Population-Proportion

A merchant claimed that more than 45% of the diamonds in the world are ideally cut. Meaning to say that 45% of the diamonds are perfectly cut and other 65% of the diamonds have cut quality less than those ideally cut diamonds.

To test this claim, a team of analytics wanted to run a hypothesis test based on population proportion to determine that less than 45% of diamonds are ideally cut.

Dataset

The dataset collected comes with ggplot2 packages in R, and the dimension of the dataset is 53940 rows and 10 columns. The distribution of the cuts is as follows:
Fair- 2.9 %, good-9%, very good-22%, premium-25.5%, ideal-40%.

Hypothesis statement

Since the data collected have only 40% ideal cut diamonds, however, are they enough evidence to say that all the diamonds around the world is less than 45%. Hypothesis testing can be used in this situation.

H0: 45% of diamonds are ideally cut. Mu= 45%
H1: Less than 45% of diamonds are ideally cut. Mu< 45%

Code

```r
install.packages("ggplot2") # installing the package
library(ggplot2) # loading the ggplot package
df <- diamonds # laoding the diamond dataset into a variable

total_count <- nrow(df) # counting total number of trials (n)

propertion <- (sum(df$cut=="Ideal")/total_count)*100 # calculating propertion of the item that has to be investegated

result <- prop.test(propertion,total_count,p=0.45,alternative="less") # running a population propertion test
result # printing the test result

```

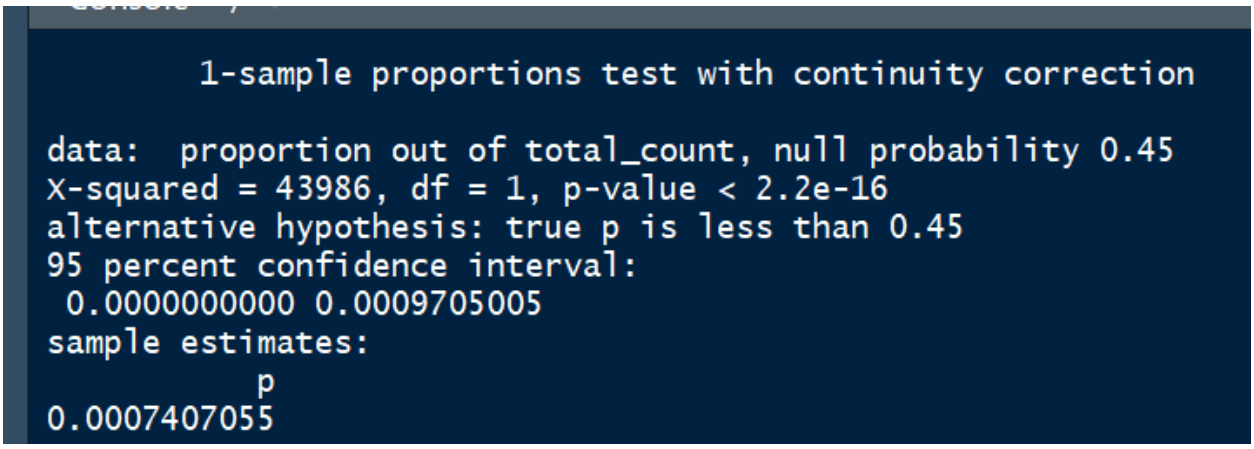
First, we have to load the diamonds dataset into a variable df. Then, since we need to know how many trails we had we need to count the number of times we have collected the data on cuts. I have saved that into the variable called total_count.

Then we also need to store a variable that can hold value for the proportion of desired outcome, that is the proportion of diamonds having ideal cut and I stored it into the variable called proportion.

Then I ran the proportion test where first argument is the percentage of diamonds having ideal cut, second is the number of trails, third signifies as to what percentage are, we are comparing our result to, and the last argument defines the alternative hypotheses.

Overall, we are saying that calculate the test statistics such that our alternative hypothesis is saying that the desired outcome has the probability of occurring less than 45%.

Result



The result says that the probability of our null hypotheses being true is very small. In other words, the probability of total number of diamonds being equal to 45% is very less. Hence, we reject the null hypothesis towards the alternative hypothesis that say less than 45% of diamonds are ideally cut.