**Assignment No. 9.2**

**Question No. 2**

How do you test the proportions and compare with hypothetical proportions? Test Hypothesis: Proportion of automatic cars is 40%

Answer: Proportion is another kind of estimate like mean and standard deviation of a population. It measures the number of elements in a population or sample, with a particular characteristic feature, out of the total population or sample. So the way to arrive at a proportion is to simply divide the number of elements with a particular characteristic by the total population or sample. Since one cannot work on the entire population, statistics teaches to work on samples, and then basing on the sample proportions and their known patterns of distribution, inference about the population proportion can be done.

Sample Proportion = p’ = No. of elements with desired feature / Sample Size

Standard Deviation of sample proportion (finite population) = where N = population size, n = sample size, p = is the sample proportion

Standard Deviation of sample proportion(Infinite population) =

With the above information in case we have a sample of size ‘n’ and the mean proportion of the sample is estimated as say “ p’ ”, and the hypothetical proportion of the population is say ‘p’, then if the n \* p >= 5, we can appropriate the sample distribution to normal distribution and then the usual z statistic can be calculated as follows to arrive at the z statistic, and then it can be compared with the critical value for a given level of confidence or significance level (alpha)

Z statistic = [(p’ – p) / (p(1-p) / n))^0.5].

Depending on the hypothesis higher or lower tail critical values can be picked up from the standard normal tables and the appropriate inferences can be drawn.

This is how the proportions can be tested.

Now regarding the proportion of automatic cars from the mtcars dataset, we count the number of cars with automatic transmission, then calculate the proportion of them in the total cars, then compare this proportion with 0.4 as asked in the question, for an ‘n’ of total cars in the dataset.

The z statistics will arrived using the above formula and then compared with the two tailed test critical values for 95% confidence interval. The r code will be as follows

> data(mtcars)

> table(mtcars$am)

0 1

19 13

> trans <- as.data.frame(table(mtcars$am))

> trans

Var1 Freq

1 0 19

2 1 13

> SamProp <- trans[1,2]/sum(trans$Freq)

> PopProp <- 0.4

> n <- sum(trans$Freq)

> ((19/32)-0.4)/sqrt((0.4\*(1-0.4))/32)

[1] 2.237232

> SamProp

[1] 0.59375

> 19/32

[1] 0.59375

> z <- (SamProp - PopProp) / sqrt((PopProp\*(1-PopProp))/n)

> z

[1] 2.237232

> qnorm(1-(0.05/2)) # critical Value for 95% confidence

[1] 1.959964

Now that we have the z statistic value greater than the critical value, we reject the null hypothesis that the proportion of automatic cars is 0.4.

**QUESTIONNO.1**

The p value for the null hypothesis of the proportion of cars being 0.4 is calculated as follows with the r code

> pvalue <- 2 \* pnorm(z,lower.tail=FALSE)

> pvalue

[1] 0.02527116

> 1-pnorm(z)

[1] 0.01263558

> 2 \* (1-pnorm(z))

[1] 0.02527116

> prop.test(trans[1,2],sum(trans$Freq),0.4,"two.sided",0.95,FALSE)

1-sample proportions test without continuity correction

data: trans[1, 2] out of sum(trans$Freq), null probability 0.4

X-squared = 5.0052, df = 1, p-value = 0.02527

alternative hypothesis: true p is not equal to 0.4

95 percent confidence interval:

0.4226002 0.7448037

sample estimates:

p

0.59375

Since the p value is less than the 5% significance level or 0.05, clearly we reject the null hypothesis that the proportion of automatic cars is 0.4 in the mtcars dataset.