STEVE G MWANGI
R LAB 2
01/20/2020

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).

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

submitted

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.

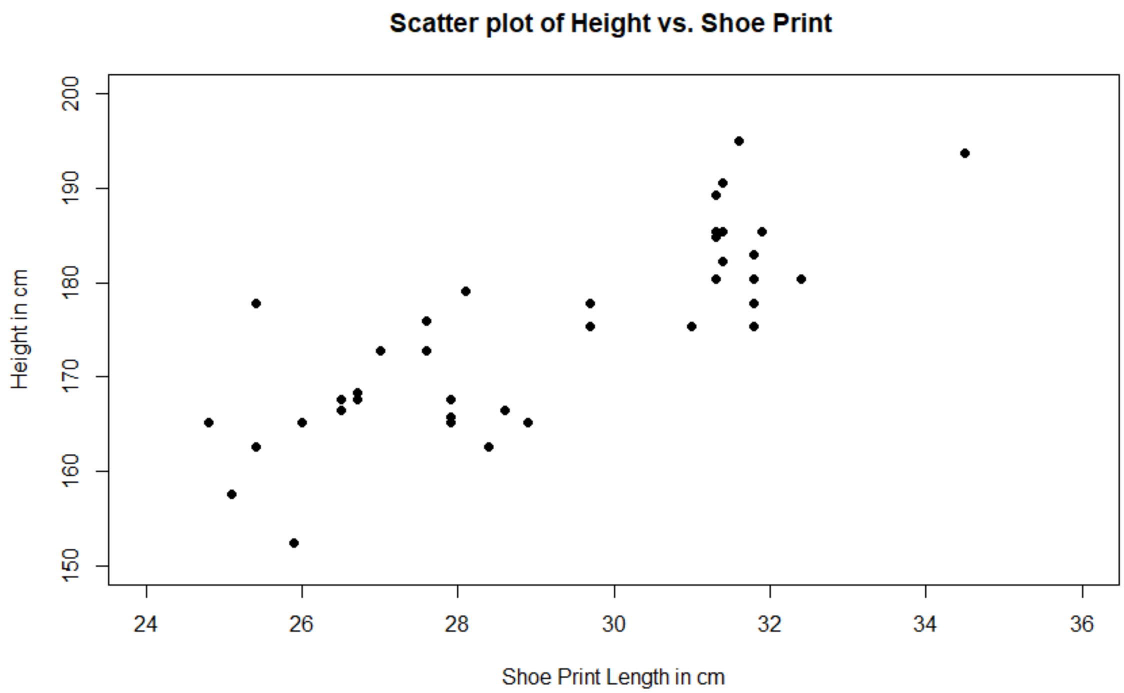
```
> #C2
> # Viewing the first 6 lines of the data frame
> head(foot.df)
  Sex Age Foot.Length Shoe.Print Shoe.Size Height
1   M  67        27.8       31.3        11  180.3
2   M  47        25.7       29.7         9  175.3
3   M  41        26.7       31.3        11  184.8
4   M  42        25.9       31.8        10  177.8
5   M  48        26.4       31.4        10  182.3
6   M  34        29.2       31.9        13  185.4
> # Determine the number of observations in the data set
> dim(foot.df)
[1] 40  6
```

40 sets of observations with 6 different data points for Sex, Age, Foot Length, Shoe Print, Shoe Size and Height of individual

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.

Predictor: Shoe Print – Since in forensics, you can be able to find the shoe print at the crime scene rather than determine/find evidence for the height of the associated individual.
Response: Height

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.

> **Linear model is reasonable in this case, since the points show a positive correlation between the predictor and response variables. Now apparent outliers that skew the points.**

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!

```
> #C6
> # Creating a linear model of Height ~ Shoe Print from the data in the data
frame
> lm(Height~Shoe.Print, data = foot.df)

Call:
lm(formula = Height ~ Shoe.Print, data = foot.df)

Coefficients:
(Intercept)   Shoe.Print
     80.930        3.219

> model1 = lm(Height~Shoe.Print, data = foot.df)
> # Getting  the summary of the model
```

```
> summary(model1)

Call:
lm(formula = Height ~ Shoe.Print, data = foot.df)

Residuals:
     Min       1Q   Median       3Q      Max 
-11.8911  -4.9409   0.3969   3.1982  15.1181 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   80.930     10.893   7.429 6.50e-09 ***
Shoe.Print     3.219      0.374   8.606 1.86e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.944 on 38 degrees of freedom
Multiple R-squared:  0.6609,	Adjusted R-squared:  0.652 
F-statistic: 74.06 on 1 and 38 DF,  p-value: 1.863e-10
```

C7. (4) Identify the values of the coefficient estimates, their standard errors, null hypotheses, pvalues, 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).

Coefficient	Estimate	Std Err	Null Hypothesis	p-value	Conclusion
Intercept	80.930	10.893	7.429	6.50e-09	Reject H_0
Shoe Print	3.219	0.374	8.606	1.86e-10	Reject H_0

Table 1 Identifying values from linear model in C6

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

$$\hat{\beta}_0 - t_{\left(n-2,\frac{\alpha}{2}\right)} * \hat{\sigma}_{\hat{\beta}_0} \leq \beta_0 \leq \hat{\beta}_0 + t_{\left(n-2,\frac{\alpha}{2}\right)} * \hat{\sigma}_{\hat{\beta}_0}$$

$$\mathbf{80.930\boldsymbol{cm} - 2.0244 * 10.893\boldsymbol{cm} \leq \boldsymbol{\beta_0} \leq 80.930\boldsymbol{cm} + 2.0244 * 10.893\boldsymbol{cm}}$$

$$\mathbf{58.8782\boldsymbol{cm} \leq \boldsymbol{\beta_0} \leq 102.9818cm}$$

Intercept 95% CI: $(\mathbf{58.8782\boldsymbol{cm}, 102.9818\boldsymbol{cm}})$

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

$$\hat{\beta}_1 - t_{\left(n-2,\frac{\alpha}{2}\right)} * \hat{\sigma}_{\hat{\beta}_1} \leq \beta_1 \leq \hat{\beta}_1 + t_{\left(n-2,\frac{\alpha}{2}\right)} * \hat{\sigma}_{\hat{\beta}_1}$$

$$\mathbf{3.219\boldsymbol{cm} - 2.0244 * 0.374\boldsymbol{cm} \leq \boldsymbol{\beta_1} \leq 3.219\boldsymbol{cm} + 2.0244 * 0.374\boldsymbol{cm}}$$

$$\mathbf{2.4619\boldsymbol{cm} \leq \boldsymbol{\beta_1} \leq 3.9761cm}$$

Slope 95% CI: $(\mathbf{2.4619\boldsymbol{cm}, 3.9761\boldsymbol{cm}})$

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

> **Even though the expected value of height of an individual when there is no shoe print ($X = 0$) can be said to be 80.930cm, in the context of forensic science, that value does not tell us anything about the suspect's height. It's insignificant in context.**

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

> **For every 1cm change in the shoe print size, there is a corresponding 3.219cm change in height of the suspect.**