**Project 2**

**Name: Luxi Liao**

**Questions:**

1. Distribution of the given sample:
2. The distribution of the random variable X:

For individual trial:

For the experiment:

1. Numerical Methods:

Sample mean: 0.16

Sample median: 0

10% trimmed mean: 0.075

Range: 1-0 = 0

Sample Variance: 0.1371429

Sample Standard Deviation: 0.370328

1st Quantile: 0

3rd Quantile: 0

1. Graphical Method:

A picture containing chart

Description automatically generated

Graph1: Barplot

1. Given here are 8 Heads and 42 Tails in the given sample:
2. Prove by contradiction:

Assume John is right and 40% of the result have Heads.

1. Simulation is run with parameters n = 50, x = 1, p = 0.4 and total experiment number of N = 1000. It generated 1000 instances of the experiment (a data frame of 1000 rows and 50 columns). The generated dataset is exported into a CSV file and included in the appendix.

The simulation is run using the R statement below:

**virtual <- replicate(1000, rbinom(50, 1, 0.4))**

1. The sample proportion of heads is generated with:

**phat50s = colSums(virtual)/50**

The statement adds the number of heads in each trial and divides the sum by sample size to reach the proportion.

The first 10 instances of the result are shown below:

**> head(phat50s, n=10)**

**[1] 0.38 0.38 0.30 0.38 0.40 0.40 0.42 0.54 0.40 0.40**

1. Using the R statement for plotting the histogram:

**hist(phat50s, prob = TRUE, main = "Histogram of Phat50s")**

Chart, histogram

Description automatically generated

Graph2: histogram of sample proportions phat50s

The label is marked based on the generated histogram using:

**points(phatstar, 0, col = "red", pch = 20)**

Chart, histogram

Description automatically generated

Graph3: histogram of phat50s with label

1. The relative frequency that from the observation is larger than the :

**mean(phat50s < phatstar)**

This generates the result of

**[1] 0**

1. Given the analysis result above, the following function is used to get the p-value:

**pvalues=function(p0,phatstar,n,N){**

**result=replicate(N,mean(rbinom(n,1,p0)))**

**pvalue=mean(result<phatstar)**

**return(pvalue)**

**}**

With input parameter p0 = 0.4, phatstar = 0.16, n = 50, N = 1000

The result is or **0** in some circumstances.

The p-value is smaller than 5%. So, John will accept Jack’s objection according to the rule.

1. (1) Given that follows Binomial Distribution,

With this approximation, =

**binomialcdf(N = 50, p = 0.4, x = 8)** .

The result is approximately **.**

(2) Using the Central Limit Theorem:

Given that the Number of RVs N > 30, the binomial distribution has an approximate Normal distribution: , μ = 0.4, σ = = 0.069282.

The approximate probability can be given by

**normalcdf(0,0.16,0.4,0.069282)** .

The result is.

1. The sequence of guesses is traversed in R and the guesses with corresponding p-value less than 0.05 are rejected. The acceptable guesses are listed below.

**[1] 0.00 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.10 0.11 0.12 0.13 0.14 0.15 0.16 0.17 0.18 0.19**

**[21] 0.20 0.21 0.22 0.23 0.24**

The maximum acceptable guess is 0.24, with a p-value of 0.0588. The minimum guess is 0.00, with a p-value of 1.0000.

The R code used to solve the question can be found in the Appendix.

1. The rejection is implemented in R. The sequence of guesses is traversed, and the guesses that satisfy the rejection rules are rejected. The accepted values are listed below:

**[1] 0.09 0.10 0.11 0.12 0.13 0.14 0.15 0.16 0.17 0.18 0.19 0.20 0.21 0.22 0.23 0.24 0.25 0.26 0.27 0.28**

The maximum acceptable guess is 0.28. The minimum acceptable guess is 0.09.

The R code used to solve the question can be found in the Appendix.

**Appendix:**

The simulation result “virtual”: [virtual.csv](https://github.com/RichardBrowning/MachineLearningWithR/blob/main/350_project_2/virtual.csv)

R code used for the analysis on GitHub: [project2.R](https://github.com/RichardBrowning/MachineLearningWithR/blob/main/350_project_2/project2.R)