Assignment - 2

① In each of the following situations, state whether it is a correctly stated hypotheses

1. $H_0: \mu = 25$ $\quad H_1: \mu \neq 25$
2. $H_0: \sigma > 25$ $\quad H_1: \sigma = 10$
3. $H_0: \bar{x} = 50$ $\quad H_1: \bar{x} \neq 50$
4. $H_0: p = 0.1$ $\quad H_1: p = 0.5$
5. $H_0: s = 30$ $\quad H_1: s > 30$

① $H_0: \mu = 25$ $\quad H_1: \mu \neq 25$

Ans: YES

H_1 contradicts H_0

Null $\rightarrow$ $\mu = 25$

Alternate $\rightarrow$ μ is either > 25 or < 25 – two failed

② $H_0: \sigma > 25$, $\quad H_1: \sigma = 10$ $\quad$ Ans: NO

$H_1 \rightarrow$ should be $\sigma \leq$ Value.

③ $H_0: \bar{x} = 50$ $\quad H_1: \bar{x} \neq 50$ $\quad$ Ans: NO

- Hypotheses are always statements about population.
- Here $\bar{x}$ is the sample mean.

④ $H_0: p = 0.1$ $\quad H_1: p = 0.5$ $\quad$ Ans: NO

- p Values are different

⑤ $H_0: s = 30$ $\quad H_1: s > 30$ $\quad$ Ans: NO

- Here s is the sample Variance, not population Variance.

② The college book store tells perspective students that the average cost of its text books is Rs: 52 with std deviation of Rs 4.50. A group of smart statistics students think that the average cost is higher. To test the book stores claim against their alternative, the student will select a random sample size of size 100. Assume that mean from their random sample is 52.80. Perform hypothesis test at 5% significance levels.

$\mu = 52$ $\sigma = 4.50$

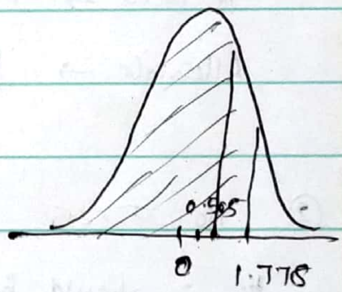
H_0: Average cost higher $\mu > 52$

H_1: $\mu <= 52$

$n = 100$ $\bar{x} = 52.80$

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{4.5}{\sqrt{100}} = \underline{\underline{0.45}}$$

$$Z = \frac{X - \mu_0}{\sigma} = \frac{52.80 - 52}{0.45} = \underline{\underline{1.778}}$$



$\alpha = 5\% \rightarrow 0.05$

$Z(0.05) = 0.505$

Accept ~~Reject~~ Null hypothesis.

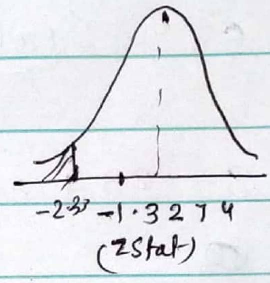
③ A certain chemical pollutant in River has been constant for several years with $\mu = 34$ ppm and std deviation $\sigma = 8$ ppm. A group of factory representatives whose companies discharge liquids into the river is now claiming that they have lowered the average with improved filtration devices. A group of environmentalists will test to see if this is true at 1% level of significance. Assume sample size 50 & mean 32.5 ppm

$\mu = 34$; $\sigma = 8$ $n = 50$ $\bar{x} = 32.5$

$H_0: \mu < 34$ (Factory's claim is true)

$H_1: \mu >= 34$ (" False)

$SE = 8/\sqrt{50} = 1.13$



$Z_{stat} = \frac{32.5 - 34}{1.13} = -1.3274$

$Z_{Test} (\alpha = 0.01) = -2.33$

$\rightarrow$ ~~Reject~~ Accept Null hypothesis.

④ Based on population figures and other general information on the US population, suppose it has been estimated that on average a family of 4 in US spend \$1135 annually on dental expenditure. To test this accuracy 22 families of 4 are randomly selected from the population in the area of the country and a log is kept of the family's dental expenditure for one year. The resulting data are given below. $\alpha: 0.05$, test the hypothesis

1008, 812, 1117, 1323, 1308, 1415, 831, 1021, 1287, 851, 930, 730, 699, 872, 913, 944, 954, 987, 1695, 995, 1003, 994

$\mu = 1135$, $n = 22$

$$\bar{x} = \frac{22689}{22} = \underline{\underline{1031.31}}$$

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1} = 55\,153.580$$

$s = 234.85$

~~SD~~

$H_0: \mu = 1135$

$H_1: \mu \neq 1135$

$$SE = \frac{234.85}{\sqrt{22}} = \underline{\underline{50.06}}$$

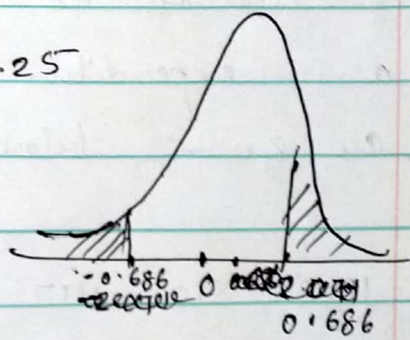
Sample size < 30, σ not given } t-test

$$T_{stat} = \frac{1031.31 - 1135}{50.06} = \underline{\underline{-2.071}}$$

$dof = 22 - 1 = \underline{\underline{22}}$

$\alpha = 0.5 \rightarrow$ Two tail test $\alpha/2 = 0.25$

$T(\alpha/2 = 0.25) = \underline{\underline{0.686}}$



~~Accept Ho~~ Reject Ho

⑤ In a report prepared by the economic research department, the average annual family income is \$48,432. What do you conclude about the validity of the report if a random sample of 400 families shows and average income \$48,574 with Std dev 2000 ?

$\mu = 48,432$

$n = 400$

$\bar{x} = 48,574$

$S = 2000$

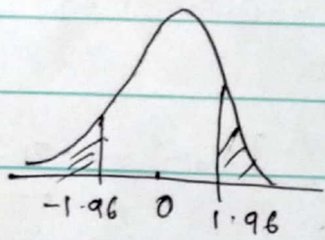
$H_0 : \mu = 48\,432$

$H_1 \quad \mu \neq 48\,432$

Let $\alpha = 5\%$ $\qquad \alpha/2 = 0.025$

dof $= 400 - 1 = 399$

$\boxed{\text{Sample size} < 30 \} \text{ T test}}$



$$T_{stat} = \frac{48574 - 48\,432}{100} = \underline{\underline{1.42}}$$

$T(\alpha/2) = 1.96$

~~Reject Ho~~ . Accept H_0

⑥ Suppose that in past three years the average price per square foot of warehouse in US has been \$32.28. A national real estate investor wants to determine whether that figure changed now. The investor here's a researcher who randomly samples 19 warehouses and find that mean price per square foot is $31.67 with std dev $1.29. At 5% significance level what is the conclusion?

$\mu = 32.28$

H_0: Not changed $\quad \mu = 32.28$

H_1: Changed $\quad \mu \neq 32.28$

$n = 19$

$\bar{x} = 31.67$

$S = 1.29$

$\left.\begin{array}{l} n < 30 \\ \sigma \text{ not given} \end{array}\right\}$ T test.

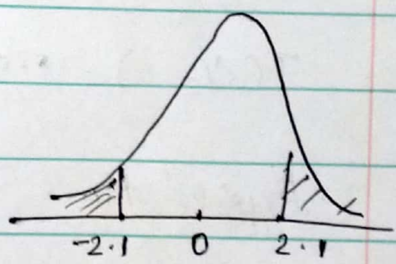
$\alpha = 5\%$ $\quad \alpha/2 = 0.025$

$SE = 1.29 / \sqrt{19} = \underline{\underline{0.296}}$

$$T_{stat} = \frac{31.67 - 32.28}{0.296} = \underline{\underline{-2.1034}}$$

dof = 18

$T(\alpha/2 = 0.025) \rightarrow 2.101$



~~Accept~~ Reject Ho

⑦ Accept Region | Sample size | α | β at $\mu=52$ | β at $\mu=50.5$

Accept Region	Sample size	α	β at $\mu=52$	β at $\mu=50.5$
$48.5 < x < 51.5$	10	?	?	μ

α = probability of TYPE I error → Reject H_0 when it is true.

Let $\mu = 50$, $\sigma = 2.5$

H_0: $\mu = 50$

H_1: $\mu \neq 50$

$$SE = \sigma/\sqrt{n} = \frac{2.5}{\sqrt{10}} = \underline{\underline{0.79}}$$

$$\alpha = P(\bar{x} < 48.5 \,;\, \mu=50) + P(x > 51.5 \,;\, \mu=50)$$

$$Z_1 = \frac{48.5-50}{0.79} \qquad Z_2 = \frac{51.5-50}{0.79}$$

$$Z_1 = -1.90 \qquad Z_2 = 1.90$$

$$\alpha = P(Z < -190) + P(Z > 1.90)$$

$$= 0.0287 + (1 - 0.9712)$$

$$= \underline{\underline{0.0574}}$$

β = probability of TYPE II error

$$\beta_{(\mu=52)} = P(48.5 \le x \le 51.5 \,;\, \mu=52)$$

$$Z_1 = \frac{48.5-52}{0.79} = -4.43 \qquad Z_2 = \frac{51.5-52}{0.79} = \underline{\underline{-0.63}}$$

$\beta = P(-4.43 \leq z \leq -0.63)$

$= P(z \leq -0.63) - P(z \leq -4.43)$

$= 0.2643 - 0.0000$

$= \underline{\underline{0.2643}}$

$\underline{\beta \text{ at } \mu = 50.5}$

$\beta = P(48.5 \leq \bar{x} \leq 51.5 ; \mu = 50.5)$

$z_1 = \dfrac{48.5 - 50.5}{0.79}$ $\qquad$ $z_2 = \dfrac{51.5 - 50.5}{0.79}$

$= \underline{\underline{-2.53}}$ $\qquad$ $= \underline{\underline{1.27}}$

$\beta = P(-2.53 \leq z \leq 1.27)$

$= P(z < 1.27) - P(z \leq -2.53)$

$= 0.8980 - 0.0057$

$= \underline{\underline{0.8923}}$

(8) Find the t-score for a sample size 16 taken from population with mean 10 when the sample mean is 12 and sample std dev is 1.5

$n = 16$ $\qquad$ $s = 1.5$

$\mu = 10$

$\bar{x} = \cancel{10}\ 12$

$SE = \frac{1.5}{\sqrt{16}} = \underline{\underline{0.375}}$

$T_{stat} = \frac{12-10}{0.375} = \underline{\underline{5.33}}$

(9) Find the t-score below which we can expect 99% of sample means will ~~fail~~ fall if samples of size 16 are taken from normally distributed population.

$\alpha = 1\% = 0.01 \qquad \alpha/2 = 0.005$

$dof = 15$

$t(\alpha/2) = \underline{\underline{2.947}}$

(10) If random sample of size 25 drawn from normal population gives a mean of 60 and a standard deviation of 4, find the range of t-scores where we can expect to find the middle 95% of all sample means. Compute probability that $(-t_{0.05} < t < t_{0.10})$

$n = 25$

$\bar{x} = 60 \qquad dof = 24$

$s = 4 \qquad \alpha = 0.05 \quad \alpha/2 = 0.025$

$SE = 4/\sqrt{25} = \underline{\underline{0.8}}$

$t_{stat} (\alpha/2 = 0.025) = 2.064$

$\underline{P(-t_{0.05} < t < t_{0.10})}$

$t_{0.05} \rightarrow 2.064$

$t_{0.10} \rightarrow 1.711$

$$P = 2.064 - 1.711 = \underline{0.353}$$

(11) Is there any evidence to conclude that the number of people travelling from bangalore to chennai is different from the no: of people travelling from Bangalore to Hosur in a week.

<u>population 1</u>: Bangalore to chennai.

$n_1 = 1200$

$x_1 = 452$

$S_1 = 212$

<u>population 2</u>: Banglore to Hosur

$n_2 = 800$

$x_2 = 523$

$S_2 = 185$

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

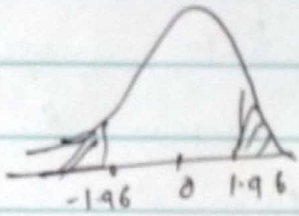
$$t_{stat} = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

$$= \frac{452 - 523}{\sqrt{\frac{212^2}{1200} + \frac{185^2}{800}}} = -7.93$$

$T_{table}(\alpha = 0.025, df = ?)$.

$$df = \frac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\frac{s_1^2/n_1}{n-1} + \frac{s_2^2/n_2}{n-1}}$$

$$= \frac{80.23^2}{3.461} = 1859.8 \approx 1000.$$



$t(0.025, df = 1000) = \underline{\underline{1.9612}}$

Reject H_0

(12) Is there any evidence to conclude the no: of people preferring duracell battery is different from no: of people preferring Energizer battery, given the following.

Population 1: Duracell

$n_1 = 100$

$x_1 = 308$

$S_1 = 84$

population 2: Energizer

$n_2 = 100$

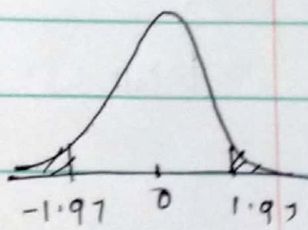
$x_2 = 254$

$S_2 = 67$

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

$$t_{stat} = \frac{308 - 254}{\sqrt{84^2/100 + 67^2/100}} = \underline{\underline{5.02}}$$



$df = 188$

$t(0.025\ df = 188) = 1.97 \rightarrow$ Reject H_0

(3) Pooled estimate of population variance.

Does the data provide sufficient evidence to conclude that average percentage increase in the price of sugar ~~of~~ differs when it is sold at two different prices?

<u>Population 1:</u> ~~percent~~

<u>Population 2</u>

price of Sugar = 27.50

$n_1 = 14$

$x_1 = 0.317\%$

$S_1 = 0.12\%$

price of Sugar = 20.00

$n_2 = 9$

$x_2 = 0.21\%$

$S_2 = 0.11\%$

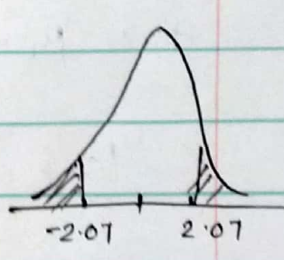
$H_0: \mu_2 > \mu_1$

$H_1: \mu \leq \mu_1$

$$S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}} = \sqrt{\cancel{14 \times 8 \times 4^2}}$$

$$= \sqrt{\frac{13 \times 0.12^2 + 8 \times 0.11^2}{21}} = \sqrt{0.013} = 0.114$$

$$t_{stat} = \frac{0.317 - 0.21}{\sqrt{\frac{0.114^2}{14} + \frac{0.114^2}{9}}} = 2.27$$



$t(\alpha = 0.025\ df = 21) = 2.0796$

Reject H_0

14 The manufacturers of compact disk players want to test whether a small price reduction is enough to increase sales of their product. Is there any evidence that the small price reduction is enough to increase sales of compact disk players.

Population 1: Before reduction

$n_1 = 15$

$x_1 = 6598$ Rs

$S_1 = 844$ Rs

Population 2: After reduction

$n_2 = 12$

$x_2 = 6870$ Rs

$S_2 = 669$ Rs.

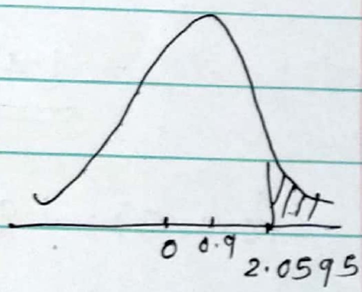
$H_0: \mu_2 > \mu_1$

$H_1: \mu_2 \leq \mu_1$

$$S_p = \sqrt{\frac{14 \times 844^2 + 11 \times 669^2}{25}} = 771.90$$

$$t_{stat} = \frac{6870 - 6598}{\sqrt{\frac{771.9^2}{15} + \frac{771.9^2}{12}}} = \underline{\underline{0.909}}$$

$t(\alpha = 0.05, df = 25) = 2.0595$



Accept H_0

15 Comparison of two population proportions when the hypothesized difference is zero. Carry out a two tailed test of the equality of banks share of the car loan market in 1980 and 1995.

Population 1 : 1980

$n_1 = 1000$

$x_1 = 53$

$\hat{p}_1 = 0.53$

Population 2 : 1985

$n_2 = 100$

$x_2 = 43$

$\hat{p}_2 = 0.43.$

$H_0 : \mu = \mu_2$

$H_1 : \mu \neq \mu_2$

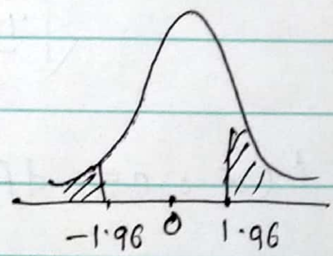
Combined population $(\hat{p}) = \dfrac{53+43}{200} = \underline{\underline{0.48}}$

$$t_{stat} = \frac{(\hat{p}_1 - \hat{p}_2) - (\pi_1 - \pi_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$\pi_1, \pi_2 \rightarrow$ population proportion

$$= \frac{0.53 - 0.43}{\sqrt{0.48(1-0.48)\left(\frac{1}{100} + \frac{1}{100}\right)}}$$

$$= \underline{\underline{1.42}}$$



$t(0.025) = 1.96$

- Accept H_0

⑥ Carry out a one tailed test to determine whether the population proportion of traveler's check buyer's who buy at least $2500 in checks when Sweepstakes prizes are offered as at least 10% higher than the proportion of such buyers when no Sweepstakes are on.

population 1: With Sweepstakes

$n_1 = 300$

$x_1 = 120$

$\hat{p}_1 = 0.40$

population 2: No sweepstakes

$n_2 = 700$

$x_2 = 140$

$\hat{p}_2 = 0.20$

Hypothesis

$H_0 : \pi_1 > \pi_2 \quad (M_1 - M_2 > 0.1)$ Given π 10% ↑

$H_1 : \pi_1 \leq \pi_2 \quad (M_1 - M_2 \leq 0.1)$

Combined proportion $\frac{120 + 140}{300 + 700} = \frac{260}{1000} = \underline{\underline{0.26}}$

$$Z_{Stat} = \frac{(0.40 - 0.20) - 0.1}{\sqrt{0.26(1-0.26)\left(\frac{1}{300} + \frac{1}{700}\right)}}$$

$$= \frac{0.1}{\sqrt{0.00090}} = \underline{\underline{3.33}}$$

$Z_{(\alpha = 0.05)} = -1.65$

Reject H_0

(17) A die is thrown 132 times with the following results. Number turned up: 1, 2, 3, 4, 5, 6.
Frequency: 16, 20, 25, 14, 29, 28
Is it the die unbiased? Consider degrees of freedom as $\hat{p} - 1$.

	observed	Expected $\left(\frac{132}{6} = 22\right)$	O - E
1	16	22	-6
2	20	22	-2
3	25	22	3
4	14	22	-8
5	29	22	7
6	28	22	6

$H_0 = O = E$ (die unbiased)
$H_1 : O \neq E$

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{-6^2 + -2^2 + 3^2 + -8^2 + 7^2 + 6^2}{22}$$

$$= \underline{9}$$

$$dof = \hat{p} - 1$$

$$\hat{p} = \frac{x}{n} = \frac{16 + 20 + 25 + 14 + 29 + 28}{6} = 3.67$$

$dof = 2.67 \sim 3$.
$\alpha = 5\%$ $F(\alpha = 0.05, df = 3) = \underline{7.81}$

Accept $\rightarrow$ Die is unbiased

(18) In a certain, there are about one million eligible voters. A simple random sample of 10,000 ~~elig~~ eligible voters was chosen to study the relationship between gender and participation in the last election. The results are summarized in the following 2X2 contigency table.

	Men	Women
Voted	2792	3591
Not Voted	1486	2131

Check whether being a man or woman is independent of having voted in the last ~~column~~ election. Is gender and voting independent?

	Men	Woman	
Voted	2792	3591	6383 (Total Voted)
Not voted	1486	2131	3617 (Total Not Voted)
Total	4278	5722	

Expected (Men who voted) $= \frac{\text{all who voted} \times \text{Total men}}{\text{Total number}}$

$= \cancel{68}\ \frac{6383 \times 4278}{10,000} = 2730.6474$

Expected (women who voted) $= \cancel{\frac{3617 \times 4278}{10,000}} =$

$= \frac{6383 \times 5722}{10,000} = 3652.3526$

Scanned with CamScanner

Expected Men (Not Voted) $= \frac{3617 \times 4278}{10,000} = 1547$

Expected women (Not voted) $= \frac{3617 \times 5722}{10,000} = 2070$.

$$\chi^2 = \Sigma \frac{(O-E)^2}{E}$$

$$= \frac{(2792-2731)^2}{2731} + \frac{(3591-3652)^2}{3652} + \frac{(1486-1547)^2}{1547} + \frac{(2131-2070)^2}{2070}$$

$$= \underline{\underline{6.607}}$$

df = (no: of rows − 1) × (no: of columns − 1)
$= (2-1)(2-1) = \underline{1}$

$\chi^2 (\alpha = 0.05, df = 1) = 3.84$

Reject Ho: Voting depend on gender.

Scanned with CamScanner

(19) A Sample of 100 voters are asked which of four candidates they would vote for an election. The number of supporting each candidate is given below.

Higgins	Reardon	White	Charlton
41	19	24	16.

Do the data suggest that all candidates are equally popular? (Chi-square = 14.96, with 3 df, $p < 0.05$)

$\chi^2_{stat} = 14.96$

dof = 3

$p < 0.05$

$$\frac{41+19+24+16}{4} = \frac{100}{4} = 25$$

Observed	Expected	$(O-E)^2$
41	25	16^2
19	25	-6^2
24	25	-1^2
16	25	-9^2

$$\chi^2_{stat} = \frac{(O-E)^2}{E} = \frac{256+36+1+81}{25} = \underline{14.96}$$

$df = 4-1 = 3$ $\quad P < 0.05$ → $0.025 \rightarrow 9.35$

$0.1 \rightarrow 11.34$

Reject Ho: Candidates are not equally popular.

(20) Children of three ages are asked to indicate their preferences for three photographs of adults. Do the data suggest that there is a significant relationship between age and photograph preferences?

Chi square = 29.6, df = 4 $\hat{p} < 0.05$

		A	B	C
Age	5 - 6 yrs	18	22	20
	7 - 8 yrs	2	28	40
	9 - 10 yrs	20	10	40

Ans:

	A	B	C	Row total
5 - 6 y	18	22	20	60
7 - 8 y	2	28	40	70
9 - 10 y	20	10	40	70
column total	40	60	100	200

Expected $E = \dfrac{\text{row total} \times \text{colum total}}{200}$

Expected

12	18	6
14	21	35
14	21	35

$(O - E)^2$

6^2	4^2	14^2
-12^2	7^2	5^2
6^2	-11^2	5^2

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \underline{58.93}$$

Given $\chi^2 = 29.6$

$P < 0.05$ — $0.001 \longrightarrow 18.467$

$\downarrow$

9.48

Reject hypothesis Ho

There is a significant relationship between age of child and photograph preference.

(21) A study of conformity using Asch paradigm involved two conditions: One where one confederate supported the true judgement and another where no confederate gave the correct response.

	Support	No support
Conform	18	40
Not conform	32	10

Is there any significant diff: between the support and no support conditions in the frequency with which individuals are likely to conform
(χ^2 - 19.87 df = 1 $p < 0.05$)

$p = 0.05 \rightarrow 3.841$

$p = 0.001 \rightarrow 10.828$

Reject → There is difference b/w support and no support cond^n.

(22) We want to test whether short people differ with their leadership qualities.

$\chi^2 = 10.71$ $df = 2$ $p < 0.01$

	Short	Tall
Leader	12	32
Follower	22	14
Unclassifiable	9	6

$P = 0.01 \rightarrow 9.210$

$$P<0.01 \begin{cases} P = 0.005 \rightarrow 10.587 \\ P = 0.002 \rightarrow 12.429 \\ P = 0.001 \rightarrow 13.816 \end{cases}$$

$P \leq 0.002$ Accept.

At 0.01 significant level, we conclude that there is a relationship b/w height & leadership qualities.

(23) Each respondent in the current population Survey of March 1993 was classified as employed, unemployed, or outside the labor force. The results for men in California age 35-44 can be cross tabulated by marital status as follows.

Scanned with CamScanner

	Married	Widowed/separated	Never Married	RS
Employed	679	103	114	896
Unemployed	63	10	20	93
Not in labor force	42	18	25	85
col sum	784	131	159	1074

Men of different marital status seem to have different distribution of labor force status.

Expected

654	109	133
68	11	14
62	10	13

$(O-E)^2$

25^2	-6^2	-19^2
-5^2	-1^2	6^2
20^2	-8^2	-10^2

$\chi^2 = 28.51$

$\chi^2 (\alpha = 0.05, df = 4) = \underline{9.487}$

Reject Null Hypothesis

There is a relationship between row & columns.