Data 1204: Statistical and Predictive Modelling

Assignment #2

Hypothesis Testing



Name: **Raj Dholakia**

Student Number: **100813041**

Lecturer(s): **Ritwick Dutta**

Table of Contents

[1. Introduction 1](#_Toc63569316)

[2. Hypothesis Statement 1](#_Toc63569317)

[3. Process to Test the Hypothesis 2](#_Toc63569318)

[4. Testing the Hypothesis in R 3](#_Toc63569319)

[4.1 About the notebook 3](#_Toc63569320)

[4.2 Preparing the Environment and Loading Data 3](#_Toc63569321)

[4.3 Calculating Key Statistics 4](#_Toc63569322)

[4.4 Testing the Null Hypothesis (p-value calculation): One-tail Test 5](#_Toc63569323)

[Conclusion 6](#_Toc63569324)

[References iii](#_Toc63569325)

# 1. Introduction

In this assignment we are looking at a small blood pressure dataset consisting of 25 data points. The dataset tries to better understand the impact of regular exercise on blood pressure of the participants. The experiment was created on 25 individuals, hence the **25 data points**. The dataset has **three columns**:

1. **Subject**: this is an identification number for the subjects that were part of the experiment. The variable **ranges from 1 to 25.**

2. **Before**: this is the **blood pressure of the subject** at the beginning of the experiment.

3. **After**: this is the **blood pressure of the subject after 30 days** into the experiment. These 30 days the participant is expected to exercise on a regular basis.

The goal of the experiment, as defined in the assignment, is to determine if exercising regularly brings any changes in the blood pressure of the subject.

A null hypothesis testing will be carried out to meet the goal of the experiment in this assignment. The results obtained from the experiment are saved in **BloodPressure.xls**.

# 2. Hypothesis Statement

A null hypothesis can be defined to determine if there is exercising regularly has any impact on blood pressure. Assuming exercising has no impact on blood pressure, not change in blood pressure after 30 days of regular exercise, can help us create a null hypothesis. Hence,

H0 (null hypothesis): Regular exercise does **not** have any impact on blood pressure. Mathematically: **difference in blood pressure before and after should be equal to 0**.

Ha (Alternate hypothesis): Regular exercise brings changes in blood pressure.

# 3. Process to Test the Hypothesis

The following steps will be carried out to check if the null hypothesis is true or false.

1. Find the **absolute percentage change (percent\_change)** in blood pressure before and after the experiment for each data point. Here we take the absolute difference because we looking for a change both ways. It can be negative or positive.

2. If the null hypothesis is true, we expect percent\_change to have a mean that is close to 0. In this step we calculate the mean and standard error in the 25 data points present. As absolute percentage change is being used, we eliminate the chances of the mean averaging out 0 due to opposite signs. For example, if some of the differences are negative (mean = -10) and some are positive (mean = 10), overall, a mean of 0 would be obtained. This would give a false indication of there being no change in blood pressure due to regular exercise. Hence, using the absolute value, eliminates this possibility. Furthermore, using the percentage values gives a better indication on how big of a change it is.

Note: As the sample size is **less than 30**, we will be using the t-test and **not** the z-test to test the null hypothesis.

3. We use the mean of **0** and compare it to the **mean of percent\_change**. We calculate the **p-value** using the **one-tail test**. We use the one-tail test as we are using absolute values, which cannot be negative and only has to be greater than a certain value. The p-value is calculated by through the t-distribution table (Glen, n.d.).

4. If the p-value found is **lesser than 0.05 or 5%**, the null hypothesis can be assumed false and we can say that there is a 95% probability that regular exercise does bring about a change in blood pressure. If the value is greater than 0.05, the null hypothesis cannot be dis-proved and nothing can be concluded.

# 4. Testing the Hypothesis in R

## 4.1 About the notebook

In this notebook we will test the following hypothesis:

The process followed to determine if the null hypothesis is true is described in the word document. As the sample size is less than 30, the t-distribution table is used to determine the p-value.

## 4.2 Preparing the Environment and Loading Data

Install required packages.

# install.packages(c("lattice", "readr", "readxl"))

Loading the required libraries.

library(lattice)  
library(readxl)  
library(readr)  
  
options(digits=3)

Loading blood pressure data.

bp\_data <- read\_excel("BloodPressure.xls")  
bp\_data

## # A tibble: 25 x 3  
## Subject Before After  
## <dbl> <dbl> <dbl>  
## 1 1 135 127  
## 2 2 142 145  
## 3 3 137 131  
## 4 4 122 125  
## 5 5 147 132  
## 6 6 151 147  
## 7 7 131 119  
## 8 8 117 125  
## 9 9 154 132  
## 10 10 143 139  
## # … with 15 more rows

## 4.3 Calculating Key Statistics

Calculating the absolute percentage change in blood pressure of the subjects before and after the 30-day experiment.

bp\_data$percent\_change <- abs((bp\_data$After - bp\_data$Before)/bp\_data$Before)  
bp\_data

## # A tibble: 25 x 4  
## Subject Before After percent\_change  
## <dbl> <dbl> <dbl> <dbl>  
## 1 1 135 127 0.0593  
## 2 2 142 145 0.0211  
## 3 3 137 131 0.0438  
## 4 4 122 125 0.0246  
## 5 5 147 132 0.102   
## 6 6 151 147 0.0265  
## 7 7 131 119 0.0916  
## 8 8 117 125 0.0684  
## 9 9 154 132 0.143   
## 10 10 143 139 0.0280  
## # … with 15 more rows

Calculating the mean, standard deviation and mean standard error of the sample provided. This will be used to calculate the t-value, which will further be used to calculate the p-value.

mean.percent\_change <- mean(bp\_data$percent\_change)  
sd.percent\_change <- sd(bp\_data$percent\_change)  
SE.percent\_change <- sd.percent\_change / sqrt(length(bp\_data$percent\_change))  
mean.percent\_change

## [1] 0.0677

sd.percent\_change

## [1] 0.0547

SE.percent\_change

## [1] 0.0109

## 4.4 Testing the Null Hypothesis (p-value calculation): One-tail Test

Null hypothesis:

Alternate hypothesis:

Calculating the t-value.

# State the Ho value and calculate the z-score  
Ho <- 0  
t <- (mean.percent\_change - Ho) / SE.percent\_change  
t

## [1] 6.19

Calulating the p-value.

p\_value <- pt(-t, df=length(bp\_data$percent\_change)-1)  
p\_value

## [1] 1.06e-06

p\_value < 0.05 # Check if the p-value is less than 0.05 or 5%.

## [1] TRUE

# Conclusion

From the above analysis, the p-value was found to be 1.06e-06. This is way smaller than 0.05 we had decide at the beginning of the analysis. Hence, it can be confidently concluded that the null hypothesis is false. It can be rejected and the alternate hypothesis, which states that there is an impact on the blood pressure due to regular exercise, can be deemed true. **30 days of regular exercise does bring changes to blood pressure.**

# References

Glen, S. (n.d.). T-Distribution Table (One Tail and Two-Tails). Retrieved February 7, 2021, from From StatisticsHowTo.com: Elementary Statistics for the rest of us! website: https://www.statisticshowto.com/tables/t-distribution-table/