Homework 12

Alizeh Azhar

11/8/2021

Problem 1: OIS Exercise 8.28, page 330 (3 pts)

Identify the outliers in the scatterplots shown below and determine what type of outliers they are. Explain your reasoning.

In (a),There is an outlier in the top left. Since it is far from the center of the data, it is a point with high leverage. It is also an influential point since, without that observation, the regression line would have a very different slope.

In (b),There is an outlier in the bottom left. Since it is far from the center of the data, it is a point with high leverage. However, it does not appear to be affecting the line much, so it is not an influential point

In (c), The observation is in the center of the data (in the x-axis direction), so this point does not have high leverage. This means the point won't have much effect on the slope of the line and so is not an influential point.

Problem 2: OIS Exercise 8.30, page 330 (2 pts)

Exercise 8.12 introduces data on the average monthly temperature during the month babies first try to crawl (about 6 months after birth) and the average first crawling age for babies born in a given month. A scatterplot of these two variables reveals a potential outlying month when the average temperature is about 53 degrees F and average crawling age is about 28.5 weeks. Does this point have high leverage? Is it an influential point?

The outlier is located almost at the center of the data (in the x-axis direction), horizontally not far from the center of the other points, so this point does not have high leverage. It is not an influential point since excluding this point from the analysis would not greatly affect the slope of the regression line

Problem 3: OIS Exercise 8.32a-d, page 335 (10 pts)

(a) Describe the relationship between the number of cans of beer and BAC (2 pts).

The relationship is positive, moderate- to -strong, and linear. There are no outliers and therefore no points that are influential. An increase in number of cans of beer increases the BAC.

(b) Write the equation of the regression line. Interpret the slope and intercept in context (3 pts).

The regression line is:

$$\hat{BAC} = -0.0127 + 0.0180 \times beers$$

Slope: For each additional number of cans of beer, the model predicts the BAC increases by 0.0180.

Intercept: According to the model, students who took 0 cans of beer are expected to measure -0.0127 grams.This is obviously not possible. Here, the y-intercept only to adjusts the height of the line and is meaningless by itself

(c) Do the data provide strong evidence that drinking more cans of beer is associated with an increase in blood alcohol? State the null and alternative hypotheses, report the p-value, and state your conclusion(3 pts).
H_0 : The true slope coefficient of beers is equal to $0(\beta_1 = 0)$ vs H_1 : The true slope coefficient of beers is not equal to $0(\beta_1 \neq 0)$. The p-value for the two-sided alternative hypothesis is incredibly small.Therefore, we reject H_0. The true slope parameter is indeed greater than 0.

Problem 4: OIS Exercise 8.32, extended (5 pts)

(a) Using the link https://www.openintro.org/stat/data/csv/bac.csv, load the blood alcohol dataset into R

```
bac <- read.csv("https://www.openintro.org/stat/data/csv/bac.csv")
```

(b) Rebuild the linear model predicting BAC from beers (1 pt).

```
bac_model <- lm(bac~beers, data = bac)
summary(bac_model)
```

```
##
## Call:
## lm(formula = bac ~ beers, data = bac)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.027118 -0.017350  0.001773  0.008623  0.041027
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.012701   0.012638  -1.005    0.332
## beers        0.017964   0.002402   7.480 2.97e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02044 on 14 degrees of freedom
## Multiple R-squared:  0.7998, Adjusted R-squared:  0.7855
## F-statistic: 55.94 on 1 and 14 DF,  p-value: 2.969e-06
```

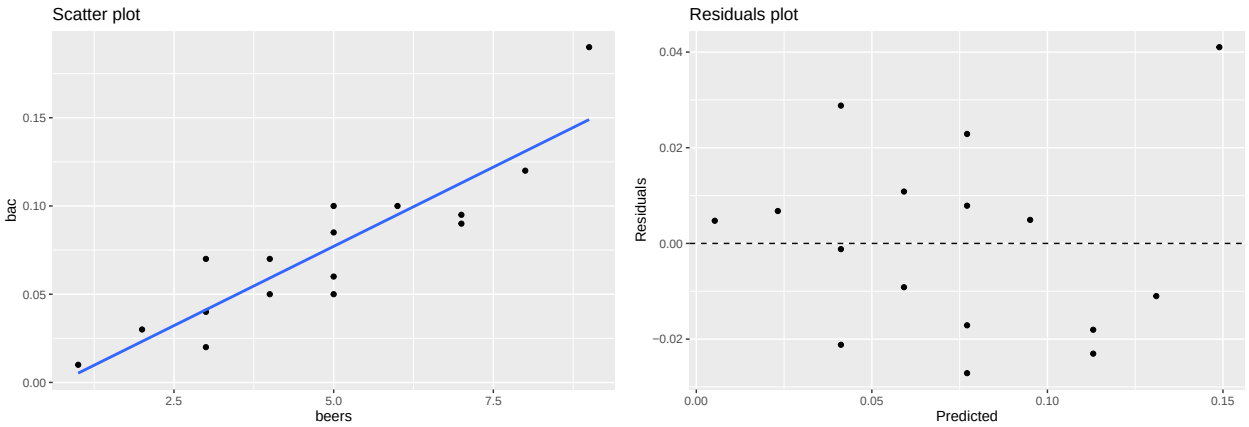
From the above output, the regression line is:

$$\hat{BAC} = -0.0127 + 0.0180 \times beers$$

(c) Create the residual plot for the model. Are there any patterns present (4 pts)?

```
library(ggplot2)
ggplot(data = bac, aes(x=beers, y=bac))+
  geom_point()+
  geom_smooth(method = "lm", se=F)+
  ggtitle("Scatter plot")

ggplot(data = bac_model, aes(x= .fitted, y= .resid))+
  geom_point()+
  geom_hline(yintercept = 0, linetype="dashed")+
  xlab("Predicted")+
  ylab("Residuals")+
  ggtitle("Residuals plot")
```



A scatter plot is shown on the left and the residual plot is on the right.There is a strong linear relationship. The linear model is appropriate and fits the data well. A linear model might be the best for the data because the negative and positive residuals are almost equal.