Biostat 250B HW2 Q6

Elvis Cui

January 28, 2021

1 Linear Model

The model is

$$\text{arsnails} = \beta_0 + \beta_1 * \text{age} + \beta_2 * \text{drinkuse} + \beta_3 * \text{cookuse} + \beta_4 * \text{arswater}$$

The fitted values and t-statistics are

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.452972   0.418132   1.083    0.296
age         -0.001290   0.003444  -0.374    0.713
sexMale     -0.145038   0.107448  -1.350    0.197
drinkuse    -0.011719   0.047010  -0.249    0.807
cookuse     -0.027471   0.082861  -0.332    0.745
arswater    13.195586   1.639792   8.047 8.01e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2302 on 15 degrees of freedom
Multiple R-squared:  0.8323,    Adjusted R-squared:  0.7764
F-statistic: 14.89 on 5 and 15 DF,  p-value: 2.339e-05
```

Figure 1: Fitted values

The figure on studentized residuals vs. leverage is

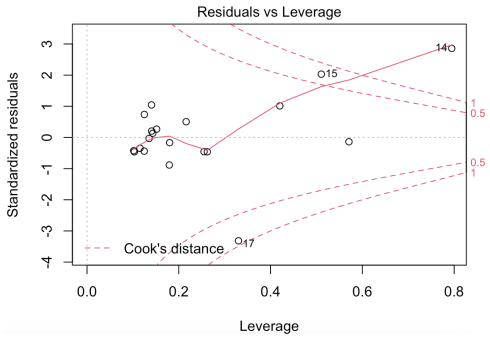

Figure 2: Studentized residuals

The Cook's distances are

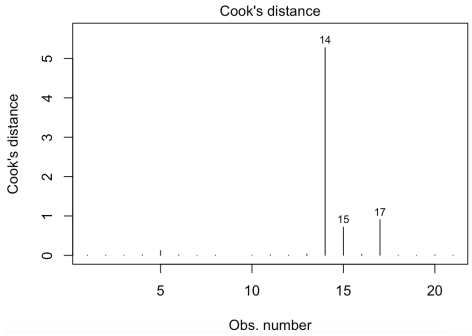

Figure 3: Cook's distance

The diagnostics on heteroscedasticity is

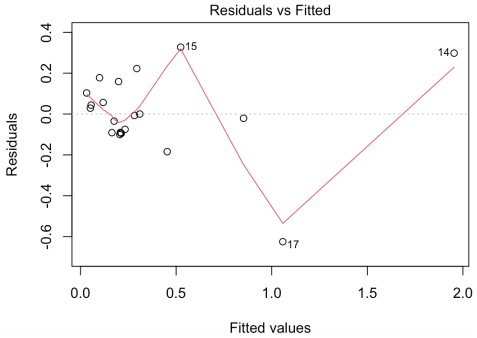

Figure 4: Fitted values

The mean-shift model outputs 2 outliers:

```
> outlierTest(mod)
    rstudent unadjusted p-value Bonferroni p
17 -6.209763         2.2762e-05   0.00045525
14  4.092296         1.0985e-03   0.02197100
```

Figure 5: Mean-shift model

The codes are

```r
setwd("~/Desktop/UCLA_Study/Bio250B/HW2")
rm(list=ls())

library(haven)
library(broom)
library(tidyverse)
theme_set(theme_classic())
data <- read_dta("arsenic.dta")

mod <- lm(arsnails~., data=data)
summary(mod)
plot(mod)

model.diag.metrics <- augment(mod)

# Cook's distance
plot(mod, 4)
# Residuals vs Leverage
plot(mod, 5)
model.diag.metrics %>%
  top_n(3, wt = .cooksd)

library(car)
outlierTest(mod)

```

Figure 6: R code