Negin Dehghanian Mohammad Movahedi Rohan Ahmed ALY6010 - Probability Theory and Introductory Statistics Northeastern University

Analyzing Cats’ Bodyweight and also Effects of Mediation on Sleep Quality Data

R Practice 4



**Part 1**

**Overview**

This part of the assignment is about how we can use a T-test to answer the question: do male, and female cat samples have the same bodyweight?

First, we import libraries and data

 install.packages("MASS")

# Module 4

#1 Call Libraries

install.packages('FSA')

install.packages("magrittr")

install.packages("dplyr")

install.packages("plyr")

install.packages("tidyverse")

install.packages("ggplot2")

install.packages('outliers')

library(FSA)

library(magrittr)

library(dplyr)

library(plyr)

library(tidyverse)

library(outliers)

library(ggplot2)

library(psych)

Then we clean the dataset. Fortunately, this dataset has no outliers

 # first of all, we delete duplicate rows

cats <- cats[!duplicated(cats), ]

#then we delete outliers

boxplot(cats$Bwt)$out

d <- boxplot(cats$Bwt, plot=FALSE)$out

# no outlier data, so we proceed

Then we create subsets and look at them to see if they look normal or not

 males <- cats$Bwt[cats$Sex == "M"]

females <- cats$Bwt[cats$Sex == "F"]

# we see the summary and description of the subseetd

summary(males)

describe(males)

summary(females)

describe(females)

hist(males)

hist(females)

the results are good so that we can proceed with the t-test

 > summary(males)

Min. 1st Qu. Median Mean 3rd Qu. Max.

2.000 2.600 2.950 2.938 3.300 3.900

> describe(males)

vars n mean sd median trimmed mad min max range skew kurtosis se

X1 1 90 2.94 0.46 2.95 2.94 0.52 2 3.9 1.9 0.06 -0.79 0.05

> summary(females)

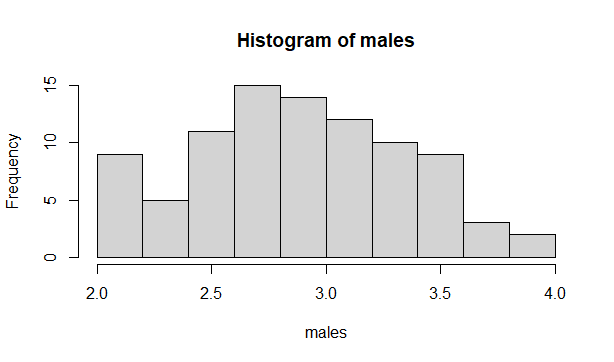
Min. 1st Qu. Median Mean 3rd Qu. Max.

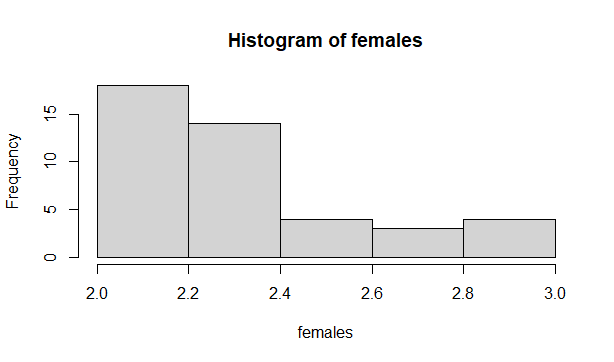
2.000 2.100 2.300 2.344 2.450 3.000

> describe(females)

vars n mean sd median trimmed mad min max range skew kurtosis se

X1 1 43 2.34 0.27 2.3 2.31 0.3 2 3 1 0.95 0 0.04





Now we run the two sample T-test, mainly because we want to prove our claim in two independent samples. Our null hypothesis is there is not differences between average of male and female cats’ bodyweight.

 > t.test(males,females)

Welch Two Sample t-test

data: males and females

t = 9.3133, df = 125.25, p-value = 5.367e-16

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

0.4674527 0.7197308

sample estimates:

mean of x mean of y

2.937778 2.344186

The cat dataset consists of 144 records, out of which 47 are females, and 97 are males. The average bodyweight of the entire sample is 2.724, whereas the average height is 10.63.

To determine if the bodyweight of male cats is the same as female cats, we subset weight the dataset into male and female cats. Running the summary of the new dataset reveals that the average weight of a male cat is 2.9, which is higher than the average weight of female cats (2.36). Female cats are relatively lighter as the weight varies from 2.0 to 3.0, whereas the weight of male cats varies from 2.0 to 3.9.

Also, the results of the t-test show that the p-value is less than α= 0.05, so we can reject the null hypothesis. It means there is a significant difference between the body weight of male and female cats.

**Part 2**

**Overview**

This assignment explores the effects of mediation on sleep quality in the second part. The data was collected from 10 students recruited for a meditation workshop. Participants wore sleeping evaluators to monitor their sleep quality. Sleeping quality is rated from 0 to 10 (the higher, the better).

Our two-sample data consists of the average sleeping quality scores before and after the workshop.

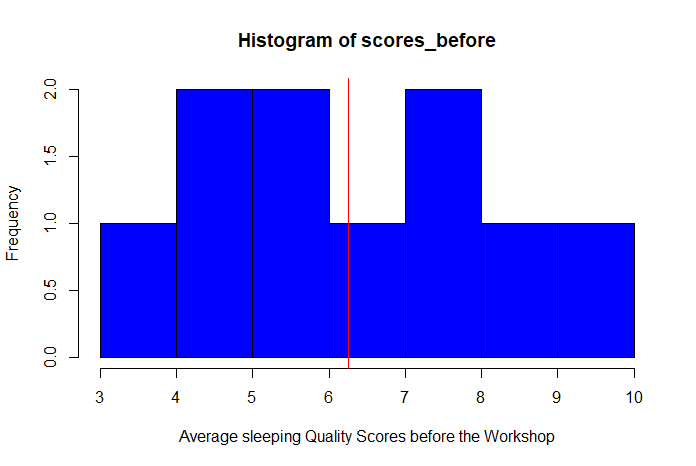
This data will address two main questions:

* Does meditation improve sleep quality?
* What is the effect of varying the significance level?

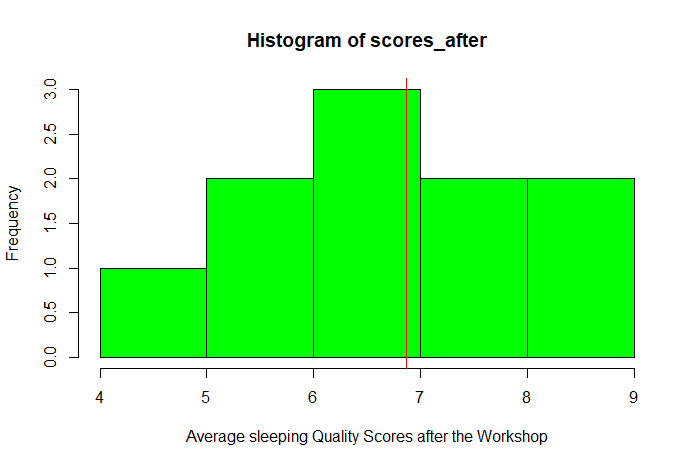
**Visualization**

Before making our hypotheses, we drew histograms and box plots to get a general overview. Using these plots, we can better define our null hypothesis. Additionally, we should check the normality of the data distribution before formatting our hypothesis.

Plot 1:

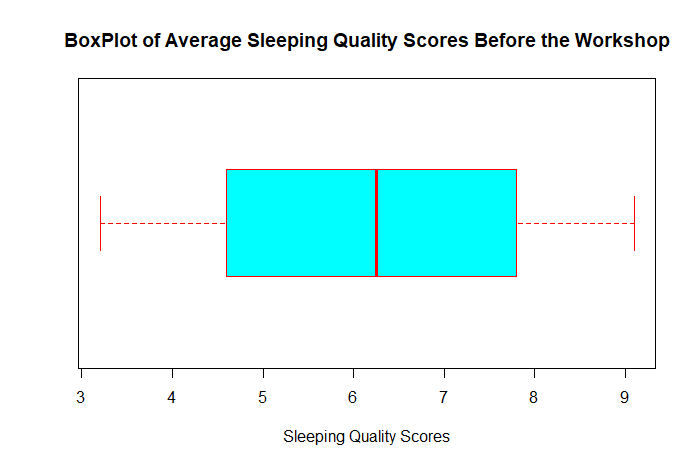


Plot 2:

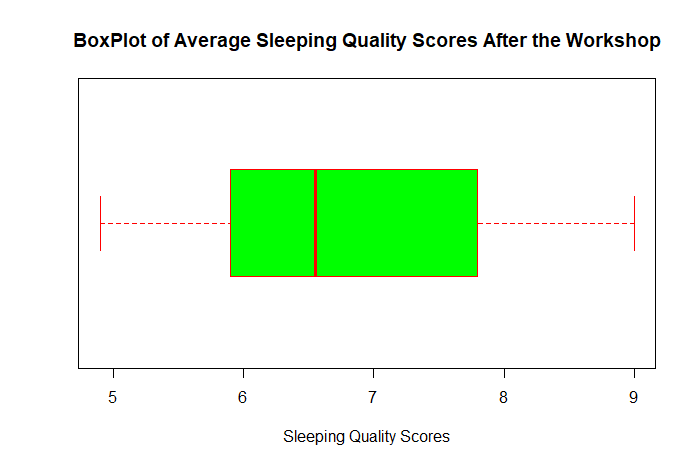


In graphs 1 and 2, the red lines indicate the mean of sleeping quality scores before and after the meditation workshop. Here, we see a slight difference in means, and a t-test should be run to see if this is significant.

Plot 3:



Plot 4:



In plots 3 and 4, it can be seen that the medians of the data for sleeping quality before and after the workshop were quite similar.

**Inferential Statistics**

We use inferential statistics to test our hypothesis after exploring the statistics of our data. Considering the paired or matched nature of subjects in our samples, we can think of them as dependent. Before and after the mediation workshop, samples were taken from the same issue: sleeping quality scores. The samples are connected. Subjects who scored highly on the pretest will usually score well on the post-test. In the same way, those with lower pretest scores will also have lower posttest scores. To control for this effect, we should use a t-test using the difference between the pretest and posttest values. This way, only changes in values are compared (Bluman, 2009).

Since the number of samples in both groups is less than 30, we had to check for normality. Figures 1 and 2 illustrate the average sleeping quality scores distribution before and after the meditation workshop. Both samples of data are considered approximately normal.

Due to the dependent nature of the samples, the t-test for dependent means was used. In this test, differences between matched pairs are calculated. We started by formulating our hypothesis, which is a two-tailed test that includes the following features:

H0=µD=0, D=X1-X2

H1= µD≠0

D is the differences in the values of the pairs of data, and μD is the symbol for the expected mean of the difference of the matched pairs.

In this test, the Null Hypothesis states no difference between the means. The Alternate Hypothesis states that there is a difference between the means of sleeping quality scores before and after the mediation workshop.

The next step is to determine the critical value. Here, the degrees of freedom are 10-1=9, and α= 0.05, so the essential values for the two tails are ±2.262.

The value of the test has been calculated by calling the "t.test" function in RStudio; it is -1.9481.

 > t.test(scores\_before,scores\_after,pair=TRUE)

Paired t-test

data: scores\_before and scores\_after

t = -1.9481, df = 9, p-value = 0.08322

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-1.33995222 0.09995222

sample estimates:

mean of the differences

-0.62

The null hypothesis **was not rejected** since the **test value was -1.9481** and in the noncritical range. Our analysis found that the zero value of the mean difference (null hypothesis) falls within the confidence intervals. Furthermore, since the **p-value is more significant than α =0.05**, we cannot reject the null hypothesis, meaning that with 95% confidence, the mean values for both groups are equal.

To assess the impact of the level of significance (α) on our hypothesis test, we changed the value of α from 0.05 to 0.1 and repeated the previous steps.

 > t.test(scores\_before,scores\_after,pair=TRUE,conf.level = 0.90)

Paired t-test

data: scores\_before and scores\_after

t = -1.9481, df = 9, p-value = 0.08322

alternative hypothesis: true difference in means is not equal to 0

90 percent confidence interval:

-1.20340497 -0.03659503

sample estimates:

mean of the differences

-0.62

Changing the significance level (α) resulted in no change in the test value or p-value, but the confidence interval changed from -1.203 to -0.036.

**Summary**

In summary, A male cat's average body weight is 2.9 kg, which is more than a female cat's typical weight (2.36). Female cats are lighter than male cats, weighing between 2.0 and 3.0 kg, while male cats weigh between 2-3.9. The t-test results indicate a substantial difference between male and female cats.

In the second part, Because the test value was -1.9481 and in the noncritical range, the null hypothesis was not rejected. We cannot reject the null hypothesis since the p-value is more significant than =0.05, implying that the mean values for both groups are approximately equivalent with 95 percent confidence. We modified the value from 0.05 to 0.1 and repeated the previous procedures to see how the significance level affected our hypothesis test. The test value and p-value did not change when the significance level was adjusted, but the confidence interval changed from -1.203 to -0.036.

**References**

Bluman, A. G. (2009). *Elementary statistics: A step by step approach*. New York; McGraw-Hill Higher Education.

Zach (2020). How to Calculate the P-Value of a T-Score in R - Statology. [online] Statology. Available at: https://www.statology.org/p-value-of-t-score-r/ [Accessed 22 Mar. 2022].

**Appendix**

 install.packages("MASS")

# Module 4

#1 Call Libraries

install.packages('FSA')

install.packages("magrittr")

install.packages("dplyr")

install.packages("plyr")

install.packages("tidyverse")

install.packages("ggplot2")

install.packages('outliers')

library(FSA)

library(magrittr)

library(dplyr)

library(plyr)

library(tidyverse)

library(outliers)

library(ggplot2)

library(psych)

### PART1

#Importing dataset

library(MASS)

cats

data(cats)

# first of all wwe delete duplicate rows

cats <- cats[!duplicated(cats), ]

#then we delete outliers

boxplot(cats$Bwt)$out

d <- boxplot(cats$Bwt, plot=FALSE)$out

# no outlier data so we proceed

summary(cats)

library(psych)

describe(cats)

# we create subsets

males <- cats$Bwt[cats$Sex == "M"]

females <- cats$Bwt[cats$Sex == "F"]

# we see the summary and description of the subseetd

summary(males)

describe(males)

summary(females)

describe(females)

hist(males)

hist(females)

#the T-test

t.test(males,females)

###part2

scores\_before=c(4.6, 7.8, 9.1, 5.6, 6.9, 8.5, 5.3, 7.1, 3.2, 4.4)

scores\_after= c(6.6, 7.7, 9.0, 6.2, 7.8, 8.3, 5.9, 6.5, 5.8, 4.9)

### Visualization

#11 Histogram plot for Female's Height

hist(scores\_before, col = "blue", xlab="Average sleeping Quality Scores before the Workshop")

m<-mean(scores\_before)

abline(v = m,col="red")

hist(scores\_after, col = "green", xlab="Average sleeping Quality Scores after the Workshop")

m<-mean(scores\_after)

abline(v = m,col="red")

#### Average sleeping Quality Scores before the Workshop Boxplot

boxplot(scores\_before, col= "cyan", border= "red",

main="BoxPlot of Average Sleeping Quality Scores Before the Workshop",

xlab="Sleeping Quality Scores", horizontal = T)

#### Average sleeping Quality Scores after the Workshop Boxplot

boxplot(scores\_after, col= "green", border= "red",

main="BoxPlot of Average Sleeping Quality Scores After the Workshop",

xlab="Sleeping Quality Scores",

horizontal = T)

t.test(scores\_before,scores\_after,pair=TRUE)

## changing alpha to 0.1

t.test(scores\_before,scores\_after,pair=TRUE,conf.level = 0.90)

**‌**