2. Problem Statement

Problem Statement 1: [50 marks]

Two-tailed test for difference between two population means

Is there evidence to conclude that the number 2. Problem Statement

Problem Statement 1: [50 marks]

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 Statement 2: [50 marks]

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 Statement 3: [100 marks]

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 4: [100 marks]

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

The manufacturers of compact disk players want to test whether a small price reduction

is enough to increase sales of their product. Randomly chosen data on 15 weekly

sales totals at outlets in a given area before the price reduction show a sample mean

of $6,598 and a sample standard deviation of $844. A random sample of 12 weekly

sales totals after the small price reduction gives a sample mean of $6,870 and a sample

standard deviation of $669. Is there evidence that the small price reduction is

enough to increase sales of compact disk players?

**EXAMPLE 8–6**

This is a one-tailed test, except that we will reverse the notation 1 and 2 so we can

conduct a right-tailed test to determine whether reducing the price increases sales

(if sales increase, then \_2 will be greater than \_1, which is what we want the alternative

hypothesis to be). We have H0: \_1 \_ \_2 \_ 0 and H1: \_1 \_ \_2 \_ 0. We assume an

equal variance of the populations of sales at the two price levels. Our test statistic has

a *t* distribution with *n*1 \_ *n*2 \_ 2 \_ 15 \_ 12 \_ 2 \_ 25 degrees of freedom. The computed

value of the statistic, by equation 8–7, is

*t* =

(6,870 - 6,598) - 0

A

(14)(844)2 + (11)(669)2

25

a 1

15

+

1

12

b

= 0.91

This value of the statistic falls inside the nonrejection region for any usual level of

significance.

Problem Statement 5: [100 marks]

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

p̂1 = 0.53

Population 2: 1985

n2 = 100

x2 = 43

p̂2= 0.53

Problem Statement 6: [100 marks]

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

p̂1 = 0.40

Population 2: No sweepstakes

n2 = 700

x2 = 140

p̂2= 0.20

The manager wants to prove that the population proportion of traveler’s check buyers

who buy at least $2,500 in checks when sweepstakes prizes are offered is at least 10%

higher than the proportion of such buyers when no sweepstakes are on. Therefore, this

should be the manager’s alternative hypothesis. We have H0: *p*1 \_ *p*2 0.10 and H1:

*p*1 \_ *p*2 0.10. The appropriate test statistic is the statistic given in equation 8–13:

Z=

*p*$1 - *p*$2 - *D*

2*p*$1 (1 - *p*$1 )>*n*1 + *p*$2(1 - *p*$2)>*n*2

=

120>300 - 140>700 - 0.10

2[(120>300)(180>300)]>300 + [(140>700)(560>700)]>700

=

(0.4 - 0.2) - 0.1

2(0.4)(0.6)>300 + (0.2)(0.8)>700

= 3.118

This value of the test statistic falls in the rejection region for \_\_0.001 (corresponding

to the critical point 3.09 from the normal table). The *p-*value is therefore less than

0.001, and the null hypothesis is rejected. The manager is probably right. Figure 8–10

shows the result of the test.

Confidence Intervals

When constructing confidence intervals for the difference between two population

proportions, we do not use the pooled estimate because we do not assume that the

two proportions are equal. The estimated standard deviation of the difference

between the two sample proportions, to be used in the confidence interval, is the

denominator in equation 8–13.

The Comparison of Two Populations 327

=

(0.4 - 0.2) - 0.1

2(0.4)(0.6)>300 + (0.2)(0.8)>700

= 3.118

=

120>300 - 140>700 - 0.10

2[(120>300)(180>300)]>300 + [(140>700)(560>700)]>700

*z* =

*p*$1 - *p*$2 - *D*

2*p*$1 (1 - *p*$1 )>*n*1 + *p*$2(1 - *p*$2)>*n*2

A large-sample (1 \_ \_) 100% confidence interval for the difference between

two population proportions is

*p*$1 - *p*$2\_ *z* (8–14) a*/*2 A

*p*$1(1 - *p*$1)

*n*1

+

*p*$2(1 - *p*$2)

*n*2

In the context of Example 8–8, let us now construct a 95% confidence interval for

the difference between the proportion of BankAmerica traveler’s check buyers who

buy more than $2,500 worth of checks during sweepstakes and the proportion of

Problem Statement 7: [100 marks]

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 n − 1.

* Let us hypothesize that the die is unbiased. If that is so, the probability of obtaining any one of the six numbers is 1/6 and as such the expected frequency of any one number coming upward is 132 x 1/6 = 22.
* Σ[*(O*i –*E*i)2/*E*i ]= 9 [198/22]
* Hence, the calculated value of *X*2 = 9.
* Degrees of freedom in the given problem is

(n-1) = (6-1) = 5.

* The table value of *X*2 for 5 degrees of freedom at 5% level of significance is 11.071. Comparing calculated and table values of *X*2, we find that the calculated value is less than the table value and as such could have arisen due to fluctuations of sampling.
* The result, thus, supports the hypothesis and it can be concluded that the die is unbiased

Problem Statement 8: [100 marks]

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 = 6383

Not voted 1486 2131 = 3617

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”?

Ans

Now we have the observed table and the expected tble under the null hypothesis of independence. Now we need to compute X2

* (O – e)2

E

* (2792 – 2731)2  = 1.36

2731

(3591 – 3652)2  = 1.0

3652

Etc . X2 =  1.4+1.0+2.4+1.8 = 6.6

Degrees of freedom 2 – 1 = 1

* Since X2  is 6.6 which has a p value of 1%, we have to reject the NULL hypothesis. The data supports the hypothesis that sex and voting are dependent in this town.

Problem Statement 9: [100 marks]

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].

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** |
| **(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 10: [100 marks]

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

**Question 2:**

|  |  |  |
| --- | --- | --- |
| **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:

χ2 = **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 11: [100 marks]

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].

**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:

χ2 = **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 12: [100 marks]

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

Problem Statement 13: [100 marks]

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.)