Labs 9 & 10 V2: Linear Regression

Yasmine Pascal

03/30/2022

## Instructions

This document is the template you will use to type up your responses to the lab exercises that you can find on Canvas.

To complete the lab type your BRIEF answers and the R code in the spaces provided below, according to these guidelines:

* In general, if an exercise is numbered you will submit the R code that you used to answer the question in the gray area between the three ` marks.
* If you need to submit an answer in addition to the code type this between the line that starts #### and the beginning of the gray area for the code.
* Do not ever type the View() command in a lab report. This works only in RStudio on your computer, so don’t include that command in your lab report or you will not be able to compile the document, a process called *knitting*.
* To produce a document that you can submit on Canvas just click on the Knit button in the upper left quadrant of RStudio.
* We recommend pressing the Knit button after entering each answer: it’s very common to get errors, and they’re easier to fix if you can easily isolate the code that caused the error.
* The Knit button will create two files: one is a .Rmd file (R Markdown) and the other is an .docx file. You will need to submit both these documents for the lab assignment on Canvas.

## Preliminaries

#### Exercise 1

library(tidyverse)  
library(mosaic)  
library(ggformula)  
library(data1135)

## Transform the data

#### Exercise 2

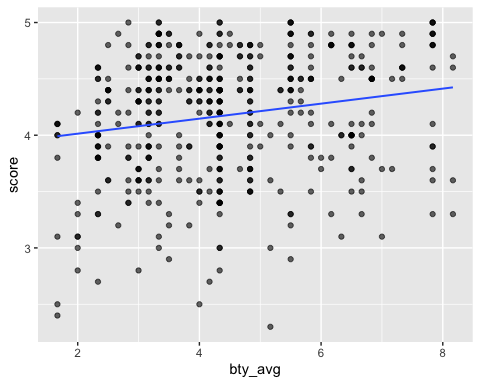
evals <- mutate(evals, bty\_avg = (bty\_f1lower +bty\_f1upper + bty\_f2upper +bty\_m1lower + bty\_m1upper +bty\_m2upper)/6)

## Simple linear regression

#### Exercise 3

According to the scatter plot, it does not seem that better looking teachers are evaluated more favorably. There seems to be a weak positive linear relationship between these two variables.

gf\_point(score~bty\_avg, alpha= .6, data= evals) %>% gf\_lm()



#### Exercise 4

m\_bty <- lm(score ~ bty\_avg, data = evals)  
msummary(m\_bty)

## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.88033 0.07614 50.96 < 2e-16 \*\*\*  
## bty\_avg 0.06664 0.01629 4.09 5.08e-05 \*\*\*  
##   
## Residual standard error: 0.5348 on 461 degrees of freedom  
## Multiple R-squared: 0.03502, Adjusted R-squared: 0.03293   
## F-statistic: 16.73 on 1 and 461 DF, p-value: 5.082e-05

#### Exercise 5

Linearity: Even tough these two variables have weak a correlation, their relationship still seems to be reasonably linear

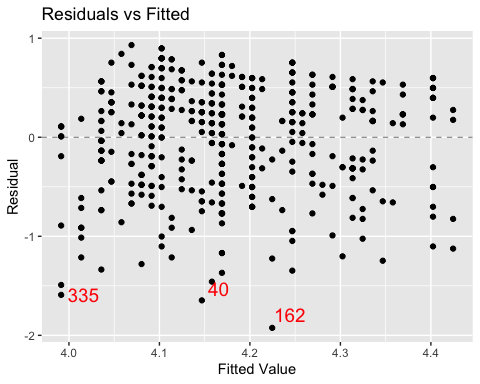
Independence: We can assume that the evaluations are independent, because it is unlikely that one student’s evaluation affected another student’s.

Equal variance: This is satisfied because there is equal scatter of points from top to bottom and there are no patterns or funnels.

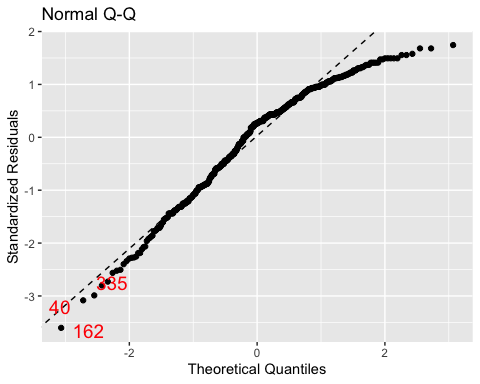
Normality of residuals: We assume that the residuals are normal because the sample size is greater than 30 ( Central Limit Theorem)

mplot(m\_bty, which = 1:2, add.smooth= FALSE)

## [[1]]



##   
## [[2]]



#### Exercise 6

Yes, average beauty score is a statistically significant predictor. The p-value is 5.08e-05.For each additional increase of average beauty score, the teacher’s score increases on average by .06664 points

#### Exercise 7

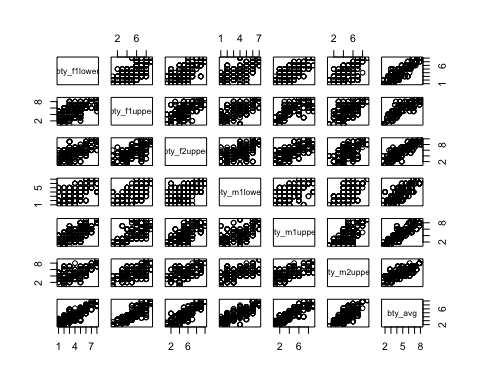
Yes, it appears to be a practically significant predictor.

## Correlated x-variables

#### Exercise 8

The correlation between bty\_avg and bty\_m1upper is 0.86.

pairs(select(evals, contains("bty")))



cor(select(evals, contains("bty"))) %>% round(digits = 2)

## bty\_f1lower bty\_f1upper bty\_f2upper bty\_m1lower bty\_m1upper  
## bty\_f1lower 1.00 0.62 0.64 0.61 0.68  
## bty\_f1upper 0.62 1.00 0.65 0.60 0.69  
## bty\_f2upper 0.64 0.65 1.00 0.54 0.65  
## bty\_m1lower 0.61 0.60 0.54 1.00 0.60  
## bty\_m1upper 0.68 0.69 0.65 0.60 1.00  
## bty\_m2upper 0.59 0.57 0.51 0.54 0.59  
## bty\_avg 0.84 0.84 0.82 0.78 0.86  
## bty\_m2upper bty\_avg  
## bty\_f1lower 0.59 0.84  
## bty\_f1upper 0.57 0.84  
## bty\_f2upper 0.51 0.82  
## bty\_m1lower 0.54 0.78  
## bty\_m1upper 0.59 0.86  
## bty\_m2upper 1.00 0.76  
## bty\_avg 0.76 1.00

## Multiple Linear Regression

#### Exercise 9

m\_bty2 <- lm(score ~ bty\_avg + gender, data = evals)  
msummary(m\_bty2)

## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.74733 0.08466 44.266 < 2e-16 \*\*\*  
## bty\_avg 0.07415 0.01625 4.563 6.48e-06 \*\*\*  
## gendermale 0.17239 0.05022 3.433 0.000652 \*\*\*  
##   
## Residual standard error: 0.5287 on 460 degrees of freedom  
## Multiple R-squared: 0.05912, Adjusted R-squared: 0.05503   
## F-statistic: 14.45 on 2 and 460 DF, p-value: 8.175e-07

#### Exercise 10

Linearity:There is a reasonable linear relationship among these variables.

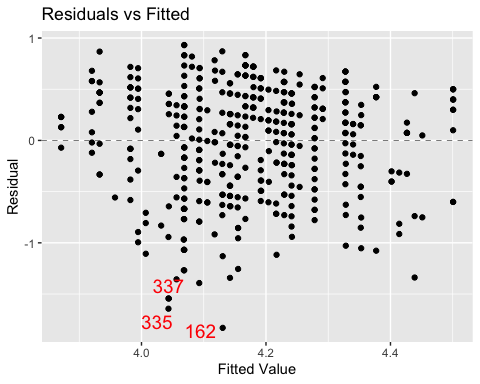
Independence:We can assume that the evaluations are independent, because it is unlikely that one student’s evaluation affected another student’s.

Equal variance:This is satisfied because there is almost/ relative equal scatter of points from top to bottom and there are no patterns or funnels.

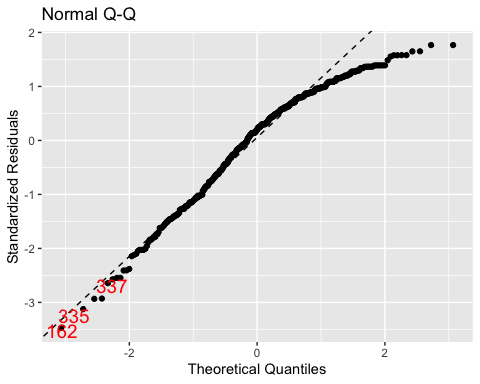
Normality of residuals: We assume that the residuals are normal because the sample size is greater than 30 ( Central Limit Theorem)

mplot( m\_bty2, which = 1:2, add.smooth = FALSE)

## [[1]]



##   
## [[2]]



#### Exercise 11

Yes bty\_avg is still a statistically significant predictor of score. The p-value is 6.48e-06.For each additional increase of beauty score average, the teacher’s score increases on average by 0.74 points, holding all other variables constant

#### Exercise 12

Yes gender is a statistically significant predictor of score. The p-value is 0.000652.The teacher’s score increases on average by 0.172 points, holding all other variables constant, when gender is male to when gender is female.

## Make your own model…

#### Exercise 13

m\_bty3 <- lm(score ~ bty\_avg + gender+ language, data = evals)  
msummary(m\_bty3)

## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.76023 0.08436 44.576 < 2e-16 \*\*\*  
## bty\_avg 0.07472 0.01616 4.623 4.93e-06 \*\*\*  
## gendermale 0.17216 0.04995 3.447 0.000619 \*\*\*  
## languagenon-english -0.25244 0.10253 -2.462 0.014175 \*   
##   
## Residual standard error: 0.5258 on 459 degrees of freedom  
## Multiple R-squared: 0.07139, Adjusted R-squared: 0.06532   
## F-statistic: 11.76 on 3 and 459 DF, p-value: 1.95e-07

#### Exercise 14

I’m investigating how beauty average, gender and language (independent variables), affect a teacher’s score ( y-variable/ dependent variable)

#### Exercise 15

All the assumptions for a linear regression are met.

#### Exercise 16

The p-value of the F statistic is 1.95e-07

#### Exercise 17

Score = 3.76 + (0.075 \* bty\_avg) + (0.172 \* gendermale) + (-0.252 \* languagenon-english)

#### Exercise 18

R squared = 0.07139

#### Exercise 19

Beauty Average Score, Gender and Language are significant. Their p-values are 4.93e-06, 0.000619, 0.014175 accordingly.

#### Exercise 20