

INTRODUCTION TO RANDOM SAMPLING

A random sample provides equal probabilities for all subsets of this data frame. Each subject in the sample has an equal probability of selection. This lowers sampling bias. This independent exercise will help you learn how to generate simple random samples, and use the t-test to examine random samples. Furthermore, you will develop an understanding of the effects of the number of sampling elements on confidence interval.

PURPOSE

This activity will help you to combine lessons from descriptive statistics with an exploration of data.

LEARNING OBJECTIVE

Use graphical analysis and scripting to explore the simple random sampling design.

REQUIRED RESOURCES

- o R, R Studio
- o PostgreSQL ODBC driver
- o Direct connection to PostgreSQL with R
- o Sample R script

TIME ALLOCATED

90 minutes in class

TASKS



A. Introduction to Random Sampling

With a sample application of R, it is easy to begin to develop more advanced numerical methods such as simulations, bootstrapping, and Monte-Carlo type analysis. The most basic function is `sample()` which samples (randomly) n elements from a data frame x . Let's apply this method to show how the mean of a dataset is just an estimate of the true value. The larger the sample, the more likely the estimate is true but there is always some element of chance involved with drawing conclusions on statistically tested samples.

This activity is accompanied with a script that contains a simple loop, where the sample size has been set to 150. In each t-test, 100 random samples are selected separately from *Set1* and *Set2*, then a t-test of the equality of the means are returned. The 100 p-values of each t test are plotted in Figure 71 as well as a histogram. It indicates in most tests, the null hypothesis is rejected or, the difference in means of two random samples is not zero.

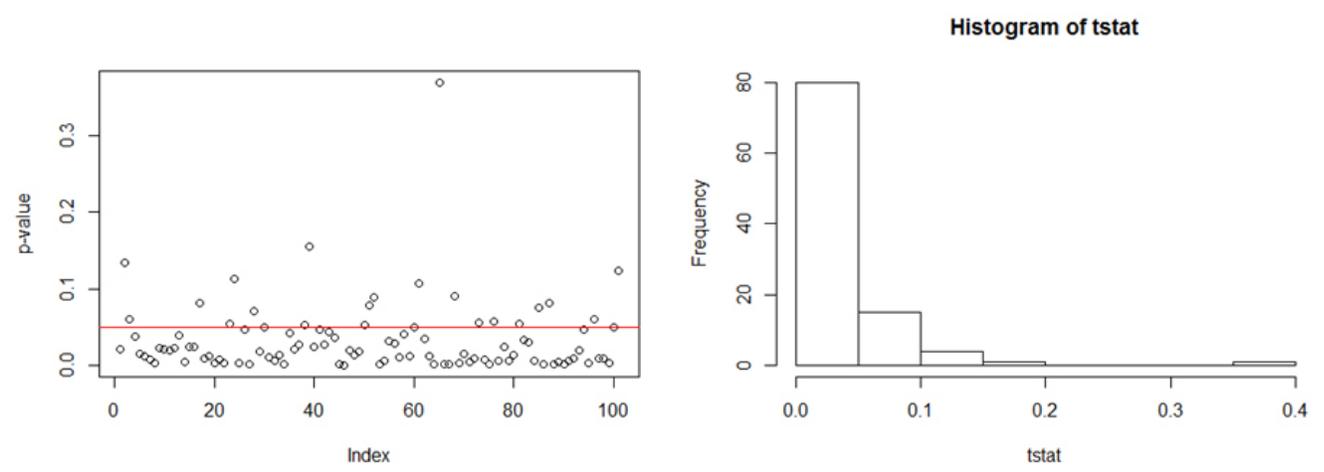


Figure 71 Random sample graphics of p-values for 100 t-tests

Since the sampling is random, your plot will not necessarily look like those in Figure 74. Rerun the code to see the plots changing. Reduce the sample size and see how this dramatically changes the outcome. This is primarily because the difference we are trying to detect is fairly small (between the means) and with smaller samples, the confidence interval becomes larger and more tests lead to the conclusion that there is no difference of the mean.

B. Effect of n on confidence interval

Use the sampling procedure shown above for the data from *set 1* to demonstrate the effect of n , or the number of sampling elements, on the confidence interval estimate. Write a loop that calculates the mean and 95th confidence interval for samples of the size 1:249. The last calculation (of all elements in set 1) will be what is shown in Figure 71. Make a plot with the x-axis being the number of elements in the sample and the upper interval, mean, and lower interval plotted on the y-axis. Comment on your observation and interpretation of this plot.

DELIVERABLE



Prepare a short write up of your discovery. Submit printed copy and attach graphs.

ASSESSMENT



Participation points (10)

Activity 38 Grading Rubric

	Excellent (10)	Good (8)	Poor (6)
Graphic	Graphics are complete and reflect the discussion regarding the effect of sample size on the confidence interval.	The graphic lacks visual components that assist in the viewer understanding the purpose of the graphic.	Graphics cannot be understood and do not reflect the learning concepts of the activity.
Discussion/Commentary	Insightful discussion or commentary relating to the question at hand demonstrating student understanding of the task.	Discussion or commentary was partially incomplete.	Minimal to no discussion or commentary.