Problem Statement 1:

In each of the following situations, state whether it is a correctly stated hypothesis testing problem and why?

1. H0: μ = 25, H1: μ ≠ 25

2. H0: σ > 10, H1: σ = 10

3. H0: x = 50, H1: x ≠ 50

4. H0: p = 0.1, H1: p = 0.5

5. H0: s = 30, H1: s > 30

Problem 1

(a) $\underbrace{H_0 : \mu = 25}_{\text{null Hypothesis}}, \underbrace{H_1 : \mu \neq 25}_{\text{Alternative Hypothesis}}$

Yes, Alternative Hypothesis specified ~~real~~ values of μ that could be either greater or less than 25, it is called a two-sided alternate hypothesis

(b) $H_0 : \sigma > 10$, $H_1 : \sigma = 10$

No, Ho claim equality and when H_1 stated with < sign - so, Ho can be taken as ≥.

(c) $H_0 : x = 50$, $H_1 : x \neq 50$

NO, Hypothesis always statements about population/distribution under study, not statements about sample

(d) $H_0 : p = 0.1$, $H_1 : p = 0.5$

NO, Values in both Hypothesis are different

(e) $H_0 : s = 30$, $H_1 : s > 30$

NO, Hypotheses are always statements about population under study, not statements about sample.

Problem Statement 2:

The college bookstore tells prospective students that the average cost of its textbooks is Rs. 52 with a standard deviation of Rs. 4.50. A group of smart statistics students thinks that the average cost is higher. To test the bookstore's claim against their alternative, the students will select a random sample of size 100. Assume that the mean from their random sample is Rs. 52.80. Perform a hypothesis test at the 5% level of significance and state your decision.

Problem 2

8

$H_0 : \mu = 52$

$H_1 : \mu \neq 52$

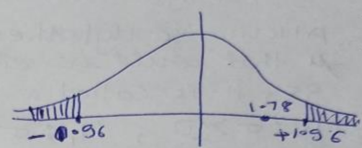
significance level = 5%

$n = 100$

$\mu = 52$

$S = 4.5$

$x = 52.8$



$$Z = \frac{x - \mu}{S_{\bar{x}} / \sqrt{n}} = \frac{52.8 - 52}{4.5/\sqrt{100}} = \frac{0.8}{0.45} = 1.78$$

Critical Value of $z = \pm 1.96$

↳ Since, $z = 1.78$ falls in acceptance region

↳ We accept the H_0

↳ Hence, the mean average cost of its textbooks is \$52.

Problem Statement 3:

A certain chemical pollutant in the Genesee River has been constant for several years with mean μ = 34 ppm (parts per million) and standard deviation σ = 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 the 1% level of significance. Assume \ that their sample of size 50 gives a mean of 32.5 ppm. Perform a hypothesis test at the 1% level of significance and state your decision.

Problem 3

$H_0 : \mu = 34$

$H_1 : \mu \neq 34$

significane level = 1%

$\mu = 34$

$S = 8$

$n = 50$

$\bar{x} = 32.5$

$$Z = \frac{\bar{x} - \mu}{S_{\bar{x}} / \sqrt{n}} = \frac{32.5 - 34}{8/\sqrt{50}} = -1.33$$

Critical value of $z = \pm 2.58$

Since, $z = -1.33$ falls in acceptance region, we accept the null hypothesis.

Hence, they have lowered the average discharge with improved filteration devices is true.

Problem Statement 4:

Based on population figures and other general information on the U.S. population, suppose it has been estimated that, on average, a family of four in the U.S. spends about $1135 annually on dental expenditures. Suppose further that a regional dental association wants to test to determine if this figure is accurate for their area of country. To test this, 22 families of 4 are randomly selected from the population in that area of the country and a log is kept of the family's dental expenditure for one year. The resulting data are given below. Assuming, that dental expenditure is normally distributed in the population, use the data and an alpha of 0.5 to test the dental association's hypothesis.

1008, 812, 1117, 1323, 1308, 1415, 831, 1021, 1287, 851, 930, 730, 699,

872, 913, 944, 954, 987, 1695, 995, 1003, 994

$\boxed{\text{Problem 4}}$

$H_0: \mu = 1135$, $H_1: \mu \neq 1135$

significance level = 5%

$\mu = 1135$

$s = 240.37$

$n = 22$

$\bar{x} = 1031.32$

$$z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{1031.32 - 1135}{240.37/\sqrt{22}} = -2.02$$

Critical value of $z = \pm 1.96$

Since, $z = -2.02$ falls in rejection region

We reject the H_0.

Hence, the average dental expenses for population is not accurate for their area

Problem Statement 5:

In a report prepared by the Economic Research Department of a major bank the Department manager maintains that the average annual family income on Metropolis is $48,432. What do you conclude about the validity of the report if a random sample of 400 families shows and average income of $48,574 with a standard deviation of 2000?

Problem 5

$H_0: \mu = 48,432$, $H_1: \mu \neq 48,432$

significance level = 10%

$\mu = 48,432$

$s = 2000$

$n = 400$

$x = 48,574$

$$z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{48574 - 48432}{2000/\sqrt{400}} = 1.42$$

critical value of $z = \pm 1.645$

Since, $z = 1.42$ falls in acceptance region,
we accept the null hypothesis

Problem Statement 6:

Suppose that in past years the average price per square foot for warehouses in the United States has been \$32.28. A national real estate investor wants to determine whether that figure has changed now. The investor hires a researcher who randomly samples 19 warehouses that are for sale across the United States and finds that the mean price per square foot is \$31.67, with a standard deviation of \$1.29. assume that the prices of warehouse footage are normally distributed in population. If the researcher uses a 5% level of significance, what statistical conclusion can be reached? What are the hypotheses?

Problem 6

$H_0: \mu = 32.28$

$H_1: \mu \neq 32.28$

significance level = 5%

$\mu = 32.28$

$s = 1.29$

$n = 19$

$\bar{x} = 31.67$

$$z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{31.67 - 32.28}{1.29/\sqrt{19}} = -2.1$$

critical value of $z = \pm 1.96$

Since, $z = -2.1$ falls in rejection region,
we reject the H_0

Hence, avg price per square foot warehouse has charged now.

Problem Statement 7:

Fill in the blank spaces in the table and draw your conclusions from it.

Problem 7

Acceptance Region	Sample size	α	β at $\mu=52$	β at $\mu=50.5$
$48.5<\bar{x}<51.5$	10	0.0576	0.2643	0.8923
$48<\bar{x}<52$	10	0.0114	0.5000	0.9705
$48.81<\bar{x}<51.19$	16	0.0576	0.0966	0.8606
$48.42<\bar{x}<51.58$	16	0.0114	0.2515	0.9578

Problem Statement 8:

Find the t-score for a sample size of 16 taken from a population with mean 10 when the sample mean is 12 and the sample standard deviation is 1.5.

Problem 8

$n = 16$

$\mu = 10$

$\bar{x} = 12$

$s = 1.5$

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{12-10}{1.5/\sqrt{16}}$$

$$= 5.33$$

Problem Statement 9:

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

Problem 9

$1-\alpha = 0.99$

$\alpha = 0.01$

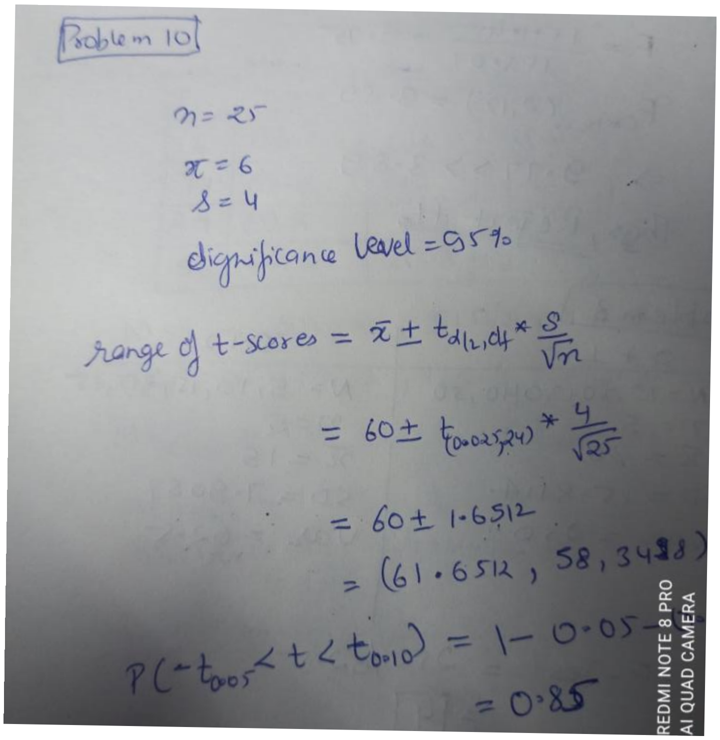
$df = n-1 = 16-1 = 15$

$t_{0.99} = -t_{0.01}$

$t_{0.99} = -t_{0.01} = -2.602$

Problem Statement 10:

If a random sample of size 25 drawn from a 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 the probability that (−t0.05 <t<t0.10).



Problem 10

$n = 25$

$x = 6$

$s = 4$

significance level = 95%

range of t-scores $= \bar{x} \pm t_{\alpha/2, df} * \frac{s}{\sqrt{n}}$

$= 60 \pm t_{(0.025, 24)} * \frac{4}{\sqrt{25}}$

$= 60 \pm 1.6512$

$= (61.6512, 58.3488)$

$P(-t_{0.05} < t < t_{0.10}) = 1 - 0.05 -$

$= 0.85$

Problem Statement 11:

Two-tailed test for difference between two population means

Is there evidence to conclude that the number of people travelling from Bangalore to Chennai is different from the number of people travelling from Bangalore to Hosur in a week, given the following:

Population 1: Bangalore to Chennai n1 = 1200

x1 = 452

s1 = 212

Population 2: Bangalore to Hosur n2 = 800

x2 = 523

s2 = 185

Problem 11

$n > 30$, hence apply z-test

$n_1 = 1200$ | $n_2 = 800$
$\bar{x}_1 = 452$ | $\bar{x}_2 = 523$
$\sigma_1 = 212$ | $\sigma_2 = 185$

$$\text{Standard Error} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{(212)^2}{1200} + \frac{(185)^2}{800}}$$
$$= 8.96$$

$$z\text{-test} = \frac{(\bar{x}_1 - \bar{x}_2)}{se} = \frac{(452-523)}{8.96} = -7.926$$

Acc to 2-tail, $\alpha = 0.05$; $\alpha/2 = 0.025$

$Z_{0.025} = -2.81$

z-test $< Z_{0.025}$, H_0 will be rejected

Problem Statement 12:

Is there evidence to conclude that the number of people preferring Duracell battery is different from the number of people preferring Energizer battery, given the following:

Population 1: Duracell

n1 = 100

x1 = 308

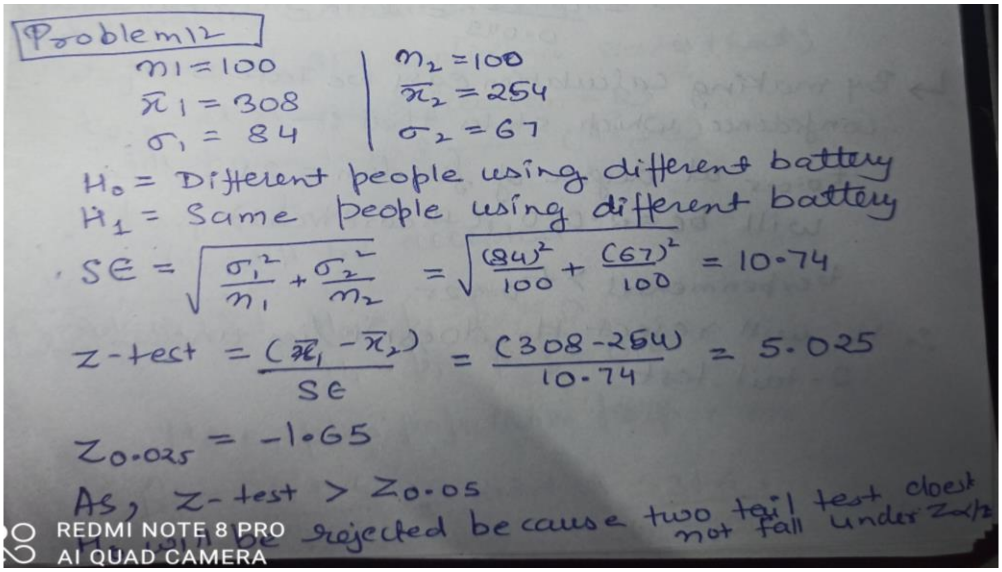
s1 = 84

Population 2: Energizer

n2 = 100

x2 = 254

s2 = 67



Problem 12

$n_1 = 100$ | $n_2 = 100$

$\bar{x}_1 = 308$ | $\bar{x}_2 = 254$

$\sigma_1 = 84$ | $\sigma_2 = 67$

H_0 = Different people using different battery

H_1 = Same people using different battery

$$SE = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{(84)^2}{100} + \frac{(67)^2}{100}} = 10.74$$

$$z\text{-test} = \frac{(\bar{x}_1 - \bar{x}_2)}{SE} = \frac{(308 - 254)}{10.74} = 5.025$$

$Z_{0.025} = -1.65$

As, z-test > $Z_{0.05}$

H_0 can be rejected because two tail test does not fall under $Z_{\alpha/2}$

Problem Statement 13:

Pooled estimate of the population variance

Does the data provide sufficient evidence to conclude that average percentage

increase in the price of sugar differs when it is sold at two different prices?

Population 1: Price of sugar = Rs. 27.50 n1 = 14

x1 = 0.317%

s1 = 0.12%

Population 2: Price of sugar = Rs. 20.00 n2 = 9

x2 = 0.21%

s2 = 0.11%

Problem 13

$n_1 = 14$ | $n_2 = 9$

$\bar{x}_1 = 0.317$ | $\bar{x}_2 = 0.21$

$s_1 = 0.12$ | $s_2 = 0.11$

H_0 = price of sugar increase

H_1 = price of sugar will not increase

$$s_{12} = \sqrt{\frac{(n_1-1)x_1^2 + (n_2-1)x_2^2}{n_1+n_2-2}}$$

$$= \sqrt{\frac{(14-1)(0.12)^2 + (9-1)(0.11)^2}{14+9-2}}$$

$$SE = S_{12} * \sqrt{(1/n_1) + (1/n_2)} = 0.049$$

$$t\text{-Stats} = \frac{(\bar{x}_1 - \bar{x}_2)}{SE} = \frac{0.317 - 0.21}{0.049}$$

$$= \frac{0.107}{0.049} = 2.183 \text{ (experimental)}$$

→ By making calculation easy we take 95% confidence which states that:—

$t_{0.025}$ at degree of freedom = 21 will be = 2.080 (theoretical)

$t_{experimental} > t_{0.025}$

∴ Ho will reject Ho doesn't lies under 2-tail test

Problem Statement 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 evidence that the small price reduction is enough to increase sales of compact disk players?

Population 1: Before reduction

n1 = 15

x1 = Rs. 6598 s1 = Rs. 844

Population 2: After reduction n2 = 12

x2 = RS. 6870

s2 = Rs. 669

Problem 14

$n_1 = 15$ | $n_2 = 12$

$\bar{x}_1 = 6598$ | $\bar{x}_2 = 6870$

$s_1 = 844$ | $s_2 = 669$

H_0 : Small price reduction is enough to inc sales

H_1 : small price reduction is not enough to inc sales

$$s_{12} = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{(15-1)844^2 + (12-1)(669)^2}{15+12-2}}$$

$$SE = S_{12} * \sqrt{1/n_\alpha + \frac{1}{n_\beta}} = 289.96$$

$$t\text{-stats} = |(\bar{x}_1 - \bar{x}_2)| / SE$$

$$= \frac{(|6598 - 6870|)}{289.96} = \frac{272}{289.96} = 0.91 \longrightarrow (exp)$$

↳ By making calculation easy we will take reference of 95% Confidence states that;

$t_{0.05}$ at degree of freedom = 25

will be = 1.708 ⟶ (theo)

$t_{exp} < t_{0.05}$

$\therefore$ H_0 will be accepted

Problem Statement 15:

Comparisons 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

n1 = 1000

x1 = 53

p 1 = 0.53

Population 2: 1985

n2 = 100

x2 = 43

p 2= 0.53

Problem 15

$H_0 = \hat{p}_1 - \hat{p}_2 = 0$

$H_1 = \hat{p}_1 - \hat{p}_2 \neq 0$

Here, $p \rightarrow$ population proportion

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{53+43}{100+100} = 0.48$$

$$z\text{-test} = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{(\hat{p})(1-\hat{p})(1/n_1)+(1/n_2)}}$$

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

$$= 1.415$$

According to H_0 this is 2-tail test. So, if experimatal value < theoretical value then accept the H_0

Problem Statement 16:

Carry out a one-tailed test to determine whether the population proportion of traveler's check buyers 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

n1 = 300

x1 = 120

p = 0.40

Population 2: No sweepstakes n2 = 700

x2 = 140

p 2= 0.20

$\boxed{\text{Problem 16}}$ $n_1 = 300$ | $n_2 = 700$

$x_1 = 120$ | $x_2 = 140$

$\hat{p}_1 = 0.40$ | $\hat{p}_2 = 0.20$

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{120 + 140}{300 + 700} = 0.26$$

$$Z\text{-test} = \sqrt{\frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\frac{(\hat{p}_1)(1-\hat{p}_1)}{n_1} + \frac{(\hat{p}_2)(1-\hat{p}_2)}{n_2}}}$$

$$= \sqrt{\frac{(0.4 - 20) - (0.10)}{\frac{(0.4)(1-0.4)}{300} + \frac{(0.2)(1-0.2)}{700}}}$$

$$= \frac{0.10}{0.032} = 3.118$$

According to Ho this is 1-tail test;

95% of significance level, $Z_{0.05} = 1.645$

$Z_{test} > Z_{0.05}$; $\boxed{\text{Accept the Ho}}$

Problem Statement 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 the die unbiased? Consider the degrees of freedom as p − .

Problem 17

Take Hypothesis that die is unbaised, If so, the probability of any one of six number is 1/6

Expected frequency of any one number $= 132 \times \frac{1}{6}$

$= 22$

Now, obsered frequency with expected fequency and value of χ^2 as below:-

No. of turned up	Observed frequency (O_i)	Expected frequency (E_i)	$O_i - E_i$	$(O_i - E_i)^2 / E_i$
1	16	22	−6	36/22
2	20	22	−2	4/22
3	25	22	3	9/22
4	14	22	−8	64/22
5	29	22	7	49/22
6	28	22	6	36/22

$$\Sigma[(O_i - E_i)^2 / E_i] = 9$$

Hence, $\chi^2 = 9$ & $DF = n - 1 = 6 - 1 = 5$

Significance level = 5%, table value of $\chi^2 = 11.071$

$\Rightarrow$ Calculated value < table value

Thus, supports hypothesis & concluded that die is unbaised

Problem Statement 18:

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

We would want to check whether being a man or a woman (columns) is independent of having voted in the last election (rows). In other words, is “gender and voting independent”?

Problem 18

H_0: Sex is independent of Voting

H_1: Sex and Voting are dependent

Expected table

	Men	Women	Total
Voted	2731	3652	6383
Didn't vote	1547	2070	3617
Total	4278	5722	10000

$\chi^2 = C_{11} + C_{12} + C_{21} + C_{22}$

$C_{11} = (2792-2731)^2/2731$

$C_{12} = (3591-3652)^2/3652$

$C_{21} = (1486-1547)^2/1547$

$C_{22} = (2131-2070)^2/2070$

$\chi^2 = 6.584$

$df = (2-1)(2-1) = 1$

from χ^2 table, $3.84 < \chi^2 < 6.64$

Thus, $1\% < \text{P-value} < 5\%$

and we reject the NULL $\therefore$

Sex and voting are dependent in this town

Problem Statement 19:

A sample of 100 voters are asked which of four candidates they would vote for in an election. The number supporting each candidate is given below:

Do the data suggest that all candidates are equally popular? [Chi-Square = 14.96, with 3 df, p 0.05 .

Problem 19

H_0 : There is no preference for any candidate

H_1 : There is preference for any candidate

$Ef = 100/4 = 25$/candidate

O	E	$(O-E)^2/E$
41	25	10.24
19	-6	1.44
24	~~36~~	0.04
16	1.44	3.24
Total		14.96

$\chi^2 = 14.96$

$df = 4-1 = 3$

↳ Critical value of χ^2 for 0.05 SL & 3 DF $= 7.82$

$\Rightarrow$ Obtain value $>$ χ^2 value critical

Thus, χ^2 value as large as ours will occur by chance only about once in 100-trials. Results ~~do~~ not by chand and voters do not prefer four candidates equally.

Problem Statement 20:

Children of three ages are asked to indicate their preference for three photographs of adults. Do the data suggest that there is a significant relationship between age and photograph preference? What is wrong with this study? [Chi-Square = 29.6, with 4 df: $p < 0.05$].

Problem 20

Calculate Expected Value :-

Expected	Data	A	B	C
age of child	5-6 yrs	12	18	30
	7-8 yrs	14	21	35
	9-10 yrs	14	21	35

$$\text{Expected Value } (\chi^2) = \frac{(18-12)^2}{12} + \frac{(2-14)^2}{14} + \frac{(20-14)^2}{14} + \frac{(22-18)^2}{18} + \frac{(28-21)^2}{21} + \frac{(10-21)^2}{21} + \frac{(20-30)^2}{30} + \frac{(40-35)^2}{35} + \frac{(40-35)^2}{35} = 29.60$$

$df = (3-1)(3-1) = 4$

At $\hat{p} < 0.05 \sim = 0.001$

Hence, $\chi^2_{0.001}$ at $df = 4$ will be 18.47

$\chi^2_{test} > \chi^2_{0.001}$

Result, Reject H_0.

There is no significance relationship between age and photograph preference

Problem Statement 21:

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

Is there a significant difference between the "support" and "no support" conditions in the frequency with which individuals are likely to conform? [Chi-Square = 19.87, with 1 df: $p < 0.05$].

Problem 21

O	E	$(\lvert O-E \rvert - 0.5)^2/E$
18	29	3.8
40	29	3.8
32	21	5.25
10	21	5.25
Total		18.10

$df = (2-1)(2-1)$
$= 1$

$\Rightarrow$ Obtain $\chi^2 >$ Critical χ^2 for 0.001 significance level

$\therefore$ there is a significante different b/w "support" and "no support"

Problem Statement 22:

We want to test whether short people differ with respect to their leadership qualities (Genghis Khan, Adolf Hitler and Napoleon were all stature-deprived, and how many midget MP's are there?) The following table shows the frequencies with which 43 short people and 52 tall people were categorized as "leaders", "followers" or as "unclassifiable". Is there a relationship between height and leadership qualities?

[Chi-Square = 10.71, with 2 df: $p < 0.01$].

Problem 22

	short	tall	total
leader	19.92	24.08	44
follower	16.29	19.71	36
Unclassifier	6.79	8.21	15
Total	43	52	95

$\chi^2 = 3.146 + 2.602 + 1.998 + 1.652$
$+ 0.720 + 0.595$
$= 10.712$

$df = 2$

$\Rightarrow$ observed $\chi^2 >$ tabular χ^2 at 0.01 SL
Thus, there seems to be a relationship between height and leadership qualities

Problem Statement 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:

Men of different marital status seem to have different distributions of labor force status. Or is this just chance variation? (you may assume the table results from a simple random sample.)

Problem23

Expected Values :-

	Married	Wid or's	Never married
Employed	654	109	133
Un Employed	68	11	14
Not in laborforce	62	10	13

$$\chi^2 = \frac{(679-654)^2}{654} + \frac{(25-13)^2}{13} = \boxed{30.96}$$

df $=(3-1)(3-1)=4$, tabular $\chi^2 = \boxed{13.28}$

Since, 30.96 >>> 13.28 and Reject H_0

Marital Status seems to be related to Job Status in this town