

Lab 3: Part 2

The aim of this part of the lab exercise is to give you practical experience in Hypothesis testing using R Studio and an R Notebook.

1 Hypothesis testing

1. A company produces synthetic diamonds that have an average weight of 0.5 carat. An experiment has been conducted to evaluate a new process for producing synthetic diamonds. Six diamonds have been generated by the new process, with recorded weights of 0.46, 0.61, 0.52, 0.48, 0.57 and 0.54 carat. It is essential that the new process produces diamonds with a weight in excess of 0.50 carat. Do the six diamond weight measurements present sufficient evidence to indicate that the average weight of the diamonds produced by the process is in excess of 0.5 carat? Test at the 1% significance level.
2. Using the data on breast cancer diagnosis from the University of Wisconsin.
 - (a) Read in the data set *breastCancer_Wisconsin.csv*. This is a sample of data on Breast Cancer.
 - (b) Explore the data and the variables. You can do this through the use of numerical summaries and graphs.
 - (c) Compute the proportion of tumours classified as benign (b) in the sample data you read in in step (a)
 - (d) The proportion of benign tumours in the population from which this sample data is drawn from is 0.6. Is the proportion in this sample data supportive of this proportion in the population? Write out the hypothesis to test and run this hypothesis test in R. Is H_0 rejected?
 - (e) Explore the `radius.mean` attribute (column) in the data. Do so numerically and graphically.
 - (f) Is there a difference between the radius mean for benign vs malignant? (Hint: use both numerical and visual clues)
 - (g) in step (c) you computed the proportion of tumours classified as benign. What is the 95% confidence interval for this proportion?

2 R commands and syntax

```
1 #reading in a csv file
2 df<-read.csv("name-of-csv-file")
3
4 #the cumulative distribution function of the t values
5 pt(test.statistic , df)
6
7 #the cumulative distribution function for the standardised Normal distribution Z
8 pnorm(test statistic)
9
10 #Testing if a variable is normally distributed
11 shapiro.wilks(variable)
12
13 #If there are two samples – test for equal variances
14 var.test(variable1 , variable2)
15
16 #If there are two samples – testing for equal means
17 # R will handle non-equal variances
18 t.test(variable1 , variable2)
```