

Class Project in Two Parts: Central Limit Theorem, Inferential Data Analysis

Paul Clark

February 5, 2017

Contents

1 Simulation Exercise (The Central Limit Theorem)	1
2 Basic Inferential Data Analysis	2

Review criteria

Did you show where the distribution is centered at and compare it to the theoretical center of the distribution?

Did you show how variable it is and compare it to the theoretical variance of the distribution?

Did you perform an exploratory data analysis of at least a single plot or table highlighting basic features of the data?

Did the student perform some relevant confidence intervals and/or tests?

Were the results of the tests and/or intervals interpreted in the context of the problem correctly?

Did the student describe the assumptions needed for their conclusions?

1 Simulation Exercise (The Central Limit Theorem)

In this project you will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with `rexp(n, lambda)` where `lambda` is the rate parameter. The mean of exponential distribution is $1/\lambda$ and the standard deviation is also $1/\lambda$. Set $\lambda = 0.2$ for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should...

Show the sample mean and compare it to the theoretical mean of the distribution.

Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

Show that the distribution is approximately normal.

In point 3, focus on the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.

Title (give an appropriate title) and Author Name

Overview: In a few (2-3) sentences explain what is going to be reported on.

Simulations: Include English explanations of the simulations you ran, with the accompanying R code. Your explanations should make clear what the R code accomplishes.

Sample Mean versus Theoretical Mean: Include figures with titles. In the figures, highlight the means you are comparing. Include text that explains the figures and what is shown on them, and provides appropriate numbers.

Sample Variance versus Theoretical Variance: Include figures (output from R) with titles. Highlight the variances you are comparing. Include text that explains your understanding of the differences of the variances.

Distribution: Via figures and text, explain how one can tell the distribution is approximately normal.

2 Basic Inferential Data Analysis

Now in the second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package.

Load the ToothGrowth data and perform some basic exploratory data analyses

Provide a basic summary of the data.

Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)

State your conclusions and the assumptions needed for your conclusions.