

Ronald A. Fisher → Father of Statistics.

ANOVA → Analysis of variance.

Data → Quantitative (Numerical)

Qualitative (Non Numerical)

Sample → Population

\bar{x}

Parameter μ

Estimated for sample s^2

s

Estimated for population σ^2

\hat{s}

Estimated for population σ

P

Estimated for population 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.

$\bar{x} = \text{mean}$

N Data

$\sum x_i = \text{total}$

x_1, x_2, \dots, x_N

Qualitative.

Quantitative.

(Numerical)

(Non numerical)

$\frac{\sum x_i}{N}$

S_1

S_2

P

P_1

Nominal

ordinal.

Discrete Continuous.

$\bar{x} = \frac{1}{N} \sum x_i = \bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_N$

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

Properties of measurement

- (1) Identity :- Each value has unique meaning.
- (2) Magnitude :- Specific order to the variable.

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

Nominal → Can be ordered or not ordered,
Ordinal → ordered

Fathers - 1000000000

Interval → Data shows both properties of nominal
and ordered.

→ Shows both order of variables and exact difference.

Ratio → Includes properties of all four scale of measurement.

LECTURE 4/8/20

Measures of central tendency:

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

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

x	f	$f \cdot x$	Mean = $\frac{\sum f \cdot x}{\sum f}$
0	3	0	
1	6	6	
2	6	12	$= \frac{31}{19}$
3	3	9	
4	1	4	$= 1.63$
$\sum f = 19$		$\sum f \cdot x = 31$	

Median: \tilde{x}

↳ odd number of measurement → middle no.

→ even number of measurement → mean of)

~~or recent suggestion) and write about too middle no.~~

~~descending~~ ~~What about Missy?~~ ~~It's~~ ~~stuck~~ ~~up~~ ~~in~~ ~~the~~ ~~air~~ ~~now~~ ~~she~~ ~~is~~ ~~ascending~~ ~~descending~~

mean = median \rightarrow Symmetric / symmetric-skewed.

mode \rightarrow 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

$$\text{Pf} = \frac{\sum f}{n} = \frac{11}{6} = 1.83 \text{ approx}$$

$$\text{Mean} = \frac{456}{12} = 41.54.$$

Median:

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

$$\text{median} = 25.$$

mode = 21 (since it has highest frequency).

Measures of variability or dispersion

\rightarrow range

\rightarrow variance

\rightarrow Standard Deviation,

Data set 1: 28, 39, 39, 39, 40, 40, 40, 40, 41, 43

$$\text{mean} = \frac{410}{10} = 41.0$$

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

$$\text{mode} = 40.$$

Data set 2: 33, 33, 36, 37, 40, 40, 41, 45, 46, 47.
 $\text{mean} = 40 = \text{median} = \text{mode}$

$$\text{Range} = X_{\max} - X_{\min}$$

X_{\max} is largest measurement observed.

X_{\min} is smallest measurement observed.

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

$$\text{range of data set 1} = 43 - 38 = 5$$

$$\text{range of data set 2} = 48 - 33 = 15$$

$$\Rightarrow \text{Sample variance} = S^2 = \frac{\sum (x - \bar{x})^2}{n-1}$$

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

$$\Rightarrow \text{sample standard deviation}$$

$$\text{Standard deviation} = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum x^2 - \frac{1}{n} (\sum x)^2}{n-1}}$$

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

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

Sample $\rightarrow n-1$
 Population $\rightarrow N$

$\rightarrow 1$ because we don't consider mean.



Data set 1:

$$\begin{array}{c} \overline{x} \\ x - \bar{x} \\ (x - \bar{x})^2 \end{array}$$

38 $\rightarrow -2$ $\rightarrow 4$

39 $\rightarrow -1$ $\rightarrow 1$

39 $\rightarrow -1$ $\rightarrow 1$

40 $\rightarrow 0$ $\rightarrow 0$

42 $\rightarrow 2$ $\rightarrow 4$

43 $\rightarrow 3$ $\rightarrow 9$

$$\sum (x - \bar{x})^2 = 20$$

$\rightarrow \text{standard deviation } \sigma = \sqrt{\frac{20}{10}} = 1.414$

OP OUT OF 100 $\rightarrow S^2 = 20 / 10 = 2.00$ (variance)

OP OUT OF 100 $\rightarrow S^2 = 20 / 9 = 2.22$ (variance)

OP OUT OF 100 $\rightarrow S = \sqrt{2.22} = 1.491$ (S.D.)

Data set 2:

$$\begin{array}{c} \overline{x} \\ x - \bar{x} \\ (x - \bar{x})^2 \end{array}$$

33 $\rightarrow -7$ $\rightarrow 49$

34 $\rightarrow -6$ $\rightarrow 36$

36 $\rightarrow -4$ $\rightarrow 16$

37 $\rightarrow -3$ $\rightarrow 9$

40 $\rightarrow 0$ $\rightarrow 0$

40 $\rightarrow 0$ $\rightarrow 0$

42 $\rightarrow 2$ $\rightarrow 4$

45 $\rightarrow 5$ $\rightarrow 25$

46 $\rightarrow 6$ $\rightarrow 36$

47 $\rightarrow 7$ $\rightarrow 49$

$$\sigma = \sqrt{224} \rightarrow 14.9$$

$$\sigma = \sqrt{224} \rightarrow 14.9$$

$$S^2 = \frac{224}{9} \rightarrow 24.8$$

$$S = \sqrt{24.8} \rightarrow 4.973$$

$$\sum (x - \bar{x})^2 = 224$$

DATA ANALYSIS

Date _____
Page _____

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

Discrete random variable.

Continuous random variable.

LECTURE 18/11/22



Three Popular Data Analysis'

(1) Stem and Leaf Diagram.

(2) Frequency Histogram

(3) Relative Frequency Histogram

(4) Dotplot

} Numerical data.

Quantitative
Data

Mark of 30 students

86	80	25	77	73	71	100	90	65
90	83	70	73	73	70	90	87	71
40	58	68	13	69	100	78	87	97
								92

→ 2 → stem (ten's place).

→ unit digit (leaf)

(1) Stem and Leaf

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

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

213

(1) 2 stem
13 leaf

(4) 21 stem
3 leaf

Histogram → Quantitative.

Date _____
Page _____

(3)

R.R. Histogram :-

80s → 5

Total 30 students

∴ 80s → 5/30 (relative frequency).

70s → 10/30 = 1/3 = 33.3%

90s → 7/30 = 23.3% (approx.)

100s → 2/30 = 1/15 = 6.6% (approx.)

60s → 3/30 = 10% (approx.)

50s → 1/30 = 3.33%

40s → 3.33%

20s → 3.33% (approx.)

⇒ Dotplot

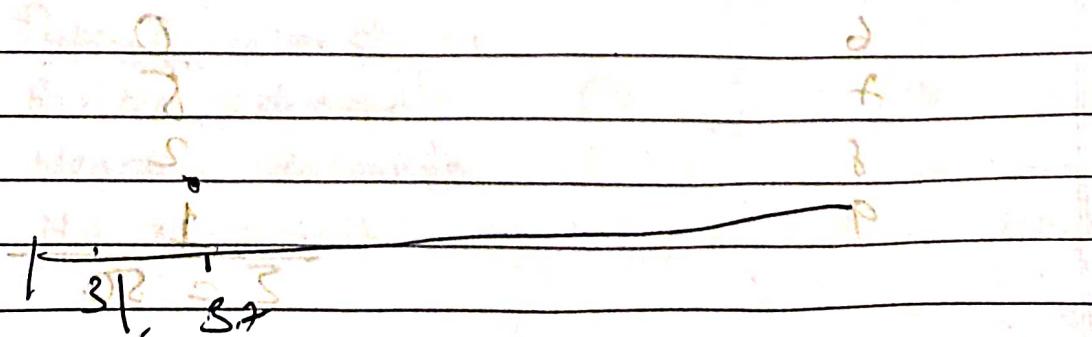
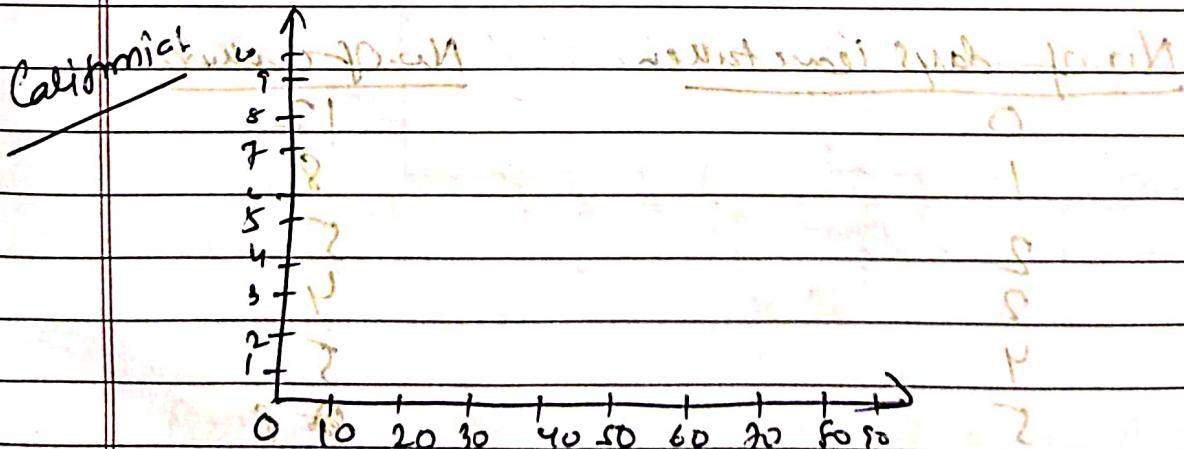
California → 20 Division I schools.] % of full

Texas → 19 Division II schools.] time freshmen

California: 64, 41, 44, 31, 37, 73, 72, 68, 35, 37, 81, 70,

Texas: 67, 21, 32, 85, 38, 37, 39, 35, 71, 82, 84, 79, 67, 66,

14, 7, 50, 63, 82, 45, 35, 29, 25, 65, 25, 24, 22, 68, 70, 53.



22/8/22

Date _____
Page _____

Qualitative / Categorical Data:

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

(number of entries) $0.52 \rightarrow 108$

\Rightarrow Relative frequency distribution

Categorical \rightarrow Bar chart / Pie chart

Quantitative \rightarrow Histogram

used to
display
numerical
data

Bar chart: height of bar \propto ref.

Pie chart: size of slice \propto ref.

Slice size \approx Block \times 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, 2, 1, 5, 4, 0, 4, 0, 1, 8, 9, 7, 0, 1, 7, 0, 5, 5, 4, 3, 3, 2, 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
	$\sum = 80$

(a) After 1 day = $72 - 18 = 200 - 38 = 162$

(b) Between 3 & 5 day leave = $5 + 8 + 4 = 17$.

(c) More than 5 days of sick leave = 8

- Symmetric Data.

- Almost symmetric Data.

- No symmetry.

LECTURE, 25/8/22

* Correlation: Relationship between two variables.

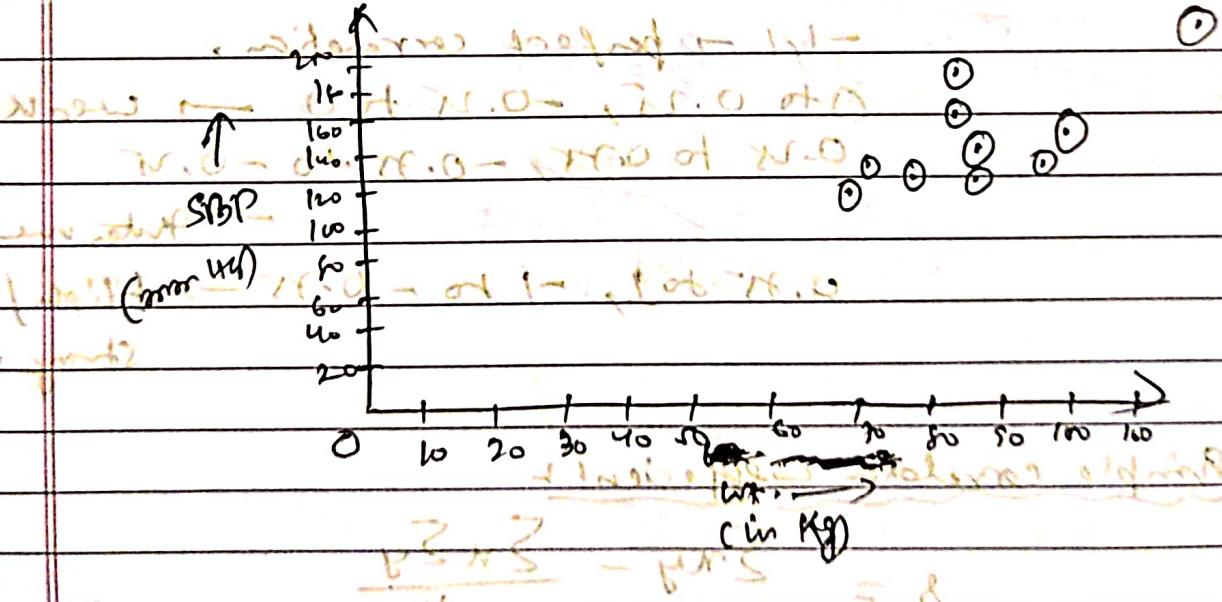
(x, y)

→ Scatter diagram for correlation.

• 1st 1 - weighted mean 'x' is median

SBP (mm Hg) n wt. (kg)

• Values on \rightarrow Y



→ Pattern of data is indicative of type of relationship.

Type of relationship

- (1) Positive relationship ($x \uparrow, y \uparrow$) / Direct relationship
- (2) Negative relationship ($x \downarrow, y \uparrow$) / Inverse relationship
- (3) No relationship. (full scattered data).

Correlation coefficient Parameter showing degree

$r = \frac{\text{PT}}{\sqrt{n}}$ r = $\frac{\text{sign of relation}}{\text{strength of relation}}$

1. Simple correlation coefficient (r) \Rightarrow Pearson's correlation

or product moment

Product moment correlation coefficient

Sign of r is +ve \Rightarrow direct relationship
(↑↑, ↓↓)

Sign of r is -ve \Rightarrow 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 0.5, -0.5 to -0.25 → moderate

0.5 to 1, -1 to -0.5 → High

Strong.

→ Simple correlation coefficient

$$r = \frac{\sum xy - \bar{x}\bar{y}}{n}$$

$$\sum xy = (\sum x)^2 (\sum y)^2$$

Intercorrelation coefficient (r_{xy})

Dimensionless coefficient

Product moment coefficient (r_{xy})

Dimensionless coefficient

Other measures of r

Dimensionless coefficient

	<u>Age (x)</u>	<u>weight (y)</u>	<u>Σxy</u>	<u>Σx^2</u>	<u>Σy^2</u>
21	18	55	84	35	10
22	19	51	85	36	11
23	20	50	86	37	12
24	21	49	87	38	13
25	22	48	88	39	14
26	23	47	89	40	15
27	24	46	90	41	16
28	25	45	91	42	17
29	26	44	92	43	18
30	27	43	93	44	19
31	28	42	94	45	20
32	29	41	95	46	21
33	30	40	96	47	22
34	31	39	97	48	23
35	32	38	98	49	24
36	33	37	99	50	25
37	34	36	100	51	26
38	35	35	101	52	27
39	36	34	102	53	28
40	37	33	103	54	29
41	38	32	104	55	30
42	39	31	105	56	31
43	40	30	106	57	32
44	41	29	107	58	33
45	42	28	108	59	34
46	43	27	109	60	35
47	44	26	110	61	36
48	45	25	111	62	37
49	46	24	112	63	38
50	47	23	113	64	39
51	48	22	114	65	40
52	49	21	115	66	41
53	50	20	116	67	42
54	51	19	117	68	43
55	52	18	118	69	44
56	53	17	119	70	45
57	54	16	120	71	46
58	55	15	121	72	47
59	56	14	122	73	48
60	57	13	123	74	49
61	58	12	124	75	50
62	59	11	125	76	51
63	60	10	126	77	52
64	61	9	127	78	53
65	62	8	128	79	54
66	63	7	129	80	55
67	64	6	130	81	56
68	65	5	131	82	57
69	66	4	132	83	58
70	67	3	133	84	59
71	68	2	134	85	60
72	69	1	135	86	61
73	70	0	136	87	62
74	71	-1	137	88	63
75	72	-2	138	89	64
76	73	-3	139	90	65
77	74	-4	140	91	66
78	75	-5	141	92	67
79	76	-6	142	93	68
80	77	-7	143	94	69
81	78	-8	144	95	70
82	79	-9	145	96	71
83	80	-10	146	97	72
84	81	-11	147	98	73
85	82	-12	148	99	74
86	83	-13	149	100	75
87	84	-14	150	101	76
88	85	-15	151	102	77
89	86	-16	152	103	78
90	87	-17	153	104	79
91	88	-18	154	105	80
92	89	-19	155	106	81
93	90	-20	156	107	82
94	91	-21	157	108	83
95	92	-22	158	109	84
96	93	-23	159	110	85
97	94	-24	160	111	86
98	95	-25	161	112	87
99	96	-26	162	113	88
100	97	-27	163	114	89
101	98	-28	164	115	90
102	99	-29	165	116	91
103	100	-30	166	117	92
104	101	-31	167	118	93
105	102	-32	168	119	94
106	103	-33	169	120	95
107	104	-34	170	121	96
108	105	-35	171	122	97
109	106	-36	172	123	98
110	107	-37	173	124	99
111	108	-38	174	125	100
112	109	-39	175	126	101
113	110	-40	176	127	102
114	111	-41	177	128	103
115	112	-42	178	129	104
116	113	-43	179	130	105
117	114	-44	180	131	106
118	115	-45	181	132	107
119	116	-46	182	133	108
120	117	-47	183	134	109
121	118	-48	184	135	110
122	119	-49	185	136	111
123	120	-50	186	137	112

$$\Sigma r = \frac{461 - \frac{66 \times 41}{10}}{10} = \frac{461 - 266.6}{10} = \frac{194.4}{10} = 19.44$$

$$\sqrt{(291 - \frac{461}{10}) (242 - \frac{66}{10})} = \sqrt{10.833 \times 16}$$

$$\frac{26}{15.16} = 0.759$$

$(\bar{r} - 1)^2 \approx 0.5$ \Rightarrow High correlation.

$(\bar{r} - 1)^2 \approx 0.5$ \Rightarrow strong direct

relation w/ 30 women = \bar{r}

relation w/ 30 women = \bar{r}

relation w/ additional 10 girls = \bar{r}

\bar{x}	\bar{y}	n^2	$\sum xy$	ny
10	2	100	4	20
8	3	64	100	24
2	9	4	8	18
14	7	11	48	2
5	6	25	36	30
6	5	36	21	30
$\Sigma x = 32$	$\Sigma y = 23$	230	204	129
11	10	11	2	
18	11	81	1	
$\Sigma x^2 = 195$	$\Sigma y^2 = 128 - \frac{32 \times 32}{6}$	$\frac{114}{6}$		
			$\sqrt{(230 - \frac{32^2}{6})(204 - \frac{32^2}{6})}$	
			$\frac{114}{6} = 19$	
			$= -41.67$	
			$(19 - 54) / \sqrt{54 \times 19} = -0.937$	
			-0.937	
			b) Strong indirect relationship.	

→ ① Pearson Product moment correlation

$$\text{Ansatz} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

$$\text{with } \bar{x} = \frac{1}{n} \sum x_i \quad \bar{y} = \frac{1}{n} \sum y_i$$

\bar{x} = mean of x variable

\bar{y} = mean of y variable

$x_i, y_i \rightarrow$ sample of variable x, y

Outlier! Extreme point of data set



Value of correlation r (-1 to +1)

LECTURE:- 26/8/22

Partial Correlation

To control effect of one or more variables

⇒ 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 keeping weather condition 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 relationships

Assumption 5: No significant outliers.

Assumption 6: Approximately normally distributed.

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

Date _____
Page _____

→ Partial Correlation coefficient formula

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

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

↳ Pearson Product formula of

<u>x</u>	<u>y</u>	<u>z</u>	<u>$x - \bar{x}$</u>	<u>$y - \bar{y}$</u>	<u>$z - \bar{z}$</u>
10	29	17	-1	1.5	0
13	33	23	-6	3.5	3
16	47	29	-3	6.5	6
13	51	32	-6	7.5	6
21	43	41	2	4	4
23	31	39	4	2	6
29	49	47	10	10	10
27	71	43	8	14	2

so two different heights are different

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

9. total number

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

$$\bar{z} = \frac{292}{9} = 33$$

if third or more are to find Product

with two conflicting sets Product

so which of them is better Product

My 20.33

Date
Page
P

$$\text{My. 2 } 2 \quad 0.532 - 0.584 \times 0.792$$

$$\sqrt{(-0.184)(1-0.884)}$$

$$\rightarrow 0.062$$

$$\sqrt{0.2328}$$

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

$\bar{Y} - \bar{Y}$	$(\bar{Y} - \bar{Y})^2$	$\bar{Z} - \bar{Z}$	$(\bar{Z} - \bar{Z})$	$(\bar{Y} - \bar{Y})(\bar{Y} - \bar{Y})$	$(\bar{Y} - \bar{Y})(\bar{Z} - \bar{Z})$	$(\bar{Z} - \bar{Z})(\bar{Y} - \bar{Y})$
-14.9	222.0	-16	208	134.1	236.4	144
-10.9	118.8	-10	100	65.4	109	60
-2.9	8.4	-12	144	0	34.8	0
+3.1	9.6	-4	16	-9.3	-12.4	12
2.1	4.0	4	16	-42.6	28.4	-24
-0.9	0.81	8	64	-1.8	-2.26	16
-12.9	166.4	6	36	-57.6	-22.4	24
5.1	25.01	14	196	57	27.4	140
27.1	734.7	-10	100	216.8	271	80
ΣY	22			$\Sigma = 362$	$\Sigma = 652$	$\Sigma = 452$
	1336.84	Σ^2	Σ^2			482
	556.84	928				

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

$$\sqrt{346 \times 1336.84}$$

$$r_{YZ} = \frac{652}{\sqrt{1336.84 \times 928}} \rightarrow 0.589$$

$$r_{ZM} = \frac{452}{\sqrt{928 \times 346}} \rightarrow 0.712$$

LECTURE 27/8/22

Date _____
Page _____

Def. of Multiple Correlation: If we have three or more variables.

Situation between three or more variables.

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

(x,y,z)	x	y	z
10	29	18	
13	33	25	

(x,y,z)	x	y	z	r _{xy}	r _{xz}	r _{yz}	r _{xyz}
10	29	18					
13	33	25					
16	41	34					
13	51	42					
21	43.45	41					
23	51.51	49					
29	49.55	42					
27	51.5 - 43	38.1 -					
45	51.66 -	41.2 -					
40	51.58	42					
61	51	49					
22	52.2 -	42.2 -					

$$R_{xyz} = \sqrt{0.747^2 + 0.574^2 - 2 \times 0.532 \times 0.581 \times 0.25}$$

$$= \sqrt{0.532 \times 48.822} = 0.632$$

$$= \sqrt{0.632} = 0.8204 = 0.8204$$

$$\frac{1}{\sqrt{0.632}} = \sqrt{0.632} = 0.8204$$

$$\frac{1}{\sqrt{0.632}} = \frac{1}{\sqrt{0.632}} = 0.8204$$

x	y	\bar{z}	$\bar{x}-\bar{y}$	$(x-\bar{y})^2$	$(y-\bar{y})^2$	$\bar{z}-\bar{x}$	$(z-\bar{x})^2$	$(x-\bar{y})(z-\bar{x})$	$(y-\bar{y})(z-\bar{x})$
15	6	25	-0.3	0.09	1.2	1.44	-0.8	0.64	0.36 0.24
18	3	20	2.4	7.29	-1.8	3.24	3.2	10.24	4.86 6.48
13	8	27	-2.3	5.29	3.2	10.24	1.2	1.44	3.36 2.16
14	6	24	-1.3	1.69	1.2	1.44	-1.8	3.24	1.56 2.34
19	2	30	3.7	13.69	-28	7.84	4.2	17.64	10.34 13.34
11	3	21	-4.3	18.49	-1.8	3.24	-4.8	23.04	2.34 3.34
17	4	26	1.7	2.89	-0.8	0.64	0.2	0.04	1.36 1.04
20	4	31	4.7	22.09	-0.8	0.64	5.2	27.04	3.26 2.04
210	5	20	-5.1	26.09	0.2	0.04	-5.8	31.64	1.06 3.06
16	7	25	0.7	0.49	2.2	4.84	-0.8	0.64	1.54 0.84
							33.6		

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

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

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

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

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

$$\sum(z-\bar{z})^2 = 112.6$$

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

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

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

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

$$r_{yz} = \frac{-15.4}{\sqrt{33.6 \times 112.6}} = -0.245$$

$$r_{xz} = \frac{99.6}{\sqrt{100.1 \times 112.6}} = \cancel{0.86} \cancel{0.82} 0.718$$

$$(29.7 \pm 2.1) = 0.369 + 0.245 \times 0.918$$

$$\sqrt{(1 - 0.245)(1 - 0.918)}$$

$$= -0.384$$

$$\text{Lny}_2 = 0.736$$

$$0.163$$

$$= 0.823$$

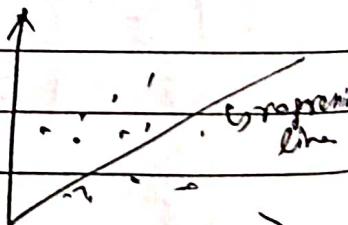
LECTURE:- 29/8/22

Regression Analysis

- 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 { $x \rightarrow \text{predictor}$ } $\text{Slope} = \beta_1 = r \cdot \frac{s_y}{s_x}$
 $y \rightarrow \text{riterion}$

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

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

$$\hat{y} = a + bx$$

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

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

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

$$\sum n^2 - (\sum n)^2$$

x	y	x^2	ny
7	12	49	84
6	8	36	48
8	12	64	96
5	10	25	50
6	11	36	66
9	13	81	117
$\sum n = 41$	$\sum y = 66$	$\sum n^2 = 291$	$\sum ny = 461$

$$\bar{x} = \frac{41}{6} = 6,83 \quad \bar{y} = 11$$

$$b_1 = \frac{41}{6} = 6,83 \quad \bar{x} = 6,83 \quad \bar{y} = 11$$

$$b_0 = 11 - 6,83 \times 6,83$$

$$b_0 = 11 - 46,609 \quad \rightarrow \quad b_0 = -35,609$$

$$b_0 = -35,609 \quad \rightarrow \quad b_0 = -35,609$$

$$b_0 = -35,609 \quad \rightarrow \quad b_0 = -35,609$$

$$b_0 = -35,609 \quad \rightarrow \quad b_0 = -35,609$$

$$b_0 = -35,609 \quad \rightarrow \quad b_0 = -35,609$$

$$b_0 = -35,609 \quad \rightarrow \quad b_0 = -35,609$$

$$b_0 = -35,609 \quad \rightarrow \quad b_0 = -35,609$$

$$b_0 = -35,609 \quad \rightarrow \quad b_0 = -35,609$$

$$b_0 = -35,609 \quad \rightarrow \quad b_0 = -35,609$$

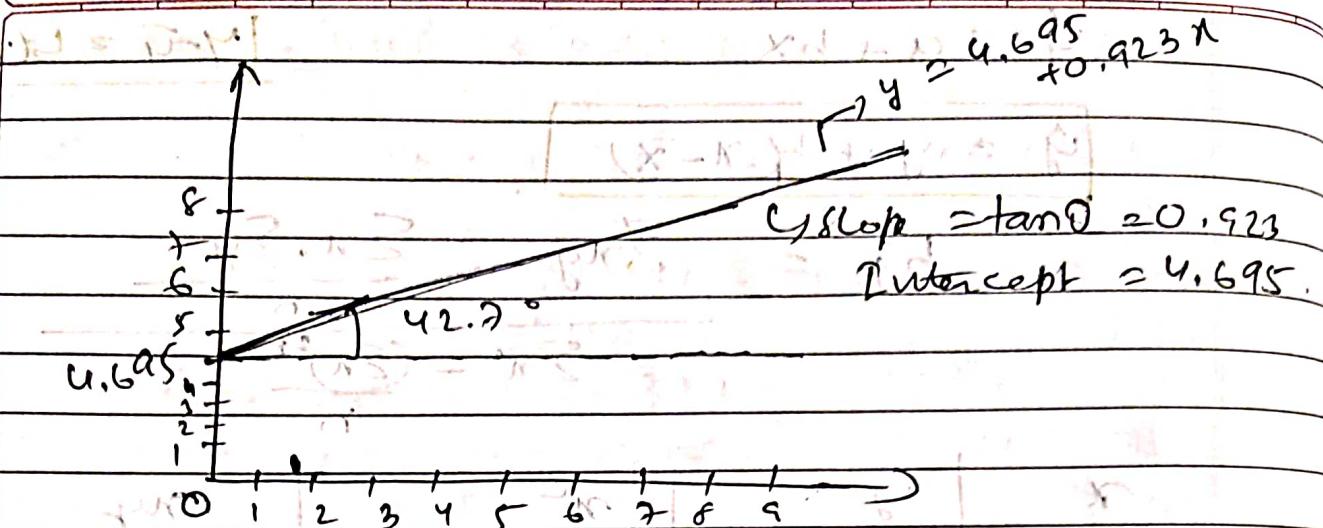
$$b_0 = -35,609 \quad \rightarrow \quad b_0 = -35,609$$

$$b_0 = -35,609 \quad \rightarrow \quad b_0 = -35,609$$

$$b_0 = -35,609 \quad \rightarrow \quad b_0 = -35,609$$

$$b_0 = -35,609 \quad \rightarrow \quad b_0 = -35,609$$

Date _____
Page _____



x	y	x^2	y^2	$x - \bar{x}$
20	120	400	14400	-22.6
43	128	1849	16384	0.4
63	141	3969	19881	20.4
26	126	676	15876	-16.6
83	134	2809	17956	10.4
31	128	961	16384	-11.6
58	136	3364	18496	15.4
46	132	2116	17424	3.4
58	140	3364	19600	15.4
30	144	900	20736	28.4
46	128	2116	16384	3.4
53	136	2809	18496	10.4
60	146	3600	21316	18.4
20	124	400	15376	-22.6
63	143	3969	20449	20.4
43	130	1849	16900	0.4
26	124	676	15376	-16.6
19	121	361	14641	-23.6
31	126	961	15876	-11.6
23	123	529	15129	-19.6
$\sum n = 852$	$\sum y = 2630$	$\sum w = 41,698$	$\sum y^2 = 3,770,80$	$\Sigma =$

$$\bar{x} = \frac{852}{20} = 42.6$$

$$\bar{y} = \frac{2630}{20} = 131.5$$

$\sum (x - \bar{x})(y - \bar{y})$ = $\sum xy - n\bar{x}\bar{y}$ is called regression equation (1)

$$(\sum (x^2 - \bar{x}^2)) (\sum (y^2 - \bar{y}^2))$$

$x - \bar{x}$	$(x - \bar{x})(y - \bar{y})$	xy	$(x - \bar{x})^2$	$(y - \bar{y})^2$
-11.5	259.9	2400	510.25	132.25
-3.5	-1.4	5504	0.16	12.25
9.5	193.8	8883	416.16	90.25
-5.5	91.3	3276	275.56	30.25
2.5	26	= 702	108.16	6.25
-3.5	40.6	3968	134.56	12.25
4.5	69.3	7888	237.16	20.25
0.5	1.7	6022	11.56	0.25
8.5	130.9	8120	237.16	72.25
12.5	342.5	10080	350.76	156.25
-3.5	-11.9	5888	11.56	12.25
4.5	46.8	7208	108.16	20.25
14.5	252.3	8760	202.76	58.25
-7.5	169.5	2480	50.76	132.25
11.5	234.6	9009	416.16	132.25
-15.0	-0.6	5590	0.16	2.25
-7.5	124.5	3224	56.25	58.25
-10.5	247.8	22993	536.96	110.25
-5.5	63.8	3906	134.56	30.25
-8.5	166.6	-2829	384.16	72.25
Σ	$\Sigma = 2448$	$\Sigma ny = 114486$	$\Sigma =$	$\Sigma = 1235$
			5382.8	

$$b = 0.484$$

$$\therefore y = 131.5 + 0.484(x - 42.6)$$

$$\because n = 25,$$

$$\Rightarrow y = 123.5$$

(d) Simple correlation = $\frac{\sum xy - \frac{\sum x \cdot \sum y}{n}}{\sqrt{(\sum x^2 - \frac{(\sum x)^2}{n})(\sum y^2 - \frac{(\sum y)^2}{n})}}$

x	y	xy	x^2	y^2
25.00	21.301	538.2	625.00	459.28
25.01	21.001	525.03	625.01	441.00
25.02	20.701	517.40	625.04	424.89
25.03	20.401	508.03	625.09	408.16
25.04	20.101	500.04	625.16	392.01
25.05	19.801	492.05	625.25	376.09
25.06	19.501	484.06	625.36	360.25
25.07	19.201	476.07	625.49	344.49
25.08	18.901	468.08	625.64	328.81
25.09	18.601	460.09	625.81	313.21
25.10	18.301	452.10	626.00	297.64
25.11	18.001	444.11	626.21	282.16
25.12	17.701	436.12	626.44	266.76
25.13	17.401	428.13	626.69	251.44
25.14	17.101	420.14	626.96	236.16
25.15	16.801	412.15	627.25	221.01
25.16	16.501	404.16	627.56	205.96
25.17	16.201	396.17	627.89	190.91
25.18	15.901	388.18	628.24	175.84
25.19	15.601	380.19	628.61	160.81
25.20	15.301	372.20	629.00	145.80
25.21	15.001	364.21	629.41	130.81
25.22	14.701	356.22	629.84	115.84
25.23	14.401	348.23	630.29	100.89
25.24	14.101	340.24	630.76	85.96
25.25	13.801	332.25	631.25	71.025
25.26	13.501	324.26	631.76	56.16
25.27	13.201	316.27	632.29	41.36
25.28	12.901	308.28	632.84	26.64
25.29	12.601	300.29	633.41	12.01
25.30	12.301	292.30	634.00	0.01

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

25.01	21.301	538.2	625.00	459.28	25.01
25.02	21.001	525.03	625.01	441.00	25.02
25.03	20.701	517.40	625.04	424.89	25.03
25.04	20.401	508.03	625.09	408.16	25.04
25.05	19.801	492.05	625.16	376.09	25.05
25.06	19.501	484.06	625.36	360.25	25.06
25.07	19.201	476.07	625.49	344.49	25.07
25.08	18.901	468.08	625.64	328.81	25.08
25.09	18.601	460.09	625.81	313.21	25.09
25.10	18.301	452.10	626.00	297.64	25.10
25.11	18.001	444.11	626.21	282.16	25.11
25.12	17.701	436.12	626.44	266.76	25.12
25.13	17.401	428.13	626.69	251.44	25.13
25.14	17.101	420.14	626.96	236.16	25.14
25.15	16.801	412.15	627.25	221.01	25.15
25.16	16.501	404.16	627.56	205.96	25.16
25.17	16.201	396.17	627.89	190.91	25.17
25.18	15.901	388.18	628.24	175.84	25.18
25.19	15.601	380.19	628.61	160.81	25.19
25.20	15.301	372.20	629.00	145.80	25.20
25.21	15.001	364.21	629.41	130.81	25.21
25.22	14.701	356.22	629.84	115.84	25.22
25.23	14.401	348.23	630.29	100.89	25.23
25.24	14.101	340.24	630.76	85.96	25.24
25.25	13.801	332.25	631.25	71.025	25.25
25.26	13.501	324.26	631.76	56.16	25.26
25.27	13.201	316.27	632.29	41.36	25.27
25.28	12.901	308.28	632.84	26.64	25.28
25.29	12.601	300.29	633.41	12.01	25.29
25.30	12.301	292.30	634.00	0.01	25.30

$$\Rightarrow b_0 = 1144.86 - \frac{882 \times 26.30}{20}$$

$$416.78 - \frac{(882)}{20}$$

$$(882) \frac{2448}{20} = 0.454$$

$$(882) \frac{5382}{20} = 259.1$$

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

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

$$\text{when } x = 25, y = 123.57$$

LECTURE 1 / 9/22

Hypothesis Testing:

\Rightarrow A manufacturer asserts that a respirator it makes delivers pure air 95 minutes on average.

Population parameter - mean, sample, standard deviation
 is about 10 minutes of 18.05 min with

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

Wishes to know if H_0 is true

Statement about population parameter assumed to be true.

H_0 : Null hypothesis.

H_a : Alternative hypothesis.

Null hypothesis is always true or

statement which is contradicting

population parameter

contradicting to

other known facts

if null hypothesis is

accepted as true only if there is evidence in favour of it.

→ 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: - \neq , $<$, $>$, $\neq \mu_0$, $\mu_1 < \mu_0$

Null: $\mu = \bar{x}$

Alternative: $\neq \bar{x}$

- 2 distinct measurements

$\rightarrow \mu \neq \bar{x}$ since if medical equipment

indicates no error in test then sample will suppose more than standard deviation of the \bar{x} . There is no harm.

Example 1: - \bar{x} = sample mean - estimated mean

$H_0: \mu = \$12.35 \rightarrow$ evidence mean is same as $H_a: \mu > \$12.35$ at greater than $\$12.35$.

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
 at first. If value of \bar{x} would be highly unlikely to occur if H_0 were true, but favours truth, then the hypothesis H_a is assumed to be true.

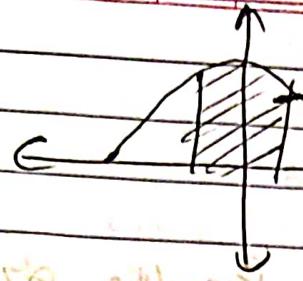
is south of H_0

in direction

\bar{x} is unusual

Normal Distribution

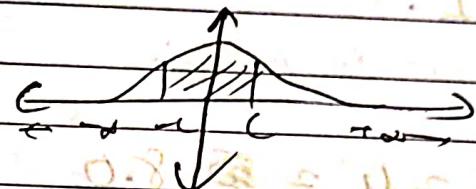
$$\mu = \mu_0.$$



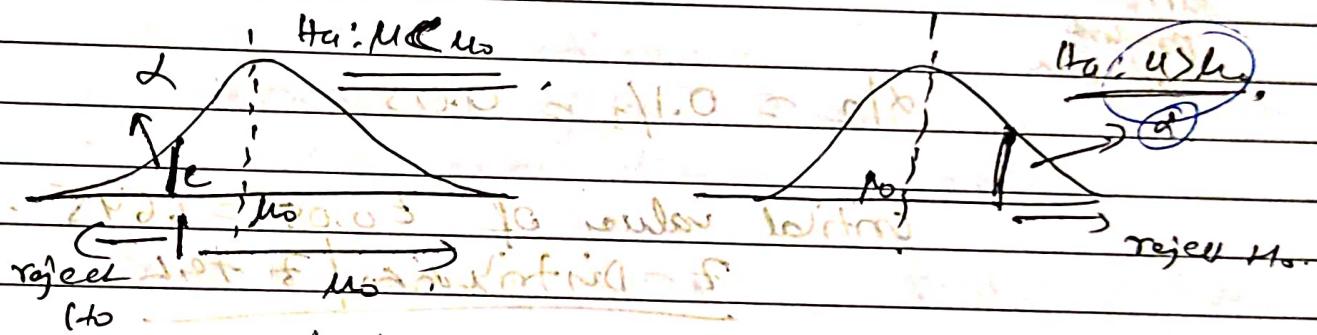
If the area is near μ_0 is true,

if farther, H_0 is false, H_a is true.

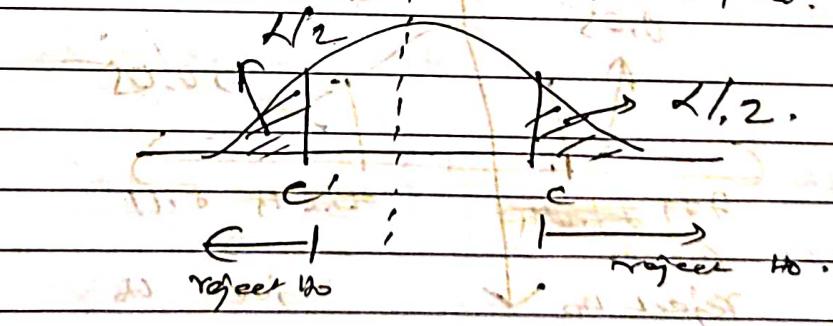
$H_a: \mu < \mu_0$, rejection area is $(-\infty, -c]$,
 $H_a: \mu > \mu_0$, rejection area is $[c, \infty)$.



$H_a: \mu < \mu_0$, rejection area $= (-\infty, c] \cup [c, \infty)$



$H_a: \mu \neq \mu_0$.



Pf. rare event, rejection of Null hypothesis.

$H_A \Rightarrow$ probability of rare event.

$H_0: (\mu, \sigma^2)$ and (μ', σ'^2) are
different is introduced

Sampling distribution

Ex-3 :-

$$\sigma = 0.15$$

$$H_0: \mu = 80$$

Sample size = 25,

$$H_a: \mu \neq 80$$

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

A:-

$$\bar{M} = M \pm 1.80$$

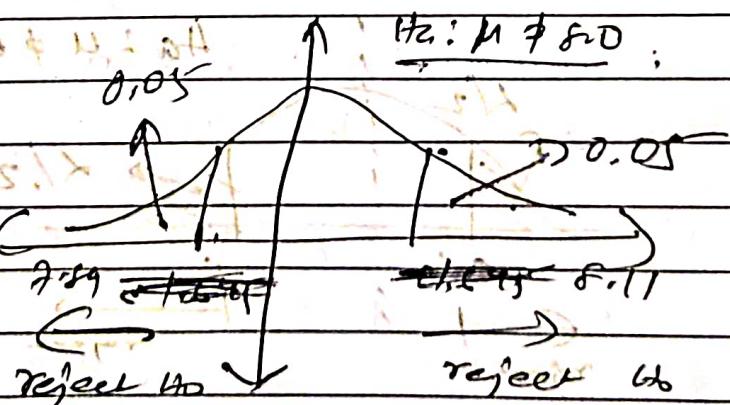
$$\text{sample standard deviation} = \frac{\sigma}{\sqrt{n}} = \frac{0.15}{\sqrt{25}} = 0.067$$

sample
standard
deviation

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

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

Z-Distribution / Z-test



$$\rightarrow CRF = (1.645)(0.067) = 2.089$$

$$\rightarrow CRF + (1.645)(0.067) = F_{11}$$

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

σ : Population standard deviation.
 s_x : Sample standard deviation.

$$s_x = \frac{\sigma}{\sqrt{n}}$$



$H_a: \mu \neq \mu_0 \rightarrow$ two tailed test.

$H_a: \mu < \mu_0 \rightarrow$ left tailed test.

$H_a: \mu > \mu_0 \rightarrow$ right tailed test.

Two types of error :-

1. reject null hypothesis in favor of alternative hypothesis H_a .

Type I, type II errors.

		Truth	
		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

$\alpha \rightarrow$ level of significance \rightarrow probability of Type I error.

\rightarrow probability of Type II error.

\Rightarrow To reduce error, increase sample size.

$$\text{Standardized test statistic}, Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

~~standard distribution~~

Symbol in H_a	Terminology	rejection region	
<	left tailed	$(-\infty, -z_{\alpha})$	Normal distribution
>	right tailed	(z_{α}, ∞)	Student t (t_{α})
\neq	two tailed	$(-\infty, -z_{\alpha/2}) \cup (z_{\alpha/2}, \infty)$	$(-z_{\alpha/2}, z_{\alpha/2})$

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 5/9/21

Q. a) Null hypothesis: $\mu = 74.5$ ft \rightarrow

Alternative: $\mu \neq 74.5$ ft \rightarrow

b) Null: $\mu = 145$ pounds

Alternative: $\mu > 145$ pounds

c) Null: $\mu = \$14,256$

Alternative: $\mu > \$14,256$

d) Null: $\mu = \$12.53$

Alternative: $\mu \neq \$12.53$

e) Null: $\mu = 69.4$ acres of land

Alternative: $\mu < 69.4$ acres of land

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 $H_a: \mu > 42.8$

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

All possible random sample:

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

(164, 160)

Sample mean \bar{x} & t test all together, what is t ?

~~(77) My health condition X~~

152 152 182

152 1551562 153 154 155 156

152 195.7160 N 156 00.0000 E

152 164 158

~~156~~ 1521 and 154 hospitalized with 15

156 51576 1576 1576 1576 1576 1576

156 160 158

156 1641 on : 160°e stopped 11M 15

160 F1521 #141 call 2015 bpd. water 410

160 158 158

160 165H \pm 10 160 \pm 10 mm

180 181 182 183

162 163 164

1640 and 1650 started at 1620 since we had

\bar{x}	152	154	156	158	160	162	164
$p(\bar{x})$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

$$\mu_{\bar{X}} = \sum x_i p(x_i)$$

$$= \frac{152}{16} + \frac{154 \times 2}{16} + \frac{156 \times 3}{16} +$$

15884 + 16048

~~total 16~~ → 162 x 2 ~~plus~~ ~~14~~

\bar{M}_x : sample mean.

σ_x : sample S.D.

Date _____
Page _____

$$\Rightarrow \bar{M}_x = 158.$$

$$\text{Var} = \sum \bar{x}^2 P(x)$$

$$= (152)^2 + 2 \times (154)^2 + 3 \times (156)^2 + 4 \times (158)^2 \\ + 3 \times (160)^2 + 2 \times (162)^2 + 164^2$$

$$\text{Var} = \frac{399584}{16} = 24,974.$$

$$\sigma_x = \sqrt{\sum \bar{x}^2 P(x) - (\bar{M}_x)^2}$$

$$= \sqrt{24974 - 158^2}$$

$$\sigma_x = \sqrt{10}$$

152

156

160

164

$$\text{mean} = 158 = \mu$$

$n = 14$

- 6

- 2

2

6

X

$$\sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{24974 - 158^2}{13}} = \sqrt{720.0}$$

$$\text{Var} \rightarrow \sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

$$\therefore \sigma^2 = 24974 - 158^2$$

$$= 24974 - 24336 = 638$$

$$= \frac{36 + 4 + 4 + 56}{4}$$

$$= \sqrt{80/4} = \sqrt{20}.$$

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

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

LECTURE - 8/9/22

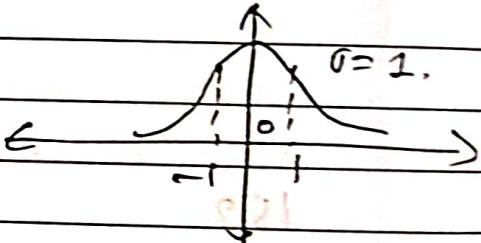
$$1. \quad \mu_{\bar{x}} = \mu = \$13,525.$$

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

Standard Normal Distribution

$\mu = 0$ standard normal random variable
 $\sigma = 1$ denoted by letter z .

$P(z < z)$: cumulative probability,



$$(a) \quad P(z < 1.40) = 0.9306 \quad (\text{from table}).$$

$$(b) \quad P(z < -0.25) = 0.4013 \quad (\text{from table}).$$

$$(c) \quad P(z > 1.6) = 1 - P(z < 1.6) \\ = 1 - 0.9452 = 0.0548$$

$$(d) \quad P(z > -1.02) = 1 - P(z < -1.02) \\ = 1 - 0.1539 \\ = 0.8461$$

$$0.5L = 1.108 L =$$

(e) $P(0.57 \leq Z \leq 1.57)$

$$= P(Z \leq 1.57) - P(Z \leq 0.57)$$

$$= 0.9418 - 0.6915 = 0.2503.$$

(f) $P(-2.5 \leq Z \leq 0.09)$

$$= P(Z \leq 0.09) - P(Z \leq -2.5)$$

$$= 0.5359 - 0.0122 = 0.0054$$

$$\approx 0.5305$$

General Normal variable:

X : normally distributed random variable

$$P(a \leq X \leq b) = P\left(\frac{a-\mu}{\sigma_x} \leq Z \leq \frac{b-\mu}{\sigma_x}\right).$$

Z : standard random variable.

' a ' can be decimal number or ' ∞ '.

' b ' can be decimal number or ' ∞ '.

(g) $P(X < 14)$

$$= P(Z < \frac{14-\mu}{\sigma}) = P(Z < 1.6)$$

$$= 0.9452$$

(h) $P(8 \leq X \leq 14)$

$$= P(-0.8 \leq Z \leq 1.6) = P(Z \leq 1.6) - P(Z \leq -0.8)$$

$$= 0.9452 - 0.2119$$

$$\approx 0.7333.$$

Q. $\mu = 37,500$ miles

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

$$\begin{aligned}
 & P(30,000 < X < 40,000) \\
 & = P\left(\frac{30,000 - 37,500}{4,500} < Z < \frac{40,000 - 37,500}{4,500}\right) \\
 & = P(-1.67 < Z < 0.56) \\
 & = P(Z < 0.56) - P(Z < -1.67) \\
 & = 0.7123 - 0.0435 \\
 & = 0.6688
 \end{aligned}$$

Central limit theorem: (after DeMoivre-Laplace)

Increasing sample size, distribution of \bar{X} , probability on lower and upper end shrink, on the middle it broadens (or increase). (decrease)

Increasing sample mean, the curves resemble bell shape curve.

Central Limit Theorem

Sample size 30 or more, sample mean is approximately normally distributed with $\mu_{\bar{X}} = \mu$ and $\sigma_{\bar{X}} = \sigma/\sqrt{n}$, where 'n' is sample size.

$$\begin{aligned}
 \text{Mean } \bar{x} &= \mu = 112 \\
 \text{S.D. of } \bar{x} &= \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{40}{\sqrt{50}} = 5.656
 \end{aligned}$$

$$(a) \text{ Mean of } \bar{x} = \mu = 112$$

$$\text{S.D. of } \bar{x} = \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{40}{\sqrt{50}} = 5.656$$

(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) $P(\bar{X} > 113)$

$$= 1 - P(\bar{X} < 113)$$

$$= 1 - P\left(Z < \frac{113 - 112}{40}\right)$$

$$= 1 - P(Z < \frac{0.18}{40})$$

$$\begin{aligned} &\times = 1 - P(Z < 0.0045) \\ &\quad \cancel{= 1 - 0.5120} \\ &\quad = 0.4880 \end{aligned}$$

$$= 0.4880$$

(d) $P(110 < \bar{X} < 114)$

$$= P\left(\frac{110 - 112}{5.658} < Z < \frac{114 - 112}{5.658}\right)$$

$$= P(-0.283 < Z < 0.353)$$

$$= P(Z < 0.353) - P(Z < -0.283)$$

$$= 0.6364 - 0.3622$$

$$= 0.2742.$$

→ If population is normally distributed ^{then} for larger ^{normally} sample size, sample mean is ^{normally} distributed.

Q:

$$\mu = 38,500$$

$$\sigma = 2500$$

$$n = 25$$

$$\bar{x} = 36000$$

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

$$= P\left(Z < \frac{36000 - 38500}{2500}\right)$$

$$= P(Z < -2.236)$$

$$= P(Z < -2.24)$$

$$= 0.0125$$

Q:

$$\mu = 50 \text{ months}$$

$$\sigma = 6 \text{ months}$$

(a)

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

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

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

(b)

$$n = 36$$

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

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

$$= P(Z < -2)$$

$$= 0.0228$$

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.

Test statistic: $\bar{x} - \mu$

$$\sigma/\sqrt{n}$$

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

$$\sigma/\sqrt{n}$$

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

Estimated S from sample

~~$\alpha = 0.05$~~

~~$Z_{\alpha/2} = 1.96$ (from table)~~

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

Alternative hypothesis: Newly developed pain relief delivers more relief quickly

$$H_0: \mu = 3.5$$

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

$$1.8 \text{ vs } 3.1 : Z_{\text{cal}} = 3.1 - 3.5$$

$$\frac{1.5}{\sqrt{50}}$$

$$= -1.886$$

D.L is left-tailed test

$$\text{Given } \alpha = 0.0505, Z = -1.64$$

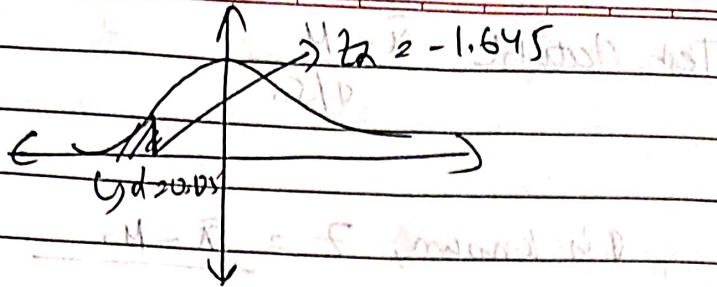
$$\text{Given } \alpha = 0.0495, Z = -1.65$$

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

$$20/1.5 =$$

$$0.0505 - 0.0495$$

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



Now, $-1.886 < -1.645$

$H_0: \mu = 8$ lies in rejection region.
 \Rightarrow Null hypothesis is rejected

∴ Newly developed pain killer delivers more relief quickly.

Q. Mean weight of oranges per jar: intended weight 8.1
 actual mean weight
 $\bar{x} = 8.2$ ounce. $n = 20$.

$n = 20$.

$\bar{x} = 8.2$ ounce. $(7.8 \text{ to } 20.0)$.

$s = 0.25$ ounce. $(7.8 \text{ to } 11.4)$

Null hypothesis: $H_0: \mu = 8.1$

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

$$\alpha/2 = 0.01 \Rightarrow 0.005$$

$$z_{\alpha/2} = -2.575 \Rightarrow -2.32$$

$$t = 2.575 \times 0.25 = 0.6437$$

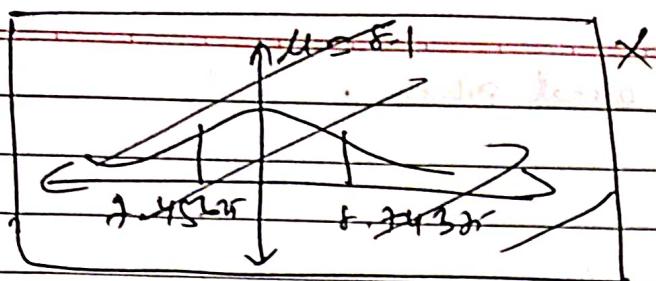
$$c_1 = 2.575 \times 0.25 = 0.6437$$

$$c_2 = 8.1 + 2.575 \times 0.25 = 8.3437$$

$$t = 2.1482 \text{ (approx)}$$

$$7.8 \text{ to } 8.2 \text{ ounce. } 21.908$$

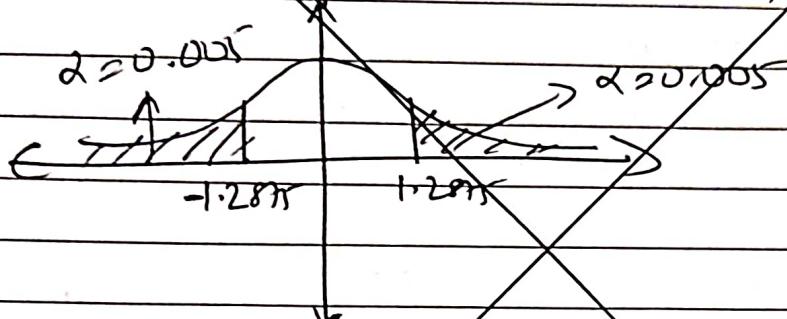
$$\frac{1}{\sqrt{20}}$$



$$z_d = -2.575$$

rejection region:

$$(-\infty, -1.2875] \cup [1.2875, \infty)$$



\therefore given z , value = 2.1808 lies in rejection region

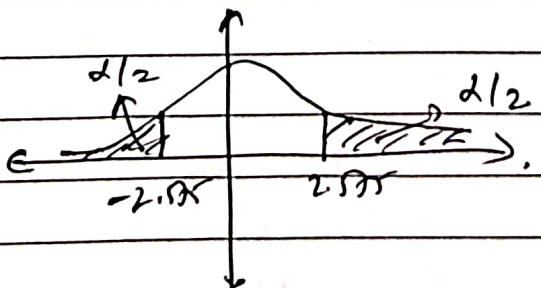
\Rightarrow Null hypothesis is rejected

\Rightarrow Alternative hypothesis is accepted

\Rightarrow Machine should be recalibrated

$$z_{d1/2} = -2.575$$

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



$$z = 2.1808$$

\hookrightarrow H_0 is accepted.

as it doesn't lie
in rejection region

\therefore Machine should
not be recalibrated

Origin Pro \rightarrow Data analysis and graphing software.
Steps to add trendline.