**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 🡪 Correct

2. 𝐻0: 𝜎 > 10, 𝐻1: 𝜎 = 10 🡪 Incorrect

\_ \_

3. 𝐻0: 𝑥 = 50, 𝐻1: 𝑥 ≠ 50 🡪 Correct

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

5. 𝐻0: 𝑠 = 30, 𝐻1: 𝑠 > 30 🡪 Correct

Because Alternate Hypothesis is reverse/opposite of Null hypothesis.

**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.

𝜇 = 52 𝜎 =4.50

n=100 **x̄** =52.80

α =0.05

H0: avg book cost<=52

H1: avg book cost>52

SE= 𝜎/ √n = 4.50/ √100 = 4.5/10=0.45

z(test) = (x̄- 𝜇)/ SE = (52.80-52)/0.45= 1.778

z(0.05) = -1.64

z(test)> z(0.05)

i.e does not lie in critical region.

Thus, we accept HO, book avg cost is not higher.

**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.

μ = 34 σ = 8

n= 50 **x̄=32.5**

α =0.01

H0: pollutant=34

H1: pollutant<34

SE= 𝜎/ √n = 8/ √50 = 1.13

z(test) = (x̄- 𝜇)/ SE = (32.5-34)/1.13= -1.33

z(0.01) = -2.33

z(test)> z(0.01)

i.e does not lie in critical region.

Thus, we accept HO, pollutant level is not lowered.

**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

μ = 1135

n= 22 **x̄=1031.32 s=240.37**

α =0.5

α/2=0.25

H0: μ = 1135

H1: μ ≠ 1135

SE= 𝜎/ √n = 240.37 / √22 = 51.25

z(test) = (x̄- 𝜇)/ SE = (1031.32-1135)/51.25 = -2.003

z(0.25) = -0.67

z(test)< z(0.25)

i.e lie in critical region.

Thus, we reject HO and avg price ≠ 1135.

**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?

μ = 48432

n= 400 **x̄=48574 s=2000**

α =0.05

α/2 =0.025

H0: μ = 48432

H1: μ ≠ 48432

SE= 𝜎/ √n = 2000/ √400 = 100

z(test) = (x̄- 𝜇)/ SE = (48574-48432)/100 = 1.42

z(0.025) = -1.96 and 1.96

-1.96<1.42<1.96

i.e does not lie in critical region.

Thus, we accept HO and avg income = 48,432.

**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?

μ = 32.28

n= 19 **x̄=31.67 s=1.29**

α =0.05

α/2 =0.025

H0: μ = 32.28

H1: μ ≠ 32.28

SE= 𝜎/ √n = 1.29 / √19 = 0.295

z(test) = (x̄- 𝜇)/ SE = (31.67-32.28)/0.295 = -2.06

z(0.025) = -1.96 and 1.96

z(test)< z(0.025)

i.e lie in critical region.

Thus, we reject HO and avg price ≠ 32.28.

**Problem Statement 7:**

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



**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.

μ = 10

n= 16 **x̄=12 s=1.5**

SE= 𝜎/ √n = 1.5/ √16 = 0.375

t = (x̄- 𝜇)/ SE = (12-10)/0.375 = 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.

1- α = 0.99 🡪 α = 0.01

df=n-1 = 16-1 = 15

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).

n= 25 **x̄=60 s=4**

df= n-1 = 25-1 = 24

1- α = 0.95 🡪 α = 0.05

two tail test 🡪

-2.064<t<2.064

probability that (−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

H0: people travelling from Bangalore to Chennai = people travelling from Bangalore to Hosur

H1: people travelling from Bangalore to Chennai ≠ people travelling from Bangalore to Hosur

SE= √ (212)sq/1200 + (185)sq/800 = √(37.45+42.78) = 8.96

z(test) = (x̄1- x̄2)/ SE = 452-523/8.96 = -7.92

Taking 5% Significance level

α = 0.05

α/2 =0.025 (Two tail test)

z(0.025) = -1.96 and 1.96

z(test)< z(0.025)

i.e lie in critical region.

Thus, we reject HO.

Therefore, people travelling from Bangalore to Chennai ≠ people travelling from Bangalore to Hosur.

**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

H0: No. of people preferring Duracell = No. of people preferring Energizer

H1: No. of people preferring Duracell ≠ No. of people preferring Energizer

SE= √ (84)sq/100 + (67)sq/800 = √(70.56+44.89) = 10.75

z(test) = (x̄1- x̄2)/ SE = 308-254/10.75 = 5.02

Taking 5% Significance level

α = 0.05

α/2 =0.025 (Two tail test)

z(0.025) = -1.96 and 1.96

z(test)> z(0.025)

i.e lie in critical region.

Thus, we reject HO.

No. of people preferring Duracell ≠ No. of people preferring Energizer

**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 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

H0: Sales before reduction = Sales after reduction

H1: Sales before reduction < Sales after reduction

Left hand tail test

SE= √ (844)sq/15 + (669)sq/12 = √(47489.1+37296.75) = 291.18

z(test) = (x̄1- x̄2)/ SE = 6598-6870/291.18 = -0.93

Taking 5% Significance level

α = 0.05

z(0.05) = -1.64

z(test)> z(0.05)

i.e does not lie in critical region.

Thus, we accept HO.

Sales before reduction = Sales after reduction

**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

D=0

H0: p1-p2=0

H1: p1-p2≠0

Two- tail test

p1=x1/n1

p1=53/1000=0.053

p2=x2/n2

p2=43/100=0.43

Comparisons of Two Population
Proportions When the Hypothesized
Difference Is Zero: Example 8-8

Slide 34

Carry out a two...

p=(x1+x2)/(n1+n2)

p=(53+43)/(1000+100)=96/1100=0.0873

Comparisons of Two Population
Proportions When the Hypothesized
Difference Is Zero: Example 8-8

Slide 34

Carry out a two...

z(test)=((p1-p2)-D)/ √ (p(1-p)(1/n1+1/n2))

z(test)=(0.053-0.43)/ √ ((0.0873)(0.9127)(0.001+0.01))=-13.28

Taking 5% Significance level

α = 0.05

α/2 =0.025 (Two tail test)

z(0.025) = -1.96 and 1.96

z(test)<z(0.025)

i.e lie in critical region.

Thus, we reject HO.

**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

H0: p1-p2<=0.1

H1: p1-p2>0.1

Comparisons of Two Population Proportions
When the Hypothesized Difference Is Not Zero:
Example 8-9

Slide 37

Carry out a...

α=0.1%

α=0.001

z(0.001)= -3.09

Therefore, H0 may be accepted.

**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.

The null hypothesis H0 : The die is unbiased.

Alternative hypothesis Ha : The die is biased.

Calculation of test statistic : On the hypothesis, that the die is unbiased we should expect the frequency of each number to be = total/6 = 132/6 =22

χ2 = ∑ (O−E)2/E = (15−22)2/22 + (20−22)2/22 + (25−22)2/22 + (15−22)2/22 + (29−22)2/22 + (28−22)2/22 = 196/12 = 8.91

Level of significance : α=0.05

Critical value : the table value of χ2 at 5% level of significance for ν=6−1=5

degrees of freedom is 11.0

Decision : Since the calculated value of |t|=8.91 is less than the table value χ2 = 11.07, the null hypothesis 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:

****

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”?

Null := `Gender is independent of Voting`:  
Alternative := `Gender 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 col\_j) = (Sum row\_i)\*(Sum col\_j) / N  The observed table with totals included is:  OBSERVED TABLE    Men Women Total  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |\_\_\_\_\_\_  Voted 2792 3591 | 6383  Didn't vote 1486 2131 | 3617  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Total 4278 5722 | 10000  Now we have the information we need to create  an expected table. Here is the equation for  calculating the expected value...  The associated expected table under the assumption that gender and voting are independent is given by  EXPECTED TABLE    Men Women Total  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |\_\_\_\_\_\_  Voted 2731 3652 | 6383  Didn't vote 1547 2070 | 3617  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Total 4278 5722 | 10000  OBSERVED Men Women  TABLE  Voted 2792 3591  Didn't vote 1486 2131  4278 5722  10000  EXPECTED Men Women  TABLE  Voted 2731...  X2:= (2792-2731)^2/2731 + + + (2131- 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 is the observed table from the expected one. The X2 statistic has as many terms as there are cells in the observed table (4 in our case):  EXPECTED Men Women  TABLE  Voted 2731 3652  Didn't vote 1547 2070  4278 5722  OBSERVED Men Women  TABLE  Voted 2792 3591  ... | | |
|  | | The X2-statistic is the sum of each of the contributions from each cell: |

 X2 := c11+c12+c21+c22;

X2 := 6.584283457

|  |  |
| --- | --- |
|  | The last part is to compute the P-value. This is done by looking under the Chi-square table with (rows-1)\*(cols-1) degrees of freedom. In the case of a 2x2 table (our case) the number of degrees of freedom is (2-1)(2-1)=1\*1=1.  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 gender 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 .

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** |
| **(O-E)2**  **---------**  **E** | **10.24** | **1.44** | **0.04** | **3.24** |

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

       (O - E)2

 -----------------   = 10.24+ 1.44 + 0.04 + 3.24 = **14.96**, with 4 - 1 = 3 degrees of freedom.

           E

 The critical value of Chi-Square for a 0.05 significance level and 3 d.f. 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 de 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:** | | |  |
| **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:

                              (row total \* column total)

            E =           --------------------------------------

                                      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**  **----------**  **E** | **3** | **0.89** | **3.33** | **10.29** | **2.33** | **0.71** | **2.57** | **5.76** | **0.71** |

Chi-squared is the sum of these:

            2 = **29.60**.

            d.f. = (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.

****

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].

**Question 3:**

Here we have a 2x2 contingency table. Chi-Square is the appropriate test to use, but since we have 1 d.f., 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**  **-----------**  **E** | **3.80** | **3.80** | **5.25** | **5.25** |

Chi-squared is the sum of these:

            2 = **18.10.**

            d.f. = (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** | **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 d.f.

            10.712 is 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 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.)

DATA TABLE

| Col 1 Col 2 Col 3 | Mean

------+-------------------------------+----------

Row 1 | 679 103 114 | 298.67

Row 2 | 63 10 20 | 31.00

Row 3 | 42 18 25 | 28.33

------+-------------------------------+----------

Mean | 261.33 43.67 53.00 | 119.33

CHI-SQUARE INDEPENDENCE TEST

Ho: Rows and columns are independent.

Ha: There is a relationship between rows and columns.

(Significance Level = 0.05)

Row | Col 1 Col 2 Col 3 | Sum

------+-------------------------------+----------

Obs 1 | 679 103 114 | 896.00

Exp 1 | 654.06 109.29 132.65 | 896.00

+-------------------------------+

Obs 2 | 63 10 20 | 93.00

Exp 2 | 67.89 11.34 13.77 | 93.00

+-------------------------------+

Obs 3 | 42 18 25 | 85.00

Exp 3 | 62.05 10.37 12.58 | 85.00

------+-------------------------------+----------

Sum | 784.00 131.00 159.00 | 1074.00

ChiSq =

(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

= 31.613103

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

Critical Value: ChiSq(0.05, 4) = 9.487729

ChiSq = 31.613103

P(>ChiSq) = 2.295E-006 < 0.05

Reject the null hypothesis at the 0.05 significance level.