1) State whether each of the following situations is a correctly stated hypothesis testing problem and why.

a. $H_0:\mu=25, H_1 \mu\neq 25$

b. $H_0:\sigma>10, H_1:\sigma=10$

c. $H_0:\bar{x}=50, H_0:\bar{x}=50$

d. $H_0:p=0.1 H_0:p=0.5$

e. $H_0:s=30, H_1:s>30$

Solution :

a. $H_0:\mu=25, H_1 \mu\neq 25$ – This statement is True

Null Hypothesis: this is a claim that is initially assumed to be true eg accepts the value when equals 25.
Alternate Hypothesis: This is a statement that contradicts null hypothesis and says don't accept when the value greater or lesser than 25.

b. $H_0:\sigma>10, H_1:\sigma=10$ - This statement is False

Null Hypothesis: This value usually takes equal to value.

Alternate Hypothesis : This usually claims greater than or lesser than value.

c. $H_0:\bar{x}=50, H_0:\bar{x}=50$ - This statement is False

Null & Alternate Hypothesis applies only to population mean and variance and not for Sample mean.

d. $H_0:p=0.1 H_0:p=0.5$ - This statement is False
Null & Alternate Hypothesis both those values are different and it is not applicable.

e. $H_0:s=30, H_1:s>30$ - This statement is False

Null & Alternate Hypothesis applies only to population mean and variance and not for Sample mean.

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.

Solution :

Null Hypothesis: H0 = 52
Alternate Hypothesis: H1 ≠ 52 (Two sided Test either greater or lesser than 52 not accepted)

Standard Deviation : $\sigma = 4.50$

Average : $\mu = 52$
Random sample size = n = 100
Xi = 52.80

Level of significance = 5% = 0.05 = >Two sided Test = 0.05/2 = 0.025 => 0.025
Level of Confidence = 95%

$Z = (x-\mu)/ \sigma$
Standard error $SE(\sigma) = SD/\sqrt{n}$
$SE = 4.50 / \sqrt{100} \Rightarrow 0.45$
$= (52.80 – 52)/0.45 \Rightarrow 1.78$

Z Test = 1.78
$P(Z<z) = 0.025$
Check for 0.025 value in Z Table
Z(0.025) = -1.96 and +1.96
Calculated value 1.78 lies between z Table value. Hence we can accept Null Hypothesis

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.

Solution :

Null Hypothesis: H0 = 34
Alternate Hypothesis: H1 ≠ 34 (Two sided Test)

Standard Deviation : $\sigma = 8$

Average : $\mu = 34$
Random sample size = n = 50
Xi = 32.50

Level of significance = 1% = 0.01= >Two sided Test = 0.01/2 = 0.005
Level of Confidence = 99%

$Z = (x-\mu)/ \sigma$
Standard error $SE(\sigma) = SD/\sqrt{n}$

$SE = 8 / \sqrt{50} \Rightarrow 1.1313$
$= (32.50 – 34)/1.1313 \Rightarrow -1.33$

Z Test = -1.33
$P(Z<z) = 0.005$
Check for 0.005 value in Z Table i.e 0.005
$Z(0.01) = -2.57 < Z < + 2.57$
Calculated value -1.33 greater than -2.57. Hence we can accept Null Hypothesis

Conclusion : we accept the Null Hypothesis at 1% level of significance.

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.

Solution :

Null Hypothesis: $H0 = 1135$
Alternate Hypothesis: $H1 \neq 1135$ (Two sided Test)

Standard Deviation : $\sigma = 240.37$

Average : $\mu = \$1135$
Random sample size = n = 22
Xi = 1031.2

$Z = (x-\mu)/ \sigma$
Standard error $SE(\sigma) = SD/\sqrt{n}$
$SE = 240.37 / \sqrt{22} \Rightarrow 51.25$
$= (1031.2 – 1135)/ 51.25 \Rightarrow -103.68/51.25 = -2.02$
The critical value of Z is -1.96 and +1.96

Conclusion: The computed value of z falls in rejection region we reject 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?

Solution :

Null Hypothesis: H0 = $48,432
Alternate Hypothesis: H1 > $48,432 (Two sided Test)

Standard Deviation : $\sigma = 2000$

Average: μ = $48,432
Random sample size = n = 400
Xi = $48,574

Level of significance = 5% = 0.05= > If level of significance not given consider 5% significance
Two sided Test 0.05/2 = 0.025
Level of Confidence = 95%

$Z = (x-\mu)/ \sigma$
Standard error $SE(\sigma) = SD/\sqrt{n}$
$SE = 2000 / \sqrt{400} \Rightarrow 100$
$= (48574 – 48432)/100 \Rightarrow 1.42$

Z Test = 1.42
$P(Z<z) = 0.025$
Check for 0.975 value in Z Table
$Z(0.025) = -1.96 < Z < + 1.96$
Calculated value 1.42 is in the Acceptance region. Hence we can accept Null Hypothesis.

Conclusion : we accept the Null Hypothesis at 5% level of significance.

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?

Solution :

Null Hypothesis: H0 = \$32.28
Alternate Hypothesis: H1 < > 32.28 (Two sided Test)

Standard Deviation : $\sigma = 1.29$

Average: μ = \$32.28
Random sample size = n = 19
Xi = \$31.67

Level of significance = 5% = 0.05= > If level of significance not given consider 5% significance
Two sided Test 0.05/2 = 0.025
Level of Confidence = 95%

$Z = (x-\mu)/\sigma$
Standard error $SE(\sigma) = SD/\sqrt{n}$
$SE = 1.29 / \sqrt{19} \Rightarrow 4.35$
$= (31.67 - 32.38)/4.35 \Rightarrow -0.14$

Z Test = -0.14
$P(Z<z) = 0.025$
Check for 0.975 value in Z Table
$Z(0.025) = -1.96 < Z < +1.96$
Calculated value -0.14 is in the Acceptance region. Hence we can accept Null Hypothesis.

Conclusion : we accept the Null Hypothesis at 5% level of significance.

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			

Solution :

== > Standard deviation is not given to calculate Z test

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.

Solution:

n = 16
μ = 10
Sample mean x = 12
S D = 1.5
t = [x - μ] / [s / sqrt(n)]
t = [12-10]/[1.5/√16]
t = 2/0.375
t = 5.333
degree of freedom = 16 -1 = 15

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.

Solution:

N = 16
Significance level = 99% = 0.99
Combined significance level = σ = 0.01/2 = 0.005
Degrees of freedom = 16-1 = 15
t-table score = -2.947 < T < 2.947

Problem Statement 10:

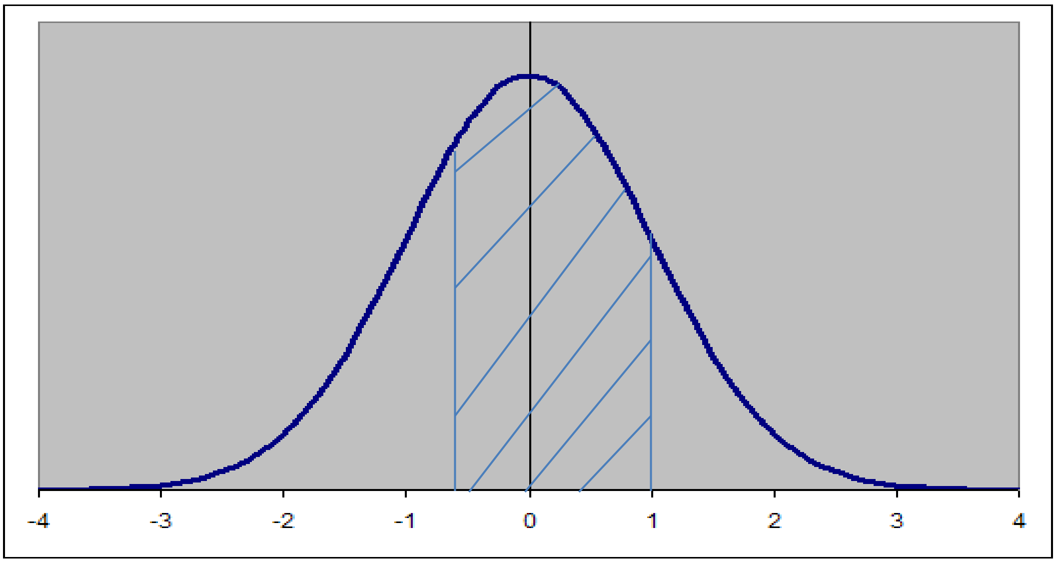
If a random sample of size 25 drawn from a normal population gives a mean of 60 and a standard deviation of 4, find the range of t-scores where we can expect to find the middle 95% of all sample means. Compute the probability that (−□0.05 <□<□0.10).

Solution:

n = 25
Sample mean x = 60
S D = 4
t = [x - μ] / [s / sqrt(n)]
t = [60- μ]/[4/√25]
t = [60- μ]/0.8
degree of freedom = 25 -1 = 24

`normalcdf:` (lower value, upper value, mean, standard error of the mean)
The parameter list is abbreviated (lower value, upper value, $\mu,\sigma/\sqrt{n}$)

`normalcdf: (-0.05 , 0.10, 0.95,4/√25) = -0.353(from T table value)`
$P(-t0.05 < t < t0.10)$. = 1.711 – 2.064 = -0.353



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

Solution: Two population samples are independent to each other

Comparing two independent population means
Sample 1: Size n1 = 1200
Sample Mean X1 = 452
Standard Deviation = 212

Sample 2: Size n2 = 800
Sample Mean X2 = 523
Standard Deviation = 185

Null Hypothesis : H0: µ1= µ 2 both the Population means are equal
Alternate Hypothesis : H1: µ1 ≠ µ 2 both the Population means are not equal

Two Tailed Test using a Z statistic and 5 % level of significance ;Reject H0 if z<= -1.960 or if z >= 1.960

Here n values are greater than 30 : n1 = 1200 ; n2 = 800 => **Z Test**

$$z = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$S_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$$

Sp => Pooled estimate of the common standard deviation

$$S_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$$

Standard Deviation = $\sqrt{(((1200-1)} * 212^2 + (800-1) * 185\text{^}2)/1200 + 800 - 2)$

= √((5,38,87,856 + 2,73,45,775)/(1998))

SD = 201.64

Z = -7.715

Conclusion : we Reject H0 because -7.715<= -1.960

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

Solution: Two population samples are independent to each other

Comparing two independent population means
Sample 1: Size n1 = 100
Sample Mean X1 = 308
Standard Deviation = 84

Sample 2: Size n2 = 100
Sample Mean X2 = 254
Standard Deviation = 67

Null Hypothesis : H0: μ1= μ 2 both the Population means are equal
Alternate Hypothesis : H1: μ1 ≠ μ 2 both the Population means are not equal

Two Tailed Test using a Z statistic and 5 % level of significance ;Reject H0 if z<= -1.960 or if z >= 1.960

Here n values are greater than 30 : n1 = 100 ; n2 =100 => **Z Test**

$$z = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

$$S_p = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$$

Sp => Pooled estimate of the common standard deviation

$$S_p = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$$

Standard Deviation = $\sqrt{(((100-1)} * 84^2 + (100-1) * 67\text{^}2)/100 + 100 - 2)$

= √((698544 + 444411)/(198))

SD = 75.97

Z = 54/75.97√(1/100 + 1/100)

Z = 5.02

Conclusion : we Reject H0 because 5.02 > 1.960

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%

Solution:

Null Hypothesis : H0: $\mu_1 = \mu_2$ both the Population means are equal
Alternate Hypothesis : H1: $\mu_1 \neq \mu_2$ both the Population means are not equal

Two Tailed Test using a Z statistic and 5 % level of significance ;Reject H0 at 5% level of significance

Here n values are lesser than 30 : n1 = 14 ; n2 =9 => **T Test**

Degree of Freedom = n1 + n2 -2 => 23-2 = 21

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$S_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$$

Sp => Pooled estimate of the common standard deviation

$$S_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$$

T = 0.107/√0.00247

T = 2.154
Critical point : t at 0.025 is 2.080

Conclusion : we Reject H0 because T value falls in the rejection region.

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

Solution :

Population 1: Before reduction
n1 = 15
x1 = Rs. 6598 s1 = Rs. 844

Population 2: After reduction n2 = 12

x2 = RS. 6870
s2 = Rs. 669

Null Hypothesis: H0:µ1 - µ2 <= 0

Alternate Hypothesis : H1 :µ1 - µ2 > 0

degree of freedom : n1 + n2 – 2 = 25

Since sample size is lesser than 30 calculating T Statistics

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

$$S_p = \sqrt{\frac{\left(n_1-1\right)s_1^2+\left(n_2-1\right)s_2^2}{n_1+n_2-2}}$$

Sp => Pooled estimate of the common standard deviation

$$S_p = \sqrt{\frac{\left(n_1-1\right)s_1^2+\left(n_2-1\right)s_2^2}{n_1+n_2-2}}$$

T = 272/√89375.25 = 272/298.96 = 0.91

Critical point at 5% level of significance t0.05 = 2.060

Ho is accepted at 5% 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

If p1 == p2 ie 0.53 == 0.53 then z = 0

Considering z = 0.43 as per x2 = 43

Population Proportions p1 and p2
Null Hypothesis : Ho: p1-p2 = 0
Alternate Hypothesis : H1 : p1 – p2 ≠ 0

Population = p
Sample population = p̂

p̂ = x /n

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

p̂ = (53 + 43)/(100 + 100) = 0.48

z = (0.53 – 0.43) – 0 / √(0.48 * (1 – 0.48) *(1/100 + 1/100))
z = 1.415

critical point at 10% level of significance : Two tailed Test : t 0.05 => z table = 1.64

Conclusion : H0 is accepted at 10% level of significance as Z vale is at the acceptance region.

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 1= 0.40
Population 2: No sweepstakes n2 = 700
x2 = 140
p 2= 0.20

Solution :

Population 1: With sweepstakes
n1 = 300`
x1 = 120
p 1= 0.40

Population 2: No sweepstakes n2 = 700
x2 = 140
p 2= 0.20

Null Hypothesis: Ho : p1 –p2 <= 0.10
Alternate Hypothesis : H1 : p1 – p2 > 0.10

Population = p
Sample population = p̂

p̂ = x /n

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - D}{\sqrt{\left(\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}\right)}}$$

Z = (0.40 – 0.20) – 0.10 / √(0.40(1-0.40)/300)+0.20(1-0.20)/700)
Z = 0.10/0.032
Z = 3.118
Z test at 5% level of significance = z at 0.05 = 1.64

Conclusion: The calculated z test statistic falls outside the acceptance and hence we are rejecting the null hypothesis and we may conclude that the proportion of customers buying at least $2500 of traveller's checks is at least 10% higher when sweepstakes are on.

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 p̂ -1.

Solution :

Step 1 :

Null hypothesis : (H_0) The die is unbiased.
Alternative Hypothesis (H_A) The die is biased.

Step 2 : Test Statistics

On the hypothesis that the die is unbiased.

total ==> 132/6 =22

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

O = Observed Frequency
E = Expected Frequency

χ2 = Randomly selected value from the population.

Chi Squared Test :

- The χ2 assumes that the data for the study is obtained through random selection, i.e. they are randomly picked from the population
- The categories are mutually exclusive i.e. each subject fits in only one category. For e.g.- from our above example – the number of people who lunched in your restaurant on Monday can't be filled in the Tuesday category
- The data should be in the form of frequencies or counts of a particular category and not in percentages
- The data should not consist of paired samples or groups or we can say the observations should be independent of each other
- When more than 20% of the expected frequencies have a value of less than 5 then Chi-square cannot be used. To tackle this problem: Either one should combine the categories only if it is relevant or obtain more data

Observed frequency (O)	Expected frequency(E)	$(O-E)^2$
15	22	49
20	22	4
25	22	9
15	22	49
29	22	49
28	22	36
	Total	**196**

χ2 = 196 / 22 = 8.91

At 5% level of significance => 0.05
Critical value = degree of freedom : 6 – 1 = 5
Chi squared Table value = 11.070

The calculated value chi squared => 8.91 and it lies within the acceptance region

Conclusion:
Null Hypothesis is accepted and 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"?

Solution:

The chi squared test can also be used to test independence.

Hypothesis :
Null Hypothesis H0: gender and voting independent -> Sex is independent of voting
Alternate Hypothesis : HA : Sex and voting are dependent.

Observed Values:

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

Expected Values are calculated using :((Total (Men) * Total(voted))/10000

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

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

Col1 = (2792 – 2731)^2/2731 = 1.363
Col2 = (3591-3652)^2/3652 = 1.019
c21 := (1486-1547)^2/1547 = 2.405
c22 := (2131-2070)^2/2070 = 1.797

χ2 = 6.583

Degrees of Freedom: (2-1)*(2-1) = 1

At 5% significant level = 3.841

Conclusion:

Our calculated Chi squared value greater than the critical value and hence we reject the Null Hypothesis.

Problem Statement 19:

Good ness of fit test : is a one variable chi squared test

A test of independence = is a two variable chi squared test

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:

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 .

Solution:

Good ness of fit test : is a one variable chi squared test , as it deals with one variable here.

Null Hypothesis: No preference for any of the candidates and hence equal number of voters to support each other.

Total = 100/4 => 25

	Higgings	Reardon	White	Charlton	Total
Observed	41	19	24	16	100
Expected	25	25	25	25	100
O-E	16	-6	-1	-9	
(O-E)^2	256	36	1	81	
(O-E)^2 / E	10.24	1.44	0.04	3.24	

$\chi2$ = 10.24+ 1.44 + 0.04 + 3.24 = **14.96**
df = 4-1 = 3

The critical value of Chi-Square for a 0.05 significance level and 3 d.f. is 7.82.

Conclusion : we reject the Null Hypothesis.

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

Solution :

Null hypothesis : Age and Photograph preference are independent

Alternate hypothesis : Age and Photograph preference are dependent

Observed :

Age	A	B	C	Total
5-6	18	22	20	60
7-8	2	28	40	70
9-10	20	10	40	70
Total	40	60	100	200

Expected:

Age	A	B	C	Total
5-6	12	18	30	60
7-8	14	21	35	70
9-10	14	21	35	70
Total	40	60	100	200

O-E	6	4	-10	-12	7	5	6	-11	5
(O-E)^2	36	16	100	144	49	25	36	121	25
(O-E)^2 / E	3	0.89	3.33	10.29	2.33	0.71	2.57	5.76	0.71

Chi Squared value = 29.60

Degree of Freedom = (rows-1)* (Columns – 1) => 4

Calculated table value : 9.488 at 5 % level of significance.

Conclusion:

The critical value is lesser than the chi squared value and hence rejecting Null Hypothesis and accepting the Alternate Hypothesis.

Problem Statement 21:

A study of conformity using the Asch paradigm involved two conditions: one where one confederate supported the true judgment 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:

$p < 0.05$].

Solution :

Null hypothesis : No significant difference between Support and no support

Alternate hypothesis : There are significant difference between Support and no support

Observed :

	Support	No Support	Total
conform	18	40	58
Not Conform	32	10	42
Total	50	50	100

Expected:

	Support	No Support	Total
conform	29	29	58
Not Conform	21	21	42
Total	50	50	200

O-E	-11	11	11	-11
(O-E)^2	121	121	121	121
(O-E)^2 / E	4.17	4.17	5.76	5.76

Chi Squared value = 4.17 + 4.17 +5.76+5.76 => 19.86

Degree of Freedom = (rows-1)* (Columns – 1) => 1

Calculated Chi squared table value: 3.841 at 5 % level of significance.

Conclusion:

The critical value is lesser than the obtained chi squared value and hence rejecting Null Hypothesis and accepting the Alternate Hypothesis. There is a significant difference between the “support” and “no support” conditions.

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

Solution:

Null hypothesis: No significant relationship between height and leadership qualities

Alternate hypothesis: There are significant relationships between height and leadership qualities

Observed:

	Short	Tall	Total
Leader	12	32	44
Follower	22	14	36
Unclassifiable	9	6	15
Total	43	52	95

Expected:

	Short	Tall	Total
Leader	20	24	44
Follower	16.29	19.70	36
Unclassifiable	6.78	8.21	15
Total	43	52	95

O-E	-8	8	6	--6	2	-2	
(O-E)^2	64	64	36	36	4	4	
(O-E)^2 / E	3.2	2.67	2.21	1.8	0.59	0.49	

Chi Squared value =10.96

Degree of Freedom = (rows-1)* (Columns – 1) => 2

Calculated Chi squared table value: 9.210 at 1 % level of significance.

Conclusion:

The critical value is lesser than the obtained chi squared value and hence rejecting Null Hypothesis and accepting the Alternate Hypothesis. There is a significant difference between the 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:

	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 labor force status. Or is this just chance variation? (you may assume the table results from a simple random sample.)

Solution:

Ho: There are no relationships between values on the rows and columns.
Ha: There is a relationship between rows and columns.
(Significance Level = 0.05)

Observed :

	Married	Widowed,Divorced	Never Married	Total
Employed	679	103	114	896
Un Employed	63	10	20	93
Not in labour force	42	18	25	95
Total	784	131	159	1074

Expected :

	Married	Widowed,Divorced	Never Married	Total
Employed	654.06	109.29	132.65	896
Un Employed	67.89	11.34	13.77	93
Not in labour force	62.05	10.37	12.58	95
Total	784	131	159	1074

Chi squared =

(O-E)^2	(679 – 654.063315)^2	(103 – 109.288641)^2	(114 – 132.648045)^2	(63 – 67.888268)^2	(10 – 11.343575)^2	(20 – 13.768156)^2	(42 – 62.048417)^2	(18 – 10.367784)^2	(25 – 12.583799)^2
(O-E)^2/E	(679 – 654.063315)^2 / 654.063315	(103 – 109.288641)^2 / 109.288641	(114 – 132.648045)^2 / 132.648045	(63 – 67.888268)^2 / 67.888268	(10 – 11.343575)^2 / 11.343575	(20 – 13.768156)^2 / 13.768156	(42 – 62.048417)^2 / 62.048417	(18 – 10.367784)^2 / 10.367784	(25 – 12.583799)^2 / 12.583799

Chi squared Value = 31.61

Degree of Freedom = (3-1)(3-1) = 4

Critical value at 0.05 , 4 = 9.49

Conclusion:

The calculated chi squared value is greater than critical value and hence we reject Null Hypothesis.There is a relationship between rows and columns with 5% level of significance at degree of freedom = 4.