ITMD/ITMS/STAT 514 Homework 0

Guruteja Kanderi

Due Date 9/19

Part I. Changing the author field and file name. (10 points)

(a) Change the `author:` field on the Rmd document

(b) Rename this file to "HW0_YourFirstInitialYourLastName.Rmd", (e.g. HW0_YDing.Rmd).

Part II. Course Related Questions (60 points)

Question 1. What is the instructor's name? Please write about her office and office hours this semester.

> Your Answer: Yuhan Ding. Tuesday and Thursday 1 to 2 pm or by appointment at Retaliata Engineering 116b

Question 2. Please list the acts of academic dishonesty in the syllabus.

> Your Answer: Stealing credit: Taking another's ideas as your own. Shortcut to success: Using banned aids to boost exam or assignment performance. Fakery in research: Inventing or tampering with data or information. Teaming up, against the rules: Unpermitted collaboration on coursework. Double dipping: Submitting the same work for multiple courses without approval. Helping hand, not allowed: Getting unauthorized assistance on academic tasks. Paying for a pass: Hiring someone to do your work for grades.

Question 3. What's the policy for late homework submission? When can you request an extension?

> Your Answer: No late homework will be accepted unless you request an extension before the deadline. You can only request a two-day extension, and a 10% penalty will be applied for submissions after the deadline but within the extended period. Late submissions will be disregarded if they are received less than two days after the original deadline.

Question 4. How many exams will we have? Will we have a final exam? What is the evaluation of the final project?

> Your Answer: 1. We will only have a single midterm exam, and there won't be a final exam. 2. The final project is a group assignment for teams with three members or fewer. Each group is responsible for conducting its own analysis on a given dataset. The analysis should encompass at least one hypothesis test and one regression model, culminating in a 15-minute oral presentation (followed by a 5-minute Q&A). The final project entails exploring a dataset using the tools acquired in class and should consist of three main parts: Data Preparation: Provide details about the dataset, including the number of observations, attributes, data types, missing values, etc. Data Exploration (EDA): Utilize data visualization techniques, such as histograms, boxplots, and overlay histograms, along with a conclusion. Data Analysis Hypothesis Testing: Formulate null and alternative hypotheses, apply the appropriate test, and draw conclusions. Build a linear

regression model with subset selection. Identify all significant attributes, evaluate the model, and make predictions on a test dataset. The presentation, whether in slides or HTML format, should adhere to the 15-minute timeframe, including a 5-minute session for questions and answers. The final project grade is determined by multiplying your team grade (based on scoring rubrics) by your personal points (from the Peer evaluation form) and dividing the result by the highest team member points (from the Peer evaluation form).

Question 5. What is the assessment of this class? What is the grade range of A?

Your Answer: Homework-20% Midterm-40% Final Project-40%. The grade range for A is 85% and above

Question 6. Is there a second chance to resubmit your assignment or exam after grading? Are there any extra assignments or work to improve your grade?

Your Answer: No Second chance to resubmit the assignments or exam once it is graded. No Extra assignemnets and work to improve the grades

I fully understand the rules in the syllabus and will follow them.

Your Name Here As Signature: Guruteja Kanderi

Part III. Basic R Syntax (30 points)

Question 1. Hello World! (5 points) Here's an R code chunk that prints the text 'Hello world!'.

```r
print("Hello world!")
```

```
## [1] "Hello world!"
```

```r
print("Guruteja Kanderi!")
```

Modify the code chunk below to print your name

```
## [1] "Guruteja Kanderi!"
```

Question 2. Play with Dataset (25 points) The following is a small data to store health information, where the unit for height is centimeters and weight is kilogram.

```r
BMI <- data.frame(
    gender = c("Male", "Male","Female"),
    height = c(152, 171.5, 165),
    weight = c(81,93,78),
    Age = c(42,38,26)
)
print(BMI)
```

```
##   gender height weight Age
## 1   Male  152.0     81  42
## 2   Male  171.5     93  38
## 3 Female  165.0     78  26
```

a) Please calculate BMI for each person and add that information to the dataframe and display the new dataframe. (10 points) BMI Formula: Weight (kg)/Square of Height(meters)

```
BMI$BMI <- BMI$weight / (BMI$height / 100)^2
print(BMI)

##   gender height weight Age      BMI
## 1   Male  152.0     81  42 35.05886
## 2   Male  171.5     93  38 31.61948
## 3 Female  165.0     78  26 28.65014
```

```
BMI[BMI$BMI > 30, c("gender", "Age")]
```

b) Please display gender and age for the people with BMI great than 30. (10 points)

```
##   gender Age
## 1   Male  42
## 2   Male  38
```

```
BMI[BMI$BMI > 25 & BMI$Age < 30, "gender"]
```

c) Please display gender for the people with BMI great than 25 and age less than 30. (5 points)

```
## [1] "Female"
```