MATH 254 – Statistical Modeling and Applications - HW1

Tural Sadigov

From Textbook 2

Introductory Statistics with Randomization and Simulation 1st Edition, by D. Diez, C.D. Barr, M. Cetinkaya-Rundel,

https://www.openintro.org/book/isrs/

PROBLEMS

Note that the label (R Studio) means that you will need to use R Studio for those problems. But even if there is no such label, you are encouraged to use it whenever you can. Use R files in Blackboard as your reference.

First, load the libraries you will need.

```
library(tidyverse)
```

1. (R Studio) Read Problem 1.5 on page 45 and type the following into R Studio console to print and view the top of the dataset iris.

```
iris %>%
  head()
```

```
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1         3.5          1.4         0.2  setosa
2          4.9         3.0          1.4         0.2  setosa
3          4.7         3.2          1.3         0.2  setosa
4          4.6         3.1          1.5         0.2  setosa
```

| 5 | 5.0 | 3.6 | 1.4 | 0.2 | setosa |
| 6 | 5.4 | 3.9 | 1.7 | 0.4 | setosa |

Choose one of the variables, Petal.Length, from iris data, and create the dotplot, the boxplot and the histogram of it.

- What does the shape look like? Is it symmetric, is it skewed?

- Is it unimodal or bimodal?

- Are there any outliers?

- Is the middle 50% of the data symmetric? Why or why not?

- Out of three plots you have, which one gives you a better understanding of the variable? Why?

2. Problem 1.6 – page 45 (R Studio)

The dataset is at the following link:https://www.openintro.org/data/index.php?data=smoking

After downloading R Data file (i.e., file with .rda), you can load it back to R Studio and view using the following codes. Make sure that the downloaded file is in your working directory.

```
load('smoking.rda')
smoking %>%
  head()
```

```
# A tibble: 6 x 12
  gender    age marital_st~1 highe~2 natio~3 ethni~4 gross~5 region smoke amt_w~6
  <fct>  <int> <fct>        <fct>   <fct>   <fct>   <fct>   <fct>  <fct>   <int>
1 Male      38 Divorced     No Qua~ British White   2,600 ~ The N~ No         NA
2 Female    42 Single       No Qua~ British White   Under ~ The N~ Yes        12
3 Male      40 Married      Degree  English White   28,600~ The N~ No         NA
4 Female    40 Married      Degree  English White   10,400~ The N~ No         NA
5 Female    39 Married      GCSE/O~ British White   2,600 ~ The N~ No         NA
6 Female    37 Married      GCSE/O~ British White   15,600~ The N~ No         NA
# ... with 2 more variables: amt_weekdays <int>, type <fct>, and abbreviated
#   variable names 1: marital_status, 2: highest_qualification, 3: nationality,
#   4: ethnicity, 5: gross_income, 6: amt_weekends
# i Use `colnames()` to see all variable names
```

3. Problem 1.10 – page 46
4. Problem 1.12 – page 47

5. Problem 1.14 – page 47
6. Problem 1.24 – page 50; Explain your plot, and come up with/draw a data cloud around your plot that could reflect a real dataset.
7. Problem 1.26 – page 51
8. Problem 1.34 – page 53 (R Studio) – Create dotplots then comment on comparisons.
9. Problem 1.36 – page 53
10. Problem 1.40 – page 55