**STAT 462 – Applied Regression Analysis**

**Fall 2017, Lab 2**

Prepare a short report with relevant output, your comments, and answers to the questions (this does not need to be exhaustive or polished, but should contain enough to show that you completed all tasks and analyses).

Submit the report at the end of the lab session.

The dataset *spaghetti.txt* contains the weight (in oz) of 20 spaghetti boxes of a famous pasta brand.

* Load the dataset *spaghetti.txt* in R, using the function *read.table.*

> spaghetti=read.table("spaghetti.txt",header=TRUE)

> spaghetti

spaghetti

1 15.31335

2 15.28379

3 15.90502

4 16.75127

5 15.89350

6 14.04193

7 16.95084

8 14.99399

9 15.02057

10 16.19778

11 14.95565

12 16.61717

13 14.45850

14 16.12587

15 15.60412

16 14.60336

17 14.16071

18 16.40069

19 15.98820

20 14.53060

The company that produces pasta wants to check if the mean box weight is the nominal one (16 oz) or if they need to inspect the packing machine. Assume the box weight is normally distributed.

|  |  |
| --- | --- |
| |  | | --- | |  | |

* Compute the sample mean and the standard deviation of the box weight.

> spaghetti=as.matrix(spaghetti)

> mean(spaghetti)

[1] 15.48985

> sd(spaghetti)

[1] 0.8821747

**Thus, the mean is 15.48985 oz, standard deviation is 0.8821747 oz.**

* Perform a suitable test to verify if the mean of the box weight is equal to the nominal one (16 oz) or not.
* What is your null hypothesis H0?

**H0: the mean of the box weight is equal to 16 oz.**

* What is your alternative hypothesis H1?

**H1: the mean of the box weight is not equal to 16 oz.**

* What distribution can you assume for the test statistic (type of distribution and parameters), and why?

**I will assume the distribution of the test statistic is t-distribution with 19 degree of freedom, because we assume the box weight is normally distributed and the sample size is smaller than 30, which is not large enough to assume normal distribution for the test statistic.**

* Compute the test statistic.

> t=(mean(spaghetti)-16)/(sd(spaghetti)/sqrt(20))

> t

[1] -2.586199

**Test statistic = -2.586199**

* Compute the test p-value.

> 2\*pt(-abs(t),19)

[1] 0.01811035

**Thus, p-value = 0.01811035**

* Do you reject the null hypothesis, at significance level 5%?

**Since 0.01811035 < 0.05, we reject H0.**

**Thus, we reject H0 at significance level of 5%.**

R code:

setwd("//udrive.win.psu.edu/Users/j/q/jql5883/Desktop/math462")

getwd()

spaghetti=read.table("spaghetti.txt",header=TRUE)

spaghetti

spaghetti=as.matrix(spaghetti)

mean(spaghetti)

sd(spaghetti)

t=(mean(spaghetti)-16)/(sd(spaghetti)/sqrt(20))

t

2\*pt(-abs(t),19)

After run:

> setwd("//udrive.win.psu.edu/Users/j/q/jql5883/Desktop/math462")

> getwd()

[1] "\\\\udrive.win.psu.edu/Users/j/q/jql5883/Desktop/math462"

>

> spaghetti=read.table("spaghetti.txt",header=TRUE)

> spaghetti

spaghetti

1 15.31335

2 15.28379

3 15.90502

4 16.75127

5 15.89350

6 14.04193

7 16.95084

8 14.99399

9 15.02057

10 16.19778

11 14.95565

12 16.61717

13 14.45850

14 16.12587

15 15.60412

16 14.60336

17 14.16071

18 16.40069

19 15.98820

20 14.53060

>

> spaghetti=as.matrix(spaghetti)

> mean(spaghetti)

[1] 15.48985

> sd(spaghetti)

[1] 0.8821747

>

> t=(mean(spaghetti)-16)/(sd(spaghetti)/sqrt(20))

> t

[1] -2.586199

The following part will NOT be considered in grading the lab report.

* Compute the 99% two-sided confidence interval for the mean of the box weight.
* Do you reject the null hypothesis H0: mu=16 oz in favor of H1: mu!=16 oz at significance level 1%?

A consumers’ association would like to sue the company, affirming that the mean box weight is lower than the nominal one (16 oz). To be sure about their statement, they ask you to perform a suitable test with level 1%.

* Perform the test for the consumers’ association.
* What is your null hypothesis H0?
* What is your alternative hypothesis H1?
* What distribution can you assume for the test statistic (type of distribution and parameters), and why?
* Compute the test statistic.
* Compute the test p-value.
* Do you reject the null hypothesis, at significance level 1%?

The following part will NOT be considered in grading the lab report.

Consider again the dataset *record.txt* of Lab1, that contains running records obtained from athletes from different countries in various types of athletics events (sprints and middle-distance).

* Load the dataset *record.txt* in R, using the function *read.table.*
* Draw a scatterplot for the variable x=m400 and the variable y=m800.
* Compute the correlation between m400 and m800.

|  |  |
| --- | --- |
| |  | | --- | |  | |