) it cannot be in in decimals only integer value

All measurables are continuous varaables (high, length, size)

- PD of discrete variables

1. Binomial Distribution
2. Poisson Distribution

- PD of continuous variables
1. Normal Distribution
2. Exponential Distribution

1. Binomial Distribution -
In binomial distribution, there are only 2 possible outcomes for each trial.

→ No of trials 'n' is finite
→ Mutually exclusive outcomes

Success = P(S) p constant
Failure = P(F) = (1 − p) = Q.

prob of x successes in n trial

$$P(x) = \frac{n!}{(n-x)!\,x!} \times \left(p^{x} q^{n-x}\right)$$

$4! = 4 \times 3 \times 2 \times 1 = 24$

$a = 1$

Ex - A coin is tossed 10 times What is the prob of (H) 6 times out of 10 times

$$6 = \frac{10!}{(10-6)!\,6!} \times \left(0.5 \times 0.5^{10-6}\right)$$

$$\frac{10!}{4!\,6!} \times 0.5^{6} \times 0.5^{4}$$

$$= \frac{10 \times 9 \times 8 \times 7 \times 6!}{4 \times 3 \times 2 \times 1 \times 6!} = 210$$

$$= 210 \times 0.5^{6} \times 0.5^{4}$$

$$= 0.204$$

Two parametrics of binomial distribution is

1. n (no of trials)
2. p (prob of success)

Q' = $n=10$ $p=0.5$
$x=4$ $q=0.5$
What is the prob of 4 tails in 10 toss of the coin

x can never be greater than n

Q. A co. manufacture's radios It takes sample of 5 units past records indicates that the quality of defective pieces is 0.15 What is prob. of finding 3 out of 5 radios set defectively

$N=5$, $x=3$, $p=0.15$ $q=0.85$

$$3 = \frac{5!}{(5-3)2!} \times \left(0.15^3 \times 0.25^{5-3}\right)$$

$$3 = \frac{5 \times 4 \times 3 \times 2 \times 1}{2 \times 1 \times 3 \times 2 \times 1}$$

$$3 = \frac{20}{2}$$

$$= 10 \times 0.15^3 \times 0.85^2$$
$$= 0.2438$$

Q. $$4 = \frac{10!}{6!\,4!} \times \left(0.5^4 \times 0.5^6\right)$$

$$4 = \frac{10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{6!\,4 \times 3 \times 2 \times 1}$$
$$= 210 \times 0.5^4 \times 0.5^6 = 0.204$$

21st Dec 2020

Poisson Distribution.

→ It is the probability distribution of rare events
→ There is only one parameter in poisson probability of distribution
i.e.
λ (greek letter) average occurrence of rare events

probability of x where x is (no. of road accidents on the highway per month

$$P(x) = \frac{\lambda^x \cdot e^{-\lambda}}{!x}$$

$$e = 2.718$$

Q. The past data indicates that on an average in a month 5 accidents take place on a highway. Find the probability of a accidents. Find prob. of 1, 2, 3 accide[nts]

1. prob of 0 accidents $= \frac{(5)^0 \times 2.718^{-5}}{0!} = 0.006739$

2. prob of 1 accidents $\frac{(5)^1 \times 2.718^{-5}}{1!}$

$\frac{5 \times 0.00673}{1} = 0.03$

3 prob of 2 accidents $= \frac{(5)^2 \times 2.718^{-5}}{2!}$

$$\frac{25 \times 0.06737}{2 \times 1} = 0.08$$

4. prob of 3 accidents $= \frac{5^3 \cdot e^{-5}}{3!}$

$$\frac{125 \cdot 0.006735}{3 \times 2 \times 1}$$

$$= 0.14$$

→ 3 or less

$= p0 + p1 + p2 + p3$

→ Less than 3

$= p0 + p1 + p2$

→ More than 3 (Not possible but)

$= 1 - x($ p3 or less$)$

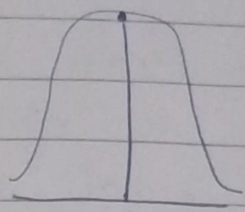
prob of either 1 or 2 accidents happening

$= (p1) + (p2)$

4 Jan. 2021

Normal Distribution

This is the distribution of continuous prob. distribution variable.



→ It is a symmetrical curve

→ There are 2 parameters of Normal Distribution : Mean & σ

→ Area of the curve is directly propotional to probability value.

→ The whole area has probability value of 1

→ Both the tails never touch x-axis

→ It is uni model

→ Mean, Median & mode are at the centre of the circle.

Standard Normal Distribution

$\mu \pm 1\gamma = 68\%$

$\mu \pm 2\gamma = 95.5\%$

$\mu \pm 3\gamma = 99.73\%$

Larger the area greater the probability

6th Jan 2021

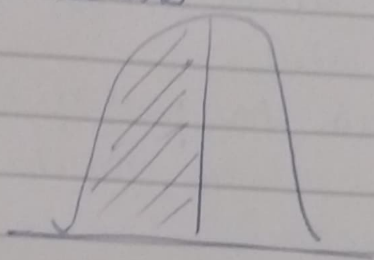
Standard Score

$$z = \frac{|x-\mu|}{\sigma}$$

z = Standard score

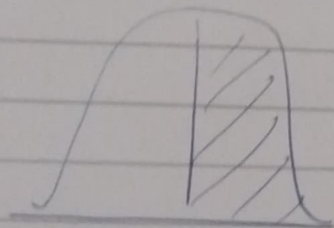
Q. Mean life of a tyre is 100 hrs with σ of 10 hrs. A tyre is randomly selected what is the probability that its age is between :-

1) less than 100 hrs



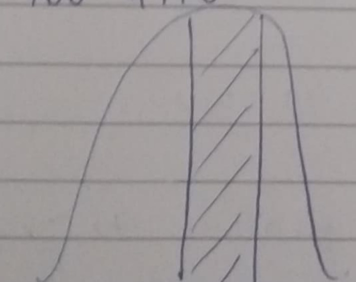
$\mu = 100$
$\sigma = 10$
50%

2. More than 100 hrs



50%

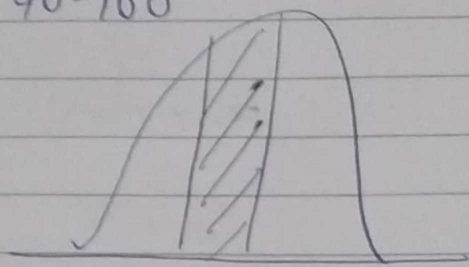
3. between 100 & 110



x value = 110

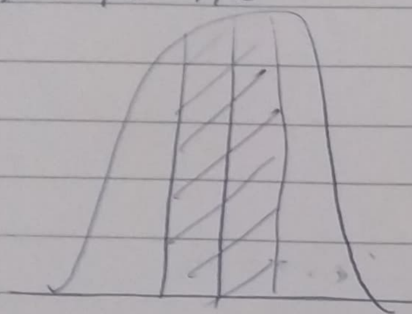
$$\frac{110-100}{10} = 1$$

4 between 90-100



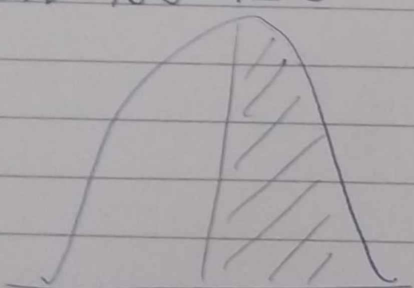
0·34

5. between 90-110



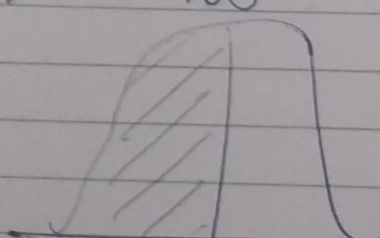
68%.

6. between 100-120

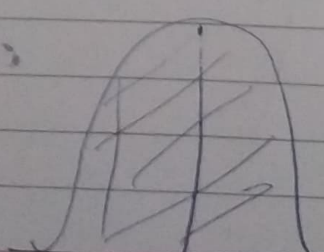


0·4775

7. between 80-100

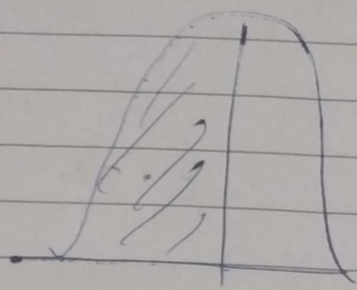


8. 80-120 ∴



95·5%

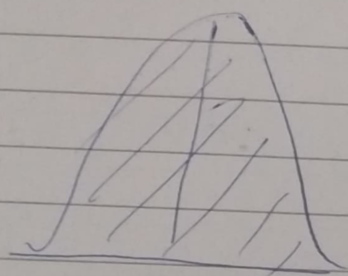
9. between 70-100



$x = 130$

$z = 3$

10. between 70-130



0.9973

11. between 100-115

$z = 1.5$

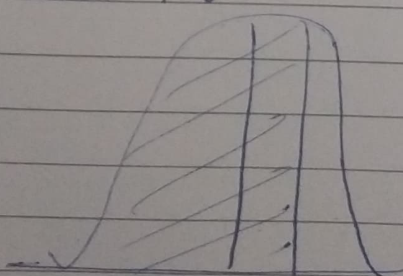
0.4332

12. between 100-117

$z = 1.7$

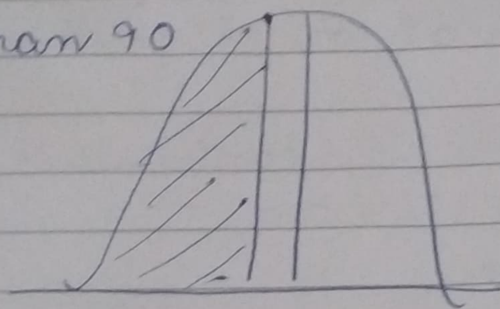
1.75 0.4599

13. less than 110



0.8423

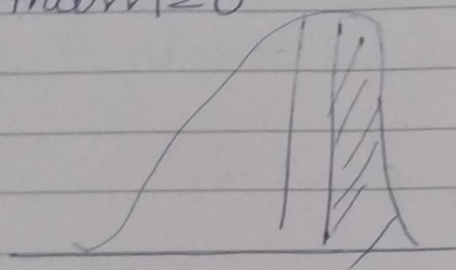
14. less than 90



$x = 90$

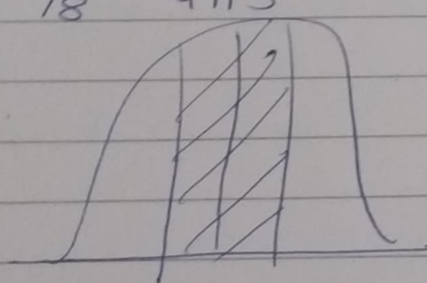
0.16

15. More than 120



$x = 120$

16. between 78 & 115



$x = 115$

Steps in applying normal distribution

1. Ensure that variable is continuous
2. Find out parameters μ or σ
3. Find out value of x
4. For each question draw a separate curve
5. Calculate Z value
6. Draw Normal Distribution Table No. 6 given on pg 691 Table No. 7 pg 692

NOTE: Probability of point is 0

th Jan. 2021

Estimation

(Central Unit Theorem)

Sampling Distribution

If from a population a no. of samples are drawn & their sample means are calculated.

Grand Mean - It is the mean of all sample means is called grand mean

$$\overline{\overline{x}} = \frac{\overline{x}_1 + \overline{x}_2 + \overline{x}_3 + \overline{x}_4 + \overline{x}_5}{5}$$

1. Acc. to CUT grand mean is equal to the population mean.

$$(\text{grand mean})\ \overline{\overline{x}} = \mu\ (\text{population mean})$$

2. Standard Error - It is the standard deviation of sample means in a sampling distribution.

$$SE = \sigma_{\overline{x}}$$

$$\sigma_{\overline{x}} = \sigma$$

It is the good approximation of the population standard deviation.

$$(SE)\quad \sigma_{\overline{x}} = \frac{s\ (\text{sample SD})}{\sqrt{n}}$$

Sta Estimation There are two types of estimation.

1. Point Estimation
2. Interval Estimation

Point Estimation - In point estimation we don't know population mean. Therefore we used hypothesised population mean.

Interval Estimation - In this first we have to estimate confidence limits. They are also known as confidence intervals.

$$z = \frac{x - \mu}{\sigma}$$

$$z\sigma = x - \mu$$

$$x = \mu \pm z \sigma \bar{x}$$

$$\boxed{\bar{x} = \bar{\bar{x}} \pm z \sigma \bar{x}}$$

$$\bar{x} = \bar{\bar{x}} \pm z \sigma \frac{s}{\sqrt{n}}$$

8th Jan 2021

Q. A quality control department of a co. selects some parts for quality inspector past records indicates that strength of their part is normally distributed with SD of 200 kgs. A random sample of 64 parts was selected & mean came out to be 6200 kgs. Find the 95% confidence limits & also the confidence level.
Draw the diagram to acceptance region and rejection region?

Sol: $\sigma = 200kg$ $(\bar{x}) = 6200kg$ SS $\overset{(n)}{=}$ 64 parts

95% confidence level $Z = 1.96\%$
99% confidence level $Z = 2.58\%$
90% confidence level $Z = 1.23\%$
98% confidence level $Z = 2.33\%$

Upper Control Limit (UCL) $= \bar{\bar{x}} \pm Z\sigma_{\bar{x}}$

$$UCL = 6200 + 1.96\,\sigma_{\bar{x}}$$

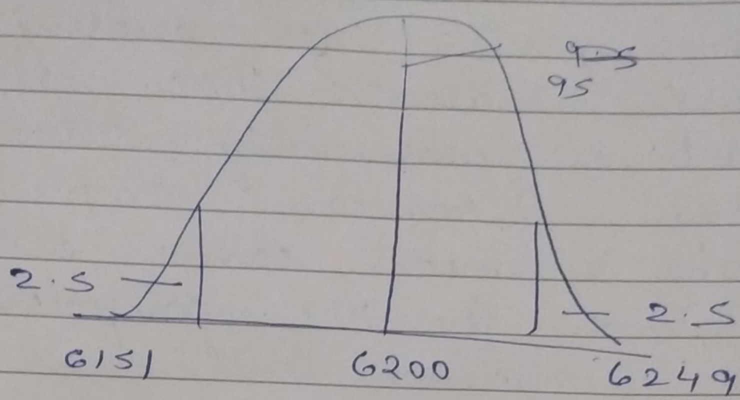
$$6200 + 1.96\left(\frac{200}{\sqrt{64}}\right)$$

$$6200 + 1.96\left(\frac{200}{8}\right)$$

$$6200 + 1.96 \times 25$$

$$= 6249$$

Lower Control ~~Unit~~ Limit = 6151



Finite Population Correction Factor

$$\bar{x} \pm z\frac{\sigma}{\sqrt{n}}\left(\sqrt{\frac{N-n}{N-1}}\right)$$