Stats Assignment 2

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

Answer:-

Yes: - The statement H0: μ = 25 is called the null hypothesis. This is a claim that is initially assumed to be true. The statement H1: μ ≠ 25 is called the alternative hypothesis and it is a statement that contradicts the null hypothesis. Because the alternative hypothesis specifies values of μ that could be either greater or less than 25, it is called a two- sided alternative hypothesis.

2. H0: σ > 10, H1: σ = 10

Answer:-

No: - We will always state the null hypothesis as an equality claim. However when the alternative hypothesis is stated with the < sign, the implicit claim in the null hypothesis can be taken as ≥. And when the alternative hypothesis is stated with the > sign, the implicit claim in the null hypothesis can be taken as ≤.

3. H0: x = 50, H1: x ≠ 50

Answer: -

No it is important to remember that hypothesis is always statement about the population or distribution under study, not statement about the sample.

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

Answer:-

No values in both hypotheses are different.

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

Answer:-

No hypothesis is always statements about the population or distribution under study, not statements about the 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.

Answer:-

- **Hypothesis**
  H0: $\mu = 52$, H1: $\mu \neq 52$

- **Significance level**
  The Significance level is 5%

- **Determination of a suitable test static**
  $\mu = 52$
  S=4.50
  n=100
  x=52.8

  $z = (x - \mu) / (\sigma / \sqrt{n})$
  $=52.8-52/4.5/\sqrt{100}$
  $=.8/.45 \quad =1.78$

The critical value of z is -1.96 and +1.96

The critical value is z=± 1.96 for a two-tailed test at 5% level of significance. Since, the computed value of z=1.78 falls in acceptance region, we accept the null hypothesis. Hence, the mean average cost of its textbooks is Rs 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.

Answer:-

- **Hypothesis**
  H0: μ = 34, H1: μ ≠ 34

- **Significance level**
  The Significance level is 1%

- **Determination of a suitable test static**
  μ = 34
  S= 8
  n=50
  x=32.5

  $z = (x - \mu) / (\sigma / \sqrt{n})$
  =32.5-34/8/√50
  =-1.5/1.13 =-1.33

The critical value of z is -2.58 and +2.58

The critical value is z=± 2.58 for a two-tailed test at 1% level of significance. Since, the computed value of z=-1.33 falls in acceptance region, we accept the null hypothesis. Hence, there claim that they have lowered the average discharge with improved filtration 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

Answer:-

- **Hypothesis**
  H0: μ = 1135, H1: μ ≠ 1135

- **Significance level**
  The Significance level is 5%

- **Determination of a suitable test static**
  μ = 1135
  S=240.37
  n=22
  x=1031.32

  $z = (x - \mu) / (\sigma / \sqrt{n})$
  =1031.32-1135/240.37/√22
  =-103.68/51.25 =-2.02

The critical value of z is -1.96 and +1.96

The critical value is z=± 1.96 for a two-tailed test at 5% level of significance. Since, the computed value of z=2.57 falls in acceptance region, we reject the null hypothesis. Hence, the average dental expenses for the 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?

Answer:-

- **Hypothesis**
  H0: μ = 48,432, H1: μ ≠ 48,432

- **Significance level**
  The Significance level is 10%

- **Determination of a suitable test static**
  μ = 48,432
  S=2000
  n=400
  x=48,574

$z = (x - \mu) / (\sigma / \sqrt{n})$
$=48,574-48,432/2000/\sqrt{400}$
$=142/100 \quad =1.42$

The critical value of z is -1.645 and +1.645

The critical value is z=± 1.645 for a two-tailed test at 5% level of significance. Since, the computed value of 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?

Answer:-

- **Hypothesis**
  H0: $\mu = 32.28$, H1: $\mu \neq 32.28$

- **Significance level**
  The Significance level is 5%

- **Determination of a suitable test static**
  $\mu = 32.28$
  S=1.29
  n=19
  x=31.67

  $z = (x - \mu) / (\sigma / \sqrt{n})$
  $=31.67-32.28/1.29/\sqrt{19}$
  $=-0.61/.29 \quad =-2.1$

The critical value of z is -1.96 and +1.96

The critical value is z=± 1.96 for a two-tailed test at 5% level of significance. Since, the computed value of z=-2.1 falls in 43 rejection region, we reject the null hypothesis. Hence, the average price per square foot for warehouses has changed now.

Problem Statement 7:

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

Acceptance region	Sample size	α	β at $\mu = 52$	β at $\mu = 50.5$
$48.5 < \bar{x} < 51.5$	10			
$48 < \bar{x} < 52$	10			
$48.81 < \bar{x} < 51.9$	16			
$48.42 < \bar{x} < 51.58$	16			

Answer:-

Acceptance Region	Sample Size	α	β at μ = 52	β at μ = 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.

Answer:-

```
import math
n=16
s=1.5
μ=10
x¯=12

t=(x¯-μ)/(s/math.sqrt(n))
t
```

```
out>> 5.333333333333333
```

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.

Answer:-

$1-\alpha = 0.99$ $df = n-1$

$\alpha = 0.01$ $df = 15$

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

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 $(-t_{0.05} < t < t_{0.10})$.

Answer:-

$n=25$

$\mu=60$

$s=4$

$t = \bar{x} - \mu/(s/\sqrt{n})$

$-t_{0.05} < t < t_{0.10} = (\bar{x}-60)/(4\sqrt{25})$

$0.985 = \bar{x}-60/4/5$ => $0.985 = \bar{x}-60/0.8$ => 60.785 (Sample Means)

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

Answer:-

Population 1: Bangalore to Chennai n1 = 1200

x1 = 452

s1 = 212

Population 2: Bangalore to Hosur n2 = 800

x2 = 523

s2 = 185

H0: 800 ≤ μ1- μ2 ≥1200

H1: 800 > μ1- μ2 <1200

$Z=(x1-x2)-(\mu1-\mu2)/\sqrt{(\sigma_1^2/n1)+(\sigma_2^2/n2)}$

$= (452-523)-(212-185)/\sqrt{(212^2/1200)+(185^2/800)}$

$= -71-27/\sqrt{37.45+42.78} \Rightarrow -44/8.95$

= -4.91

Critical p(-4.91>z+4.91)=0.201

H0 may not be rejected at any common level of significance.

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

Answer:-

Population 1: Duracell

n1 = 100

x1 = 308

s1 = 84

Population 2: Energizer

n2 = 100

x2 = 254

s2 = 67

H0: μ1- μ2 ≤45

H1: μ1- μ2 >45

$Z=(x1-x2)-(\mu1-\mu2)/\sqrt{(\sigma_1^2/n1)+(\sigma_2^2/n2)}$

$= (308-254)-(45)/\sqrt{(84^2/100)+(67^2/100)}$

$=-9/\sqrt{115.45}$

$= 9/10.75 \quad =0.838$

Critical pz+0.838)=0.201

H0 may not be rejected at any common level of significance.

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%

Answer:-

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%

H0: μ1- μ2 =0

H1: μ1- μ2 ≠0

Df=(n_1+n_2-2) = 14+9-2 = 21

$Z=(x1-x2)-(\mu1-\mu2)/\sqrt{(n_1-1)s_1^2+(n_2-1)s_2^2/n_1+n_2-2)(1/n_1+1/n_2)}$

= (0.107)/ √0.00247

=0.107/0.0497

=2.154

Critical point: $t_{0.025}$=0.08

H0 may not be rejected at the 5% level of significance.

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

Answer:-

Population 1: Before reduction

n1 = 15

x1 = Rs. 6598 s1 = Rs. 844

Population 2: After reduction n2 = 12

x2 = RS. 6870

s2 = Rs. 669

H0: $\mu 1 - \mu 2 \leq 0$

H1: $\mu 1 - \mu 2 > 0$

$Df=(n_1+n_2-2) = 15+12-2 = 25$

$Z=(x2-x1)-(\mu 2-\mu 1)/\sqrt{(n_1-1)s_1^2+(n_2-1)s_2^2/n_1+n_2-2)(1/n_1+1/n_2)}$

$= (272)/\sqrt{89375.25}$

$=272/298.96$

$=0.91$

Critical point: $t_{0.10}=1.316$

H0 may not be rejected even at the 10% level of significance.

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

Answer:-

Population 1: 1980

n1 = 1000

x1 = 53

p 1 = 0.53

Population 2: 1985

n2 = 100

x2 = 43

p 2= 0.53

H0: p1-p2 =0

H1: p1- p2 ≠ 0

$P^{\wedge}=(x_1+x_2)/n_1+n_2$ = 53+43/100+100 = 0.48

$Z=(p_1-p_2)- / \sqrt{p^{\wedge}(1- p^{\wedge})((1/n_1)+(1/n_2))}$

= (0.53-0.43)/ √(.48)(.52)((1/100)+(1/100))

=0.1/√0.004992

=0.1/0.07065 = 1.415

Critical point: $z_{0.05}$=1.645

H0 may not be rejected even at the 10% level of significance.

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

Answer:-

Population 1: With sweepstakes

n1 = 300

x1 = 120

p = 0.40

Population 2: No sweepstakes n2 = 700

x2 = 140

p2= 0.20

H0: p1- p2 ≤0.1

H1: p1- p2 >0.1

$Z=(p_1\hat{}-p_2\hat{})-D / \sqrt{(p_1\hat{}(1- p_1\hat{})/n1)(p_2\hat{}(1- p_2\hat{})/n_2)}$

= (0.1)/ 0.03207

=3.118

Critical point: $z_{0.001}$=3.09

H0 may not be rejected at any common level of significance.

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^−1.

Answer:-

Observed Frequency(O)	Expected Frequency (E)	$(O-E)^2$
16	22	36
20	22	4
25	22	9
14	22	64
29	22	49
28	22	36
Total		**196**

Step 1:

Null Hypothesis (H0): The die is unbiased

Alternative Hypothesis (HA): The die is not unbiased

Step 2: Test Statistics

On the hypothesis that the die is unbiased we should expect the frequency of each number to be 132/6 =22

: - Xcal2= $\Sigma(0-E)^2/E$ =198/22=0.91

Step 3:

L.O.S (**α**) =0.05

Degree of freedom =n-1= 6-1 =5

Step 4:

Critical value $x(\alpha)^2$**=11.0705**

Step 5: Decision

Since $x^2cal < x(\alpha)^2$

H0 is accepted.

The die is unbiased.

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:

	Men	Women
Voted	2792	3591
Not voted	1486	2131

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”?

Answer:-

H0: 'Sex is independent of voting'

H1: 'Sex and voting are dependent'

After specifying the Null hypothesis we need to compute the expected table under the assumption that rows and columns are in fact independent. To compute the expected table we use the product rule for chances:

Chance of (row i ,col j) = (chance row i)* (chance col j)

From here we deduce that the expected number of counts in (row i, col j) is given by:

N*(chance row i)*(chance j) = (sum row i)*(sum col j)/N

The observed table with totals included is:

OBSERVED TABLE:-

	Men	Women	Total
Voted	2792	3591	6383
Not voted	1486	2131	3617
Total	4278	5722	10000

The associated expected table under the assumption that sex and voting are independent is given by

EXPECTED TABLE:-

	Men	Women	Total
Voted	2731	3652	6383
Not voted	1547	2070	3617
Total	4278	5722	10000

We now have the observed table and the expected table under the null hypothesis of independence. After that we need to compute the X2 statistic. The X2 statistic measures how far away the observed table from the expected one is. The X2 statistic has as many terms as their cells in the observed table (4 in our case):

- $C11 = (3591-3652)^2/3652$
- $C12 = (1486-1547)^2/1547$
- $C21 = (2792-2731)^2/2732$
- $C22 = (2131-2070)^2/2070$

The X2- Statistic is the sum of each of the contributions from cell:

X2 =c11+c12+c21+c22 =6.584

Since the observed X2 =6.58 and thus.

3.84 < X2 <6.64

We conclude that:

1% < P-value <5%

And we reject the NULL. The data supports the hypothesis that sex and voting are dependent in this town.

Problem Statement 19:

A sample of 100 voters is asked which of four candidates they would vote for in an election. The number 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].

Answer:-

A Chi-Squared Goodness-of-Fit test is appropriate here. The null hypothesis is that there is no preference for any of the candidates: if this is so, we would expect roughly equal numbers of voters to support each candidate. Our expected frequencies are therefore 100/4 = 25 per candidate.

O	**41**	**19**	**24**	**16**
E	**25**	**25**	**25**	**25**
(O-E)	**16**	**-6**	**-1**	**-9**
$(O-E)^2$	**256**	**36**	**1**	**81**
$\frac{(O-E)^2}{E}$	**10.24**	**1.44**	**0.04**	**3.24**

Adding together the last row gives us our value of c^2 :

$$\sum \frac{(O-E)^2}{E} = 10.24+ 1.44 + 0.04 + 3.24 = \mathbf{14.96}$$, with 4 - 1 = 3 degrees of freedom.

The critical value of Chi-Square for a 0.05 significance level and 3 df. is 7.82. Our obtained Chi-Square value is bigger than this, and so we conclude that our obtained value is unlikely to have occurred merely by chance. In fact, our obtained value is bigger than the critical Chi-Square value for the 0.01 significance level (13.28). In other words, it is possible that our obtained Chi-Square value is due merely to chance, but highly unlikely: a Chi-Square value as large as ours will occur by chance only about once in a hundred trials. It seems more reasonable to conclude that our results are not to chance, and that the data do indeed suggest that voters do not prefer the 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: $p < 0.05$].

###		Photograph		
		A	B	C
Age of child	5 – 6 years	18	22	20
	7 – 8 years	2	28	40
	9 – 10 years	20	10	40

Answer:-

	photograph:			
age of child:	**A:**	**B:**	**C:**	**row totals:**
5-6 years	**18**	**22**	**20**	**60**
7-8 years	**2**	**28**	**40**	**70**
9-10 years	**20**	**10**	**40**	**70**
column totals:	**40**	**60**	**100**	**200**

(a) Work out the row, column and grand totals (as shown in the shaded parts of the table, above).
(b) Work out the expected frequencies, using the formula:

$$E = \frac{(\text{row total} * \text{column total})}{\text{grand total}}$$

For each cell of the above table, this gives us:

O:	**18**	**22**	**20**	**2**	**28**	**40**	**20**	**10**	**40**
E:	**12**	**18**	**30**	**14**	**21**	**35**	**14**	**21**	**35**

Next, work out (O - E):

(O-E):	**6**	**4**	**-10**	**-12**	**7**	**5**	**6**	**11**	**5**

Square each of these, to get $(O - E)^2$:

$(O - E)^2$: 36 16 100 144 49 25 36 121 25

Divide each of the above numbers by E, to get $(O - E)^2 / E$:

$(O - E)^2$ 3 0.89 3.33 10.29 2.33 0.71 2.57 5.76 0.71

E

Chi-squared is the sum of these:

c^2 = **29.60**.

df. = (rows - 1) * (columns - 1) = 2 * 2 = 4.

The critical value of Chi-Square in the table for a 0.001 significance level and 4 d.f. is 18.46. Our obtained Chi-Square value is bigger than this: therefore we have a Chi-Square value which is so large that it would occur by chance only about once in a thousand times. It seems more reasonable to accept the alternative hypothesis, that there is a significant relationship between age of child 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.

	Support	No support
Conform	18	40
Not conform	32	10

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: $p < 0.05$].

Answer:-

Here we have a 2x2 contingency table. Chi-Square is the appropriate test to use, but since we have 1 df., we will modify the formula to include "Yates' correction for continuity".

	support	**no support**	**row totals:**
conform:	**18**	**40**	**58**
not conform:	**32**	**10**	**42**
column totals:	**50**	**50**	**100**

(a) Calculate the row, column and grand totals.

(b) Calculate the expected frequency for each cell of the table, by multiplying together the appropriate row and column totals and then dividing by the grand total.

(c) Subtract each expected frequency from its associated observed frequency; but then apply Yates' correction, by subtracting 0.5 from the absolute value of each O-E value. (The vertical bars in the formula mean "ignore any minus signs").

O: **18** **40** **32** **10**
E: **29** **29** **21** **21**

Next, work out (O - E):

(|O-E|- 0.5): **10.5** **10.5** **10.5** **10.5**

Square each of these, to get $(O - E)^2$:
(|O-E|- 0.5)^2: **110.25** **110.25** **110.25** **110.25**

Divide each of the above numbers by E, to get $(O - E)^2 / E$:
$(O - E)^2$ **3.80** **3.80** **5.25** **5.25**

E
Chi-squared is the sum of these:

c^2 = **18.10.**

df. = (rows - 1) * (columns - 1) = 1 * 1 = 1.

Our obtained value of Chi-Squared is bigger than the critical value of Chi-Squared for a 0.001 significance level. In other words, there is less than a one in a thousand chance of obtaining a Chi-Square value as big as our obtained one, merely by chance. Therefore we can conclude that there is a significant difference between the "support" and "no support" conditions, in terms of the frequency with which individuals conformed.

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: $p < 0.01$].

##	Height	
	Short	Tall
Leader	12	32
Follower	22	14
Unclassifiable	9	6

Answer:-

Expected frequencies are in brackets:

	height:		
	short	**tall**	**row totals:**
leader:	**12 (19.92)**	**32 (24.08)**	**44**
follower:	**22 (16.29)**	**14 (19.71)**	**36**
unclassifiable:	**9 (6.79)**	**6 (8.21)**	**15**
column totals:	**43**	**52**	**95**

Chi-Square = 3.146 + 2.602 + 1.998 + 1.652 + 0.720 + 0.595 = **10.712**, with 2 df.

10.712 are bigger than the tabulated value of Chi-Square at the 0.01 significance level. We would conclude that there seems to be a relationship between height and leadership qualities. Note that we can only say that there is a relationship between our two variables not that once causes the other. There could be all kinds of explanations for such a relationship.

Problem Statement 23:

Each respondent in the Current Population Survey of March 1993 was classified as employed, unemployed, or outside the labour force. The results for men in California age 35-44 can be cross tabulated by marital status, as follows:

	Married	Widowed, divorced or separated	Never married
Employed	679	103	114
Unemployed	63	10	20
Not in labor force	42	18	25

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

Answer:-

```
> Obs_table := matrix(3,3,[679,103,114,63,10,20,42,18,25]);

                    [679   103   114]
                    [                ]
        Obs_table := [ 63    10    20]
                    [                ]
                    [ 42    18    25]
> R1 := 679+103+114:R2:=63+10+20:R3:=42+18+25:
> C1:=679+63+42:C2:=103+10+18:C3:=114+20+25:N:=evalf(R1+R2+R3):
> Exp_table := matrix(3,3,(i,j)-> round(R.i*C.j/N));
```

$$\text{Exp\_table} := \begin{bmatrix} 654 & 109 & 133 \\ 68 & 11 & 14 \\ 62 & 10 & 13 \end{bmatrix}$$

$$X2 := \frac{(679 - 654)^2}{654} + \ldots + \frac{(25 - 13)^2}{13}$$

> X2 := 30.96:

Looking at the table of the Chi-square distribution with (3-1)(3-1)=2*2=4 degrees of freedom we get:

Degrees of freedom	99% ...	10%	5%	1%
1		0.00016	2.71	3.84	6.64
2		0.020	4.60	5.99	9.21
3		0.12	6.25	7.82	11.34
4		0.30	7.78	9.49	13.28
5		0.55	9.24	11.07	15.09

since 30.96 > > 13.28 we conclude from the table that:

P < < 1%

So we reject with all confidence. Conclusion: Marital Status seems to be related to Job Status in this town.