TMATH 410 On-Computer Assignment 2

Lectures 2-4, Chapter 2

For each assignment there will be questions from your textbook and from your instructor to be completed without the help of R (called "On paper", numbered P1-…), and a section to be completed using R (called "On computer", labeled C1-…). Note, you are allowed to submit your "On paper" section using a computer program such as MS Word, just not using R. For any questions that ask for an interpretation or explanation, you are to provide your answer in full sentences. Any R commands given in the assignment should be typed into R, not copied from the text.

Data

For this assignment we will investigate data collected by a student to try to infer a person's stature from their footprints, for forensic purposes. The goal would be to identify the stature of an unseen suspect from evidence (such as a foot print) left at a crime scene. (Rohren, B. 2006. Estimation of Stature from Foot and Shoe Length: Applications in Forensic Science, obtained from Triola Elementary Statistics.)

The variables in the data set are the biological sex of the individual (M, F), their foot length (cm), length of shoe from shoe print (cm), reported shoe size, and individual height (cm).

On computer

First we will conduct and exploratory data analysis (EDA) of the data, then we will estimate a linear model that predicts a person's height based on their footprint. Download the dataset FOOT.csv from the Canvas page into the directory on your computer you have created for this class. Open a new R session, change the working directory to your class directory, and read the data into R. Call your data frame object foot.df.

```
foot.df=read.csv(file.choose())
```

C1. (4) Submit the R script you used for this assignment.

C2. (3) Use the head(object) command to view the first 6 lines of the data frame, and the dim(object) to determine the number of observations in this dataset. **Copy and paste the results of both** into your assignment. Explicitly identify how many observations were made.

C3. (2) We will investigate whether a linear relationship is reasonable between shoe print length and height. For forensic purposes, **which is the response and which is the predictor variable? Explain why.**

C4. (2) **Produce a publication-quality scatter plot** with shoe print length (Shoe.Print) on the x-axis and height (Height) on the y-axis. Make sure to label your axes!

C5. (3)**Comment on whether you think a linear model is reasonable given your graph in C4.** Explain why or why not

C6. (2) Use the lm function to fit a linear model between shoe print length (Shoe.Print) and height (Height) and use the summary function to produce a summary of that model. **Copy and paste the R code and output of the summary function** to your assignment document. Make sure you correctly identify the explanatory variable and the response variable!

C7. (4) **Identify the values of the coefficient estimates, their standard errors, null hypotheses, p-values, and the conclusions of hypothesis tests for those coefficients.** Organize your results in a publication-quality table with rows for each coefficient and a column for the estimate, a column for the standard error, a column for each null hypothesis, a column for the p-value, and a column for the hypothesis test conclusion (reject or fail to reject). See the table below for an example (replace the NA values with your answers)

Coefficient	Estimate	StdErr	NullHypothesis	pVal	Conclusion
Intercept	NA	NA	NA	NA	NA
Shoe Print	NA	NA	NA	NA	NA

C8. (3) **Calculate and report a 95% confidence interval for the intercept** Include units!

C9. (3) **Calculate and report a 95% confidence interval for the slope.** Include units!

C10. (2) **Give an interpretation of the intercept, in the context of the original problem.**

C11. (2) **Give an interpretation of the slope, in the context of the original problem.**

Note: This regression model is not yet complete! We still have to assess if we meet the regression assumptions and we need to perform model diagnostics (Chapter 4 in textbook)!