due to < then sign Negative / left area shaded
due to > then / Right area shaded

NORMAL PROB DISTRIBUTION:

Example:

X = Voltages $\quad\quad \mu = 220$ Volts

$\sigma = 25$ Volts

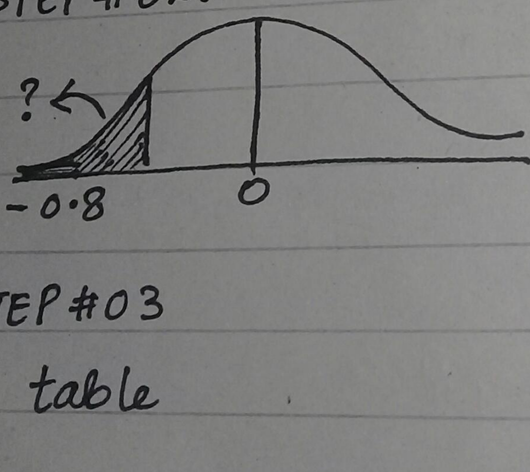
a- $P(x<200) = ?$

STEP #01:

$$z = \frac{x-\mu}{\sigma} = \frac{200-220}{25} = -0.80$$

$P(z < -0.80) = ?$

STEP #02:



STEP #03

table

II decimal

z	0.00
-0.8	0.2119

$P(z<-0.80) = 0.2119$

$P(x<200) = 0.2119$

* In rule #01 direct table value is our answer

b) P(more

STEP #01

z

P(z

Note: P

Step #02

STEP

(b) $P(\text{more than } 250) = P(x > 250)$

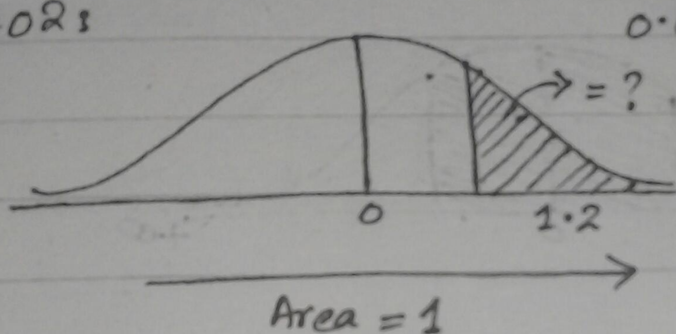
STEP #01:

$$z = \frac{x - \mu}{\sigma} = \frac{250 - 220}{25} = 1.20$$

$P(z > 1.20) = ?$

Note: $P(z > 1.20) = 1 - P(z < 1.20)$

Step #02:



Area under the curve = 1

STEP #03:

$$P(z > 1.20) = 1 - P(z < 1.20)$$

$$= 1 - 0.8849$$

$$= 0.1106 \quad 0.1151$$

$$P(X > 250) = 0.1106 \quad 0.1151$$

Jaweria Asif (9442)

Q3-

a- $P(X < 45) = ?$ $\quad \because \mu = 30$

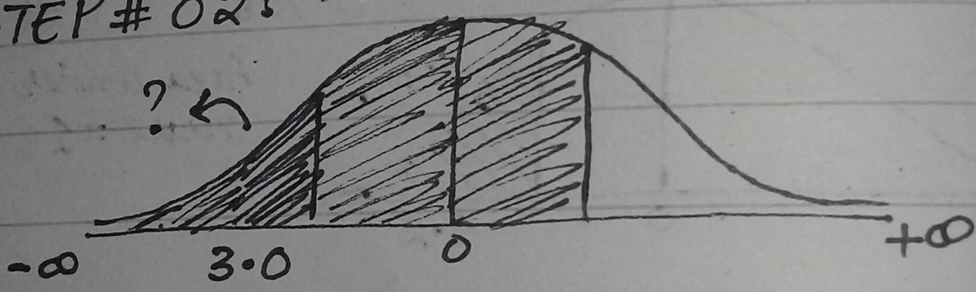
$\because \sigma = 5$

STEP # 01:

$$z = \frac{x - \mu}{\sigma} = \frac{45 - 30}{5} = 3$$

$P(z < 3.0) = ?$

b- STEP # 02:



c- STEP # 03:

$P(z < 3.0) = 0.9987$

$P(X < 45) = 0.9987$

Jaweria Asif (9442)

Q3- $P(X > 20) = ?$

$\because \mu = 30$

a- STEP #01:

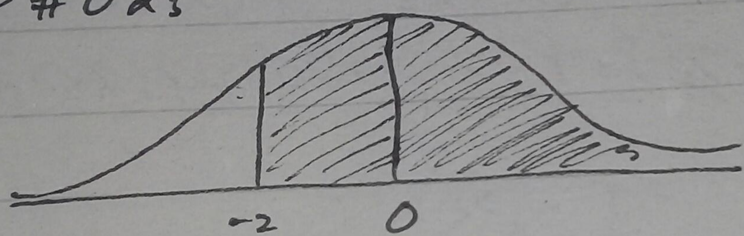
$\sigma = 5$

$$z = \frac{x - \mu}{\sigma} = \frac{20 - 30}{5} = -2$$

$P(z > -2.0) = ?$

Note: $P(z > -2.0) = 1 - P(z < -2.0)$

b- STEP #02:



c- STEP#03:

$$P(z > -2.0) = 1 - P(z < -2.0)$$
$$= 1 - 0.0228$$
$$= 0.9772$$

$$P(z > -2.0) = 0.9772$$

c- P(between 200 to 250) = P(200 < x < 25

STEP #01:

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{200-220}{25} = -0.80$$

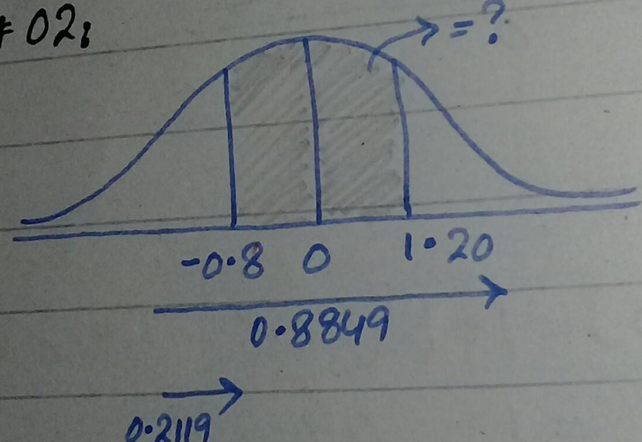
$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{250-220}{25} = 1.20$$

Rule #3:

$P(z_1 < z < z_2) = P(z < z_2) - P(z < z_1)$

$P(-0.80 < z < 1.20) = P(z < 1.20) - P(z < -0.80)$

STEP # 02:



STEP#03

$P(-0.80 < z < 1.20) = 0.8849 - 0.2119$

$= 0.673$

$P(200 < x < 250) = 0.675$

(c) $P(23 < x < 39) = ?$ $\mu = 30$ $\sigma = 5$

STEP #01:

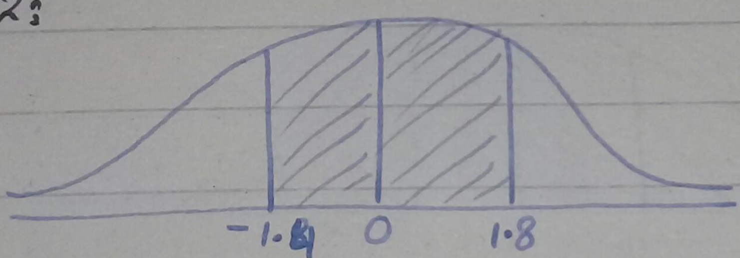
$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{23 - 30}{5} = -1.4$$

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{39 - 30}{5} = 1.8$$

$P(z_1 < z < z_2) = P(z < z_2) - P(z < z_1)$

$P(-1.4 < z < 1.8) = P(z < 1.8) - P(z < -1.4)$

STEP #02:



STEP #03:

$P(-1.4 < z < 1.8) = 0.9641 - 0.0808$

$= 0.8833$

$P(23 < x < 39) = 0.8833$

(d) $P(22 < x < 33) = ?$

$\mu = 30$

$\sigma = 5$

STEP#01:

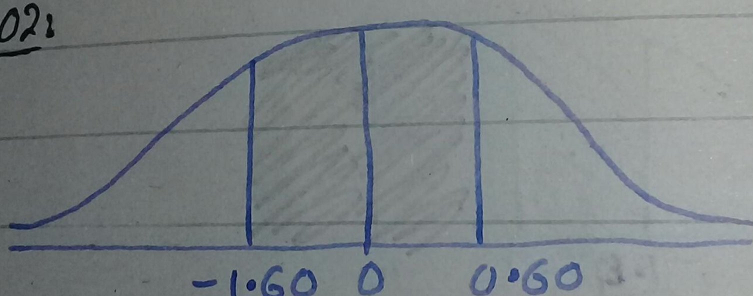
$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{22 - 30}{5} = -1.6$$

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{33 - 30}{5} = 0.6$$

$$P(z_1 < z < z_2) = P(z < z_2) - P(z < z_1)$$

$$P(-1.6 < z < 0.6) = P(z < 0.6) - P(z < -1.6)$$

STEP#02:



STEP#03:

$$P(-1.6 < z < 0.6) = 0.7257 - 0.0548$$

$$= 0.6709$$

$$P(22 < x < 33) = 0.6709$$

ⓔ $P(20 < x < 27) = ?$ $\mu = 30$

$\sigma = 5$

STEP # 01:

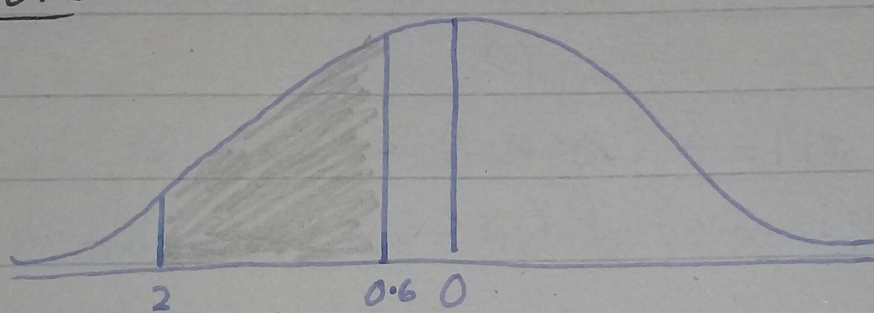
$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{20 - 30}{5} = -2$$

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{27 - 30}{5} = -0.6$$

$P(z_1 < z < z_2) = P(z < z_2) - P(z < z_1)$

$P(-2 < z < -0.6) = P(z < -0.6) - P(z < -2)$

STEP # 02:



STEP # 03:

$P(-2 < z < -0.6) = 0.2743 - 0.0228$

$= 0.2515$

$P(20 < x < 27) = 0.2515$

5- Chocolate Bar Calories: ... number of calories in a 1.5 ounce choclate bar is 225. Suppose that the distribution of Calories is approximately normal with $\sigma = 10$. Find

a- Between 200 & 220 calories

STEP #01:

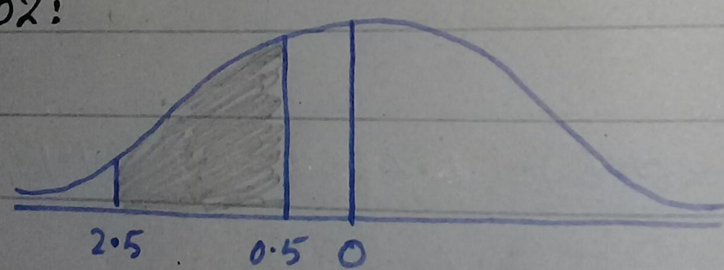
$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{200 - 225}{10} = -2.5$$

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{220 - 225}{10} = -0.5$$

$$P(z_1 < z < z_2) = P(z < z_2) - P(z < z_1)$$

$$P(-2.5 < z < -0.5) = P(z < -0.5) - P(z < -2.5)$$

STEP #02:



STEP #03:

$$P(-2.5 < z < -0.5) = 0.3085 - 0.0062$$

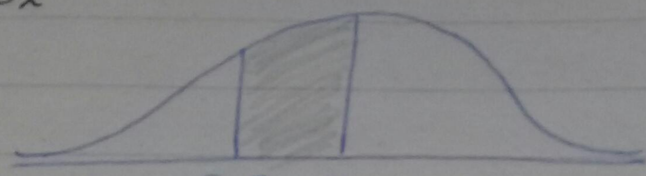
$$= 0.3023$$

$$P(200 < x < 220) = 0.3023$$

b- Less then 200 Calories

$$z = \frac{x - \mu}{\sigma} = \frac{200 - 225}{10} = -2.50$$

STEP #02



STEP #03

$P(z < -2.50) = 0.0062$

$P(x < 200) = 0.0062$

8- Doctorial Student Salaries

a- The student makes more than \$ 15,000

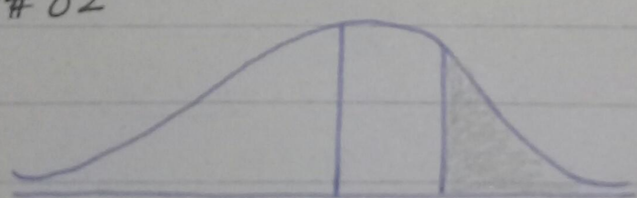
STEP #01:

$\mu = 12837$

$$z = \frac{x - \mu}{\sigma} = \frac{15000 - 12837}{1500} \qquad \sigma = 1500$$

$z = 1.442$

STEP # 02



STEP #03

$P(z > 1.442) = 1 - P(z < 1.442)$

$1 - 0.9251$

$P(x > \$15,000) = 0.0749$

Jaweria Asif

b- The student makes between \$ 13,000 & \$ 14,000

Step #01:

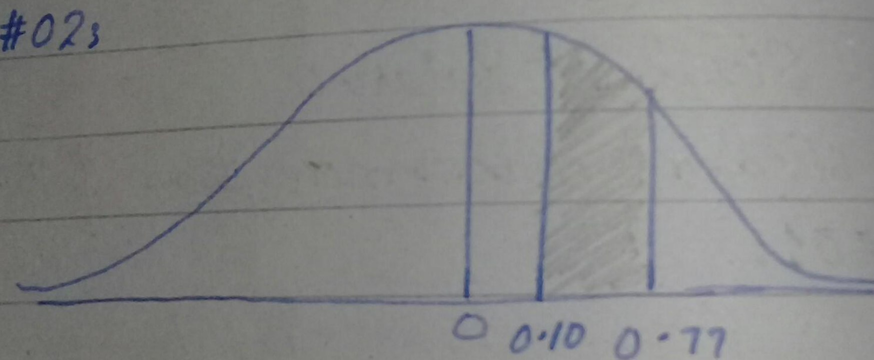
$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{13000 - 12837}{1500} = 0.10$$

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{14000 - 12837}{1500} = 0.77$$

$$P(z_1 < z < z_2) = P(z < z_2) - P(z < z_1)$$

$$P(0.10 < z < 0.77) = P(z < 0.77) - P(z < 0.10)$$

Step #02:



Step #03:

$$P(0.10 < z < 0.77) = P(z < 0.77) - P(z < 0.10)$$

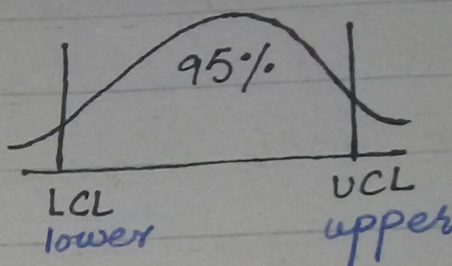
$$= 0.7580 - 0.5398$$

$$P(13000 < X < 14000) = 0.2182$$

ESTIMATION:

Estimate the population parameter value (μ, σ, etc) by the help of sample statistics

Construct the 95% confidence interval for population mean (n<30)



→ Confidence level = 95

maybe 90, 95, 99

$$\bar{x} - t\frac{s}{\sqrt{n}} < \mu < \bar{x} + t\frac{s}{n}$$

Construct the 95% interval for Population Mean

ہم t distribution اس وقت استعمال کرتے ہیں جب ہمارے پاس sample کا size less than 30 ہو -

Example:

X = marks of students

Sample size ⇒ n = 15

$\bar{x} = 72$ S = 12

t = ?

t = t distribution

t = t table

CL = 95 Page 786

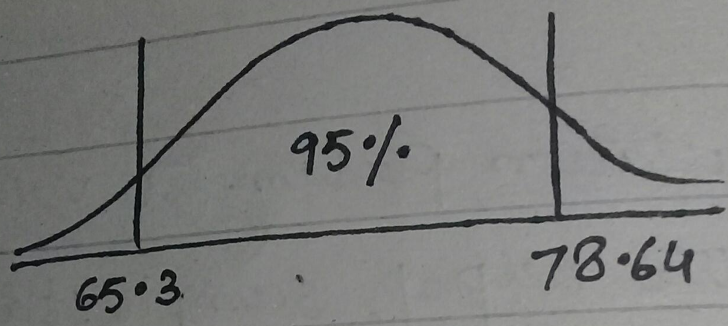
CI	95%
df (degree of Freedom) = n−1 = 15−1 = 14	→ 2.145

$$\bar{x} - t\frac{s}{\sqrt{n}} < \mu < \bar{x} + t\frac{s}{\sqrt{n}}$$

$$72 - 2.145\frac{12}{\sqrt{15}} < \mu < 72 + 2.145\frac{12}{\sqrt{15}}$$

$$65.353 < \mu < 78.64$$

Graph:



Comment:

95% Confidence Interval for the problem mean marks lies from 65.35 to 78.64 in whole class

$n > 30$ then we will use z distribution.

Assignment #02 (submit on LMS)

For Example:

x = weight of device (Kg)

$n = 36 \rightarrow \bar{x} = 5\text{ kg}$ $CL = 95\%$

$s = 2$

Step #01: since $n > 30$ use z value

$z = ?$

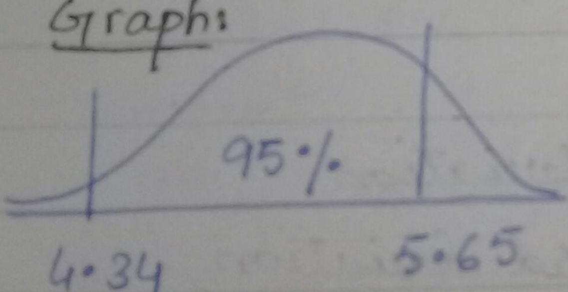
CI	95% Pg 786
$z=$	1.96

$$\bar{x} - z\frac{s}{\sqrt{n}} < \mu < \bar{x} + z\frac{s}{\sqrt{n}}$$

$$5 - 1.96\left(\frac{2}{\sqrt{36}}\right) < \mu < \bar{x} + 1.96\left(\frac{2}{\sqrt{36}}\right)$$

$$4.34 < \mu < 5.65$$

Graph:



Comment:

Weight lies from 4.34 Kg to 5.65 Kg

CONFIDENCE INTERVAL FOR POPULATION MEAN (μ)

CASE 1: $n \leq 30 \rightarrow$ use t distribution

$$\bar{x} - t\frac{s}{\sqrt{n}} < \mu < \bar{x} + t\frac{s}{\sqrt{n}}$$

CASE 2: $n > 30 \rightarrow$ use z distribution

$$\bar{x} - z\frac{s}{\sqrt{n}} < \mu < \bar{x} + z\frac{s}{\sqrt{n}}$$

CASE 3: $n < 30$ or $n \geq 30$, σ mean

$\rightarrow$ use z distribution

$$\bar{x} - z\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z\frac{\sigma}{\sqrt{n}}$$

$\bar{x}$ = Sample Mean

S = Sample Standard deviation

σ = Population Standard deviation

n = Sample Size

5. Visits to Networking Sites: A sample of 10 networking sites for a specific month has a mean of 26.1 & a std of 4.2. Find 99% confidence interval of the true mean

Data: $\bar{x} = 26.1$

Sample Size → $n = 10$

std $= 4.2$

$t = ?$

because, Sample size is < 30

CL = 99

CI	99%
df $= n-1 = 10-1 = 9$	$= 3.250$

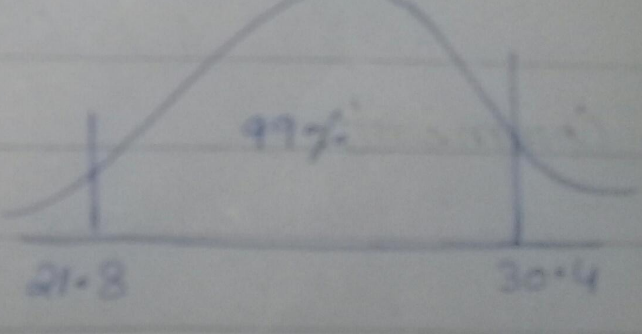
$$\bar{x} - t\frac{s}{\sqrt{n}} < \mu < \bar{x} + t\frac{s}{\sqrt{n}}$$

$$26.1 - 3.250\left(\frac{4.2}{\sqrt{10}}\right) < \mu < 26.1 + 3.250\left(\frac{4}{\sqrt{10}}\right)$$

$$21.8 < \mu < 30.4$$

Comment

99% confidence interval of networking site lie between 21.8 to 30.4



Name: ...

6. Digital Camera Prices: The prices for a particular model of digital camera with 6.0 megapixels & an optical 3X zoom lens are shown below for 10 online retailors. Estimate the True mean price for this particular model with 95% confidence. Assume the variable is Normally distributed

225 240 215 206 211 210 193 250 225 202

CL = 95%	
CI	95%
n-1	
10-1 =9	=2.262

Data:

$\bar{X} = 217.7$

Sample size → n = 10

std = 17.486

t = ?

because, Sample size is < 30

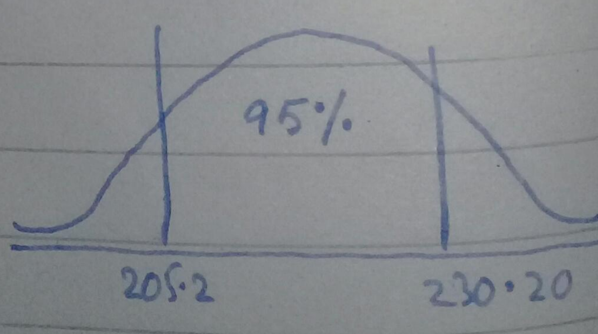
$$\bar{x} - t\frac{s}{\sqrt{n}} < \mu < \bar{x} + t\frac{s}{\sqrt{n}}$$

$$217.7 - 2.262\left(\frac{17.486}{\sqrt{10}}\right) < \mu < 217.7 + 2.262\left(\frac{17.486}{\sqrt{10}}\right)$$

$$205.2 < \mu < 230.20$$

Comment:

95% confidence of digital camera lies between 205.2 to 230.20



16. Number of Farms: A random sample of the number of farms (in thousands) in various states follows. Estimate the mean number of farms per state with 90% confidence. Assume $\sigma = 31$

47 95 54 ...... 40 50 29

CL = 90

CI	90%
df = n-1 = 31-1 = 30	= 1.645

Data:

$\bar{x} = 43.45$

$\sigma = 31$

$s = 31.266$

because sigma is given

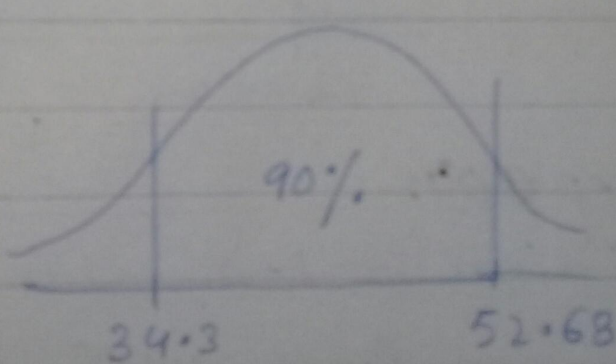
$$\bar{x} - z\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z\frac{\sigma}{\sqrt{n}}$$

$$43.45 - 1.645\left(\frac{31}{\sqrt{31}}\right) < \mu < 43.45 + 1.645\left(\frac{31}{\sqrt{31}}\right)$$

$$34.3 < \mu < 52.68$$

Comment:

90% farms per state lies in between 34.3 to 52.68



HYPOTHESIS TES

Introduction:

Verify claim (value about population) by the Help of sample data.

Claim → Accept / Reject

We have a procedure to take decission either we accept or reject claim

HYPOTHESIS TESTING FOR POPULATION MEAN (μ)

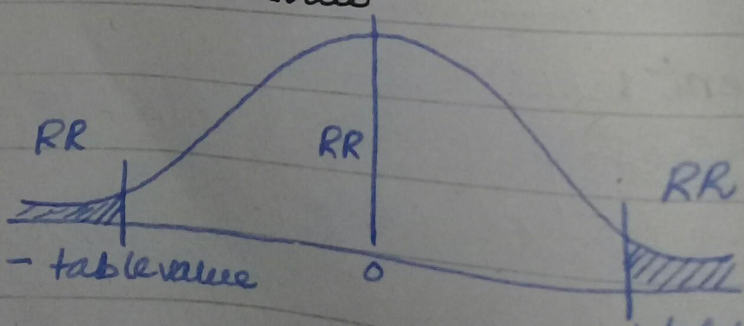
Step#01: $H_0 : \mu = \mu_0$ $\qquad H_A : \mu \neq \mu_0$

H_0: Null Hypothesis $\qquad H_A$: Alternative Hy

Step#02: $t_{cal} = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$ or $z_{cal} = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$

$n < 30 \qquad n \geq 30$

Step#03: Decission Rule



Step #04:

If calculated value lies in acceptance region than accept H_0 otherwise reject H_0.

* less than * More than * ~~eq~~Notequals to

Two tails means $\neq$

Pg# 786

X= voltages

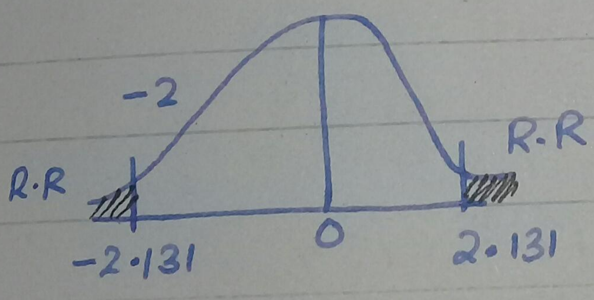
Population average voltages ⇒ $\mu_0 = 220$ select a Sample of 16 recordings &

STEP#01: $H_0: \mu = 220$ $H_A: \mu \neq 220$

STEP#02: use tcal because $n<30$

$$t_{cal} = \frac{200-220}{40/\sqrt{16}} = -2.0$$

STEP#03: Decision Rule



Two Tailed

two	$\alpha = 0.05$
df n-1 16-1	
15	2.131

STEP#04: CONCLUSION

Since tcal lies in acceptance region therefore accept Ho.

Example 2:

X = Voltages

Data remains same as Example 1

$n = 16 \rightarrow \bar{x} = 200 \quad S = 40$

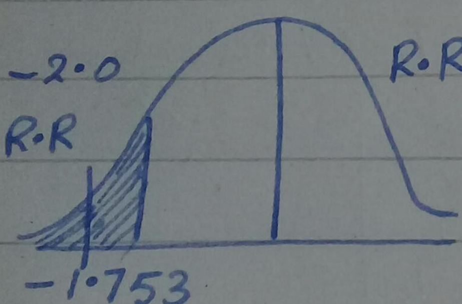
Test the Hypothesis that population average voltage is less than 220 at $\alpha = 0.05$

STEP #01: $H_0: \mu = 220$ $\quad H_n: \mu < 220$

STEP #02: Use t_{cal} because $n < 30$.

$$t_{cal} = \frac{200 - 220}{40/\sqrt{16}} = -2.0$$

STEP #03: Decission Rule: < (one tailed)



one	$\alpha = 0.05$
df	
16-1	
15	1.753

Conclusion:

Since t_{cal} lies in

Example 3:

X = voltages

Data remains same from example 1

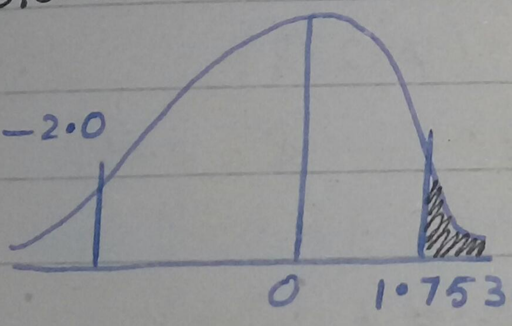
$n = 16 \rightarrow \bar{x} = 200 \cdot S = 40$

Test the Hypothesis that population average voltages is more than 220 at $\alpha = 0.05$

STEP # 01: $H_0: \mu = 220$ $\quad H_n: \mu > 220$

STEP # 02: $t_{cal} = \dfrac{200 - 220}{40/\sqrt{16}} = -2.0$

STEP # 03: Decission rule > (one tailed)



one	$\alpha = 0.05$
df	
16 - 1	
15	

selection of cities are shown below

14 55 165 9 15 66 23 30 150

22 12 13 54 73 55 41 78

Construct 90% confidence interval for the true population mean cost for a 30-second advertisment cable network

For Mean: $14+55+165+9+15+66+23+30+$

$150+22+12+13+54+73+55+41+78$

$\bar{x} = 875/17$

$\bar{x} = 51.470$

X = Cost of cable networks

Confidence level = 90%

$n = 17$

For Standard Deviation: $\dfrac{\sum_i (x_i-\bar{x})^2 + \dots + (x_n-\bar{x})^2}{n-1}$

$$s^2 = \frac{\begin{gathered}(14-51.4)^2+(55-51.4)^2+(165-51.4)^2+\\(9-51.4)^2+(15-51.4)^2+(66-51.4)^2+\\(23-51.4)^2+(30-51.4)^2+(150-51.4)+\\(22-51.4)^2+(12-51.4)^2+(13-51.4)^2+\\(54-51.4)^2+(73-51.4)^2+(55-51.4)^2+\\(41-51.4)^2+(78-51.4)^2\end{gathered}}{17-1}$$

$s^2 = 2114.52$ $\qquad \dfrac{33832.32}{16} = 2114.52$

$s = 45.9839$

n	CI = 90%
df	
n − 1	
17 − 1	
16	= 1.746

$s = 46$

$t = ?$

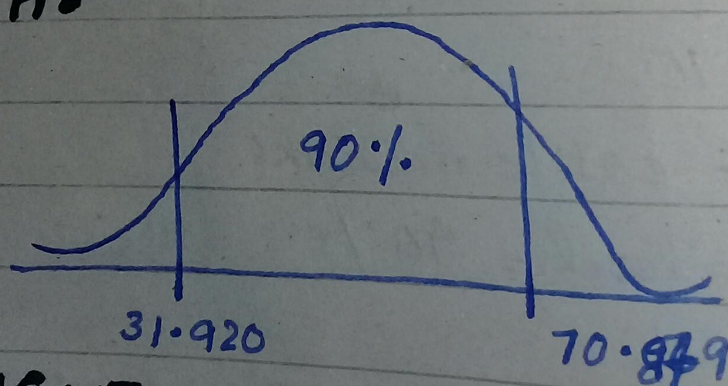
FORMULA:

$$\bar{x} - t\frac{s}{\sqrt{n}} < \mu < \bar{x} + t\frac{s}{\sqrt{n}}$$

$$51.4 - 1.746\frac{46}{\sqrt{17}} < \mu < 51.4 + 1.746\frac{46}{\sqrt{17}}$$

$$31.920 < \mu < 70.879$$

GRAPH:



COMMENT:

90% class interval for advertisment on cable network lies between 31.920 to 70.879

CORELATION AND REGRESSION:

- Statisticians use a measure called the corelation coefficient to determine two quantitative variables.
- The corelation co-efficient computed from the sample data measures the strength & direction of a linear relationship between two quantitative variables.
  +ve, −ve
- The symbol for the sample corelation coefficient is r. The symbol for the population corelation coefficient is ρ (Greek letter rho).

{Pearson Co-efficient of Correlation}

direction → +ve / −ve

PEARSON COEFFICIENT OF CORRELATION FORMULA & RANGE

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

where r is the number of data pairs.

n = sample size

Weak R

←　|　→

−1　strong Negative Linear Relationship　0　strong Positive Linear Relationship　+1

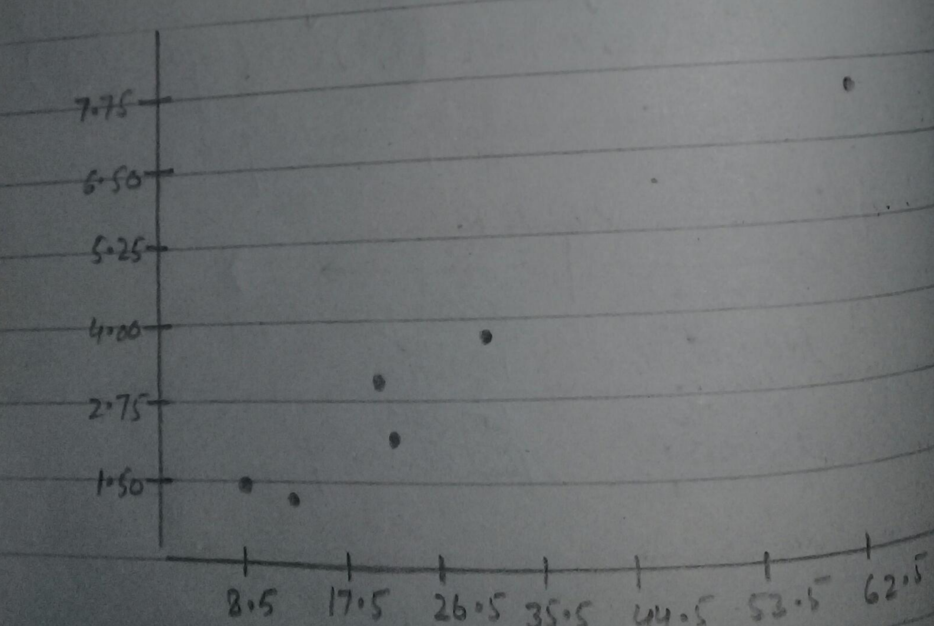
The most important thing is "Scatter" diagram

SCATTER PLOT:

→ is a graph of the ordered pair x, y of number consisting of the independ variable x & dependant variable Y

Construct a scatter plot for the data shown for car rental companies in United

Company	Cars (in thousands)	Revenue (in billion
A	63.0	$7.0
B	29.0	3.9
C	20.8	2.1
D	19.1	2.8
E	13.4	1.4
F	8.5	1.5



When x increase

The trend of th

linear relat

Company	Cars (x) in thousand
A	63.0
B	29.0
C	20.8
D	19.1
E	13.4
F	8.5
	$\Sigma x = 153$

Substitute in t

$r =$

$\sqrt{[n(\Sigma}$

$r = \frac{6 (68}{\sqrt{(6(}}$

$r = 0.982$

Positive

if r value is m

LINEAR REGRE

$f(x) =$ line

Simple linear

Pg # 536

When x increase Y also increases → POSITIVE
The trend of the data is considered to be linear relationship between x & y

Company	Cars (x) in thousands	Revenue (Y) in billions	xy	x^2	y^2
A	63.0	7.0	441.00	3969.00	49.00
B	29.0	3.9	113.10	841.00	15.21
C	20.8	2.1	43.68	432.64	4.41
D	19.1	2.8	53.48	364.81	7.84
E	13.4	1.4	18.76	179.56	1.96
F	8.5	1.5	12.75	72.25	2.25
	$\Sigma x = 153.8$	$\Sigma y = 18.7$	$\Sigma xy = 682.77$	$\Sigma x^2 = 5859.26$	$\Sigma y^2 = 80.67$

Substitute in the formula & solve for r

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

$$r = \frac{6(682.77) - (153.8)(18.7)}{\sqrt{(6(5859.26) - (153.8)^2)(6(80.67) - (18.7)^2}}$$

$r = 0.982$

Positive strong relation

if r value is negative → strong negative

LINEAR REGRESSION:

f(x) = linear relationship

Simple linear regression equation

1- Scatter Plot
2- Comment
3- Coefficient of correlation
4- Interpret
5- Simple linear regression
6- Predict

$y = \overset{\nearrow \text{slope}}{m}x + c$

$y' = \underset{\text{intercept}}{\underline{\underline{a}}} + \underline{\underline{b}}x$ slope

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

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

$$a = \frac{(18.7)(5859.26) - (153.8)(682.77)}{(6)(5859.26) - (153.8)^2}$$

$a = 0.396$

$$b = \frac{6(682.77) - (153.8)(18.7)}{6(5859.26) - (153.8)^2}$$

$b = 0.106$

$y' = a + bx$ is

$y' = 0.396 + 0.106x$

Let $x = 15$

$y' = 0.396 + 0.106(15)$

$y' = 1.986$

3- Absenc

to see

in her

The da

No of abse

Final g

x
10
12
2
0
8
5

$\Sigma x = 3$

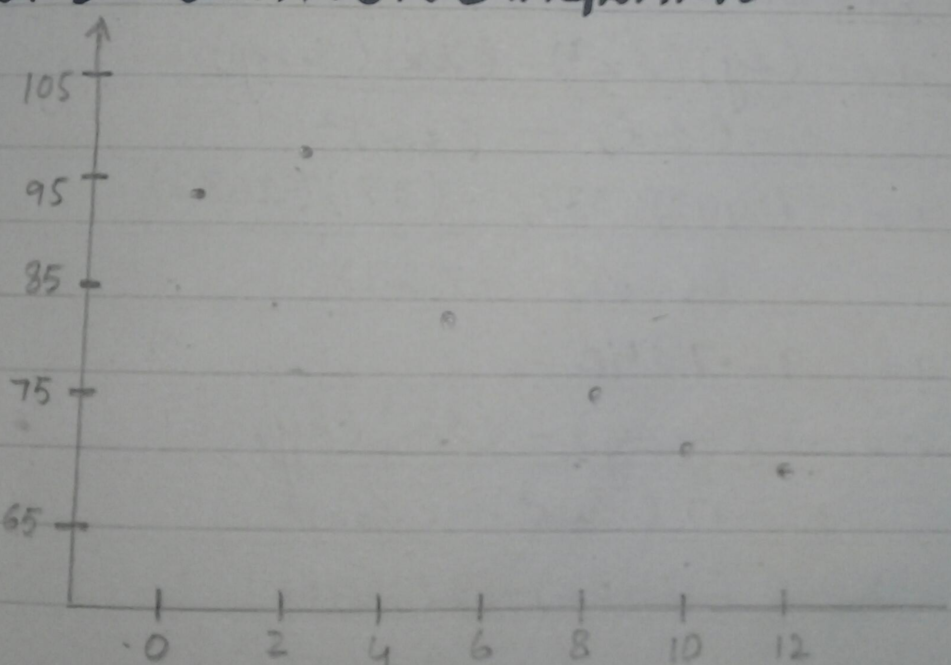
STE

33- Absences and Final Grades: An educator wants to see how the number of absences for a student in her class effects the students' final grade. The data obtained from a sample are shown

No of absences X	10	12	2	0	8	5
Final grade Y	70	65	96	94	75	82

x	Y	xy	x^2	y^2
10	70	700	100	4900
12	65	780	144	4225
2	96	192	4	9216
0	94	0	0	8836
8	75	600	64	5625
5	82	410	25	6724
$\Sigma x=37$	$\Sigma y=482$	$\Sigma xy=2682$	$\Sigma x^2=337$	$\Sigma y^2=39526$

STEP 1: SCATTER DIAGRAM:



STEP#02: INTERPRET CO-EFFICIEN OF CORELATION:

Scatter plot indicicates strong negative relationship between variables

For r:

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

$$r = \frac{6(2682) - (37)(482)}{\sqrt{[6(337) - (37)^2][6(39526) - (482)^2}}$$

$$r = \frac{-1742}{\sqrt{(653)(4832)}}$$

$$r = -0.9806$$

3-FOR SIMPLE LINEAR REGRESSION:

$$Y' = a + bx$$

Here, $$a = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$a = \frac{(482)(337) - (37)(2682)}{6(337) - (37)^2}$$

$$a = 96.7840$$

Here, $$b = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$b = \frac{6(2682) - (37)(482)}{6(337) - (37)^2}$$

$$b = -2.6676$$

Now, $Y' = a -$

$Y' = 9$

$Y' =$

4- PREDICT:

→ Final grade would be

Here $x = 7$

5- STANDAR

Sest =

Sest

Sest

Sest