Testing of Hypothesis - Proportion test

Introduction

A two sample Z test of proportions determines whether two populations differ significantly on specific characteristics. In other words, compare the proportion of two different populations that have some single characteristic. Based on the difference between the proportions of the population, the range of values is calculated.

Procedure:

- Import the data set
- Determine the critical value and sample statistic using R functions
- Conclude the problem using R functions

Problem no 1:

In a CBSE School X, 89 from a sample of 600 students had scored 100% marks in class 10. In the other CBSE School Y,92 from a sample of 900 students had the same marks. We want to study whether there is any significant difference between the two proportions at the 5% level of significance.

Code and Results:

```
# Perform two-proportion z-test
prop.test(x=c(89,92),n=c(600,900),alternative="two.sided")
```

```
##
##  2-sample test for equality of proportions with continuit
y correction
##
## data:  c(89, 92) out of c(600, 900)
## X-squared = 6.7859, df = 1, p-value = 0.009188
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.01007348 0.08214875
```

```
## sample estimates:
##    prop 1    prop 2
## 0.1483333 0.1022222
```

Interpretation:

X-squared: value of the Pearson's chi-squared test statistic.

df: the degree of freedom of the approximate chi-squared distribution of the test statistic.

p-value: p-value corresponding to a statistic. Here the p-value is0.009188.

alternative: Alternative hypothesis is used for the z-test. In this case, an alternative hypothesis is population proportions are not equal, i.e. two-tailed.

95 percent confidence interval: shows us the 95% confidence interval for the true proportion. 95% confidence interval is [0.01007348,0.08214875].

Sample estimates: Shows the sample proportions. prop1=0.148 &prop2=0.102

Since the p-value[0.009188] is less than the level of significance α = 0.05, we can reject the null hypothesis.

This means that the population proportions are not equal.

Problem no:2

Right-tailed two proportion test

Out of a sample of 1000 people, 300 watched movies in theatres prior to the pandemic. In a sample of 1200 persons, it was discovered that 350 of them watched movies in theatres after the outbreak. We wish to investigate whether there is a decline in theatre attendance following the pandemic at the 5% level of significance.

Code:

```
# Perform two-proportion z-test
prop.test(x=c(800,900),n=c(1000,1200),alternative="greater")
```

```
##
##  2-sample test for equality of proportions with continuit
y correction
##
## data:  c(800, 900) out of c(1000, 1200)
## X-squared = 7.4826, df = 1, p-value = 0.003115
## alternative hypothesis: greater
## 95 percent confidence interval:
##  0.01983221 1.00000000
## sample estimates:
## prop 1 prop 2
##   0.80   0.75
```

Interpretation:

X-squared: It's the value of Pearson's chi-squared test statistic.

df: It is the degree of freedom of the approximate chi-squared distribution of the test statistic.

p-value: p-value corresponding to a statistic. the p-value is 0.0031

alternative: Alternative hypothesis used for the z-test. In this case, an alternative hypothesis is population proportions for people watching movies in theatre after the pandemic is less,i.e. right-tailed.

95 percent confidence interval: It shows a 95% confidence interval for the true proportion. Here the 95% confidence interval is [0.0198,1.0000].

sample estimates: It shows the sample proportions. In this case, prop1= 0.80 & prop2= 0.75

Since the p-value is less than the level of significance $\alpha = 0.05$, we reject the null hypothesis.

This means that the population proportions of people watching movies in theatres after pandemic decreased.