Statistics → Descriptive (organize & summarize particular set of measurements)
→ Inferential
(data gathered from sample to make inferences)

Ronald A Fisher → Father of Statistics.
ANOVA → Analysis of variance.
Data → Quantitative (Numerical)
Qualitative (Non Numerical).

Sample	Population
$\bar{x}$	μ
s^2	σ^2
s	σ
$\hat{p}$	p

Limitation of Statistics:-
→ Not suited to study of qualitative phenomenon.
→ Doesn't study individual.
→ Statistical laws are not exact. (Doctor - Surgery example).
→ Statistics is liable to be misused.

Data
→ Quantitative (Numerical)
→ Discrete
→ Continuous.
→ Qualitative (Non numerical).
→ Nominal
→ ordinal.

Four scales of measurement → nominal, ordinal, interval and ratio.

Properties of measurement :-
(i) Identity :- Each value has unique meaning.
(ii) Magnitude :- Specific order to the variable.

Scanned with CamScanner

(iii) Equal Intervals: Data points along the scale are equal.

Nominal → can be ordered or not ordered.
Ordinal → ordered
Interval → Data shows both properties of nominal and ordered.
→ Shows both order of variables and exact difference
Ratio → Include properties of all four scale of measurement.

LECTURE :- 4/8/22

Measures of central tendency :-

$$\text{Sample mean} = \bar{x} = \frac{\sum x_i}{n}$$

$$\text{Population mean} = \mu = \frac{\sum x}{N}$$

x	f	$f \cdot x$
0	3	0
1	6	6
2	6	12
3	3	9
4	1	4
	$\Sigma f = 19$	$\Sigma fx = 31$

$$\text{Mean} = \frac{\Sigma f \cdot x}{\Sigma f} = \frac{31}{19} = 1.63$$

Median : $\tilde{x}$
↳ odd number of measurement → middle no.
→ even number of measurement → mean of two middle no.
⇓
ascending / Descending order

Scanned with CamScanner

Date
Page

mean = median → Symmetric / symmetric-skewed.

mode → highest frequency.
Highest point on histogram (highest R.F.).

Any set of data has 1 mean and 1 median, not true in case of mode.

Q. 21, 23, 23, 54, 67, 21, 25, 21, 54, 72, 75:

$$\text{Mean} = \frac{456}{11} = 41.54.$$

Median:-
21, 21, 21, 23, 23, 25, 54, 54, 67, 72, 75.

median = 25.

mode = 21 (Since it has highest frequency).

Q. Measure of variability of dispersion:-
→ range
→ variance
→ Standard Deviation.

Data set 1:- 38, 39, 39, 39, 40, 40, 40, 40, 42, 43

$$\text{mean} = \frac{400}{10} = 40.$$

$$\text{median} = \frac{40+40}{2} = 40.$$

mode = 40.

Data set 2:- 27, 34, 33, 36, 37, 40, 40, 42, 45, 46, 47.

mean = 40 = median = mode

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

x_{max} : largest measurement
x_{min} : smallest.

range indicates size of interval over which data points are distributed.

range of data set 1 = 43 − 38 = 5.
range of data set 2 = 47 − 33 = 14.

⇒ Sample variance = $s^2 = \frac{\sum (x-\bar{x})^2}{n-1}$

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

⇒ Sample standard Deviation :

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

⇒ Population variance = $\sigma^2 = \frac{\sum (x-\mu)^2}{N}$

⇒ Population standard deviation = $\sigma = \sqrt{\frac{\sum (x-\mu)^2}{N}}$

Sample variance : $s = \frac{\sum (x-\bar{x})^2}{n-1}$

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

DATA ANALYSIS

Date
Page

(i) Random variable: variable that assumes numerical values associated with random outcome of an experiment.

Discrete random variable:
continuous random variable.

LECTURE 18/7/22

* Three Popular Data Analysis:-
(1) Stem and Leaf Diagram.
(2) Frequency Histogram
(3) Relative frequency Histogram.
(4) Dotplot
} Numerical data.

Marks of 30 students:-

86	80	25	77	73	76	100	90	69
90	83	70	73	73	70	90	83	71
40	58	68	69	100	78	87	97	92

93
95
74

→ 2 → stem (ten's place).
→ unit digit (leaf)

(1) Stem and Leaf

2	5
3	—
4	0,
5	8
6	9, 8, 9,
7	7, 3, 6, 3, 3, 0, 1, 4, 8
8	6, 0, 3, 3, 7
9	0, 0, 0, 7, 3, 5, 2
10	0, 0

Stem → consist of beginning digits
Leaf → last part of number / final digit

213
(i) 2 stem
13 leaf
(ii) 21 stem
3 leaf

Histogram → Quantitative.

(3) R.F. Histogram :-

80s → 5

Total 30 students

∴ 80s → 5/30 (relative frequency).

70s → 10/30 = 1/3 = 33.3%.

90s → 7/30 = 23.3%.

100s → 2/30 = 1/15 = 6.6%.

60s → 3/30 = 10%.

50s → 1/30 = 3.33%.

40s → 3.33%.

20s → 3.33%.

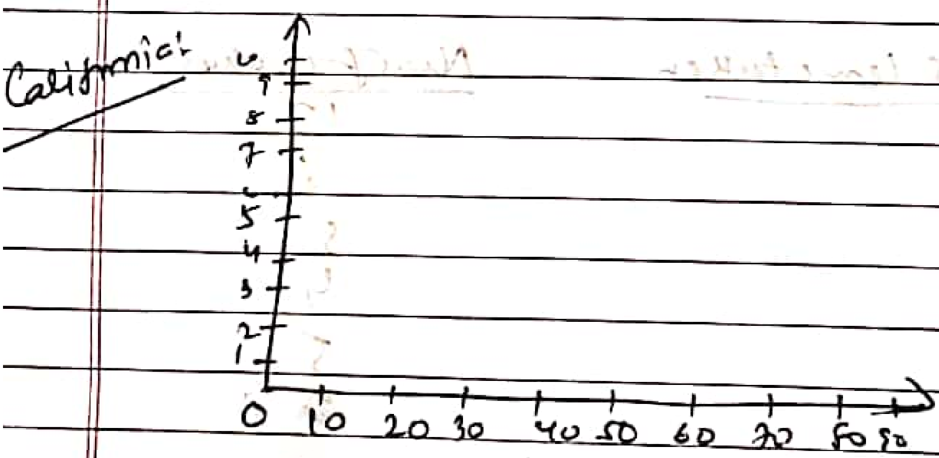
⇒ Dotplot (Measures of Variability).

California → 20 Division I schools.] % of full time freshman

Texas → 19 Division I schools.

California: 64, 41, 44, 31, 37, 73, 72, 68, 35, 37, 81, 90, 82, 74, 79, 67, 66, 66, 70, 63

Texas: 67, 21, 32, 88, 38, 71, 39, 35, 71, 63, 12, 42, 35, 39, 25, 65, 25, 24, 22.



22/8/22

Qualitative / Categorical Data :-

$$R.f. = \frac{\text{frequency}}{\text{No of obs in data set}}$$

⇒ Relative frequency Distribution

Categorical → Bar chart / Pie chart
Quantitative → Histogram.

used to display numerically categorical data
[Bar chart: Height of bar ∝ r.f.
[Pie chart: size of slice ∝ r.f.

Slice size = 360 × relative frequency

Question:

Number of days of sick leave taken by each of 50 workers, 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, 7, 5, 1

No. of days leave taken	No. of workers
0	12
1	8
2	5
3	4
4	5
5	8
6	0
7	5
8	2
9	1
	Σ = 50

50 - 12 = ?

(a) Atleast 1 day = ~~12+8~~ = ~~20~~ 38
(b) Between 3 & 5 day leave = ~~13~~ 5+8+4 = 17.
(c) More than 5 days of sick leave = 8

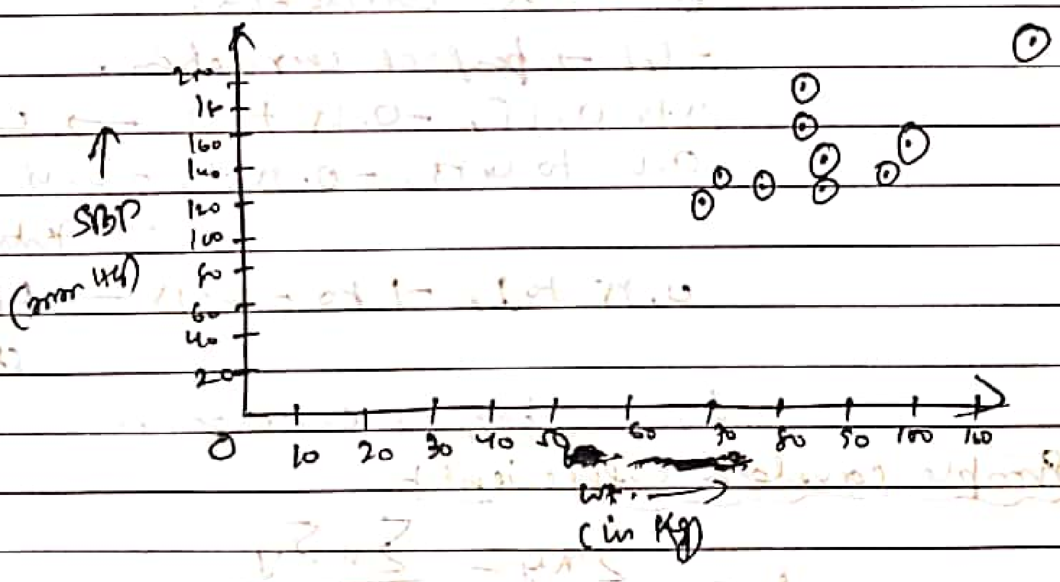
- Symmetric Data.
- Almost symmetric Data.
- No symmetry.

LECTURE : 25/8/22

Correlation : Identify relation between two variables.

→ Scatter diagram for correlation.

SBP (mm Hg) vs wt. (kg).



→ Pattern of data is indicative of type of relationship.
Type of relationship :
(1) Positive relationship ($x\uparrow, y\uparrow$) / Direct relationship
(2) Negative relationship ($x\downarrow, y\uparrow$) / Inversed relationship
(3) No relationship. (full scattered data).

Date
Page

Correlation coeffecient Parameter showing degree of relation.

(1.) Simple correlation coeffecient (r) ⇒ ~~Pearson's correlation or product moment correlation coeffecient~~

Sign of 'r' is +'ve ⇒ direct relationship (↑↑, ↓↓).

Sign of 'r' is −'ve ⇒ indirect relationship (↑↓, ↓↑).

Value of 'r' ranges between −1 to 1.

'0' → no correlation.
−1, 1 → perfect correlation.
0 to 0.25, −0.25 to 0 → weak
0.25 to 0.75, −0.75 to −0.25 → Intermediate
0.75 to 1, −1 to −0.75 → High/Strong.

r = 1, → perfect correlation.

⇒ Simple correlation coeffecient

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

n = 6.

Age (x)	weight (y)	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
$\Sigma x = 41$	$\Sigma y = 66$	$\Sigma xy = 461$	$\Sigma x^2 = 291$	$\Sigma y^2 = 742$

$$\therefore r = \frac{461 - \frac{66 \times 41}{6}}{\sqrt{\left(291 - \frac{(41)^2}{6}\right)\left(742 - \frac{(66)^2}{6}\right)}}$$

$$= \frac{10}{\sqrt{10.833 \times 16}}$$

$$= \frac{10}{13.16} = 0.759$$

↳ High correlation.
or
strong direct.

x	y	x^2	y^2	xy
10	2	100	4	20
8	3	64	9	24
2	9	4	81	18
1	7	1	49	7
5	6	25	36	30
6	5	36	25	30
$\Sigma X = 32$	$\Sigma Y = 32$	230	204	129

$$\therefore r = \frac{129 - \frac{32 \times 32}{6}}{\sqrt{\left(270 - \frac{32^2}{6}\right)\left(204 - \frac{32^2}{6}\right)}}$$

$$= \frac{-41.67}{\sqrt{59.33 \times 37.33}}$$

$$\cancel{0.937} = -0.937$$

↳ Strong indirect relationship.

→ ② Karan Product moment correlation

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

$\bar{x}$ = mean of 'x' variable

$\bar{y}$ = mean of 'y' variable

x_i, y_i → sample of variable x, y.

Outlier: Extreme point of data set

value of correlation b/w -1 and 1.

LECTURE:- 26/8/22

Partial Correlation

⇒ Two continuous variables, relationship → partial correlation, controlling effect of one or more variable.

⇒ More than two variables.

Ex: (1) Studying relationship between fertilizer and crop ~~yield~~ yield keeping weather conditions constant.

(2) Relationship between anxiety level and academic achievement keeping intelligence level constant.

Assumption 1:- one dependent variable and one independent variable, measured on continuous scale.

Assumption 2: one or more continuous variable, covariates

Assumption 3: Control variables measured on continuous scale.

Assumption 4: Linear relationship

Assumption 5: No significant outliers.

Assumption 6: Approximately normally distributed.

Partial correlation: Studying correlation b/w two variables keeping one variable const.

Date
Page

→ Partial ~~coef.~~ ~~corr.~~ correlation coeffecient formula

$x, y \to$ continuous, $z \to$ control variable

$$r_{(xy.z)} = \frac{r_{xy} - r_{yz} \times r_{zx}}{\sqrt{(1 - r_{yz}^2)(1 - r_{zx}^2)}}$$

↳ Pearson Product formula of 'r'.

~~EXAMPLE.~~

1.

x	y	z	$x - \bar{x}$	$(x-\bar{x})^2$	$y-$
10	29	17	−9	81	−14
13	33	23	−6	36	−10
19	41	21	0	0	−2
16	47	29	−3	9	+3
13	51	37	−6	36	7.
21	43	41	2	4	−0.
23	31	39	4	16	−1
29	49	47	10	100	5.
27	71	43	8	64	27
				Σ 346	

$$\bar{x} = \frac{171}{9} = 19$$

$$\bar{y} = \frac{395}{9} = 43.9$$

$$\bar{z} = \frac{297}{9} = 33$$

$r_{xy} = 0.53$

$r_{yz} = 0.17$

$r_{zx} = 0.79$

$$r_{xy.z} = \frac{0.532 - 0.589 \times 0.792}{\sqrt{(1-0.589^2)(1-0.792^2)}}$$

$$= \frac{0.062}{\sqrt{0.2348}}$$

$$= 0.127 \text{ (ans)}$$

$y-\bar{y}$	$(y-\bar{y})^2$	$z-\bar{z}$	$(z-\bar{z})^2$	$(x-\bar{x})(y-\bar{y})$	$(y-\bar{y})(z-\bar{z})$	$(z-\bar{z})(x-\bar{x})$
−14.9	222.0	−16	256	134.1	238.4	144
−10.9	118.8	−10	100	65.4	109	60
−2.9	8.41	−12	144	0	34.8	0
+3.1	9.61	−4	16	−9.3	−12.4	12
7.1	50.4	4	16	−42.6	28.4	−24
−0.9	0.81	8	64	−1.8	−7.2	16
−12.9	166.4	6	36	−51.6	−77.4	24
5.1	26.01	14	196	51	71.4	140
27.1	734.7	10	100	216.8	271	80
	$\Sigma =$ 1336.84		$\Sigma =$ 928	$\Sigma = 362$	$\Sigma = 656$	$\Sigma = 452$

$$r_{xy} = \frac{362}{\sqrt{346 \times 1336.84}} = 0.532$$

$$r_{yz} = \frac{656}{\sqrt{1336.84 \times 928}} = 0.589$$

$$r_{zx} = \frac{452}{\sqrt{928 \times 346}} = 0.792$$

LECTURE - 27/8/22

Multiple Correlation:-

Situation between three or more variables.

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

x	y	z
10	29	12
13	33	21
19	41	24
16	47	26
13	57	37
21	43	41
23	31	33
29	48	42
27	31	43

$$R_{x.yz} = \sqrt{\frac{(0.797^2 + 0.579^2 - 2 \times 0.532 \times 0.579 \times 0.797)}{1 - 0.532^2}}$$

$$= \sqrt{0.673} \quad \cancel{= 0.8204} \quad = 0.8204$$

$$= 0.8204$$

x	y	z	$x-\bar{x}$	$(x-\bar{x})^2$	$(y-\bar{y})$	$(y-\bar{y})^2$	$z-\bar{z}$	$(z-\bar{z})^2$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})(z-\bar{z})$	yz
15	6	25	-0.3	0.09	1.2	1.44	-0.8	0.64	$\bar{0}.36$	0.24	
18	3	29	2.7	7.29	-1.8	3.24	3.2	10.24	$\bar{4}.86$	8.64	
13	8	27	-2.3	5.29	3.2	10.24	1.2	1.44	$\bar{7}.36$	-2.76	
14	6	24	-1.3	1.69	1.2	1.44	-1.8	3.24	$\bar{1}.56$	2.34	
19	2	30	3.7	13.69	-2.8	7.84	4.2	17.64	$\bar{10}.36$	15.54	
11	3	21	-4.3	18.49	-1.8	3.24	-4.8	23.04	7.74	13.32	
17	4	26	1.7	2.89	-0.8	0.64	0.2	0.04	$\bar{1}.36$	1.04	
20	4	31	4.7	22.09	-0.8	0.64	5.2	27.04	$\bar{3}.76$	[illegible]	
10	5	20	-5.3	28.09	0.2	0.04	-5.8	33.64	$\bar{1}.06$	[illegible]	
16	7	25	0.7	0.49	2.2	4.84	-0.8	0.64	1.54	[illegible]	
						33.6					

$\bar{x} = \frac{153}{10} = 15.3$

$\bar{y} = \frac{48}{10} = 4.8$

$\bar{z} = \frac{258}{10} = 25.8$

$\Sigma(x-\bar{x})^2 = 100.1$

$\Sigma(y-\bar{y})^2 = 33.6$

$\Sigma(z-\bar{z})^2 = 117.6$

$\Sigma(x-\bar{x})(y-\bar{y}) = -21.4$

$\Sigma(y-\bar{y})(z-\bar{z}) = -15.4$

$\Sigma(x-\bar{x})(z-\bar{z}) = 99.6$

$$r_{xy} = \frac{-21.4}{\sqrt{100.1 \times 33.6}} = -0.369$$

$$r_{yz} = \frac{-15.4}{\sqrt{33.6 \times 117.6}} = -0.245$$

$$r_{zx} = \frac{99.6}{\sqrt{100.1 \times 117.6}} = 0.91$$

Scanned with CamScanner

$$(r_{xy.z}) = \frac{-0.369 + 0.245 \times 0.918}{\sqrt{(1-0.245^2)(1-0.918^2)}}$$

$$= -0.374$$

$$R_{x.yz} = \sqrt{\frac{0.736}{0.863}}$$

$$= 0.923$$

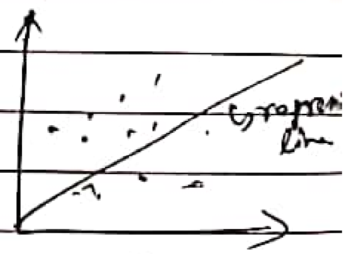
Regression = Prediction

LECTURE:- 29/8/22

Regression Analyses

→ Predicting some variable, knowing other.
→ Predicting X by Y or Y by X.

Regression line: line covering entire data in a scatter plot.
regression line / best fitting line, more closer to data points.



Predicting y by x

$$\begin{cases} x \to \text{predictor} \\ y \to \text{criterion.} \end{cases}$$

Slope:- $\beta_1 = r \cdot \frac{S_y}{S_x}$

Correlation:- Strength of linear relationship b/w two variables. / St. line

Regression:- How to draw the st. line given by correlation.

Date ______
Page ______

$\hat{y} = a + bx$

$\boxed{y - \bar{y} = b(x - \bar{x})}$

Predicting 'y' by 'x':

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

$$b_1 = \frac{\sum xy - \frac{\sum x \cdot \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$

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
$\sum x = 41$	$\sum y = 66$	$\sum x^2 = 291$	$\sum xy = 461$

$\bar{x} = \frac{41}{6} = 6.83$; $\bar{y} = 11$

$$b = \frac{461 - \frac{41 \times 66}{6}}{291 - \frac{(41)^2}{6}}$$

$$= \frac{10}{10.833} = 0.923$$

$\therefore \hat{y} = 11 + 0.923(x - 6.83) = 4.695 + 0.923x$

$\boxed{\therefore x = 8.5, \ y = 12.54}$

(i) Simple correlation = $\dfrac{\sum xy - \dfrac{\sum x \cdot \sum y}{n}}{\sqrt{\left(\sum x^2 - \dfrac{(\sum x)^2}{n}\right)\left(\sum y^2 - \dfrac{(\sum y)^2}{n}\right)}}$

$$= \frac{114486 - \dfrac{852 \times 2630}{20}}{\sqrt{\left(41678 - \dfrac{(852)^2}{20}\right)\left(347080 - \dfrac{(2630)^2}{20}\right)}}$$

$$= \frac{2448}{\sqrt{5382.8 \times 1235}}$$

$$= \frac{2448}{2578.32} = 0.949 \simeq 0.95$$

Pearson product correlation = $\dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$

$$= \frac{2448}{\sqrt{5382.8 \times 1235}}$$

$$= 0.949 \simeq 0.95$$

(ii) $b = \dfrac{\sum xy - \dfrac{\sum x \sum y}{n}}{\sum x^2 - \dfrac{(\sum x)^2}{n}}$

Date ____
Page ____

$$\Rightarrow b = \frac{114486 - \frac{852 \times 2630}{20}}{41678 - \frac{(852)^2}{20}}$$

$$= \frac{2448}{5382.8} = \underline{0.454}.$$

$$\Rightarrow y = 131.5 + 0.454(x - 42.6)$$

$$\Rightarrow y = 112.16 + 0.454x$$

when $x = 25$, $y = 123.51$.

LECTURE:- 1/9/22

Hypotheses Testing:

⇒ A manufacturer asserts that a respirator it makes deliver pure air 75 minutes on average.

Population parameter – mean, ~~sample~~ standard deviation, variance.

→ A hypothesis about the value of a population parameter is an assertion about its value.

H_0 : Null hypothesis. ⇒ Statement about population parameter assumed to be true.

H_a : Alternative hypothesis.

⇒ Statement about population parameter contradicting to null hypothesis, accepting as true only if there is evidence in favour on it.

$\boxed{\text{Null hypothesis \& always } +\text{ve or contradicting}}$

Scanned with CamScanner

→ Hypothesis testing :- choice between a null hypothesis and an alternative hypothesis.

2 possible result of hypothesis testing :-
① Reject H_0 (and therefore accept H_a).
② Fail to reject H_0 (and then reject H_a).

Null hypothesis :- Equal sign will be there.
Alternative hypothesis :- $\mu \neq \mu_0$, $\mu > \mu_0$, $\mu < \mu_0$

$\boxed{\mu_0 \rightarrow \text{sample mean}}$

Null: $\mu = 75$
Alternative: $\mu < 75$
↳ ($\mu > 75$) since if medical equipment supplies more than 75, there is no harm.

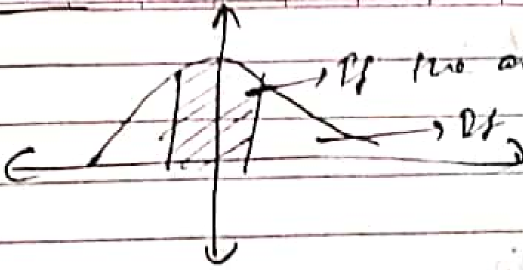
Example 1 :-
$H_0: \mu = \$12.75$
$H_a: \mu > \$12.75$
⟹ evidence mean is greater than \$12.75.

Example 2: $H_0: \mu = 8$ grams / fat serving.
$H_a: \mu \neq 8$ grams / fat serving.

→ If there is no contrary, we assume H_0 is true.

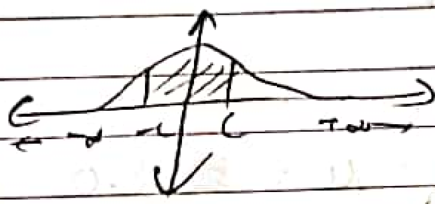
→ If value of $\bar{x}$ would be highly unlikely to occur if H_0 were true, but favours truth of H_a, H_a is assumed to be true.

Normal Distribution:- $\mu = \mu_0$.

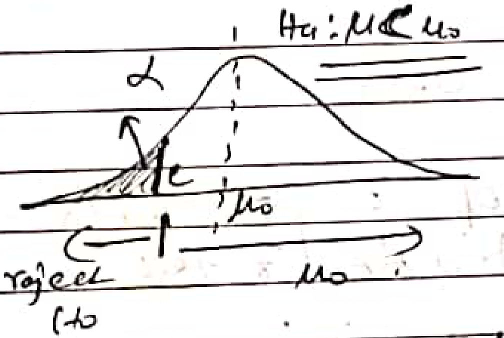


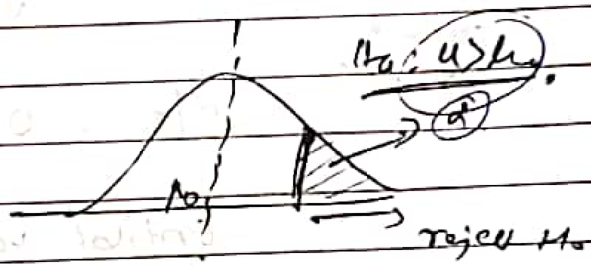
If Ho area is near Ho is true.
If farther, Ho is false, Ha is true.

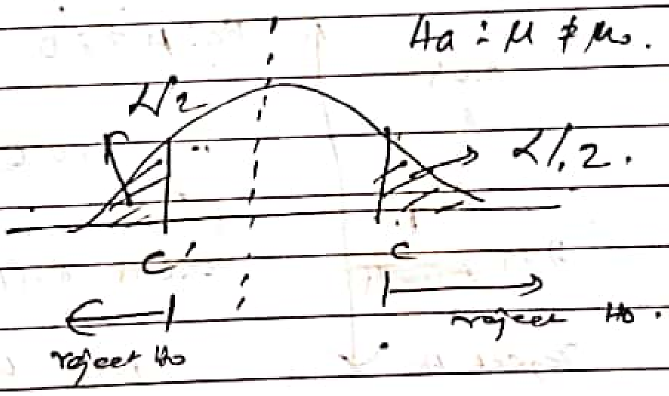
Ha: $\mu < \mu_0$, rejection area is:- $(-\infty, -c]$.
Ha: $\mu > \mu_0$, rejection area is:- (c, ∞).



Ha: $\mu \neq \mu_0$, rejection area:- $(-\infty, c] \cup [c', \infty)$







rare event: rejection of Null hypothesis.
α: probability of rare event.

α: level of significance

Scanned with CamScanner

LECTURE: 2/9/22

Date
Page

Sampling distribution

Ex-3:-

$\sigma = 0.15$

$H_0: \mu = 8.0$ Sample size = 5.

$H_a: \mu \neq 8.0$.

Construct rejection region for test of choice, $\alpha = 0.1$.

A:-

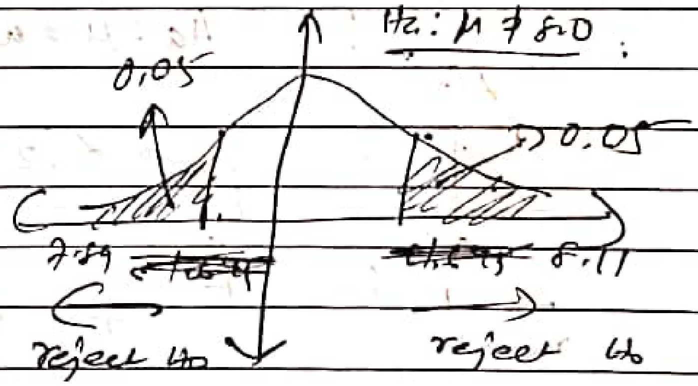
$\mu_{\bar{x}} = \mu = 8.0$

sample standard deviation $\leftarrow$ $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.15}{\sqrt{5}} = 0.067$

$\alpha/2 = 0.1/2 = 0.05$

critical value of $z_{0.05} = 1.645$.

Z - Distribution / Z test



$\rightarrow C = 8 - (1.645)(0.067) = 7.89$

$\rightarrow C' = 8 + (1.645)(0.067) = 8.11$.

from $(-\infty, 7.89]$, and $[8.11, \infty)$, null hypothesis is rejected.

σ : Population standard deviation
$\sigma_{\bar{x}}$; sample standard deviation. $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

Date
Page

$H_a: \mu \neq \mu_0$ → two tailed test.
$H_a: \mu < \mu_0$ → left tailed test.] one tailed test.
$H_a: \mu > \mu_0$ → right tailed test.

Two types of error :-
1. reject null hypothesis H_0 in favor of alternative hypothesis H_a.

Type I, type II error.

		True state of nature	
		H_0 is true	H_0 is false
Our Decision	Don't reject H_0	Correct Decision	Type II error
	Reject H_0	Type I error	Correct Decision

α - level of significance. | Type II more dangerous
→ probability of type II error.
⇒ To reduce error, increase sample size.

Standardized test statistic, $\frac{\bar{x} - 8.0}{0.067}$ | $\frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}}$

~~Normal Distribution~~ x

Symbol in H_a	Terminology	rejection region: Normal Distribution	Student t
<	left tailed	$(-\infty, -z_\alpha]$	$(-\infty, -t_\alpha]$
>	right tailed	$[z_\alpha, \infty)$	$[t_\alpha, \infty)$
≠	two tailed	$(-\infty, -z_{\alpha/2}] \cup [z_{\alpha/2}, \infty)$	$(-\infty, -t_{\alpha/2}] \cup [t_{\alpha/2}, \infty)$

Scanned with CamScanner

Procedure:-

1. Identify null and alternative hypothesis
2. Identify relevant test statistic and its distribution
3. Compute from data, the value of test statistic
4. Construct the rejection region.
5. Compare value computed in step 3 to the rejection region constructed in step 4 and make a decision.

LECTURE:- 5/9/22

4. (a) Null hypothesis : $\mu = 74.5°F$
   Alternative : $\mu > 74.5°F$

(b) Null : $\mu = 145$ ~~pound~~
   Alternative : $\mu > 145$ ~~pound~~

(c) Null : $\mu = \$14,756$.
   Alternative : $\mu > \$14,756$

(d) Null : $\mu = \$82.53$.
   Alternative : $\mu \neq \$82.53$

(e) Null : $\mu = 69.4$ ~~ans~~.
   Alternative : $\mu < 69.4$ ~~ans~~.

Correct way of writing

Null hypothesis $H_0 : \mu = 74.5$
Alternative hypothesis $H_a : \mu > 74.5$

2. (a) Null hypothesis H_0 : $\mu = 38.2$
Alternative hypothesis H_a: $\mu < 38.2$

(b) Null hypothesis H_0 : $\mu = 58,291$
Alternative hypothesis H_a: $\mu \neq 58,291$.

(c) Null hypothesis H_0 : $\mu = 133$
Alternative hypothesis H_a: $\mu > 133$

(d) Null hypothesis H_0 : $\mu = 161.9$
Alternative hypothesis H_a: $\mu \neq 161.9$.

(e) Null hypothesis H_0 : $\mu = 42.8$
Alternative hypothesis H_a: $\mu > 42.8$.

sample mean

Random variable $\bar{X}$, has a mean denoted by $\mu_{\bar{x}}$ and standard deviation $\sigma_{\bar{x}}$

All possible random sample :

(152, 152) (164, 164)
(152, 156)
(152, 160)
(152, 164)
(156, 152)
(156, 156)
(156, 160)
(156, 164)
(160, 152)
(160, 156)
(160, 160)
(160, 164)
(164, 152)
(164, 152)

$\mu_{\bar{x}} = \mu$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

LECTURE :- 8/9/22

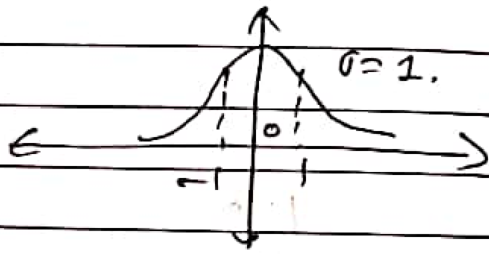
1. $\mu_{\bar{x}} = \mu = \$13,525$.

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4180}{\sqrt{100}} = \$418$.

Standard Normal Distribution:-

$\mu = 0$ Standard normal random variable
$\sigma = 1$. denoted by letter Z.

$P(Z<z)$: cumulative probabilities.



(a) $P(Z<1.48) = 0.9306$ (from table).

(b) $P(Z<-0.25) = 0.4013$ (from table).

(c) $P(Z>1.6) = 1 - P(Z<1.6)$
$= 1 - 0.9452 = 0.0548$

(d) $P(Z>-1.02) = 1 - P(Z<-1.02)$
$= 1 - 0.1539$
$= 0.8461$

(e) $P(0.5 < Z < 1.57)$

$= P(Z < 1.57) - P(Z < 0.5)$

$= 0.9418 - 0.6915$

$= 0.2503$.

(f) $P(-2.55 < Z < 0.09)$

$= P(Z < 0.09) - P(Z < -2.55)$.

$= 0.5359 - \cancel{0.0122}\ 0.0054$

$= 0.5305$.

General Normal variables:

X: normally distributed random variable

$$P(a < X < b) = P\left(\frac{a-\mu_x}{\sigma_x} < Z < \frac{b-\mu_x}{\sigma_x}\right)$$

$\mu_x = \mu$

$\sigma_x = \sigma$

Z: standard random variable.

'a' can be decimal number or $-\infty$.

'b' can be decimal number or $+\infty$.

(a) $P(X < 14)$

$= P\left(Z < \frac{14-10}{2.5}\right) = P(Z < 1.6)$

$= 0.9452$.

(b) $P(8 < X < 14)$

$= P(-0.8 < Z < 1.6) = P(Z < 1.6) - P(Z < -0.8)$

$= 0.9452 - 0.2119$

$= 0.7333$.

Q. $\mu = 37{,}500$ miles

$\sigma = 4{,}500$ miles.

Date ____
Page ____

$P(30,000 < X < 40,000)$

$= P\left(\frac{30,000 - 37,500}{4500} < Z < \frac{40,000 - 37,500}{4500}\right)$

$= P(-1.67 < Z < 0.56)$

$= P(Z < 0.57) - P(Z < -1.67)$
$= 0.7123 - 0.0475$
$= 0.6648$

~~Central Limit theorem:~~

Increasing ~~sampling~~ sampling distribution of $\bar{X}$, probability on lower and upper end shrink, on the middle it broadens (or increases).

Increasing sample mean, the curves resemble bell shape curve.

Central Limit Theorem:-
Sample size 30 or more, sample mean is approximately normally distributed with $\mu_{\bar{x}} = \mu$, $\sigma_{\bar{x}} = \sigma/\sqrt{n}$, where 'n' is sample size.

8. $\mu = 112$
$\sigma = 40$

(a.) Mean of $\bar{x} = \mu_{\bar{x}} = \mu = 112$.
S.D. of $\bar{x} = \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{40}{\sqrt{50}} = 5.656$.

Scanned with CamScanner

Date
Page

(b) $P(110 < \bar{X} < 114)$
$= P\left(\frac{110-112}{40} < Z < \frac{114-112}{40}\right)$
$= P(-0.05 < Z < 0.05)$
$= P(Z < 0.05) - P(Z < -0.05)$
$= 0.5199 - 0.4801$
$= 0.0398$

(c) $P(\bar{X} > 113)$
$= 1 - P(\bar{X} < 113)$
$= 1 - P\left(Z < \frac{113-112}{5.656}\right)$
$= 1 - P(Z < 0.18)$ $= 1 - 0.5714$
$= 0.4286$.

~~$= 1 - P(Z < 0.03)$~~
~~$= 1 - 0.5120$~~
~~$= 0.488$~~

(b) $P(110 < \bar{X} < 114)$
$= P\left(\frac{110-112}{5.656} < Z < \frac{114-112}{5.656}\right)$

$= P(-0.353 < Z < 0.353)$
$= P(Z < 0.35) - P(Z < -0.35)$
$= 0.6368 - 0.3632$
$= 0.2736$.

→ If population is normally distributed whatever be the sample size, sample mean is normally distributed.

Date ____
Page ____

Q. $\mu = 38,500$

$\sigma = 2500$

$n = 5$

$$\sigma_{\bar{x}} = \frac{2500}{\sqrt{5}} = 1118.03$$

$P(\bar{X} < 36000)$

$$= P\left(Z < \frac{36000 - 38,500}{1118.03}\right)$$

$= P(Z < -2.236)$

$= P(Z < -2.24)$

$= 0.0125.$

Q. $\mu = 50$ months.

$\sigma = 6$ months.

(a) $P(\bar{X} < 48)$.

$$= P\left(Z < \frac{48 - 50}{6}\right).$$

$= P(Z < -0.33) = 0.3707$

(b) $n = 36$.

$$\sigma_{\bar{x}} = \frac{6}{\sqrt{36}} = 1.$$

$\therefore P(\bar{X} < 48)$

$= P(Z < -2)$

$= 0.0228$

Conclusion:

→ sample size $n \geq 30$, sample mean is approximately normally distributed.

→ If population is normally distributed, whatever be the sample size, sample mean is normally distributed.

LECTURE:- 12/9/22

σ: Population
S: sample

Test statistic $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

σ is known, $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$

σ is unknown, $z = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$

Ex-4. $\alpha = 5\% = 0.05$

$\boxed{\cancel{z_\alpha = 0.5199 \text{ (from table)}}}$

Null hypothesis: Newly developed pain reliever doesn't deliver more relief.

Alternative hypothesis: Newly developed pain reliever delivers more relief quickly

$H_0: \mu = 3.5$

$H_a: \mu < 3.5$, $\alpha = 0.05$

$$z = \frac{3.1 - 3.5}{\frac{1.5}{\sqrt{50}}}$$

$= -1.886$

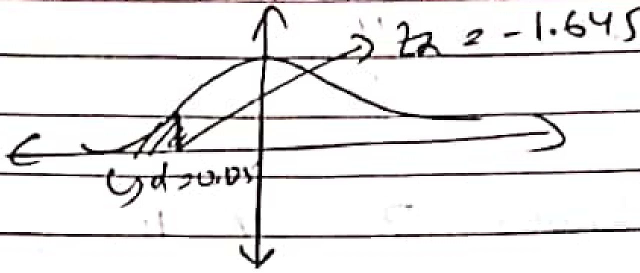
It is left-tailed test

when $\alpha = 0.0505$, $z = -1.64$

$\alpha = 0.0495$, $z = -1.65$

$$\frac{-1.64 + 1.65}{0.0505 - 0.0495} = \frac{z_\alpha + 1.65}{0.05 - 0.0495}$$

$\Rightarrow \boxed{z_\alpha = -1.645}$



Now, $-1.886 < -1.645$

-1.886 lies in rejection region.

$\Rightarrow$ Null hypothesis is rejected

$\Rightarrow$ Newly developed pain delivers more relief quickly.

Q. $\mu = 8.1$ ounce per jar.

$\sigma = 0.22$ ounce

$n = 30$.

$\bar{x} = 8.2$ ounce $\qquad \alpha = 0.01$.

$s = 0.25$ ounce.

Null hypothesis: $H_0 : \mu = 8.1$

Alternative hypothesis: $H_a : \mu \neq 8.1$.

$$\alpha/2 = \frac{0.01}{2} = 0.005$$

$$z_\alpha = -2.575 \quad \cancel{z = 2.32} \quad \cancel{2.4524}$$

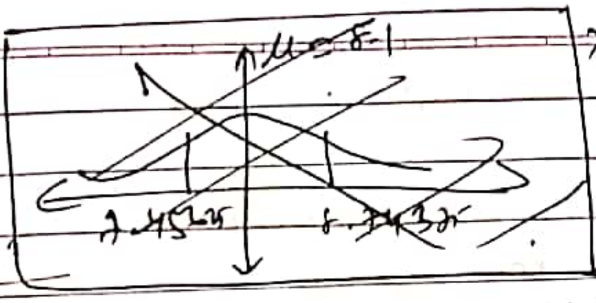
$$\cancel{c = 8.1 - 2.575 \times 0.25 = 0.64375}$$

$$\cancel{c' = 2.575 \times 0.25 = 0.64375}$$

$$\cancel{c' = 8.1 + 2.575 \times 0.25 = 8.74375}$$

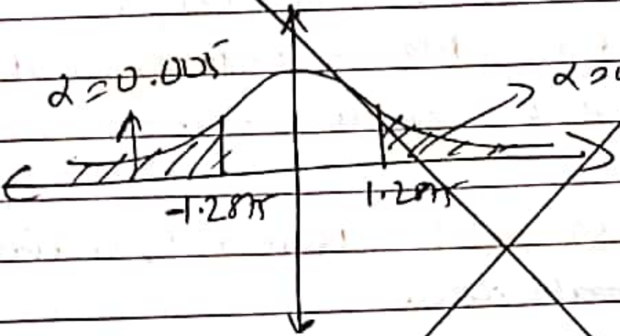
$$z = \frac{8.2 - 8.1}{\frac{0.25}{\sqrt{30}}} = 2.1908$$

Date
Page



$z_\alpha = -2.575$

rejection region : $(-\infty, -1.2875] \cup [1.2875, \infty)$



∵ given z value = 2.1905 lies in rejection region

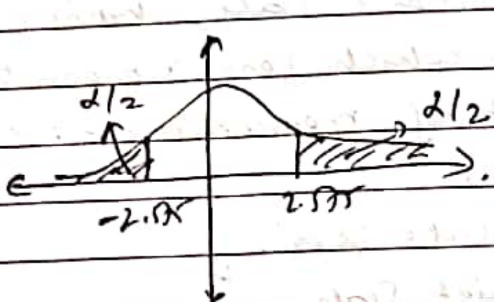
=> Null hypothesis is rejected

=> Alternative hypothesis is accepted

=> Machine should be recalibrated.

$z_{\alpha/2} = -2.575$

rejection region :- $(-\infty, -2.575] \cup [2.575, \infty)$



$z = 2.1905$

↳ H_0 is accepted. as it doesn't lie in rejection region.

∴ Machine should not be recalibrated

Origin Pro → Data analysis and graphic software.

Steps to add trendline.

Date
Page

→ Adding ~~andar~~ excel solver.

→ Descriptive statistics:- used to organize or summarize a particular set of measurements.
- It will describe set of measurements.
- Involves in organizing, displaying.

→ Inferential statistics:- Making inferences about the population from which sample was drawn.
- Test hypothesis.

→ Scales of measurement:- Nominal, ordinal, interval and ~~scale~~ ratio.

→ Nominal scale:- Categorical data and numbers used as identifiers represent nominal scale.
Ex: Yes/No, colours, jersey.
⇒ Lowest measurement level.

Ordinal scale:- ordered series or rank order of sorting data.
Ex: First, second, third rank in an exam.
Grade A, B, C, D
Sick/healthy, guilty/innocent.

Interval scale:- Numeric scale, having ~~unit interval~~ equal intervals, for which zero represents simply an additional point of measurement i.e. data can lie below zero also.
There is no "absolute zero".
Ex: Celsius/Farhenheit Scale
Measurement of sea level.

Ratio scale:- Ratio scale is also having equal intervals, but it has an absolute zero (i.e. no number

can exist below 0 (?).
Ex: Height scale, weight scale, Kelvin scale

=) Most informative scale of measurement, has properties of all the other scales.

Data Display:-
→ Stem and leaf:-
Advantages:- (i) Provide a quick display of how the data are distributed.
(ii) original data can be recovered from the stem and leaf diagram.

Disadvantages:- (i) Not practical for large data sets.

→ Frequency Histogram:-
Drawing a vertical bar for each group, whose length is number of observations in that group

=) Individual values are lost, but we know the number in each class.

Frequency (or no. of observation) vs class.

→ Relative Frequency Histogram:-
Construct a diagram, by drawing for each group, class, a vertical bar, length is the relative frequency of that group.

R.F.H is more important because labelling on the vertical axis reflects what is important.

Scanned with CamScanner

LECTURE - 7/10/22

PROBABILITY

Date ______
Page ______

1. Sample space $S = \{1, 2, 3, 4, 5, 6\}$

   Event E: an even number is rolled

   $E = \{2, 4, 6\}$

   $P(E) = \frac{3}{6} = \frac{1}{2} = 0.5.$

   Event T: number greater than two is rolled

   $T = \{3, 4, 5, 6\}$

   $P(T) = 4/6 = 2/3 = 0.67.$

2. $S = \{HH, HT, TH, TT\}.$

   $E = \{HH, TT\}$

   $P(E) = 2/4 = 1/2 = 0.5.$

   $\hookrightarrow P(E) = 1/4 + 1/4 = 1/2$

3. $P(E^c) = 1 - P(E) = 1 - 0.5 = 0.5$

   $P(T^c) = 1 - P(T) = 1 - 2/3 = 1/3 = 0.33$

   $E = \{2, 4, 6\}$; $E^c = \{1, 3, 5\}$

   $T = \{3, 4, 5, 6\}$; $T^c = \{1, 2\}$

   $S = \{1, 2, 3, 4, 5, 6\}$

4. Atleast one head will appear in five tosses of a fair coin

   Probability of Atleast one head $= 1 -$ Probability of no head

   $= 1 - \frac{1}{2^5}$

   $= \frac{31}{32}.$

   $= 0.96875$

$\frac{1}{12} + \frac{3}{12} + \frac{8}{12}$

Date
Page

5. Intersection: $A \cap B$.

$E = \{2,4,6\}$
$T = \{3,4,5,6\}$.
$E \cap T = \{4,6\}$
$P(E \cap T) = 2/6 = 1/3$.

Q.

(a) Even E: $\{2,4,6\}$
Greater than two T: $\{\cancel{2}, 3, 4, 5, 6\}$
$P(E \cap T) = 2/6 = 1/3$.

(b) $1 = \frac{1}{12} + \frac{3}{12} + 4x$

$\boxed{x = P(3) = P(4) = P(5) = P(6)}$

$\Rightarrow 4x = \frac{12-4}{12} = \frac{8}{12}$

$\Rightarrow x = \frac{2}{12} = 1/6$.

$P(4) + P(6) = \frac{1}{6} + \frac{3}{12}$

$= \frac{5}{12}$

Q. $E = \{2,4,6\}$
any three choice, mutually exclusive
$= \cancel{\{1} \ \{1,3\}, \{1,5\}, \{3,5\}, \{1\}, \{3\}, \{5\}, \phi, \{1,3,5\}$.

PROBABILITY

→ Random Experiment :- Definite outcome that cannot be predicted with certainity.
→ Sample space :- Set of all possible outcomes.
→ Event: Subset of sample space.

Date
Page

$P = 0 \rightarrow$ impossible event
$P = 1 \rightarrow$ Certain / Sure event.

LECTURE : 13/10/22

Union of Events :- $A \cup B$, word "or".

Q. $E = \{2, 4, 6\}$ E : number rolled is even
$T = \{3, 4, 5, 6\}$ T : number rolled is > than 2.
$E \cup T = \{2, 3, 4, 5, 6\}$

→ Additive rule of probability :-

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Q. (a) $E = \{4, 4\}$. $S = \{(1,1), (1,2), (1,3) \ldots \ldots (6,4), (6,1)(6,6)\}$
$P(E) = \frac{1}{36}$.

(b) No die show 4, $5 \times 5 = 25$.
$\therefore P(E) = 1 - \frac{25}{36} = \frac{11}{36}$

$$\left.\begin{array}{l} (1, \;) \rightarrow 5 \\ (2, \;) \rightarrow 5 \\ (3, \;) \rightarrow 5 \\ (5, \;) \rightarrow 5 \\ (6, \;) \rightarrow 5 \end{array}\right\} 25.$$

⇒ Atleast one die show 4's $P(A) = 1/6$.
A : Die 1 shows 4 $P(B) = 1/6$.
B : Die 2 shows 4. $P(A \cap B) = \frac{1}{36}$.

$$\therefore P(A \cup B) = \frac{1}{6} + \frac{1}{6} - \frac{1}{36}$$

$\Rightarrow P(A \cup B) = \frac{12-1}{36} = \frac{11}{36}$ (Ans).

→ ~~Conditio~~: Conditional Probability and Independent Events:

$P(A/B)$ = Probability of occurence of A after B has occured definitely.

Rule for conditional probability :-

$$\rightarrow P(A/B) = \frac{P(A \cap B)}{P(B)}$$

Q (a) A: Number rolled is five

B: number is odd

$(A \cap B) = \{5\}$.

$B = \{1, 3, 5\}$.

~~$P(A) = 1/6$~~

$P(A \cap B) = \frac{1}{6}$.

$P(B) = \frac{3}{6} = \frac{1}{2}$.

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{1/2} = 1/3.$$

(b) $\Rightarrow P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{1/6}{1/6} = 1.$

(c) A: number is odd

B: number is 5

$(A \cap B) = \{5\}$

$B = \{5\}$

$P(B) = \frac{1}{6}$.

$P(A \cap B) = 1/6$.

$\therefore P(A/B) = \frac{1/6}{1/6} = 1.$

$P(A/B) = P(A)$, $\Rightarrow$ occurence of B has no effect on likelihood of A.

→ Then $P(A \cap B) = P(A) . P(B)$

↳ the events are independent.

Hence, in a situation if we can compute all three probabilities $P(A)$, $P(B)$, $P(A\cap B)$

If $P(A\cap B) = P(A).P(B)$, then A and B are independent.

$P(A\cap B) \neq P(A).P(B)$, then A and B are dependent.

Q. $A = \{3\}$ $\quad P(A) = 1/6 \quad P(A\cap B) = 1/6$.

$B = \{1, 3, 5\}$ $\quad P(B) = 1/2$

$P(A\cap B) \neq P(A).P(B)$

$\therefore$ A & B are dependent

Q. (a) both will be positive $\quad P(A) = P(B) = 0.92$

$P(A\cap B) = P(A).P(B)$

$= (0.92)^2 = 0.8464$

independence

(b) $P(A\cup B) = P(A) + P(B) - P(A\cap B)$

$= 0.92 + 0.92 - 0.8464$

$= 0.9936$.

$\left[P(A\cap B) = P(A).P(B) = (0.92)^2 = 0.8464 \right]$

Mutually Exclusive Events, $P(A\cap B) = 0$.

→ A and B have no outcomes in common

→ For example $A - A^c$, $B - B^c$ are mutually exclusive

Date ______
Page ______

Example 18 :-

$P(A) = 0.63$ $\quad$ $P(A \cap B) = 0.27$

$P(B) = 0.34$ $\quad$ $\therefore P(A \cup B) = 0.63 + 0.34 - 0.27$

$\Rightarrow \boxed{P(A \cup B) = 0.7}$

Example 19 :-

Total no. of volunteers = 28.

(a) Speciality is medicine and speaks two or more languages

$= 2/28 = 0.0714$

(b) Speciality is medicine : A

Two or more languages : B

$P(A) = \frac{8}{28}$ $\quad$ $\therefore P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$\Rightarrow P(A \cup B) = \frac{8}{28} + \frac{6}{28} - \frac{2}{28}$

$P(B) = \frac{6}{28}$

$\Rightarrow P(A \cup B) = \frac{12}{28} = \frac{3}{7} = 0.43$

(c) Speciality is something other than medicine.

$= 1 - P(\text{speciality is medicine})$

$= 1 - \frac{8}{28} = \frac{20}{28} = 0.714$

Example 21 :-

(a) Teenager at first marriage.

$P(E) = \frac{125}{902} = 0.139$

(b) Teenager at first marriage, person is male.

~~$P(E) =$~~ $P(E/M) = \frac{P(E \cap M)}{P(M)}$

$= \frac{\frac{43}{902}}{\frac{450}{902}}$

$\Rightarrow P(E/M) = \frac{43}{450} = 0.095$

Date
Page

Example 22:-

(a) $P(H|O) = \frac{P(H \cap O)}{P(O)} = \frac{0.09}{0.11} = \frac{9}{11} = 0.818$

(b) $P(H|O^c) = \frac{P(H \cap O^c)}{P(O^c)} = \frac{0.11}{0.89} = 0.124$

(c) Yes, overweight people tend to suffer from hypertension.

Example 24:-

$P(F) = \frac{452}{902}$ $\quad$ $P(F \cap B) = \frac{82}{902}$

$P(B) = \frac{125}{902}$

$\therefore P(F \cap B) = 0.09$

$P(F) \cdot P(B) = 0.069$

$P(F \cap B) \neq P(F) \cdot P(B)$

$\Rightarrow$ The events are not independent.

Example 26:-

Specificity $\rightarrow$ test will be negative.

High specificity, low false +ve rate.

(a) P(test result +ve) = 1 − 0.89 = 0.11.

(b) Both result +ve = 0.11 × 0.11 ~~0.0121~~

= 0.0121

0.89 $\rightarrow$ test result is −ve.

Example 27:- P(Contraband detected) = ~~0.9 × 0.9 × 0.9~~

~~= 0.729~~

= 1 − P(contraband not detected)

$= 1 - P(D_1^c \cap D_2^c \cap D_3^c)$

$= 1 - P(D_1^c) \cdot P(D_2^c) \cdot P(D_3^c)$

$= 1 - (1-0.9)(1-0.9)(1-0.9)$

$\boxed{= 0.999}$ Ans.

or P(contraband detected)

$= P($contraband detected by all three dogs$)$

$+ P($contraband detected by one dog, undetected by two dogs$)$

$+ P($contraband detected by two dogs, undetected by one dog$)$

$$= P(D_1 \cap D_2 \cap D_3) + P(D_1 \cap D_2^c \cap D_3^c) \times 3 + P(D_1 \cap D_2 \cap D_3^c) \times 3$$

$$= (0.9)(0.9)(0.9) + (0.9)(0.1)(0.1) \times 3 + (0.9)(0.9)\cancel{(0.1)}(0.1) \times 3$$

$\boxed{= 0.999}$ (ans)

LECTURE: 14/10/22

Random variable :-

- Random variable denoted by 'X' or 'Z'.
- Actual values taken by random variable denoted by 'x' or 'z'.

	Experiment	Number X	Possible values
1.	Sum of dots when two die is rolled	Sum of number of dots on the top faces.	2, 3, 4, . . . , 11, 12.

→ Probability within 0 to 1, $0 \le P(x) \le 1$.

→ $\sum P(x) = 1$.

Example: Fair coin is tossed twice.

X, is Number of heads observed.

Atleast one head : $P(X \geq 1) = P(1) + P(2)$

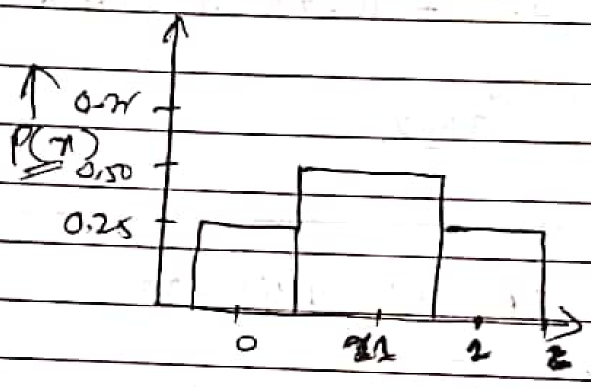
X	0	1	2
P(x)	~~1/8~~ 1/4	~~2~~ 2/4	~~3/8~~ 1/4.

(b) Probability atleast one head is observed $= 1 - 1/4 = $ ~~3/8~~

$= 3/4.$

$P(X \geq 1) = P(1) + P(2)$

$= 2/4 + 1/4 = 3/4.$



Discrete Random variable:

$\Rightarrow$ Mean of a discrete random variable $= \mu = \sum x P(x)$

Q. Mean $= -2(0.21) + 1(0.34) + 2(0.24) + (3.5)(0.21)$

$= 1.135$

$\rightarrow$ Variance of a discrete random variable

$\sigma^2 = \sum (x - \mu)^2 P(x)$

$\Rightarrow \sigma^2 = \sum x^2 P(x) - \mu^2$..

$\rightarrow$ Hence Standard Deviation,

$\sigma = \sqrt{\sum x^2 P(x) - \mu^2}$

Date
Page

Q.

x	-1	0	1	4
$P(x)$	0.2	0.5	a	0.1

$\because \sum P(x) = 1$.
$\Rightarrow a = 0.2$.

(a) $a = 0.2$
(b) $P(0) = 0.5$.
(c) $P(X>0) = P(1) + P(4) = 0.2 + 0.1 = 0.3$
(d) $P(X \geq 0) = P(0) + P(1) + P(4) = 0.8$.
(e) $P(X < -2) = 0$.
(f) $\mu = \sum x P(x) = 0.4$
(g) variance, $\sigma^2 = 0.2(1) + 0.2(1) + 0.1(16) - (0.4)^2$
$= 1.84$
(h) Standard Deviation $\sigma = 1.356$

Discrete Random Variable

Random variable: Numerical quantity generated by random experiment
Random variable denoted by capital X and Z, actual values they take denoted by 'x' or 'z'.

→ A random variable is called discrete if it has either a finite or countable number of possible values.
→ A random variable is continuous if its possible value contain whole interval of numbers.

→ Example 4, examples.

LECTURE: 17/10/22

: Binomial Distribution:

⟹ n identical and independent trials.
Two outcomes, success and failure
Probability of success = p.
Probability of failure = q.
We say that a binomial distribution X, with parameters 'n' & 'p' ~~q~~.

1. $n = 125$
Proportion of student female = 0.57
X: number of female student in sample.
Success - student is female.
$\therefore p = 0.57$.
$\therefore$ X is a binomial distribution with $n = 125, p = 0.57$

⟹ If 'X' is a binomial random variable with parameters 'n' and 'q' then
x: actual value taken by the random variable X.

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

$$q = 1-p.$$

$$\therefore \boxed{P(x) = {}^nC_x \, p^x q^{n-x}} \quad (q = 1-p)$$

Example 7:
Q1. It is a binomial distribution, since there is a success and failure.
$p = 0.17$
$q = 0.83$.

Q2. $n = 5, p = 0.17$.

$$P(x) = \frac{5!}{x!(5-x)!} (0.17)^x (0.83)^{5-x}$$

Scanned with CamScanner

Mean = np.
S.D. = $\sqrt{npq}$

$\therefore P(0) = 0.3939$
$P(1) = 0.4033$
$P(2) = 0.1652$
$P(3) = 0.0338$
$P(4) = \cancel{0.4024}$ 0.0035
$P(5) = 0.0001$

x	0	1	2
$P(x)$	0.3939	0.4034	0.1652

3	4	5
0.0338	0.0035	0.001

(b) most frequent cases, $X = 1$.
Most frequent number of cases, seen each day in which the victim knew the prepetrator is one.

(c) $\mu = \sum x P(x)$
$= 0 + 0.4033 + 2 \times 0.1652 + 3 \times 0.0338 + 4 \times 0.0035 + 5 \times 0.001$

$= 0.8496$

→ Mean = np = $0.17 \times 5 = 0.85$
S.D. = $\sqrt{5 \times 0.17 \times 0.83} = 0.8399$.

⇒ If X is a ~~random~~ binomial random variable, with parameter 'n' and 'q' then,
→ $\mu = np$, $\sigma^2 = npq$, $\sigma = \sqrt{npq}$
$q = 1 - p$.

⇒ Cumulative Probability Distribution
$P(X \le x) = P(0) + P(1) + P(2) + \dots + P(x)$

Example 9:-
Q. Question right → when it is 'true'
p: Probability that a student guessed 'true'
$\therefore p = 1/2$
$q = 1/2$

$n = 10, p = 1/2, q = 1/2$.

(a) $\therefore P(6) = \cancel{{}^{10}C_6 (p)^{1/2} (q)^{1/2}}$. ${}^{10}C_6 \left(\frac{1}{2}\right)^{10-6} \left(\frac{1}{2}\right)^6$

$\Rightarrow P(6) = {}^{10}C_6 \left(\frac{1}{2}\right)^{10}$

$\therefore \boxed{P(6) = 0.2051}$

(b) $P(X \geq 6) = P(6) + P(7) + P(8) + P(9) + P(10)$

$= \left({}^{10}C_6 + {}^{10}C_7 + {}^{10}C_8 + {}^{10}C_9 + {}^{10}C_{10}\right) \times \left(\frac{1}{2}\right)^{10}$

$= 0.3769$

Example 10:

Q. $p = 1/3, n = 5$.

(a) $P(X \leq 1) = P(0) + P(1)$

$= {}^5C_0 \left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^5 + {}^5C_1 \left(\frac{1}{3}\right)^1 \left(\frac{2}{3}\right)^4$

$= \boxed{0.4609}$

(b) $P(X \leq n) = P(0) + P(1) + P(2) + \dots + P(n)$

$P(X \leq n) \geq 0.95$.

For $n = 2$, $P(X \leq 2) = 0.7901$

$n = 3$, $P(X \leq 3) = 0.9547$

→ he has enough of the day's practice i.e. atmost condition

$\therefore$ $\boxed{\text{minimum } 3}$

Variance:- $\sigma^2 = x^2 P(x) - \mu^2$

Standard Deviation:- $\sigma = \sqrt{x^2 P(x) - \mu^2}$

Date ______
Page ______

In case of a binomial random variable i.e. discrete random variable,

mean $= \mu = np$

variance $= \sigma^2 = npq$

Standard Deviation $= \sigma = \sqrt{npq}$. $(q = 1-p)$.

Atleast means $P(X \geq x)$

Atmost $P(X \leq x)$

LECTURE :- 3/11/22. ANOVA (Analysis of Variance)

→ Single factor ANOVA: involves comparison of k population or treatment means μ_1, μ_2, μ_3. The objective is to test whether all the k population means are the same.

$H_0: \mu_1 = \mu_2 = \dots = \mu_k$

against, H_a: Atleast two of the μ's are different.

⇒ ANOVA Notation:

K = number of population or treatments being compared

(i) Population or treatment	1	2	....	K
(ii) Population or treatment mean	μ_1	μ_2	......	μ_k
(iii) Population or treatment variance	σ_1^2	σ_2^2	......	σ_k^2
(iv) Sample size	n_1	n_2	......	n_k
(v) Sample mean	$\bar{x}_1$	$\bar{x}_2$	......	$\bar{x}_k$
(vi) Sample variance	s_1^2	s_2^2	......	s_k^2

Scanned with CamScanner

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

Date
Page

→ Combined sample size.

$N = n_1 + n_2 + \cdots + n_k$ (the total number of observation in the data set).

T = grand total = sum of all N observations
$= n_1\bar{x}_1 + n_2\bar{x}_2 + \cdots + n_k\bar{x}_k$.

$\bar{\bar{x}}$ = grand mean $= \frac{T}{N} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \cdots + n_k\bar{x}_k}{n_1 + n_2 + n_3 + \cdots + n_k}$

⇒ Assumptions for ANOVA:

1. Each of the population or treatment response distribution is normal.
2. $\sigma_1 = \sigma_2 = \cdots = \sigma_k$ (The 'K' normal distribution are identical standard deviation)
3. The observations in the sample from any particular one of the k population or treatments are independent of one another.
4. When comparing population mean, 'K' random samples are selected independently of one another.

⇒ Test procedure of ANOVA:

(a) SSTr (Treatment sum of squares): measure of disparity among sample means.

$$SSTr = n_1(\bar{x}_1 - \bar{\bar{x}})^2 + n_2(\bar{x}_2 - \bar{\bar{x}})^2 + \cdots + n_k(\bar{x}_k - \bar{\bar{x}})^2.$$

(b) SSE (Error sum of squares): measure of variation within the 'K' samples.

~~$SSE = n_1$~~

$$SSE = (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_k - 1)s_k^2.$$

$$\boxed{SST_0 = SSTr + SSE}$$

One way = taking one random sample from each population under consideration.

Date
Page

Each sum of squares has an associated df:

$df_1 =$ Treatment df $= K-1$

$df_2 =$ Error df $= N-K$

} df: degree of freedom.

A mean square is sum of square divided by its df.

mean square for treatments $= MST_r = \frac{SST_r}{K-1}$

mean square for error $= MSE = \frac{SSE}{N-K}$

Number of error degrees of freedom comes from adding the number of degree of freedom associated with each of the sample variance:

$(n_1-1)+(n_2-1)+(n_3-1)+\ldots+(n_k-1)$

$=(n_1+n_2+n_3+\ldots+n_k)-K$

$= N-K.$

⇒ Simple ANOVA Factor Table :-

Source of Variation	df	Sum of Squares	means square	F
Treatment	$K-1$	SST_r	$MST_r = \frac{SST_r}{K-1}$	$F = \frac{MST_r}{MSE}$
Error	$N-K$	SSE	$MSE = SSE/(N-K)$	
Total	$N-1$	$SSTo$		

Test statistic, for testing null hypothesis that K population means are equal:

$$\boxed{F = \frac{MST_r}{MSE}}$$

The test is right tailed: H_0 is rejected at level of significance 'α' if $F \geq F_\alpha$.

LECTURE :- 4/11/22

$$F = \frac{MST_r}{MSE} \qquad df_1 = K-1$$
$$df_2 = N-K$$

F distribution always right tailed

$$\boxed{SST_0 = SST_r + SSE}$$

Example: 5% level of significance

Mathematics	English	Education	Biology
2.59	3.64	4.00	2.78
3.13	3.19	3.59	3.51
2.97	3.15	2.80	2.65
2.50	3.78	2.39	3.16
2.53	3.03	3.47	2.94
3.29	2.61	3.59	2.32
2.83	3.20	3.74	2.58
3.17	3.30	3.77	3.21
2.70	3.54	3.13	3.23
3.88	3.25	3.00	3.57
2.64	4.00	3.47	3.22
$\bar{x}_1 = 2.903$	$\bar{x}_2 = 3.335$	$\bar{x}_3 = 3.359$	$\bar{x}_4 = 3.015$

$n_1 = n_2 = n_3 = n_4 = 11$, $K = 4$

$n = 11 + 11 + 11 + 11 = 44$

$\therefore df_1 = K - 1 = 4 - 1 = 3$

$df_2 = 44 - 4 = 40$

$df_1 = 3$
$df_2 = 40$

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

$$\therefore S_1^2 = \frac{1.8806}{10} = 0.1881$$

$$S_2^2 = 0.1475$$

$S_3^2 = 0.228$ $\qquad$ $T = 138.732$; ~~A~~ $N = 44$

$S_4^2 = 0.1566$ $\qquad$ $\bar{\bar{x}} = 3.152$

$\therefore$ ~~MSTr = 10x~~

$SSTr = 11(0.157) = 1.728 \Rightarrow MSTr = 0.576$

$SSE = 7.202 \Rightarrow MSE = 0.1801$

$$\therefore f = \frac{0.576}{0.1801} = 3.198.$$

$df_1 = 3,\ df_2 = 40,\ \alpha = 0.05,\ f_\alpha = \cancel{2.8387}\ 2.8387.$

$\therefore f \geq f_\alpha \Rightarrow$ $\boxed{\text{Ho, null hypothesis is rejected.}}$

Example:

$k = 3$ $\qquad$ $\alpha = 0.01$

$N = 33$

Treatment 1	Treatment 2	Control group
71	77	81
72	67	79
75	79	73
80	78	71
60	81	84
65	72	77
63	71	62
78	84	75
75	91	
73		
72		
66.		
63		

65		
64		
71		
$n_1 = 16$	$n_2 = 9$	$n_3 = 8$

$df_2 = N - K = 30$
$df_1 = K - 1 = 3 - 1 = 2$

$\bar{x}_1 = 69.75$
$\bar{x}_2 = 77.78$
$\bar{x}_3 = 75.88$

$T = 2423.06$
$N = 33$
$\Rightarrow \bar{x} = 73.4261$

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

$\therefore S_1^2 = 517/15 = 34.46$
$S_2^2 = 421.56/8 = 52.69$
$S_3^2 = \frac{214.8752}{7} = 30.696$

$SST_r = 16(69.75 - 73.4261)^2 + 9(77.78 - 73.4261)^2 + 8(75.88 - 73.4261)^2$

$\Rightarrow SST_r = 435$

$\therefore MST_r = \frac{435}{2} = 217.5$

$SSE = 15 \times 34.46 + 8 \times 52.69 + 7 \times 30.696$

$\Rightarrow SSE = 1153.43$

$MSE = \frac{1153.43}{30} = 38.448$

$\therefore f = \frac{217.5}{38.448} = 5.656$

$\therefore f = 5.656$

$f_\alpha = 5.350$ $\alpha = 0.01,\ df_1 = 2,\ df_2 = 30$

$\therefore f > f_\alpha \Rightarrow$ Null hypothesis, H_0 is rejected.

LECTURE - 11/11/22

Non Parametric test ⇒ Sign Test is a non parametric test that is used to test whether or not two group ~~are having same age~~ are equally sized. Also called binomial sign test.

Assumption: assume data is normally distributed

(1) Data distribution :- sign test is non parametric test, don't.

(2) Two sample :- Data should be from two samples.

(3) ~~Dep~~ ~~Dat~~ Dependent sample :- ~~Dep~~ Dependent sample should be a paired sample or matched. It is 'before - after' sample.

→ One of easiest non parametric test is sign test.

⇒ Types of sign test :-

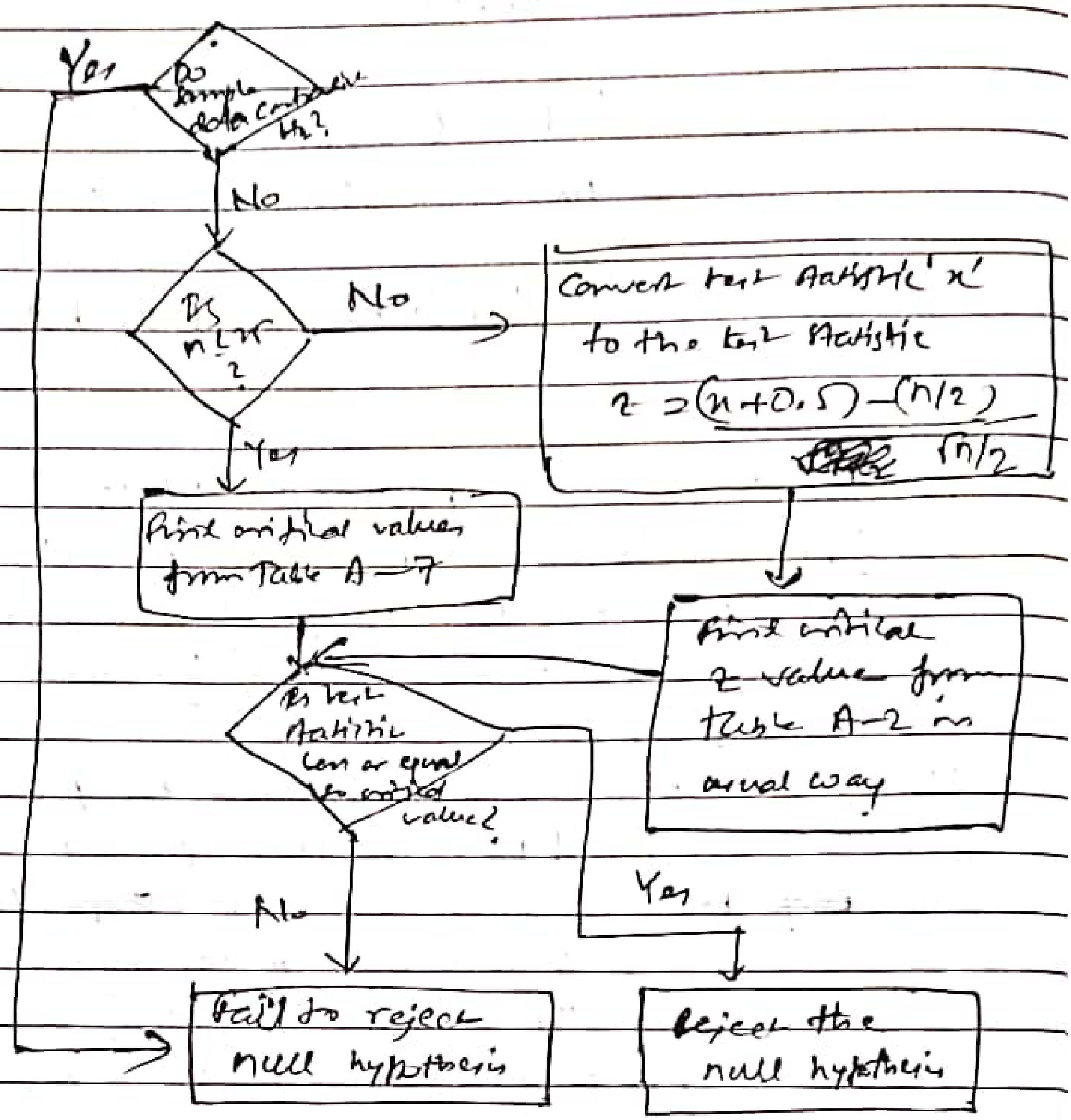
→ One sample : '+' and '-' signs are values of random variable having equal size

→ Paired sample : Alternative to paired 't'-test.

⇒ Basic concept of the sign test :- The basic idea is to analyze the frequencies of plus and minus sign to determine whether they are significantly different.

$\boxed{\text{~~Sign test procedure~~}}$

Sign Test Procedure:



⇒ **Requirements:**

(i) Sample data have been randomly selected.

(ii) No requirment that the sample data come from a population with a particular distribution, such as normal distribution.

⇒ **Notation of sign test:**

✓ x = number of times the less frequent sign occurs.

✓ n = total number of positive and negative signs combined.

Date ______
Page ______

→ Test Statistic:

$n \le 25$ — x: No. of times the less frequent sign occurs.

$n > 25$ — $$z = \frac{(x+0.5) - n/2}{\frac{\sqrt{n}}{2}}$$

→ Critical value:

$n \le 25$) critical x value in table A-7

$n > 25$, critical z value in table A-2.

⇒ Key concept underlying the use of sign test:

→ Two sets of data have equal medians, number of positive sign approximately equal to number of negative sign.

Example:

1. $n = 11 \le 25$, significance level 0.05, Two tailed test. $H_0: \mu_1 = \mu_2$.

Regular	Kiln dried	Difference
1903	2009	-'ve
1935	1915	+
1910	2011	-
2496	2463	+
2108	2180	-
1961	1925	+
2060	2122	-
1444	1482	-
1612	1542	+
1316	1443	-
1511	1535	-'ve

7 -'ve
4 +'ve

⇒ 7 -'ve
4 +'ve

Null Hypothesis :- There is difference between yeild from regular and Kiln dried seed.

Date ______
Page ______

A's

$\because$ since the sample size is (25,

$x =$ no. of times less frequent sign

$\Rightarrow x = 4$. [$\because$ +'ve nos are 4 only]

$\therefore \boxed{x = 4}$

Critical value of x, when level of significance $= 0.05$, is $x_{0.025}$ ~~$x_{0.05}$~~ $= 1$ ($\alpha/2 = 0.025$)

{ $n = 11$, two tailed $\therefore$ [$\because H_0: \mu = \mu_0$, $H_a: \mu \neq \mu_0$]
$\alpha/2 = 0.05/2 = 0.025$

~~$\therefore x < x_{0.025}$ $\Rightarrow$ Null hypothesis is rejected.~~

$\because x > x_{0.025} \Rightarrow$ fail to reject null hypothesis.

2. Out of 325 types of crops, 295 are of one type. Use the sign test at 0.05 significance level to test the claim that the method, used to identify all the crops has no effect on the selection of type of crops.

A's

$n = 325$

$x = 325 - 295 = 30$.

H_0: Method to identify the crops has effect on selection of type of crops.

H_a: Method doesn't have effect on selection

$$z = \frac{(30+0.5) - \frac{325}{2}}{\sqrt{\frac{325}{2}}}$$

$$\Rightarrow z = \frac{-132}{9.015} = -14.64$$

$$\boxed{z = -14.64}$$

$\alpha/2 = 0.025$

$z_{crit} = z_{0.025} = -1.96$

$\therefore$ ~~$z > z_\alpha$~~ $z < z_\alpha$

$\Rightarrow$ Null hypothesis rejected.

$\Rightarrow$ Doesn't have effect.

Non parametric Test :- A non parametric test is a hypothesis test that doesn't require any specific conditions concerning the shape of population or value of any population parameters.

- Non parametric test don't require normally distributed populations.
- Applied to categorical data.
- less efficient than parametric test. Stronger evidence required to reject null hypothesis.

Ex: Sign Test / Binomial Sign test.

$\Rightarrow$ Sign test is a weak test,

Paired T-Test :-

LECTURE :- 14/11/22

⇒ Paired T-test : Compare two ~~pat~~ population mean where we have two samples, in which observation in one sample can be paired with observation in the other sample.

d :- before + after = post − pre

(i) x = before the module
y = after the module
Difference, $d_i = y_i - x_i$ = After − Before

(ii) mean difference, $\bar{d}$.

(iii) S.D. of the difference = s_d
Standard error of the mean difference, $SE(\bar{d}) = \frac{s_d}{\sqrt{n}}$

(iv) Calculate the value of T using,

$$T = \frac{\bar{d}}{SE(\bar{d})}$$

$(n-1)$ degree of freedom.

(v) Compare T and t_{n-1}, this will give p value for paired t-test.

⇒ Approximately normally distributed, for test to be valid.

p-value : probability value.
Probability of occurrence of the event.

p-value small → Null hypothesis is rejected
Alternative hypothesis accepted.

→ Similarity between paired and unpaired t-test :-
Data in normal distribution.

Unpaired T-test	Paired t-test
1. Two groups independent	1. Dependent
2. Sample size not be equal	2. Data taken from subjects that have been measured twice.
3. Compares mean data of the two groups	

$$S = \sqrt{\frac{\Sigma d^2 - n(\bar{d})^2}{n-1}} = \sqrt{\frac{\Sigma (d-\bar{d})^2}{n-1}}$$

Date
Page

Student	Pre module (x_i)	Post module (y_i)	$d_i = y_i - x_i$
1	18	22	+4
2	21	25	+4
3	16	17	+1
4	22	24	+2
5	19	16	-3
6	24	29	+5
7	17	20	+3
8	21	23	+2
9	23	19	-4
10	18	20	+2
11	14	15	+1
12	16	15	-1
13	16	18	+2
14	19	26	+7
15	18	18	0
16	20	24	+4
17	12	18	+6
18	22	25	+3
19	15	19	+4
20	17	16	-1

$\Sigma d = +41$

$\therefore \bar{d} = \frac{+41}{20} = +2.05$ $\quad \Sigma d^2 = 237$

$$S_d = \sqrt{\frac{237 - 20(-2.05)^2}{19}}$$

$$= \sqrt{8.05} = 2.837$$

$$\therefore T = \frac{+2.05}{\frac{2.837}{\sqrt{20}}} = +3.232$$

~~$\therefore T = +3.232$~~

~~$\therefore \frac{3.23 - 3.20}{3.30 - 3.20} = P = 0.002$~~

~~$\therefore T = +3.32$~~

$\therefore T = +3.232$

Degree of freedom = $n-1 = 20-1 = 19$

for d.f. = 19

t_{n-1}	p
3.20	0.002
3.30	0.002

$\therefore$ when $T = +3.232$, $p = 0.002$

p value Table:

p-value	Decision
(i) p-value > 0.05	result not ~~significantly~~ statistically significant, ~~don't~~ don't reject null hypothesis.
(ii) pvalue < 0.05	result statistically significant, reject null hypothesis
(iii) p value ~~< 0.05~~ < 0.01	Highly statistically significant reject null hypothesis

One way ANOVA

LECTURE 1 - 17/11/22

0 mg	50 mg	100 mg
9	7	4
8	6	3
7	6	2
8	7	3
~~8~~ 8	8	4
9	7	3
8	6	2

$k = 3$

$N = 7 \times 3 = 21$

$\alpha = 0.05$

$df_2 = N - k = 18$

$df_1 = k - 1 = 2.$

→ $H_0 : \mu_0 = \mu_{50} = \mu_{100}$ → H_a : all are not equal.

$\bar{x}_1 = 57/7 = 8.14$

$\bar{x}_2 = 47/7 = 6.71$

$\bar{x}_3 = 21/7 = 3$

$T = 125$ ∴ $\bar{\bar{x}} = \frac{125}{21} = 5.952$

$N = 21$

∴ $SSTr = 98.5335$

$s_1^2 = \frac{\sum (x_i - \bar{x}_i)^2}{n-1} = 0.4762$

$s_2^2 = 0.5715$

$s_3^2 = 4/6 = 0.666$

$SSE = 10.2858$

→ $MSTr = \cancel{97.8444}$ 49.2667

$MSE = 0.5714$

$\boxed{F = 86.2211}$

$F_{\alpha} = 3.8846$ $(\alpha = 0.05)$

$\therefore$ $F > F_{\alpha}$ → Null hypothesis is rejected.

Minitab.

⇒ One way ANOVA :
→ Response data in different columns ⇒ 0mg, 5mg, 100mg
→ Confidence interval 95%.
→ Graphs : Interval plot.
Individual value plot.
Box plot
Histogram plot.

P value $\approx$ 0 ⇒ Null hypothesis rejected

⇒ Wilcoxon Signed Rank Test :
→ Popular non parametric test, for matched or pair data i.e. dependent data.
→ Like the sign test it is based on different scores, but in addition to analyzing the sign of differences, it also takes into account the magnitude of the observed difference.

$$Z = \frac{|R - \text{mean}|}{\text{Standard Deviation}}$$

R = Sum of ranks (~~...~~ smaller, larger, or any one).
n = Number of samples (After excluding 0 difference)

Non parametrical; d = Before − After

$$\text{Mean} = \frac{n(n+1)}{4}$$

$$S.D. = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

Q. A drug is given to 12 patients and difference in their BP were recorded to be :-

Before Drug A	After drug A
112	116
113	120
~~119~~ 118	117
~~115~~ 120	125
~~116~~ 119	126
~~117~~ 113	111
~~118~~ 110	111
~~119~~ 122	117
126	126
115	112
119	129

Use the Wilcoxon signed rank test to test the hypothesis that ~~has~~ the drug has no effect on change on BP?

Ans H_0 : Drug A doesn't effect BP.

H_a : Drug A affects BP.

Before	After	d	|d|	R
112	116	−4	4	5
113	120	−7	7	8.5
118	117	1	1	1.5
120	125	−5	5	6.5
119	126	−7	7	8.5
113	111	2	2	3
110	111	−1	1	1.5
122	117	+5	5	6.5
126	126	0	0	—

115	112	3	3	4
119	129	-10	10	10

$R^+ = 1.5 + 3 + 4 + 6.5 = 15$

R^- = rank of -'ve sign number

$= 8 + 8.5 + 6.5 + 8.5 + 1.5 + 10$

$= 40.$

$n = 10 \rightarrow$ n is value of top rank.

$$\text{mean} = \frac{10 \times (10+1)}{4} = \frac{10 \times 11}{4} = 27.5$$

$$S.D. = \sqrt{\frac{10 \times 11 \times 21}{24}} = 9.811$$

$$z = \frac{|40 - 27.5|}{9.811} = \frac{12.5}{9.811} = 1.274$$

$\therefore \boxed{z_{cal} = 1.274}$

$\alpha = 0.05$, $z_{crit} = 1.96$.

$z_{0.05} = 1.96$.

~~$z_{tab} > z_{cal}$~~

LECTURE: 18/11/22

Q.

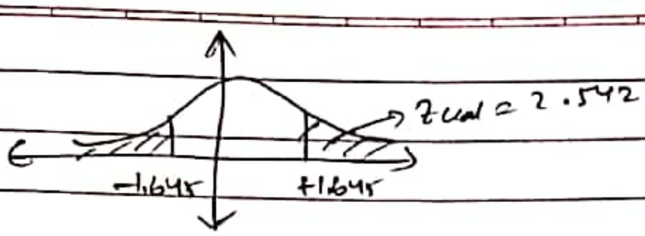
a: $H_0 : \mu = 48,750$ $\alpha = 0.1$

$H_a : \mu \neq 48,750$ $\frac{\alpha}{2} = 0.05$

(i) $$z_{cal} = \frac{x - \mu}{\sigma/\sqrt{n}} = \frac{51,505 - 48,750}{\frac{6872}{\sqrt{40}}}$$

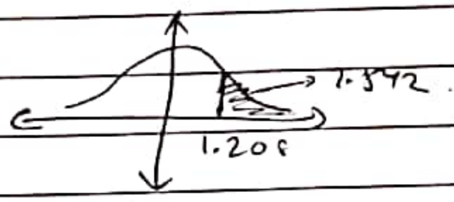
$$\Rightarrow z_{cal} = \frac{2755 \times \sqrt{40}}{6872} = 2.542$$

$z_{0.05} = -1.645$.



⇒ Null hypothesis is rejected.
⇒ $\mu \neq 48750$.

(d)) $H_0: \mu = 48,750$
$H_a: \mu > 48,750$.
$\alpha = 0.1$
$Z_{cal} = 2.542$
$Z_{0.1} = -1.2088$



⇒ Null hypothesis is rejected
⇒ $\mu > 48750$.