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

(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
	$\sum f = 19$	$\sum fx = 31$

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

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

ascending / Descending order

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.

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).

Measure of variability of dispersion

→ range
→ variance
→ Standard Deviation.

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

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

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

mode = 40.

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

mean = 40 = median = mode

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

x_{max} : largest measurment
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 → n−1 } '−1' because we don't consider
population → N }   mean.

Data set 1:

x	$(x-\bar{x})$	$(x-\bar{x})^2$
38	−2	4
39	−1	1
39	−1	1
39	−1	1
40	0	0
40	0	0
40	0	0
40	0	0
42	2	4
43	3	9

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

$S^2 = \frac{20}{9} = 2.22$ (variance)

$S = 1.491$ (S.D.)

Data set 2:

x	$x-\bar{x}$	$(x-\bar{x})^2$
33	−7	49
34	−6	36
36	−4	16
37	−3	9
40	0	0
40	0	0
42	2	4
45	5	25
46	6	36
47	7	49

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

~~$\sigma^2 = \frac{224}{10} = 22.4$~~

~~$\sigma = \sqrt{22.4} = 4.73$~~

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

$S = \sqrt{24.8} = 4.973$

DATA ANALYSIS

(1) 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:

Quantitative Data
- (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

→ 7 → 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, 0, 1, 8, 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 :-

Date
Page

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%

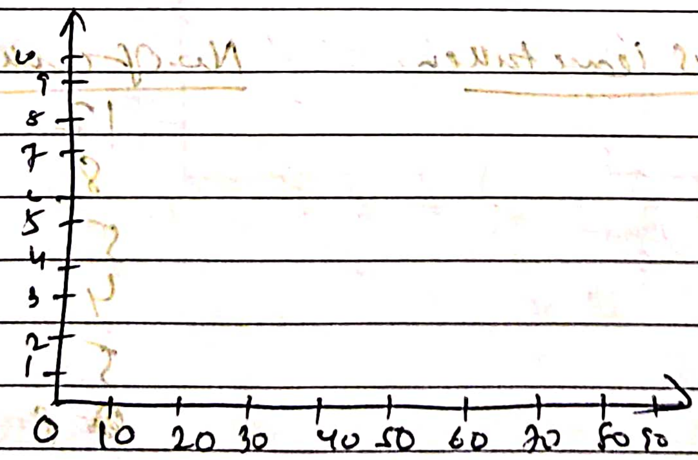
⇒ Dotplot

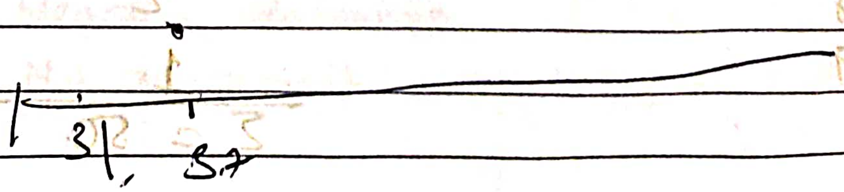
California → 20 division I school.] % of full
Texas → 19 Division I schools.] time freshmen

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, 77, 39, 35, 71, 63, 12, 45, 38, 39, 25, 65, 25, 24, 22.

~~California~~





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 non-Numerical 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 . = 2 .

(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.

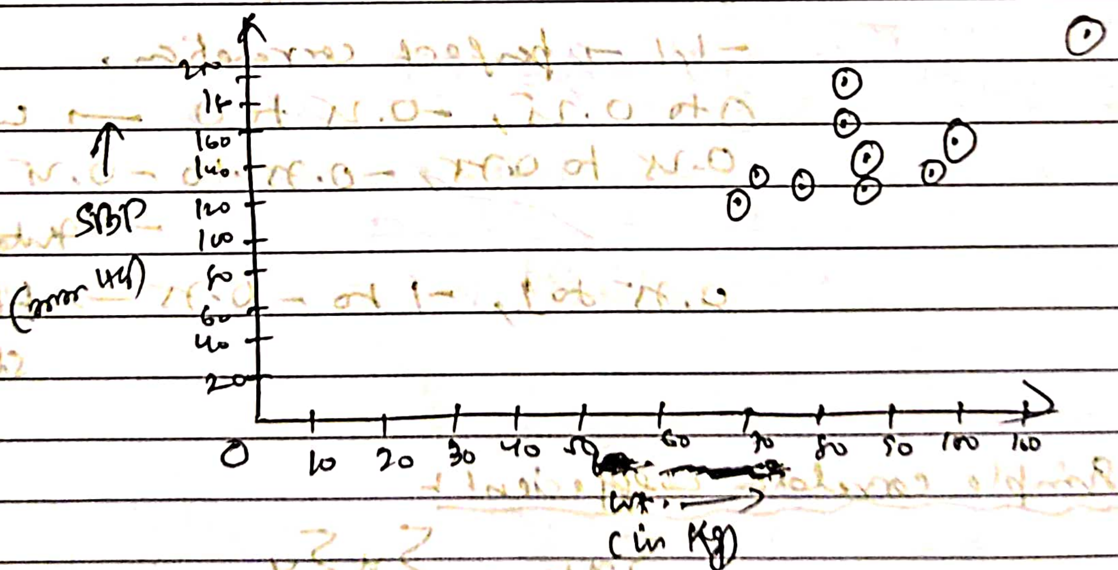
LECTURE; 25/8/22

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.

Types of relationship :

(1) Positive relationship (x ↑, y ↑) / Direct relationship

(2) Negative relationship (x ↓, y ↑) / Inversed relationship

(3) No relationship. (full scattered data).

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 ($\uparrow\uparrow, \downarrow\downarrow$).

Sign of 'r' is -'ve ⇒ indirect relationship ($\uparrow\downarrow, \downarrow\uparrow$).

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.

⇒ Simple correlation coeffecient :

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

Date ______
Page ______

$n = 6$.

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
$\Sigma x = 41$	$\Sigma y = 66$	$\Sigma xy = 461$	$\Sigma x^2 = 291$	$\Sigma y^2 = 742$

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

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

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

$\Rightarrow$ 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

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

$$= \frac{-41.67}{\sqrt{54.33 \times 33.33}}$$

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

↳ Strong indirect relationship.

→ ② Karan Product moment correlation

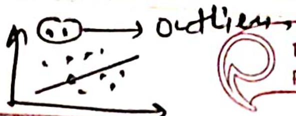
$$r = \frac{\Sigma\{(x_i - \bar{x})(y_i - \bar{y})\}}{\sqrt{\Sigma(x_i - \bar{x})^2 \, \Sigma(y_i - \bar{y})^2}}$$

$\bar{x}$ = 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 coefficient formula

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

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

↳ Pearson Product formula of 'r'.

~~EFCURE~~

1.

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

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

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

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

$r_{xy} = 0.532$
$r_{yz} = 0.589$
$r_{zx} = 0.797$

$$r_{xy.z} = \frac{0.532 - 0.589 \times 0.797}{\sqrt{(1-0.589^2)(1-0.797^2)}}$$

$$= \frac{0.062}{\sqrt{0.2398}}$$

$= 0.127$ (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	$\Sigma = 362$	$\Sigma = 656$	$\Sigma = 452$

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

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

$$r_{zx} = \frac{452}{\sqrt{928 \times 346}} = 0.797$$

LECTURE - 27/8/22

Date
Page

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	18
13	33	23
19	41	24
16	47	28
13	57	32
21	43	41
23	31	29
29	43	42
27	31	43

$$R_{x.yz} = \sqrt{\frac{(0.797^2 + 0.584^2 - 2 \times 0.532 \times 0.58 \times 0.787)}{1 - 0.532^2}}$$

$= \sqrt{0.673}$ ~~= 0.8204~~ $= 0.8204$

$= 0.8204$

x	y	z	$x-\bar{x}$	$(x-\bar{x})^2$	$(y-\bar{y})$	$(y-\bar{y})^2$	$z-\bar{z}$	$(z-\bar{z})^2$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})(z-\bar{z})$	[illegible]
15	6	25	-0.3	0.09	1.2	1.44	-0.8	0.64	0.36	0.24	
18	3	20	2.7	7.29	-1.8	3.24	3.2	10.24	4.86	8.64	
13	8	27	-2.3	5.29	3.2	10.24	1.2	1.44	7.36	-2.76	
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	-2.8	7.84	4.2	17.64	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	1.36	1.04	
20	4	31	4.7	22.09	-0.8	0.64	5.2	27.04	3.76	2.04	
10	5	20	-5.3	28.09	0.2	0.04	-5.8	33.64	1.06	30.74	
16	7	25	0.7	0.49	2.2	4.84	-0.8	0.64	1.54	-0.56	
						33.6					

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

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

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

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

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

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

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

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

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

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

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

$$r_{zx} = \frac{99.6}{\sqrt{100.1 \times 117.6}} = 0.918$$

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

$$= -0.374$$

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

$$= 0.923$$

LECTURE:- 29/8/22

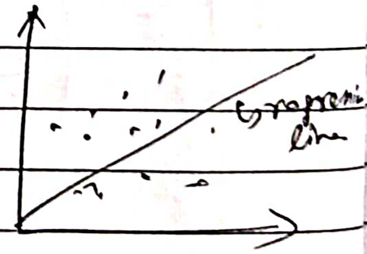
Regression Analyses

→ Predicting some variables, 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:- $\beta_1 = r.\frac{S_y}{S_x}$

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

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

Date
Page

$\hat{y} = a + bx$

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

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

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

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

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

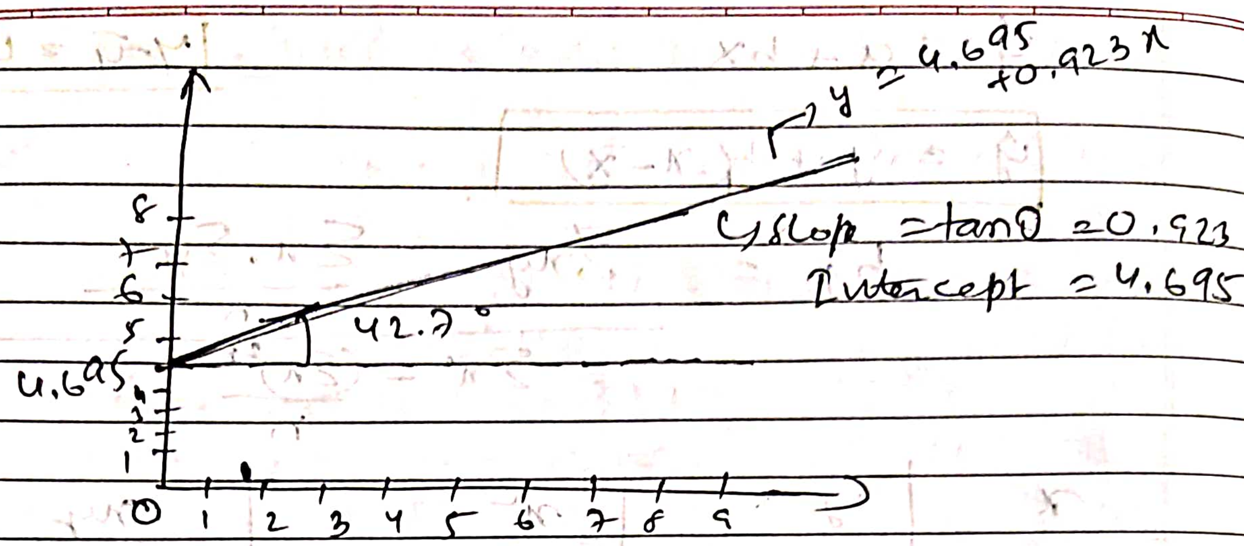
$$b = \frac{461 - \frac{41 \times 66}{6}}{291 - \frac{(41)^2}{6}}$$

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

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

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

~~x^2, y^2~~, $x-\bar{x}$, $y-\bar{y}$, $(x-\bar{x})(y-\bar{y})$, xy



x	y	x^2	y^2	$x-\bar{x}$
20	120	400	14,400	-22.6
43	128	1849	16,384	0.4
63	141	3969	19,881	20.4
26	126	676	15,876	-16.6
53	134	2809	17,956	10.4
31	128	961	16,384	-11.6
58	136	3364	18,496	15.4
46	132	2116	17,424	3.4
58	140	3364	19,600	15.4
70	144	4900	20,736	27.4
46	128	2116	16,384	3.4
53	136	2809	18,496	10.4
60	146	3600	21,316	17.4
20	124	400	18,376	-22.6
63	143	3969	20,449	20.4
43	130	1849	16,900	0.4
26	124	676	15,376	-16.6
19	121	361	14,641	-23.6
31	126	961	15,876	-11.6
23	123	529	15,129	-19.6
$\Sigma x = 852$	$\Sigma y = 2630$	$\Sigma x^2 = 41,678$	$\Sigma y^2 = 347,080$	$\Sigma =$

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

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

$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$	xy	$(x-\bar{x})^2$	$(y-\bar{y})^2$
-11.5	259.9	2400	510.76	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	7102	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	6072	11.56	0.25
8.5	130.9	8120	237.16	72.25
12.5	342.5	10,080	750.76	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	302.76	210.25 ~~56.25~~
-7.5	169.5	2480	510.76	56.25 ~~132.25~~
11.5	234.6	9009	416.16	132.25 ~~2.25~~
-1.5	-0.6	5590	0.16	2.25
-7.5	124.5	3224	275.56	56.25
-10.5	247.8	2299	556.96	110.25
-5.5	63.8	3906	134.56	30.25
-8.5	166.6	2829	384.16	72.25
Σ	$\Sigma = 2448$	$\Sigma xy = 114486$	$\Sigma = 5392.8$	$\Sigma = 1235$

$b = 0.454.$

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

$\therefore x = 25,$

$\Rightarrow y = 123.5$

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

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

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

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

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

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

$= 0.949 \simeq 0.95$

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

Date
Page

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

$$\Rightarrow b = \frac{2448}{5382.8} = 0.454.$$

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

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

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, 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 of it.

$\boxed{\text{Null hypothesis is always true or contradicting}}$

→ 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.

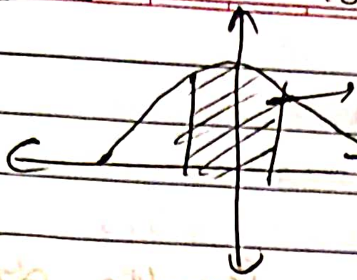
Example 1 :-
$H_0 : \mu = \$12.75$
$H_a : \mu > \$12.75$
⇒ evidence mean is greater than \$12.75.

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:- $\mu = \mu_0$.

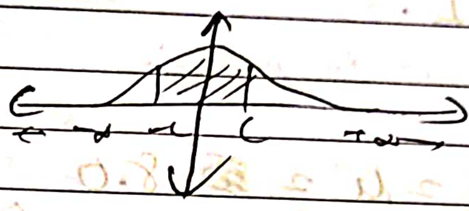


→ If the area is near, H_0 is true.

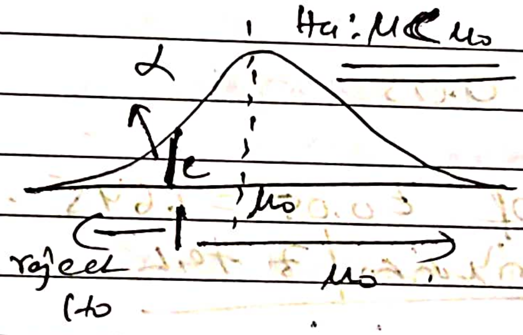
→ If farther, H_0 is false, H_a is true.

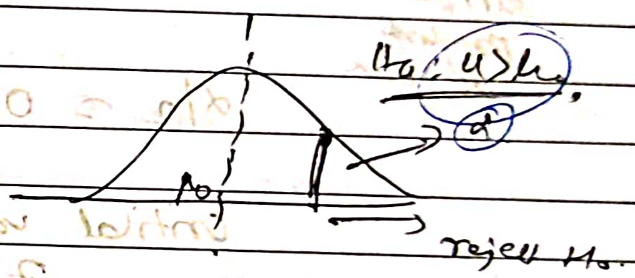
$H_a: \mu < \mu_0$, rejection area is:- $(-\infty, -c]$.

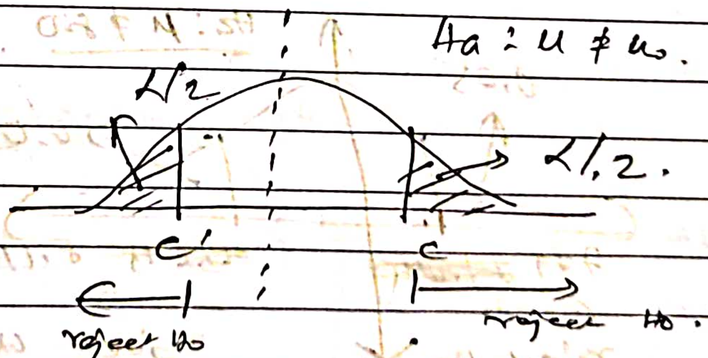
$H_a: \mu > \mu_0$, rejection area is:- (c, ∞).



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







rare event: rejection of Null hypothesis.

α: probability of rare event.

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:-

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

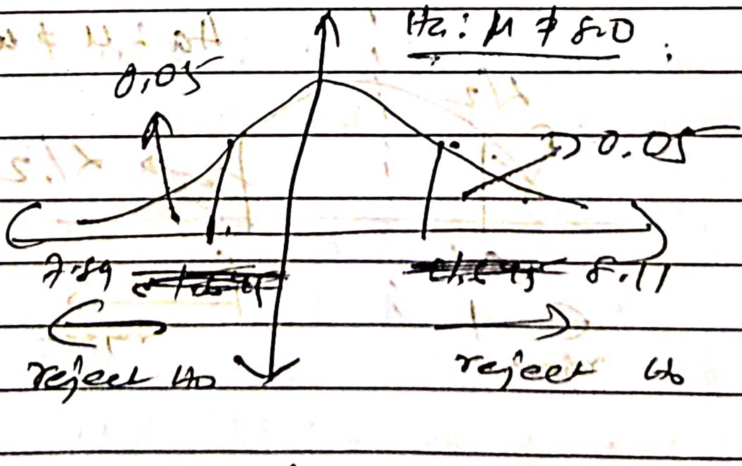
$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.15}{\sqrt{5}} = 0.067$

sample standard deviation

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

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

Z-Distribution / 2 tail



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

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

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

σ : Population standard deviation.
$\sigma_{\bar{x}}$; Sample standard deviation.

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

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

Two types of error :-
1. reject null hypothesis H_0 in favour 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.
→ probability of type I error.

Type I more dangerous

⟹ To reduce error, increase sample size.

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

~~Normal Distribution~~ ×

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	$(-\infty, -z_{\alpha/2}] \cup [z_{\alpha/2}, \infty)$	$(-\infty, -t_{\alpha/2}] \cup [t_{\alpha/2}, \infty)$

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 : $\mu = 145$ pounds
Alternative : $\mu > 145$ pounds

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

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

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

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)
(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)
(164, 164)

Sample mean $\bar{x}$ of

		$\bar{x}$
152	152	152
152	156	154
152	160	156
152	164	158
156	152	154
156	156	156
156	160	158
156	164	160
160	152	156
160	156	158
160	160	160
160	164	162
164	152	~~156~~ 158
164	156	160
164	160	162
164	164	164

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

$$\mu_{\bar{x}} = \sum \bar{x}\, P(\bar{x})$$

$$= \frac{152}{16} + \frac{154\times 2}{16} + \frac{156\times 3}{16} + \frac{158\times 4}{16} + \frac{160\times 3}{16} + \frac{162\times 2}{16} + \frac{164}{16}$$

$\mu_{\bar{x}}$: sample mean.
$\sigma_{\bar{x}}$: sample S.D.

$$\Rightarrow \mu_{\bar{x}} = 158.$$

$$= \sum \bar{x}^2 P(\bar{x})$$
$$= \frac{(152)^2 + 2\times(154)^2 + 3\times(156)^2 + 4\times(158)^2 + 3\times(160)^2 + 2\times(162)^2 + 164^2}{16}$$

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

$$\sigma_{\bar{x}} = \sqrt{\sum \bar{x}^2 P(\bar{x}) - (\mu_{\bar{x}})^2}$$
$$= \sqrt{24974 - 158^2}$$
$$= \sqrt{24974 - 24,964}$$

$$\boxed{\sigma_{\bar{x}} = \sqrt{10}}$$

152		$x - \mu$
156	mean = 158 = μ	−6
160.		−2
164.		2
		6.

$$S.D. = \sqrt{\frac{\cancel{\sum(x-\bar{x})^2} \quad \sum(x-\mu)^2}{N}}$$

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

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

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

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

LECTURE - 8/9/22

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

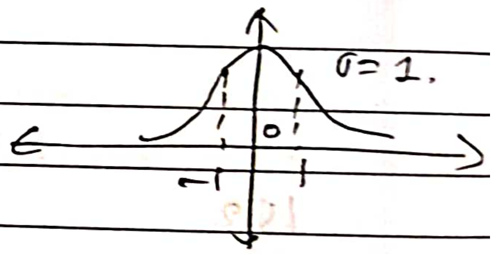
$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4180}{\sqrt{100}} = \$418$.

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) $P(Z>-1.02) = 1 - P(Z<-1.02)$

$= 1 - 0.1539$

$= 0.8461$

Date
Page

(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

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

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.

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

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

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

$= P(Z < 0.57) - P(Z < -1.67)$

$= 0.7123 - 0.0475$

$= 0.6648$

~~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.

Q. $\mu = 112$

$\sigma = 40.$

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

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

(b) $\cancel{P(110 < \bar{x} < 114)}$

$\cancel{= P\left(\frac{110-112}{40} < Z < \frac{114-112}{40}\right)}$

$\cancel{= P(-0.05 < Z < 0.05)}$

$\cancel{= P(Z < 0.05) - P(Z < -0.05)}$

$\cancel{= 0.5199 - 0.4801}$

$\cancel{= 0.0398}$

(c) $P(\bar{x} > 113)$

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

$= 1 - P\left(Z < \frac{113-112}{5.656}\right)$

$= 1 - P(Z < 0.18)$ $= 1 - 0.5714$

$= 0.4286.$

$\cancel{= 1 - P(Z < 0.03)}$

$\cancel{= 1 - 0.5120}$

$\cancel{= 0.488}$

(b) $P(110 < \bar{x} < 114)$

$= P\left(\frac{110-112}{5.656} < Z < \frac{114-112}{5.658}\right)$

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

$= P(Z < 0.35) - P(Z < -0.35)$

$= 0.6368 - 0.3632$

$= 0.2736.$

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

Date
Page

Q.
$\mu = 38,500$

$\sigma = 2500$

$n = 5$

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

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

$$= P\left(Z < \frac{36000 - 38500}{1118.03}\right)$$

$= P(Z < -2.236)$

$= P(Z < -2.24)$

$= 0.0125.$

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

Date ____
Page ____

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: $\alpha = 5\%, = 0.05$

$\boxed{\cancel{Z_\alpha = 0.5799} \text{ ~~(from table)~~}}$ ✗.

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

Alternative hypothesis: Newly developed pain reliver deliver more relief quickly

$H_0: \mu = 3.5$

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

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

$$= -1.886$$

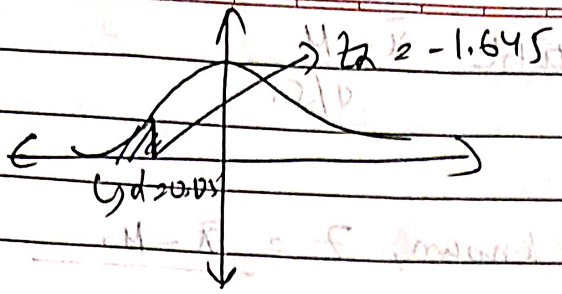
It is left-tailed test

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.

<u>Null hypothesis</u>: $H_0 : \mu = 8.1$

<u>Alternative hypothesis</u>: $H_a : \mu \neq 8.1$.

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

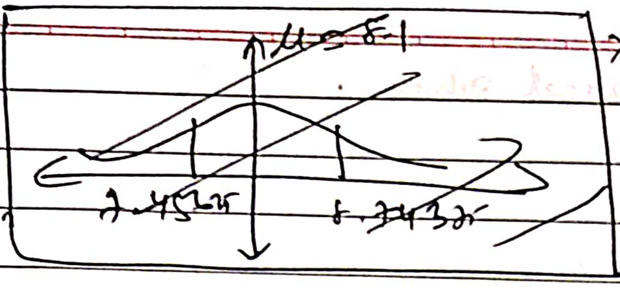
$Z_\alpha = -2.575$, ~~3.32~~ ~~7.45625~~

~~$c = 8.1 - 2.575 \times 0.25 = 0.64375$~~

~~$c' = 2.575 \times 0.25 = 0.64375$~~

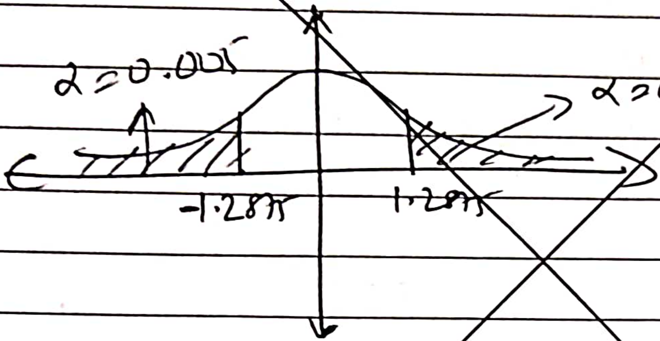
~~$c' = 8.1 + 2.575 \times 0.25 = 8.74375$~~

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



$z_\alpha = -2.575$

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

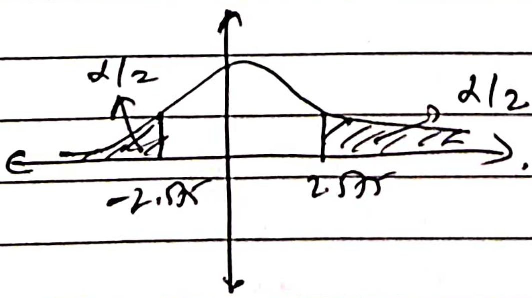


∵ given z value = 2.1908 lies in rejection region

⇒ 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 = 2.1908.
↳ Ho 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.