Page No.:
Date: 22/07/22

UNIT: 1

ORIGIN AND DEVELOPMENT OF STATISTICS:

→ Statistics, in a sense, is an old as the human society itself.
→ Its origin can be traced to the old days when it was regarded as the 'science of state-craft' and was the by product of the administrative activity of the State.
→ The word 'Statistics' seems to have been derived from the word 'status' or the italian word

→ Sir Ronald A Fisher (1890-1962), known as "Father of Statistics', placed statistics on a very sound footing by applying.

CONCEPT OF STATISTICAL POPULATION AND A SAMPLE:

→ There are millions of people in India, we are measuring price of automobile in Surat.
→ The set of those millions of vehicles is called the population of intrest, and the number attached to each one, its value is a measurement.
→ The average value is a parameter, a no. that describes a characteristic of the population, in this case monetary worth.
→ The set of 200 cars selected from the population is called sample.
→ The avg of data is called a statistic

Limitation: Statistics always gives approx. value.
Till collection of data descriptive
After that average to draw conclusion is inferential
Categorial & qualitative data are one & the same.

Statistics — Parameters
$\bar{x}$

Statistics & Planning

Statistics & Economics:
New disciplines: Economic Statistics & Econometrics
Such as wages, prices, analysis of time series and demand analysis.

Tourism:
Arrivals, departure, expenditure by the tourist, fatal accidents, facilities, etc.

Statistics & Business:
For studying the needs & desires of consumers
how many garments are to be manufactured.

Statistics & Industry:
Widely used in quality control.
Tools like inspection plans, control charts, etc are of extreme importance.

Insurance:
car, bike, life, and many more.

Limitation: Statistics always gives approx. value.
Till collection of data descriptive
After that average to draw conclusion is inferential.
Categorial & qualitative data are one & the same.

Statistics — Parameters
$\bar{x}$

Statistics & Planning

Statistics & Economics:
New disciplines: Economic Statistics & Econometrics
Such as wages, prices, analysis of time series and demand analysis.

Tourism:
Arrivals, departure, expenditure by the tourist, fatal accidents, facilities, etc.

Statistics & Business:
For studying the needs & desires of consumers how many garments are to be manufactured.

Statistics & Industry:
Widely used in quality control.
Tools like inspection plans, control charts, etc are of extreme importance.

Insurance:
Car, bike, life, and many more.

→ Premium is based on statistics.
→ Company use statistics that are collected from various homeowners, drivers, vehicle registration offices and many more. Recieve from all this and then decide.

Statistics and Mathematics :
→ Recent advancements in statistical techniques are the outcome of wide applications of advanced mathematics.
→ Statistics is a branch of Applied Mathematics which specialises in data.

→ Statistics data acquisation, analysis, explanation, interpretation and presentation.

→ Uses of statistics : In research can lead researches to summarization, proper characterization, performance and description of the outcome of the research.
→ Medical area : Recognize which drugs or interventions run best and how the individual groups respond to medicine. Conducts studies by age, race or country to identify the effect.
→ Most test of Covid 19, protocols with WHO.
→ Relating to causes and incidence diseases and results obtained. the efficacy of manufacture drug & injection.
→ Cyclones prediction.
→ Indian Meteorological Department, disaster management. Response & recovery teams always prefer statistics for getting the population data, services & information.

→ Premium is based on statistics.
→ Company use statistics that are collected from various homeowners, drivers, vehicle registration offices and many more. Recieve from all this and then decide.

Statistics and Mathematics :
→ Recent advancements in statistical techniques are the outcome of wide applications of advanced mathematics.
→ Statistics is a branch of Applied Mathematics which specialises in data.

→ Statistics data acquisation, analysis, explanation, interpretation and presentation.

→ Uses of statistics : In research can lead researches to summarization, proper characterization, performance and description of the outcome of the research.
→ Medical area : Recognize which drugs or interventions run best and how the individual groups respond to medicine. Conducts studies by age, race or country to identify the effect.
→ Most test of Covid 19, protocols with WHO.
→ Relating to causes and incidence diseases and results obtained. the efficacy of manufacture drug & injection.
→ Cyclones prediction.
→ Indian Meteorological Department, disaster management. Response & recovery teams always prefer statistics for getting the population data, services & information.

Statistics & Psychology & Education :

→ Eg Determine the reliability, validity of test.
'Factor Analysis', New subject called 'Psychometry' has come to existence.

Statistics & War :

→ In war, the theory of 'Decision Functions' can be great assistance to military.
'max. destruction with min effort'

LIMITATIONS :

→ It is not suited to the study of qualitative phenomenon.

→ Is applicable to study of enquiry which are capable of quantitative measurement. Eg. Intelligence of groups, poverty, culture, honesty, etc.

→ Statistics does not study individuals.
Eg. industry, cotton, national income, etc.

→ Statistical laws are not exact.
We talk on based probability not certainty.
Are not universally true - avg.
Eg. 20% certain surgical operation are successful doesn't mean out of 5 every day one will be saved or all will die.

→ Statistics is liable to be misused.
It is most dangerous tools in hands of the inexperts.
It is one of those whose adepts must exercise

Statistic & Psychology & Education :

→ Eg Determine the reliability, validity of test, 'Factor Analysis', New subject called 'Psychometry' has come to existence.

Statistics & War :

→ In war, the theory of 'Decision Functions' can be great assistance to military. (max. destruction with min. effort)

LIMITATIONS :

→ It is not suited to the study of qualitative phenomenon.

→ It is applicable to study of enquiry which are capable of quantitative measurement. Eg. Intelligence of group, poverty, culture, honesty, etc.

→ Statistics does not study individuals. Eg. industry outturn, national income, etc.

→ Statistical laws are not exact. We talk on based probability not certainty. Are not universally true - avg. Eg. 20% certain surgical operation are succesful doesn't mean out of 5 every day one will be saved or all will die.

→ Statistics is liable to be misused. It is most dangerous tools in hands of the inexperts. It is one of those whose adepts must exercise

the self restraint of an artist.
→ Used by inexperience & untrained persons cause fallacious conclusions.
→ Intresting examples of misinterpresention of statistical data.
→ Arguments based on incomplete data leads to fallacious conclusions.

SCALES OF MEASUREMENTS:
→ It is used more broadly and is more appropriately
→ Scales of mea, refer to ways in which variables numbers are defined and categorize.
→ Four scales: nominal — indentify
ordinal — certain, in order
interval
ratio

PROPERTIES:
→ Identity: each value having a unique meaning.
→ Magnitude: The values have an ordered relationship to one another, so there is a specific order to the variables
→ Equal intervals: The data pts along the scale are equal, so diff^{n} pt bet pts will be same.
→ A min. value of 0.: ~~no~~ Scale has a true 0 pt. Eg. Deg. can fall below 0 and still have meaning but not true for weight.

NOMINAL SCALE :

→ Property of data. Scale don't have numerical meaning. Data can be placed in categories but can't be x, ÷, +, − from one another.

→ Simplest, no numerical value for options.

→ Eg. Where do you live ?
1 Suburbs 2 City 3 Town

→ Used by research surveys & questionnaires where only variable labels hold significance.

→ While capturing nominal data, researchers conduct analysis based on associated label. Number orders doesn't matter.

→ Eg. eye colour ; country of birth.

3 categories :

- Nominal with order : Eg. 'cold, warm, hot'.
- Nominal without order : Eg. male & female
- Dichotomous : only 2 categories or level (always 2 opposite to each other).
  Eg. Yes & No

ORDINAL SCALE :

→ Variable. Simply depict order of variables and not the difference.

→ Generally non-mathematical ideas such as frequency, satisfaction, happiness.

→ 'Ordinal' sounds similar to 'Order' which is exact purpose of scale.
Eg. How satisfied are you with our services?

Very Unsatisfied → 1
Unsatisfied → 2
Neutral → 3
Satisfied → 4
Very satisfied → 5

→ Represents an ordered series of relationships or rank order.
→ Eg. Stood 1st, 2nd & 3rd.
→ You are not able to measure its just simple measurement
→ Data placed in a specific order.
→ Can't be ×, ÷, +, −
→ These are used in market research to gather reviews.
→ A step above nominal scale.
→ Order is prime imp. and so the labelling.
Eg. very pain always last & painless 1.
→ Becomes convinient for researcher.
→ Intend to obtain more info, then use ordinal.
→ This scale not only assigns values of variables but also measures the rank or order of variables.

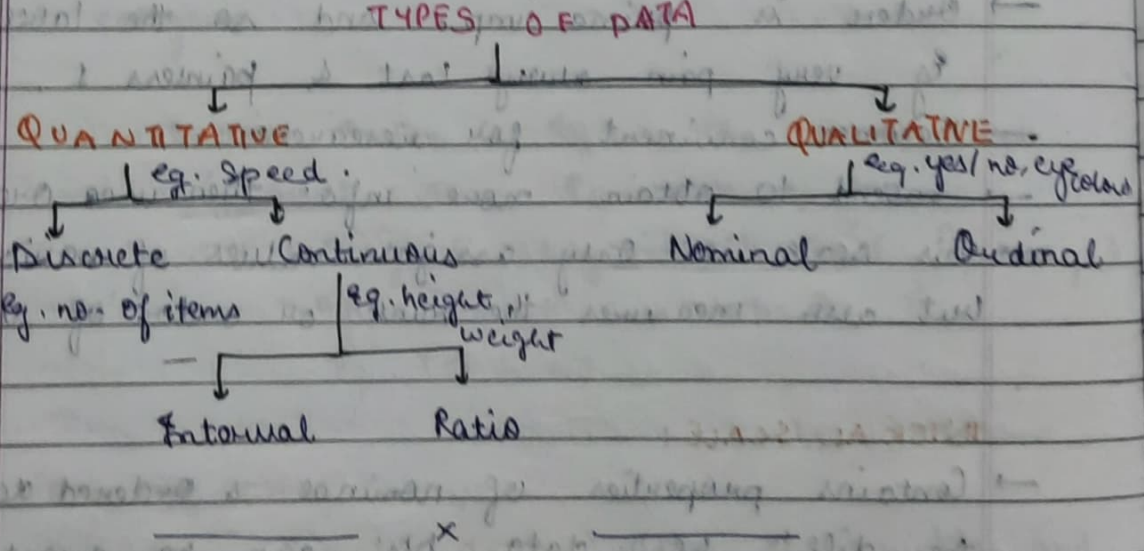
INTERVAL SCALE:

→ Contains properties of nominal & ordered data, but the diff^{n} betⁿ data pts' can be quantified.
→ Show both exact data & order along with diff.
→ +, − possible, ×, ÷ not possible for eg. Degrees.
→ No. 0 zero is existing variable. In ordinal scale 0 has no value.
→ Diff^{n} remains same.

→ Here quantitative, ordinal & nominal only qualitative.
→ They are the numeric scales.
→ Eg. Temp., Time,

RATIO SCALE:

→ Include properties of all other scales.
→ The data is nominal and defined by an identity, can be classified in order, contains intervals and can be broken down to exact value.
→ +, −, ×, ÷ can be done
→ Shares → Eg.
→ Differ from interval that has a 'true zero'.
→ The 0 means that data pt. has no variable



CH: 2

First in CRE - II book.

5/02/22 Range:

Two data sets:

Set 1	40	38	42	40	39	39	43	40	39	40
Set 2	46	37	40	33	42	36	40	47	34	45

Mean: $\frac{\Sigma x}{n} = \frac{400}{10} = 40$

Median: 40

Mode: 40

Set 2: Mean: $\frac{\Sigma x}{n} = 40$

Median: 40

Mode: 40

→ Same mean, median, mode but the variability of sets are different so new things are to be used.

→ locate its center.

RANGE:

The range of a data set is the no. R defined by formula

$$R = x_{max} - x_{min}$$

where x_{max}, is largest measurement

x_{min} is smallest measurement.

Range of Set 1: $43 - 38 = 5$

Set 2: $47 - 33 = 14$

THE VARIANCE AND STANDARD DEVIATION:

The sample variance of a sample data is the no. s^2 defined by the formula.

$$s^2 = \frac{\Sigma (x - \bar{x})^2}{n - 1}$$

which by algebra is equivalent to the formula.

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

The sample Standard Deviation:

$$s = \sqrt{\frac{\Sigma (x - \bar{x})^2}{n - 1}} = \sqrt{\frac{\Sigma x^2 - 1/n(\Sigma x)^2}{n - 1}}$$

For population, population Standard deviation.

N population data

σ → Std. Vaari deviation

σ^2 → Variance

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

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

Page No. :
Date :

Set 1: $s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$

$\therefore s^2 = \frac{0 + (-2)^2 + (2)^2 + 0 + (1)^2 + (-1)^2 + (3)^2 + 0 + (-1)^2 + 0}{9}$

$\therefore s^2 = \frac{8+3+9}{9} = \frac{20}{9} = \boxed{2.22}$

$\therefore s = \sqrt{2.22} = \boxed{1.489}$

Set 2: $s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$

$\therefore s^2 = \frac{(6)^2 + (-3)^2 + 0 + (7)^2 + (2)^2 + (-4)^2 + (0)^2 + (7)^2 + (6)^2 + (5)^2}{9}$

$s^2 = \frac{224}{9} = 24.888$

$\therefore s = \sqrt{24.888} = \boxed{4.987}$

KEY POINTS:

The R, S, S^2 give quantitative answer. How variable the data is?

CH:3
DATA ANALYSIS

RANDOM VARIABLES :

A RV is a v that assumes numerical values associated with the random outcome.

Types →
- Discrete , eg. no of steps eifel tower
- Continuous . eg , time tourist takes to reach top.

→ Data collection is an imp. step in the data analysis process. When we set out to collect information, it is imp. to keep in mind the ques. we hope to answer on the basis of the resulting data.

→ Sometimes we are interested in answering ques about characteristics of a single existing population or in comparing two or more well defined populations.

→ Ques like What happen's when . . . ? What is the effect . . . ?

For eg. an ecologist might be interested in estimating the average shell thickness of tree stem. These are eg. of studies that are observational.

→ Experimental Study.

An educator may wonder what would happen to test scores if the req. lab time for chemistry course increased 3 h to 6h per week.

→ In exp., the researcher manipulates one or more variables, called factors, to create the experimental conditions.

Statistical Studies: Observation & Experimentation.

→ A study is an

Graphs in other book:

Frequency Distributions & Bar Charts for Categorical Data

→ An appropriate graphical or tabular display of data can be

→ A freq. distribution for categorical data is a table that displays the possible categories along with the associated freq. and/or relative freq.

→ Rel. freq. = $\frac{\text{freq}}{\text{no. of observations in data set}}$

Rel. freq. total should be 1, but in some cases may be slightly off due to rounding.

→ A bar chart is one of the most widely used types of graphical displays for categorical data.

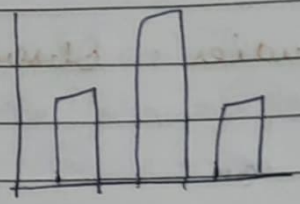
When to use: Categorical data

How to Construct: 1. Draw a — line, write category labels.

2. Draw 1 line, label freq. & rel. freq.

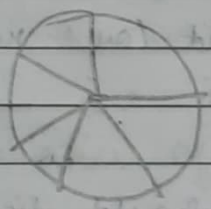
3. Place rectangular bar above each category level.

→ The height ∝ rel. freq. of data
→ A bar chart provides a visual representation of information in freq. distribution.



Pie Charts:

→ In a pie chart, a circle is used to represent the whole data set, with slices of the pie representing the possible categories.
→ Size of slice ∝ corresponding freq. or rel. freq.
→ Pie charts are most effective for summarising



When to use: When small no. of possible categories
How to construct: 1. Draw circle
2. Calculate slice size = 360 * cat rel. freq.
3. Pie charts are generated using a graphing calculator or statistical softwares.
What to look for: Categories that form large & small data sets.

Draw a comparative bar chart.

Q. Data represent the no. of days of sick leave taken by each of 50 workers of a given company over last 6 weeks :

2, 2, 0, 0, 5, 8, 3, 4, 1, 0, 0, 7, 1, 7, 1, 5, 4, 0, 4, 0, 1, 8, 9, 7, 0, 1, 7, 2, 5, 5, 4, 3, 3, 0, 0, 2, 5, 1, 3, 0, 1, 0, 2, 4, 5, 0, 5, 7, 5, 1

a) How many workers had at least 1 day of sick leave.

b) How many workers had betⁿ 3 & 5 days of sick leaves.

c) How many workers had more than 5 days of sick leaves.

a) 38 (50 − 12)

b) 5 17 (4 + 3 + 8)

c) 8 (2 + 5 + 2 + 1)

Value freq. table.

Value	freq.
0	12.
1	
2	

→ Data can be 1) Symmetric
2) Almost symmetric
3) No symmetry.

A set of data is said to be symmetric about the value x_0 if freq. of values $x_0 - c$ & $x_0 + c$ are same for all c.

Page No.:
Date :

Draw a comparative bar chart.

Q. Data represent the no. of days of sick leave taken by each of 50 workers of a given company over last 6 weeks :

2, 2, 0, 0, 5, 8, 3, 4, 1, 00, 7, 1, 7, 1, 5, 4, 0, 4, 0, 1, 8, 9, 7, 0, 1, 7, 2, 5, 5, 4, 3, 3, 0, 0, 2, 5, 1, 3, 0, 1, 0, 2, 4, 5, 0, 5, 7, 5, 1

a) How many workers had at least 1 day of sick leave.

b) How many workers had betⁿ 3 & 5 days of sick leaves.

c) How many workers had more than 5 days of sick leaves.

a) 38 (50 − 12)

b) 5 + 7 (4 + 3 + 8)

c) 8 (0 + 5 + 2 + 1)

Value freq. table.

Value	freq.
0	12
1	
2	

→ Data can be 1) Symmetric
2) Almost Symmetric
3) No symmetry.

A set of data is said to be symmetric about the value x_0 if freq. of values $x_0 - c$ & $x_0 + c$ are same for all c.

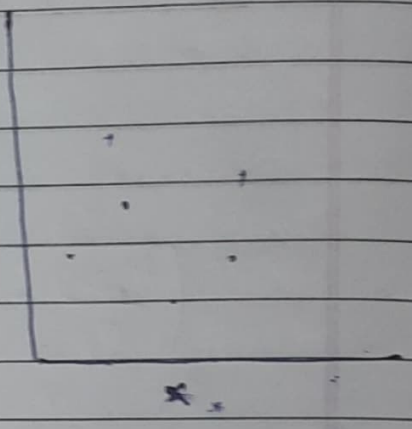
CORRELATIONS:

→ Finding the relationship bet^{n} 2 quantitative variables without being able to infer casual relationships.

→ Correlation is a statistical technique used to determine

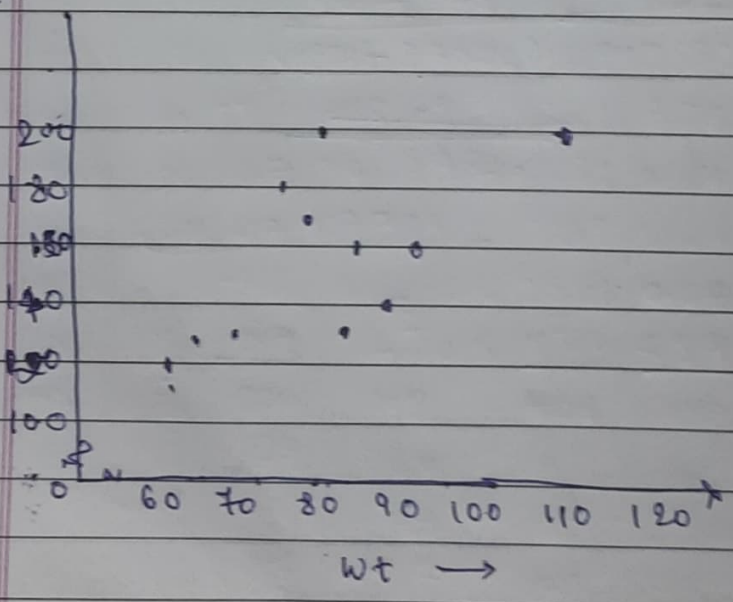
SCATTER DIAGRAM

1. Rectangular coordinate
2. Two quantitative variables.
3. One variable is called independent (x) and second is called dependent (Y)
4. Points are not joined.
5. No frequency table.



Eg. Wt (Kg) 67 69 85 83 74 81 97 92 110 85

Systolic Blood Pressure → SBP (mmHg) 120 125 140 160 130 180 150 140 200 130



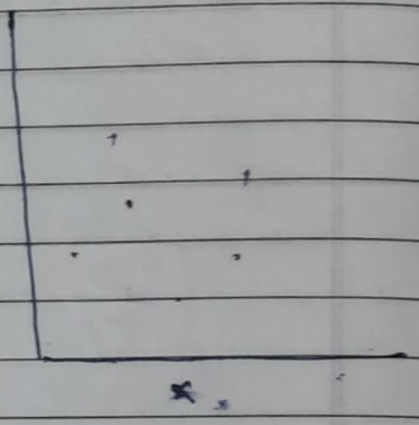
CORRELATIONS:

→ Finding the relationship bet^{n} 2 quantitative variables without being able to infer casual relationships.

→ Correlation is a statistical technique used to determine

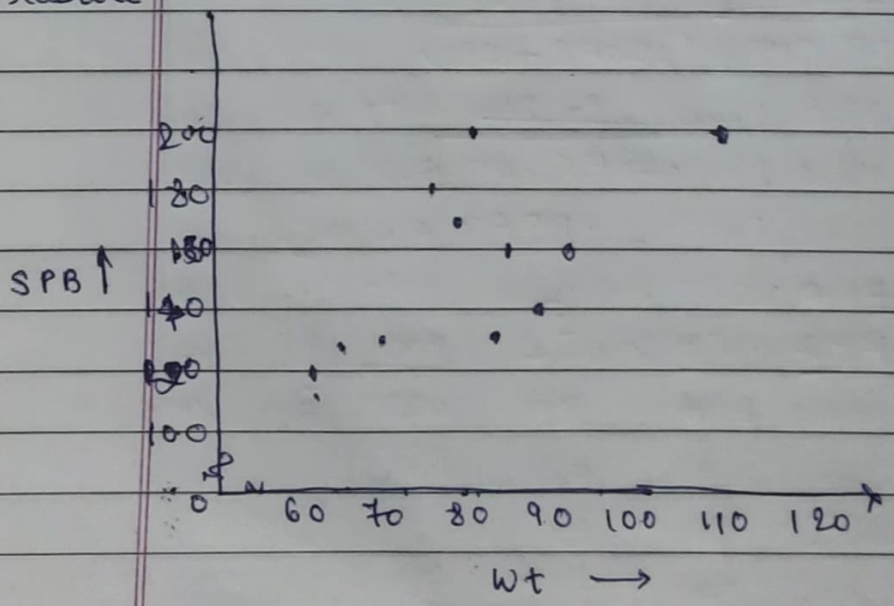
SCATTER DIAGRAM

1. Rectangular coordinate
2. Two quantitative variables.
3. One variable is called independent (x) and second is called dependent (Y)
4. Points are not joined.
5. No frequency table.



Eg. Wt (Kg) 67 69 85 83 74 81 97 92 114 85

Systolic Blood Pressure → SBP (mm Hg) 120 125 140 160 130 180 150 140 200 130



→ The pattern of data is indicative of type of relationship

1) Positive 2) Negative 3) NO relation

CORRELATION COEFFICIENT:

Simple correlation coeff. (r)

→ It is also called Pearson's corr. or product moment corr. coff.

→ It measures the nature & strength betⁿ 2 variables of quantitative.

→ Sign of r denotes nature of association.

→ +ve sign → Direct relation.

-ve sign → Inverse & Indirect

Range : -1 to +1

r denotes the strength of association.

strong weak weak weak strong

-1 -0.75 -0.25 0 0.25 0.75 1

perfect corr. | no rel. | intermediate | Perfect corr.

$$r = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\sqrt{\left(\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right)\left(\Sigma y^2 - \frac{(\Sigma y)^2}{n}\right)}}$$

→ The pattern of data is indicative of type of relationship

1) Positive 2) Negative 3) No relation

CORRELATION COEFFICIENT:

Simple correlation coeff. (r)

→ It is also called Pearson's corr. or product moment corr. coff.

→ It measures the nature & strength betn 2 variables of quantitative.

→ Sign of r denotes nature of association.

→ +ve sign → Direct relation.

-ve sign → Inverse & Indirect

Range : -1 to +1

r denotes the strength of association.

-1	-0.75	-0.25	0	0.25	0.75	1
strong		weak	weak		strong	
↑			↓			↑
perfect corr.			no rel.	Intermediate		Perfect corr.

$$r = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\sqrt{\left(\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right)\left(\Sigma y^2 - \frac{(\Sigma y)^2}{n}\right)}}$$

X	Y			
Age	wt	XY	x^2	y^2
7	12	84	49	144
6	8	48	36	64
8	12	96	64	144
5	10	50	25	100
6	11	66	36	121
9	13	117	81	169
41	66	461	291	742

$$\therefore r = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\sqrt{\left(\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right) \times \left(\Sigma y^2 - \frac{(\Sigma y)^2}{n}\right)}}$$

$$= \frac{461 - \frac{2706}{6}}{\sqrt{\left(291 - \frac{1681}{6}\right) \cdot \left(742 - \frac{4356}{6}\right)}}$$

$$= \frac{461 - 451}{\sqrt{(291 - 280.16)\,(742 - 726)}}$$

$$= \frac{10}{\sqrt{10.84 \times 16}} = \frac{10}{\sqrt{173.44}} = \frac{10}{13.169}$$

$$\therefore r = 0.759$$

X	Y	XY	x^2	y^2
10	2	20	100	4
8	3	24	64	6
2	9	18	4	81
1	7	7	1	49
5	6	30	25	36
6	5	30	36	25
32	32	129	230	201

$$\therefore r = \frac{129 - 170.66}{\sqrt{(230 - 170.66)(201 - 170.66)}}$$

$$= \frac{-41.66}{\sqrt{59.34 \times 30.34}} = \frac{-41.66}{42.43}$$

$$\therefore \boxed{r = -0.9818}$$

↳ Strong -ve

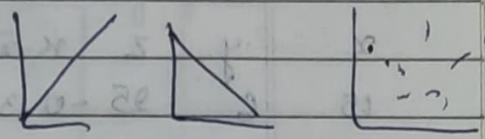
Coeff. of correlation :

$$r = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma (x_i - \bar{x})^2 (y_i - \bar{y})^2}}$$

↓ Pearson coeff.

Value bet^n -1 & +1

0 → no degree of correlation.



Strong +ve | Strong -ve | no rel.

MULTIPLE CORRELATION :

→ Multiple correlation deals with the situation in which the correlation bet^n 3 or more variables are req. to be found.

Eg :) Rel. bet^n yield of wheat per acre, amt. of rainfall and avg. daily temp.

2) Suppose a military organization wants to find the association

$$R_{x y_2} = \sqrt{\frac{r^2_{x2} + r^2_{y2} - 2 r_{xy} r_{yz} r_{zx}}{(1 - r^2_{xy})}}$$

$$\therefore r = \frac{129 - 170.66}{\sqrt{(230 - 170.66)(201 - 170.66)}}$$

$$= \frac{-41.66}{\sqrt{59.34 \times 30.34}} = \frac{-41.66}{42.43}$$

$$\therefore \boxed{r = -0.9818}$$

↳ Strong -ve

Coeff. of correlation :

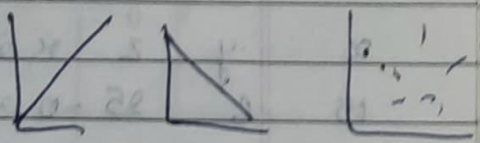
$$r = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma (x_i - \bar{x})^2 (y_i - \bar{y})^2}}$$

↓

Pearson coeff.

Value bet^n -1 & +1

0 → no degree of correlation.



Strong +ve | Strong -ve | no rel.

MULTIPLE CORRELATION :

→ Multiple correlation deals with the situation in which the correlation bet^n 3 or more variables are req. to be found.

Eg :) Rel. bet^n yield of wheat per acre, amt. of rainfall and avg. daily temp.

2) Suppose a military organization wants to find the association

$$R_{x y_2} = \sqrt{\frac{r^2_{x_2} + r^2_{y_2} - 2 r_{xy} r_{y_2} r_{2x}}{(1 - r^2_{xy})}}$$

IQ(x)	Study hr (y)	grades(z)
10	29	17
13	33	23
19	41	21
16	47	29
13	51	37
21	43	41
23	31	39
29	49	47
27	71	43
171		

Last time data.

x	y	z	$x-\bar{x}$	$y-\bar{y}$	$z-\bar{z}$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$	$(y-\bar{y})(z-\bar{z})$	$(x-\bar{x})(z-\bar{z})$	$(z-\bar{z})^2$
15	6	25	-0.3	1.2	-0.8	0.09	1.44	-0.36	-0.96	0.24	0.64
18	3	29	2.7	-1.8	3.2	7.29	3.24	-4.86	-5.76	8.64	10.24
13	8	27	-2.3	3.2	1.2	5.29	10.24	-7.36	3.84	-2.76	1.44
14	6	24	-1.3	1.2	-1.8	1.69	1.44	-1.56	-2.16	2.34	3.24
19	2	30	3.7	-2.8	4.2	13.69	7.84	-10.36	-11.76	15.54	17.64
11	3	21	-4.3	-1.8	-4.8	18.49	3.24	+7.74	8.64	20.64	23.04
17	4	26	1.7	-0.8	0.2	2.89	0.64	-1.36	-0.16	0.34	0.04
20	4	31	4.7	-0.8	5.2	22.09	0.64	-3.76	-(1.04)	24.44	27.04
10	5	20	-5.3	0.2	-5.8	28.09	0.04	-1.06	-1.16	30.74	33.64
16	7	25	0.7	2.2	-0.8	0.49	4.84	1.54	-1.76	-0.56	0.64
153	48	258	0			100.1	33.6	-21.4	(-12.28)	99.6	117.6

-15.4

$\bar{x} = 15.3$, $\bar{y} = 4.8$, $\bar{z} = 25.8$

IQ(x)	Study hr (y)	grades(z)
10	29	17
13	33	23
19	41	21
16	47	29
13	51	37
21	43	41
23	31	39
29	49	47
27	71	43
171		

Last time data.

x	y	z	$x-\bar{x}$	$y-\bar{y}$	$z-\bar{z}$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$	$(y-\bar{y})(z-\bar{z})$	$(x-\bar{x})(z-\bar{z})$	$(z-\bar{z})^2$
15	6	25	-0.3	1.2	-0.8	0.09	1.44	-0.36	-0.96	0.24	0.64
18	3	29	2.7	-1.8	3.2	7.29	3.24	-4.86	-5.76	8.64	10.24
13	8	27	-2.3	3.2	1.2	5.29	10.24	-7.36	3.84	-2.76	1.44
14	6	24	-1.3	1.2	-1.8	1.69	1.44	-1.56	-2.16	2.34	3.24
19	2	30	3.7	-2.8	4.2	13.69	7.84	-10.36	-11.76	15.54	17.64
11	3	21	-4.3	-1.8	-4.8	18.49	3.24	+7.74	8.64	20.64	23.04
17	4	26	1.7	-0.8	0.2	2.89	0.64	-1.36	-0.16	0.34	0.04
20	4	31	4.7	-0.8	5.2	22.09	0.64	-3.76	-4.16	24.44	27.04
10	5	20	-5.3	0.2	-5.8	28.09	0.04	-1.06	-1.16	30.74	33.64
16	7	25	0.7	2.2	-0.8	0.49	4.84	1.54	-1.76	-0.56	0.64
153	48	258	0			100.1	33.6	-21.4	-12.28	99.6	117.6

-15.4

$\bar{x} = 15.3$, $\bar{y} = 4.8$, $\bar{z} = 25.8$

$$r_{xy} = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma (x_i - \bar{x})^2 \Sigma (y_i - \bar{y})^2}} = \frac{-21.4}{\sqrt{100.1 \times 33.6}} = \frac{-21.4}{57.99}$$

$$= -0.369$$

$$r_{xy} \simeq -0.37$$

$$r_{yz} = \frac{\Sigma (y_i - \bar{y})(z_i - \bar{z})}{\sqrt{\Sigma (y_i - \bar{y})^2 \Sigma (z_i - \bar{z})^2}} = \frac{-15.4}{\sqrt{33.6 \times 117.6}} = \frac{-12.28}{62.85}$$

$$r_{yz} = -0.19$$

$$r_{xz} = \frac{99.6}{\sqrt{100.1 \times 117.6}} = \frac{99.6}{108.49} = 0.92$$

$$\therefore R_{xyz} = \sqrt{\frac{r_{xz}^2 + r_{yz}^2 - 2 r_{xy} r_{yz} r_{zx}}{(1 - r_{xy}^2)}}$$

$$= \sqrt{\frac{0.8464 + 0.0361 - 2(-0.37)(-0.19)(0.92)}{(1 - 0.1369)}}$$

$$= \sqrt{\frac{0.753148}{0.8631}}$$

$$R_{xyz} = 0.934 \simeq 0.93$$

$$R(xy, z) = \frac{(r_{xy} - r_{yz} * r_{zx})}{\sqrt{(1 - r^2_{yz})(1 - r^2_{zx})}}$$

$$= \frac{-0.37 - (-0.19 \times 0.92)}{\sqrt{(1 - (0.19)^2)(1 - (0.92)^2}}$$

$$= \frac{-0.1952139}{\sqrt{0.9639 \times 0.1536}} = \frac{-0.1752}{0.3847} = -0.46 \simeq -0.4$$

$$r_{xy} = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\Sigma(x_i-\bar{x})^2 \Sigma(y_i-\bar{y})^2}} = \frac{-21.4}{\sqrt{100.1 \times 33.6}} = \frac{-21.4}{57.99}$$

$$= -0.369$$

$$r_{xy} \simeq -0.37$$

$$r_{yz} = \frac{\Sigma(y_i-\bar{y})(z_i-\bar{z})}{\sqrt{\Sigma(y_i-\bar{y})^2 \Sigma(z_i-\bar{z})^2}} = \frac{-15.4}{\sqrt{33.6 \times 117.6}} = \frac{-12.28}{62.85}$$

$$r_{yz} = -0.19$$

$$r_{xz} = \frac{99.6}{\sqrt{100.1 \times 117.6}} = \frac{99.6}{108.49} = 0.92$$

$$\therefore R_{xy.z} = \sqrt{\frac{r_{xz}^2 + r_{yz}^2 - 2 r_{xy} r_{yz} r_{zx}}{(1 - r_{xy}^2)}}$$

$$= \sqrt{\frac{0.8464 + 0.0361 - 2(-0.37)(-0.19)(0.92)}{(1 - 0.1369)}}$$

$$= \sqrt{\frac{0.753148}{0.8631}}$$

$$R_{xyz} = 0.934 \simeq \boxed{0.93}$$

$$R(xy, z) = \frac{(r_{xy} - r_{yz} * r_{zx})}{\sqrt{(1 - r_{yz}^2)(1 - r_{zx}^2)}}$$

$$= \frac{-0.37 - (-0.19 \times 0.92)}{\sqrt{(1 - (0.19)^2)(1 - (0.92)^2}}$$

$$= \frac{-0.19139}{\sqrt{0.9639 \times 0.1536}} = \frac{-0.19139}{0.3847} = \boxed{-0.49}$$

REGRESSION ANALYSES :

→ Regression technique concerned with predicting some variable by knowing others.

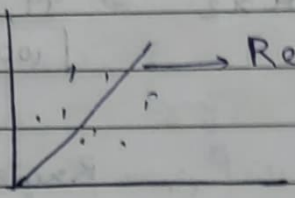
→ The process of

→ It is one step beyond correlation in identifying the relationship b/w 2 variables.
i.e. if you know x you can predict y.

→ This is done by an eqn called regression eqn.

→ We have a scatter plot, we can draw a st. line covering all pts, such line is called regression line.

→ Can be done in small no.s only manually.



→ The line is as close as possible to the data.

→ Such a line is best fitting line or Regression line.

Guidelines :

1. Use only when there is significant correl. to predict values.
2. Don't used if not signi
3. Stay within range of data.
4. The y variable is termed as criterion variable and x variable the predictor.

Variable.

The slope is often called the regression coeff.

$$\beta_1 = r \times \frac{Sy}{Sx}$$

REGRESSION ANALYSES :

→ Regression technique concerned with predicting some variable by knowing others.

→ The process of

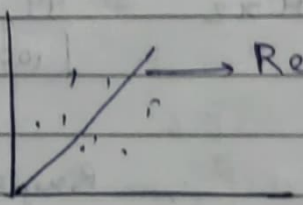
→ It is one step beyond correlation in identifying the relationship b/w 2 variables.

i.e. if you know x you can predict y.

→ This is done by an eqn called regression eqn.

→ We have a scatter plot, we can draw a st. line covering all pts, such line is called regression line.

→ Can be done if small no.s only manually.



→ The line is as close as possible to the data.

→ Such a line is best fitting line or Regression line.

Guidelines :

1. Use only when there is significant correl. to predict values.
2. Don't used if not signi
3. Stay within range of data.
4. The y variable is termed as criterion variable and x variable the predictor.

Variable.

The slope is often called the regression coeff.

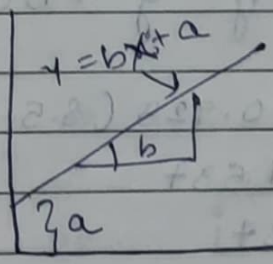
$$b_1 = r \times \frac{Sy}{Sx}$$

→ Correlation describes the strength of linear rel.
→ Regression tells us how to draw the st. line described by correlation.
→ The line (reg.) makes the sum of sq. of residuals smaller than other lines.

$$\hat{y} = a + bx \qquad \hat{y} = \bar{y} + b(x - \bar{x})$$

$$\therefore \boxed{b_1 = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}}$$

→ Regression eq^n describes the line mathematically.
→ Intercept → Slope.



Q.

Ser. No.	Age (x)	Weight (Y)
1	7	12
2	6	8
3	8	12
4	5	10
5	6	11
6	9	13

Find at 8.5ye

x	y	x^2	xy
7	12	49	84
6	8	36	48
8	12	64	96
5	10	25	50
6	11	36	66
9	13	81	117
41	66	291	461

$$\therefore b_1 = \frac{\Sigma xy - (\Sigma x \Sigma y)/n}{\Sigma x^2 - (\Sigma x)^2/n} = \frac{461 - 451}{291 - 280.167}$$

$$= \frac{10}{10.833}$$

$$\therefore b_1 = 0.923$$

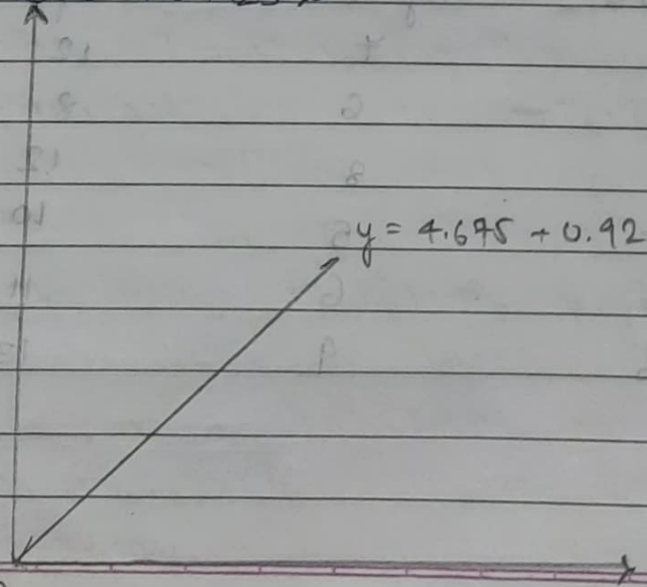
$$\therefore \hat{y} = a + bx \qquad \hat{y} = \bar{y} + b(x - \bar{x})$$

$$\therefore \hat{y} = 11 + 0.923\,(8.5 - 6.834)$$

$$= 11 + 1.537$$

$$\therefore \boxed{\hat{y} = 12.537}$$

$\hat{y} = 4.675 + 0.923x$



x	y	x^2	xy
7	12	49	84
6	8	36	48
8	12	64	96
5	10	25	50
6	11	36	66
9	13	81	117
41	66	291	461

$$\therefore b_1 = \frac{\Sigma xy - (\Sigma x \Sigma y)/n}{\Sigma x^2 - (\Sigma x)^2/n} = \frac{461 - 451}{291 - 280.167}$$

$$= \frac{10}{10.833}$$

$$\therefore b_1 = 0.923$$

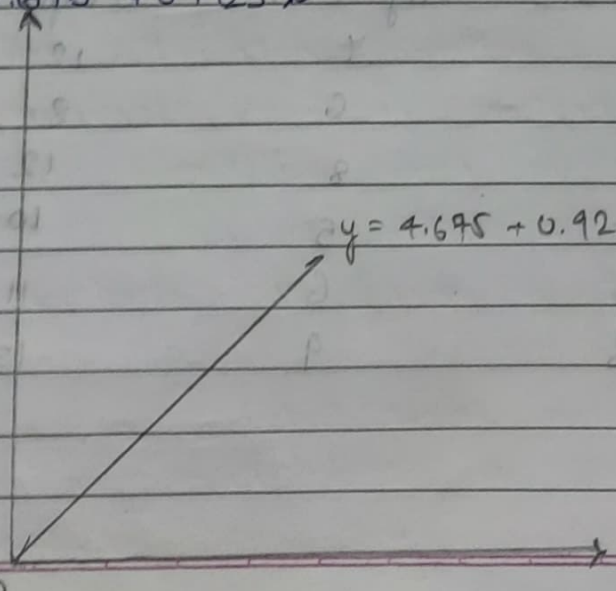
$$\therefore \hat{y} = a + bx \qquad \hat{y} = \bar{y} + b(x - \bar{x})$$

$$\therefore \hat{y} = 11 + 0.923\,(8.5 - 6.834)$$

$$= 11 + 1.537$$

$$\therefore \boxed{\hat{y} = 12.537}$$

$\hat{y} = 4.675 + 0.923x$



Age (x)	B.P. (y)	x^2	xy
20	120	400	240
43	128	1849	5504
63	141	3969	8883
26	126	676	3276
53	134	2809	7102
31	128	961	3968
58	136	3364	7888
46	132	2116	6072
58	140	3364	8120
70	144	4900	10080
46	128	2116	5888
53	136	2809	7208
60	146	3600	8760
20	124	400	8000
63	143	3969	9009
43	130	1849	5590
26	124	676	3224
19	121	361	2299
31	126	961	3906
23	123	529	2829
853	2630	41678	170744
		39230	114360

$n = 20$

$\therefore \hat{y} = \bar{y} + b(x - \bar{x}) = 131.5$

$b = \frac{170744 - 112169.5}{41678 - 36380.45} = \frac{58574.5}{5297.55}$

$b = 11.05$

$\therefore \hat{y} = 131.5 + 11.05(x - 42.65) = -339.78 + 11.05x$

Age (x)	B.P. (y)	x^2	xy
20	120	400	240
43	128	1849	5504
63	141	3969	8883
26	126	676	3276
53	134	2809	7102
31	128	961	3968
58	136	3364	7888
46	132	2116	6072
58	140	3364	8120
70	144	4900	10080
46	128	2116	5888
53	136	2809	7208
60	146	3600	8760
20	124	400	8000
63	143	3969	9009
43	130	1849	5590
26	124	676	3224
19	121	361	2299
31	126	961	3906
23	123	529	2829
853	2630	41678	170744
		39230	114360

$n = 20$

$\therefore \hat{y} = \bar{y} + b(x - \bar{x}) = 131.5$

$b = \dfrac{170744 - 112169.5}{41678 - 36380.45} = \dfrac{58574.5}{5297.55}$

$b = 11.05$

$\therefore \hat{y} = 131.5 + 11.05(x - 42.65) = -339.78 + 11.05x$

Page No.:
Date 01/09/22

HYPOTHESES TESTING:

→ In the sampling that we have studied so far the goal has been to estimate a population parameter.
→ But the sampling done by the gov. agency has a somewhat diff n objective, not so much to estimate the population mean μ as to test an assertion or a hypothesis about it, namely, whether it is as large as 75 or not.
→ The agency is not necessarily interested in the actual value of μ.

TYPES OF HYPOTHESIS:

A hypothesis about the value of population parameter is an assertion. (confident & forceful statement of fact) about its value.

H_0 → The null hypothesis about the population parameter that is assumed to be true unless there is convincing evidence to the contrary.

H_a → The alternative hypothesis, about the pop. parameter that is contradictory to the null hypothesis, and is accepted as true only if there is convincing evidence in favour of it.

→ Hyp. testing is a statistical procedure in which a choice is made betn H_0 & H_a based on information in a sample.

Page No.
Date 01/09/22

HYPOTHESIS TESTING:

→ In the sampling that we have studied so far the goal has been to estimate a population parameter.
→ But the sampling done by the govt. agency has a somewhat diff^n objective, not so much to estimate the population mean as to test an assertion or a hypothesis about it, namely, whether it is as large as 75 or not.
→ The agency is not necessarily interested in the actual value of it.

TYPES OF HYPOTHESIS:

A hypothesis about the value of population parameter is an assertion. (confident & forceful statement of fact) about its value.

H_0 → The null hypothesis about the population parameter that is assumed to be true unless there is convincing evidence to the contrary.

H_a → The alternative hypothesis, about the pop. parameter that is contradictory to the null hypothesis, and is accepted as true only if there is convincing evidence in favour of it.

→ Hyp. testing is a statistical procedure in which a choice is made betn H_0 & H_a based on information in a sample.

→ The end result of hypp. testing procedure is a choice of one of the following 2 possible conclusions.

1. Reject H_0 (∴ accept H_a) OR
2. Fail to reject H_0 (∴ fail to accept H_a).

Eg. of respirators, claim is avg. 75.

As we don't have reason to

Null Hypp. $H_0 = \mu = 75$.

Alternative Hypp. : Contradictory statement H_a

$: \mu < 75$

Compulsory to be above it's do ~~not~~ from the this.

H_0 always assertion with $=$ sign.

H_a can have all 3 $>, <, \neq$.

Eg: 1 $H_0 : \mu = 127.50$

$H_a : \mu > 127.50$

Eg.: 2 $H_0 : \mu = 8$ gm

$H_a : \mu \neq 8$ gm

Recipe not decided for less or more.

→ The null hypp. always have a form $H_0 : \mu = \mu_0$ for a specific μ_0 =

→ The test procedure is based on the initial assumption that H_0 is true.

→ The criteria for judging betn H_0 & H_a based on sample dat is :

→ The end result of hyp. testing procedure is a choice of one of the following 2 possible conclusions.
1. Reject H_0 (∴ accept H_a) or
2. Fail to reject H_0 (∴ fail to accept H_a).

Eg. of respirators, claim is avg. 75.
As we don't have reason so
Null Hyp. $H_0 = \mu = 75$.

Alternative Hyp. : Contradictory statement H_a
$: \mu < 75$
compulsory to be above it's to written like this.

H_0 always assertion with = sign.
H_a can have all 3 $>, <, \neq$.

Eg: 1 $H_0 : \mu = 127.50$
$H_a : \mu > 127.50$

Eg.: 2 $H_0 : \mu = 8$ gm
$H_a : \mu \neq 8$ gm
Recipe not decided for less or more.

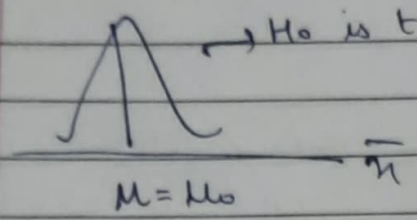
→ The null hyp. always have a form $H_0 : \mu = \mu_0$ for a specific μ_0 =

→ The test procedure is based on the intial assumption that H_0 is true.

→ The criteria for judging betn H_0 & H_a based on sample dat is :

→ If the value of $\bar{x}$ would be highly unlikely to occure if H_0 were true, but favors the truth of H_a then we reject in favor of H_a otherwise we don't

→ The Density Curve for $\bar{x}$ is a bell centered at μ_0



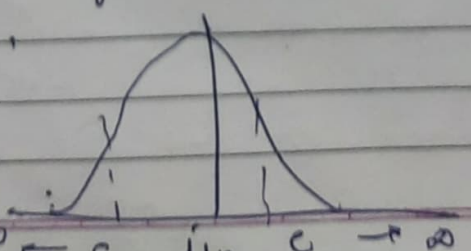
In some area only we can accept H_0, not at far away distances.

1. If H_a has form $\mu < \mu_0$ then reject H_0 if $\bar{x}$ is far from left of μ_0.
2. If $\neq \mu_1$ then reject H_0 if $\bar{x}$ is far on right of μ_0.
3. If H_a has form $H_a = \mu \neq \mu_0$, the reject H_0 if $\bar{x}$ is far away from μ_0 in either dirn.

→ We need to select rejection region.
Reject H_0 if the sample mean ($\bar{x}$) lies in the rejection region, but not reject H_0 if it does not.

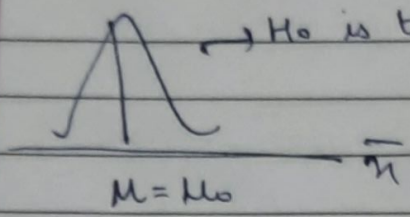
THE REJECTION REGION:

Rejection region $(-\infty, c]$ → for respirators.
Rejection region $[c, \infty)$ → for textbook.



→ If the value of $\bar{x}$ would be highly unlikely to occure if H_0 were true, but favours the truth of H_a & then we reject in favor of H_a otherwise we don't

→ The Density Curve for: a bell centered at μ_0



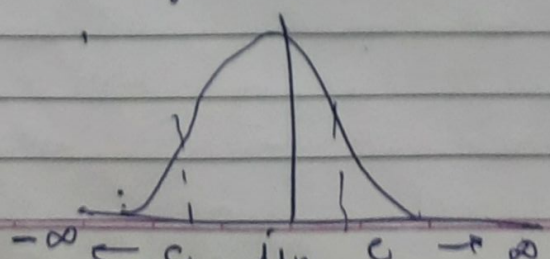
In some area only we can accept H_0, not at far away distances.

1. If H_a has form $\mu < \mu_0$ then reject H_0 if $\bar{x}$ is far from left of μ_0.
2. If ~~H_a~~ $\neq \mu_1$ then reject H_0 if $\bar{x}$ is far on right of μ_0.
3. If H_a has form $H_a = \mu \neq \mu_0$, the reject H_0 if $\bar{x}$ is far away from μ_0 in either dirn.

→ We need to select rejection region.
Reject H_0 if the sample mean ($\bar{x}$) lies in the rejection region, but not reject H_0 if it does not.

THE REJECTION REGION:

Rejection region $(-\infty, c]$ → for respirators
Rejection region $[c, \infty)$ → for textbook.

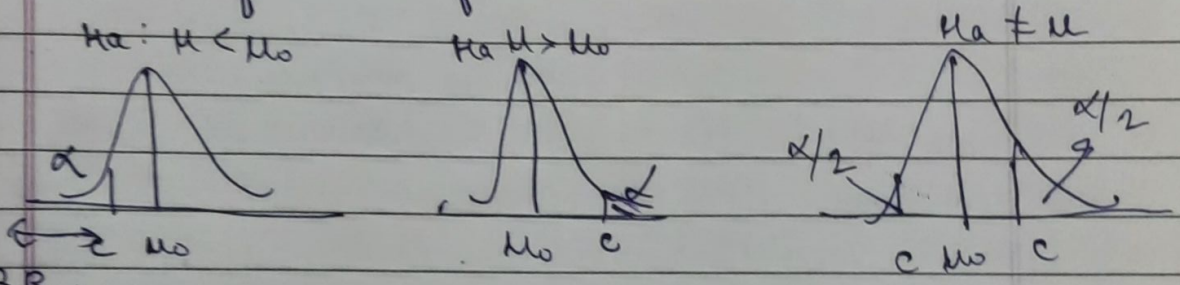


Rejection region $(-\infty, c] \cup [c', \infty)$ → for baked good.

diff^n no. possible.

c → Critical value or Critical values of the statistics.

→ Suppose the RR is single interval, so we need to select single c.
We select small pro. α say 1%. that's our definition of 'rare event'.



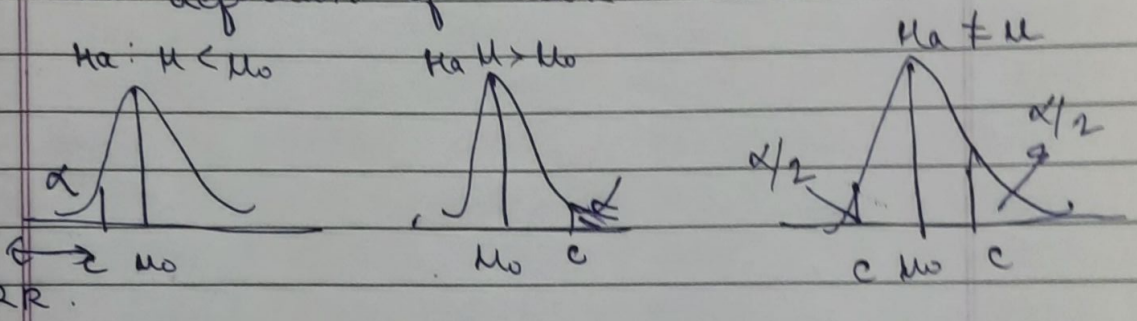
Rare event : Rejection of Null Hypothesis.

Rejection region $(-\infty, c] \cup [c', \infty)$ → for baked good.

diff^n no. posible.

c → Critical value or Critical values of the statistics.

→ Suppose the RR is single interval, so we need to select single c.
We select small pro. α say 1%. that's our definition of 'rare event'.



Rare event : Rejection of Null hypothesis.