**ALY2010**

**Project 5 100 Points**

# Overview and Rationale

This project will help you measure your understandings of basic concepts on analytics.

It will help you measure your skills on R, R Studio and R Markdown.

It will help you to measure your level of understanding of the processes of testing the difference between two means, two proportions, and two variances.

It will help you measure your skills to apply critical thinking to make meaningful observations of your data analysis results.

**Assignment**

# Part 1. Title and Introduction

Prepare your report using R Markdown and present your report using an HTML file by clicking on ( “Knit to HTML”) in R studio for HTML file.

1. **Title**: Present a title in your report.
2. **Introduction**: Present a well informative introduction section, this will measure your understanding of the topic and analytical processes for data analysis:

Your introduction needs good information and good organization. This applies for any report you make. Try to separate each topic in a paragraph.

* Hypothesis testing**:** Using your own words, talk about the significance of the hypothesis test process. Provide a practical example.
* Testing differences**:** Using your own words, talk about the significance of comparing two sample means, using experimental and control groups. Provide a practical example.

Use Bluman as a reference. Also present **at least one additional academic reference for each topic**.

# Part 2. Analysis section

Task 1. Hypothesis test using z test (n>30).

You have a population with n = **3000**, mean = **16.4**, and a standard deviation = 1.35.

From that population you obtain a sample of n = **65**, with a mean = **17**.

Test the alternative hypothesis that the population mean is different than 16.4

Use a confidence level of 99%. In this case, alpha (significance level) is 1 - 0.99 = 0.01 (**α = 0.01**).

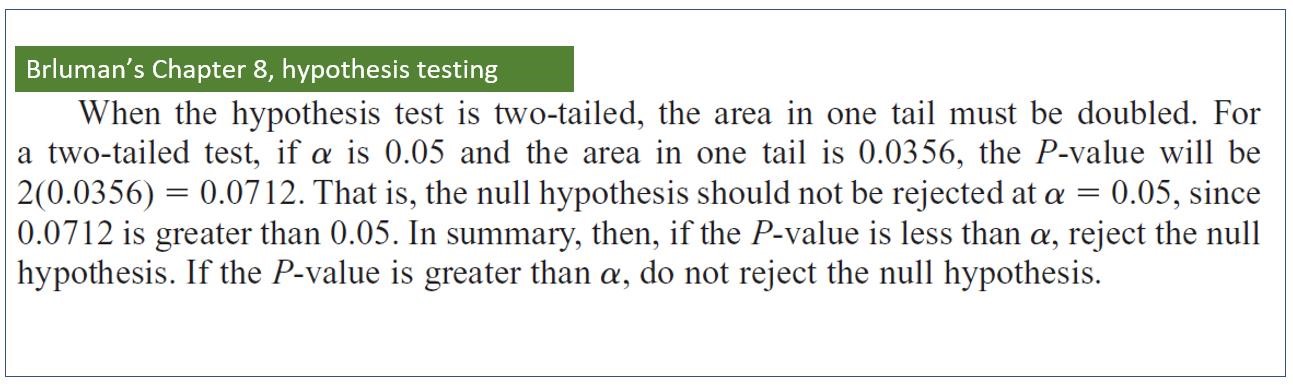
**Test your hypothesis using your Z test value.**

1. Using your data, you need to calculate your z test.
2. Using the alpha value, you need to calculate the critical value. Would you use a CV on the right or on the left of the curve?
3. Then compare your z test to the critical value.
4. To reject Ho, your Z test value must be higher than the CV on the right side of the curve, and smaller if compared in the left side. Logical questions would be: Ztest > CV (right side) or Ztest < CV (left side) , and the R outcome should be TRUE.
5. If all your calculations are correct, and the outcome to your logical question R code is FALSE, then you cannot reject Ho.

**Test your hypothesis using your p value.**

1. Using your z test value, calculate your p value. Remember than in 2-tailed test, the value you obtain must be multiplied by 2 to obtain the actual p value.
2. Then compare the p value with the alpha value.
3. To reject Ho, your p value must be smaller than the alpha value, on either side of the curve. The Logical question would be: pvalue<alpha (ether side), and the R outcome should be TRUE.
4. If all your calculations are correct, and the outcome to your logical question R code is FALSE, then you cannot reject Ho.

Write a short paragraph explaining if, based on the test results, you have enough evidence to reject Ho.



Task 2. Hypothesis testing using t test (n<30).

You have a population with mean = 16.40943.

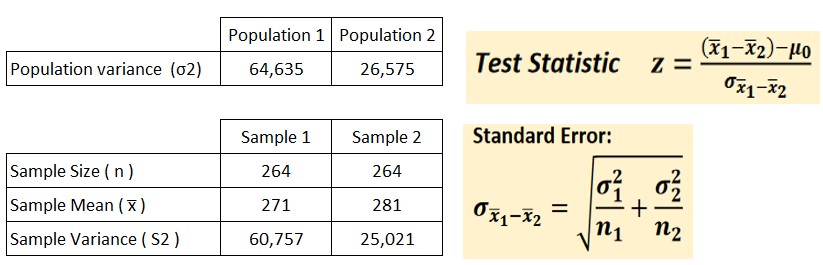
The sample is: n = 25, with mean = 16.02584 and standard deviation = 0.7432.

Using a confidence level of 99% (α = 0.01), test the alternative hypothesis that population mean is different than 16.40943.

Write a short paragraph explaining if, based on the test results, you have enough evidence to reject Ho.

Task 3. Compare two means using z test.

Using the following data:



At α = 0.05, use hypothesis testing to test your claim that the two-sample means are different.

Run your test by (a) comparing your z test with the corresponding critical value.

Also, run your test by comparing your p value with the α value.

Write a short paragraph explaining if, based on the test results, you have enough evidence to reject Ho.

Task 4. Compare two independent means using t test.

Create an R chunk for this task to enter all your codes.

Use the two vectors below to enter the data in R.

Sample\_4a=c(1.75, 1.18, 1.19, 1.52, 2.73, 1.91, 2.87, 1.39, 1.85, 1.65, 2.92, 2.40, 1.38, 2.86, 2.56, 2.85, 2.66,

2.43, 2.70, 2.27, 2.99, 2.47, 1.90, 1.49, 1.10, 2.15, 2.51, 2.66)

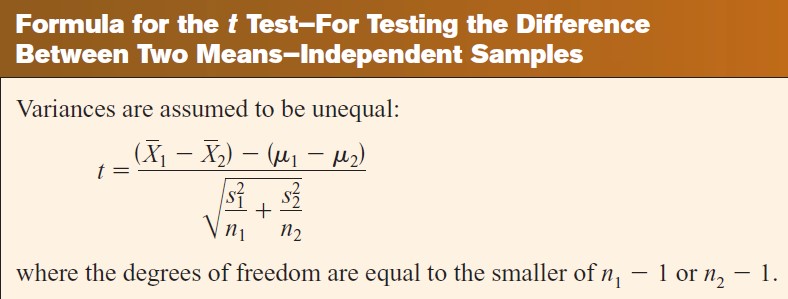
Sample\_4b=c(2.68, 2.16, 2.50, 2.18, 2.30, 3.00, 2.28, 2.00, 2.19, 2.11, 2.17, 2.57, 2.98, 2.01, 2.35, 2.23, 2.54,

2.85, 2.04, 2.23, 2.87, 2.46, 2.73, 2.43, 2.06, 2.46, 2.84, 2.32)

The n value is 28 (n=28, therefore, degrees of freedom df = 27).

Calculate the mean and standard deviations of each sample.

• At α = 0.01, use hypothesis testing to test your claim that the two means are different.



Task 5. Correlation and regression analysis

Use ?faithful in the console and read the information about faithful. This is a public data set.

1.1 Using your own words, describe the data set.

Prepare one single R Chunk and present your answers using either one Table or using Inline R Codes.

1.2 What is the coefficient of correlation between eruption time and waiting time?

1.3 Explain the meaning of the coefficient of correlation?

1.4 What is the coefficient of determination between eruption time and waiting time?

1.5 Explain the meaning of the coefficient of determination?

1.6 Create an object to obtain the linear regression model for eruption time and waiting time. Remember that the independent variable is waiting time. Do not present the values yet, just create the object.

1.7 Present a scatter plot of eruption time versus waiting time. Remember that the independent variable is waiting time. Using the object you create in 1.6, add the regression line to the graph.

Add a good title, and x- and y-axes labels.

Present data points as triangles (check codes using ?pch).

Add the regression line with a color.

Increase the thickness of the regression line.

Write the formula for the regression line inside the plot.

1.8 Explain the values for intercept and slope.

1.9 Describe the direction and strength of the regression line and explain what they tell you about your data.

Check this page for pch codes:

<http://www.sthda.com/english/wiki/r-plot-pch-symbols-the-different-point-shapes-available-in-r>