week10-pdfgen

Jin Chen

8/3/2020

Data Import and Cleaning

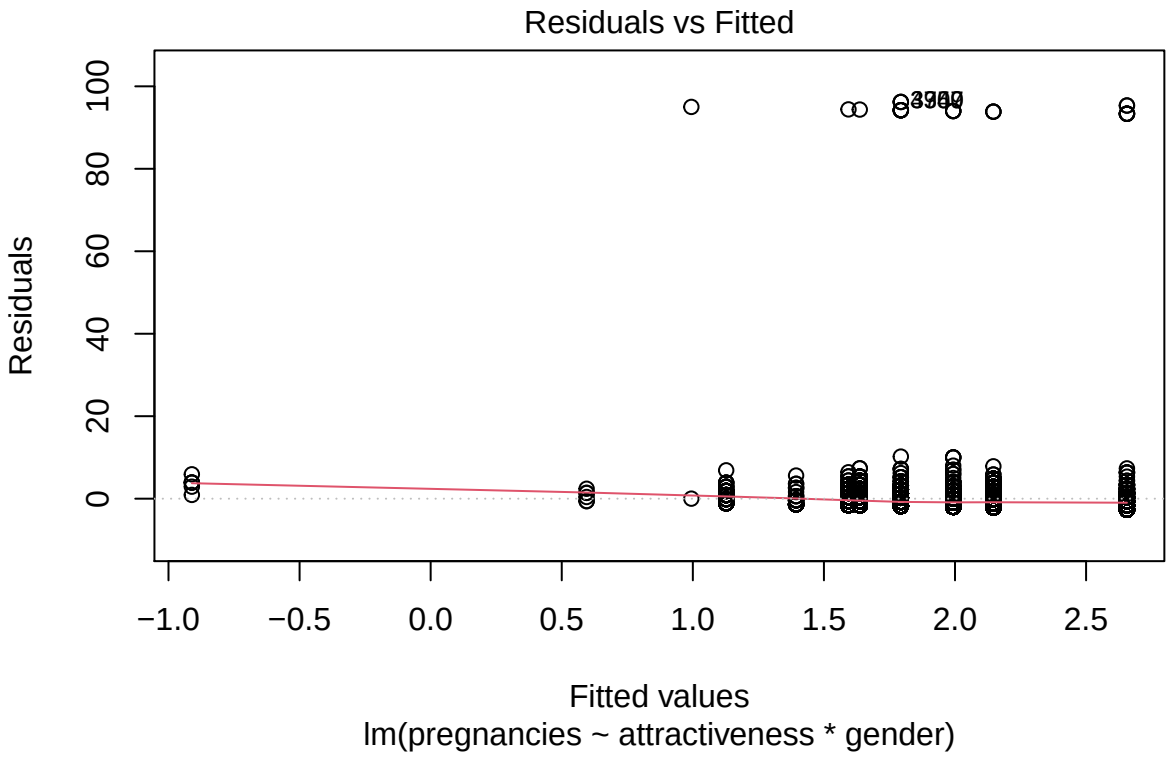
In this block I imported the National Ongitudinal Study of Adolescent to Adult Health Wave IV, 2008 dataset from the UNC Dataverse website. I selected three variables and convereted gender to a factor variable.

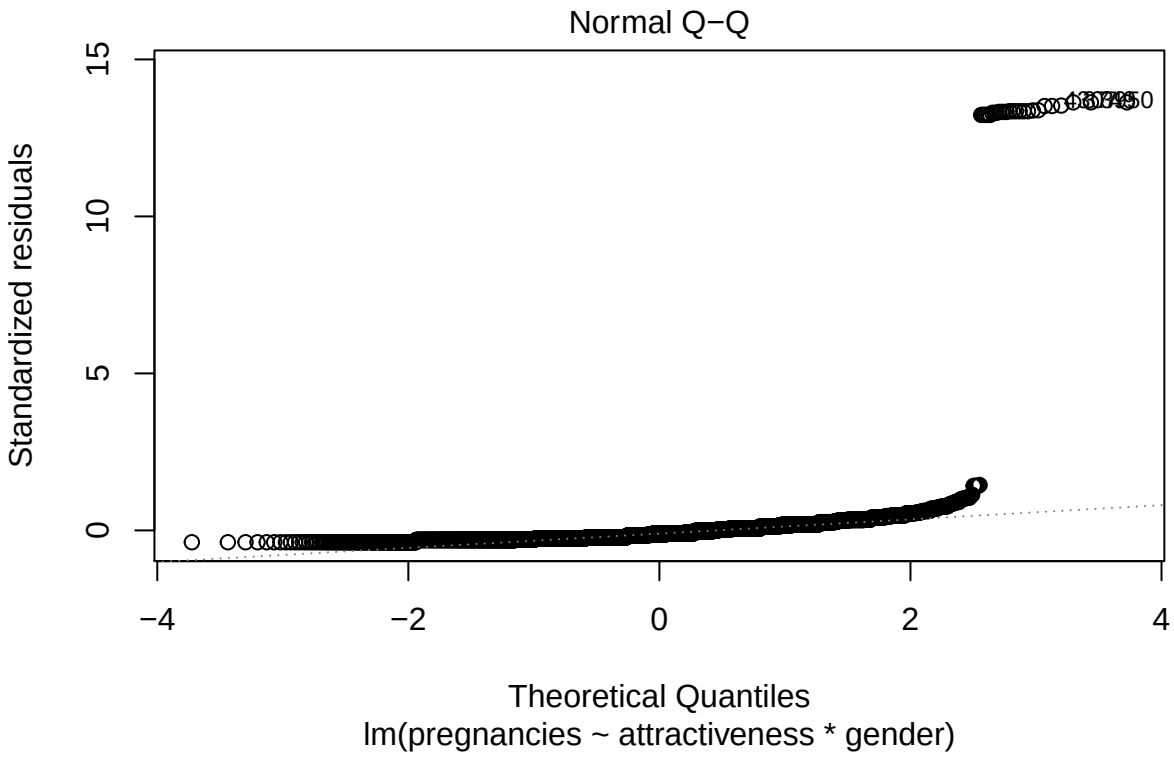
```
health_tbl <- read_tsv("../data/w4inhome_dvn.tab") %>%
  select(attractiveness = H4MH8,
         pregnancies = H4TR9,
         gender = BIO_SEX4) %>%
  mutate(gender = factor(gender, levels = c(1,2), labels = c("Male", "Female")))
```

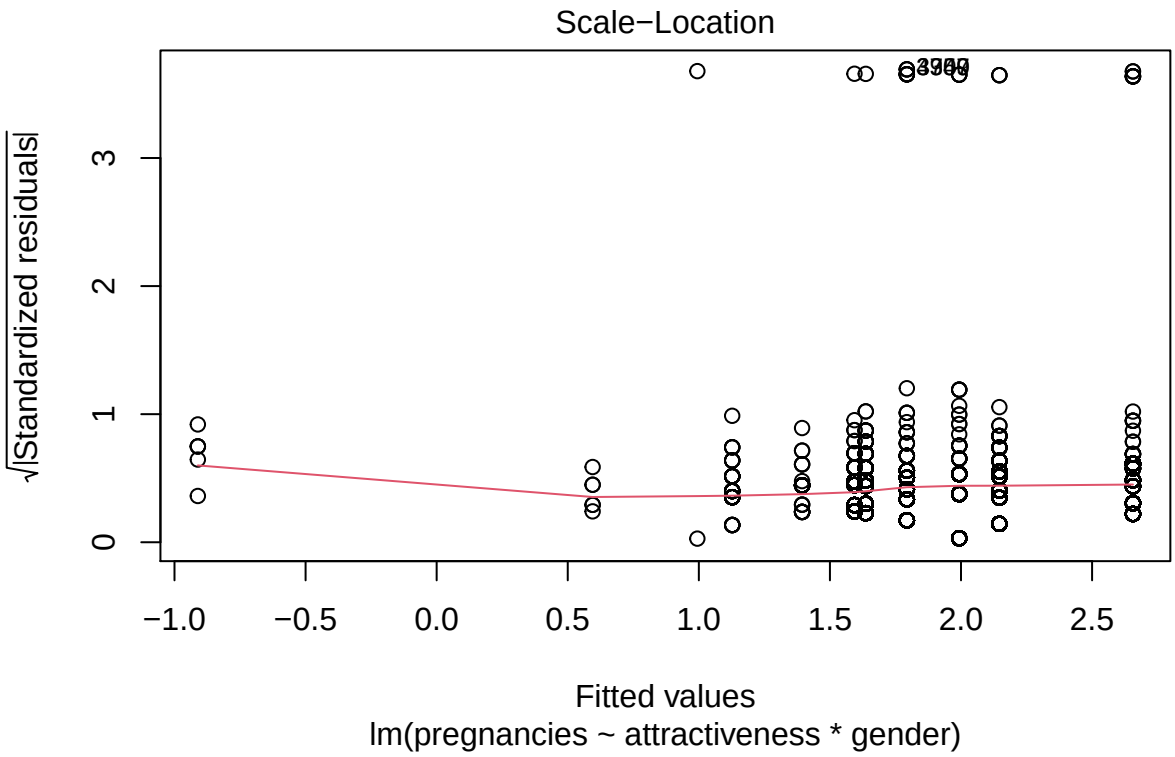
Analysis

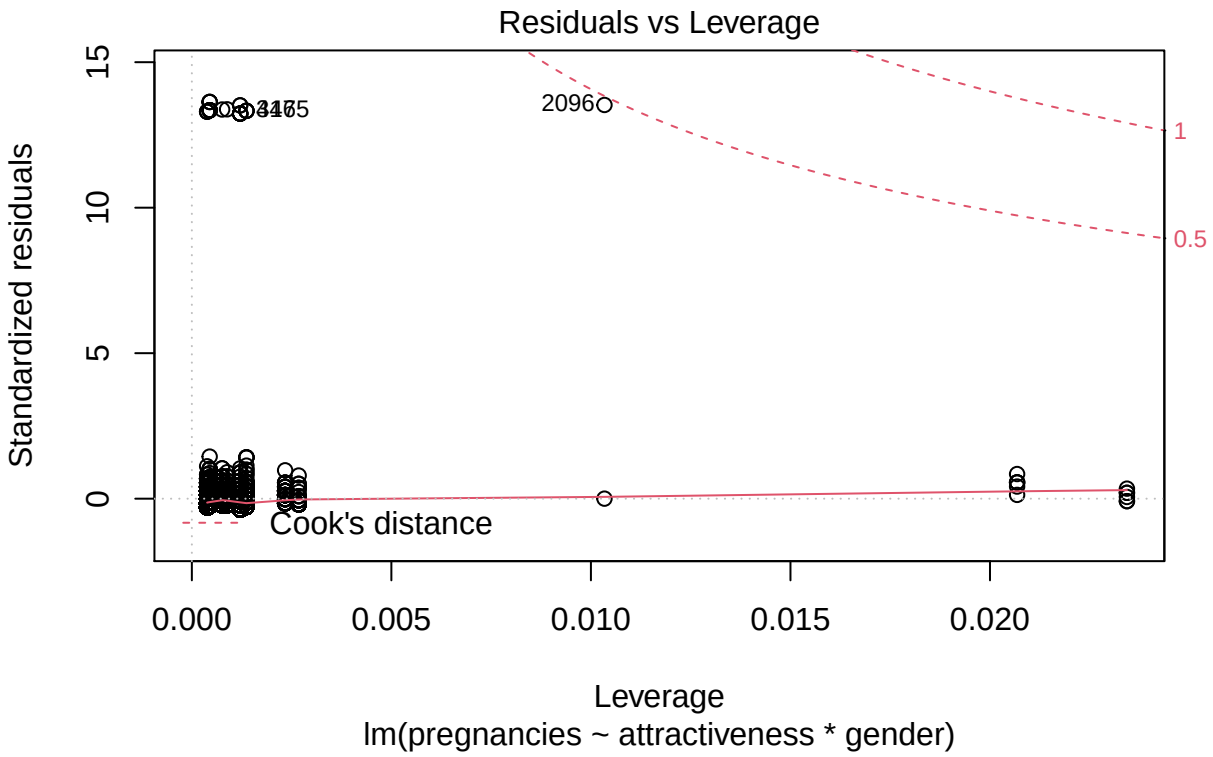
In this block I conducted a multiple linear regression of pregnancies on the two independent variables attractiveness and gender, checked assumptions and exmained a coefficients table.

```
model <- lm(pregnancies ~ attractiveness * gender, data = health_tbl) #The specification first*second i
plot(model)
```


Residuals vs Fitted
Residuals
0
20
40
60
80
100
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
Fitted values
lm(pregnancies ~ attractiveness * gender)


Normal Q−Q
Standardized residuals
0
5
10
15
−4
−2
0
2
4
Theoretical Quantiles
lm(pregnancies ~ attractiveness * gender)


Scale−Location
√|Standardized residuals|
0
1
2
3
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
Fitted values
lm(pregnancies ~ attractiveness * gender)




```
summary(model)
```

```
##
## Call:
## lm(formula = pregnancies ~ attractiveness * gender, data = health_tbl)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -2.656 -1.793 -0.793  0.363 96.207
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    2.1929     0.4263   5.145 2.78e-07 ***
## attractiveness                -0.1998     0.1839  -1.086    0.277
## genderFemale                   0.9724     0.5858   1.660    0.097 .
## attractiveness:genderFemale   -0.3098     0.2526  -1.227    0.220
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.058 on 5110 degrees of freedom
## Multiple R-squared:  0.002349,   Adjusted R-squared:  0.001764
## F-statistic: 4.011 on 3 and 5110 DF,  p-value: 0.007314
```

Attractiveness was statistically significant (t= 0.2521, p = 0.012) but gender and interaction were not.

#Visualization I visualize the interaction using the fitted variables.

```
ggplot(health_tbl, aes(x = attractiveness, y = pregnancies, color = gender)) +
  geom_smooth(method = "lm") +
  geom_point()
```