STATS 500- Homework 2

Xiaoyue Liu [xiaoyliu@umich.edu](mailto:xiaoyliu@umich.edu) UMID 28589009 Sep. 25th

Dataset used: uswages

Solutions for Problem 1 and 2 in Chapter 2 (page 30)

1.

lm(formula = uswages$wage ~ uswages$educ + uswages$exper)

Residuals:

Min 1Q Median 3Q Max

-1014.7 -235.2 -52.1 150.1 7249.2

Coefficients:

Estimate Std. Error t value

(Intercept) -239.1146 50.7111 -4.715

uswages$educ 51.8654 3.3423 15.518

uswages$exper 9.3287 0.7602 12.271

Pr(>|t|)

(Intercept) 2.58e-06 \*\*\*

uswages$educ < 2e-16 \*\*\*

uswages$exper < 2e-16 \*\*\*

---

Signif. codes:

0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 426.8 on 1964 degrees of freedom

(33 observations deleted due to missingness)

Multiple R-squared: 0.1348, Adjusted R-squared: 0.1339

F-statistic: 153 on 2 and 1964 DF, p-value: < 2.2e-16

2.

From RSS we know that only 13.39% of variation in the wage is explained by years of education and experience in this regression model.

3.

Case 1550 has the largest (positive) residual.

4.

The mean of the residuals is -1.3815348345639e-15

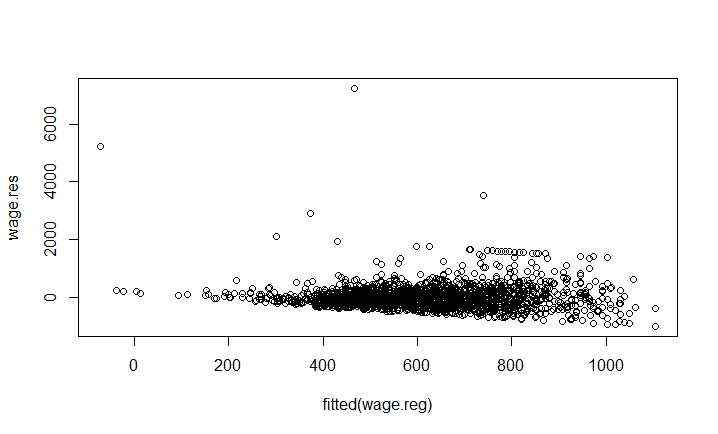
The median of the residuals is -52.1433717055574

The difference between the mean and median indicates that there are a few large wage numbers dragging mean towards positive while the median is not affected.

5.

The correlation of the residuals with the fitted values is 6.35677984435814e-17.

Plot residuals against fitted values



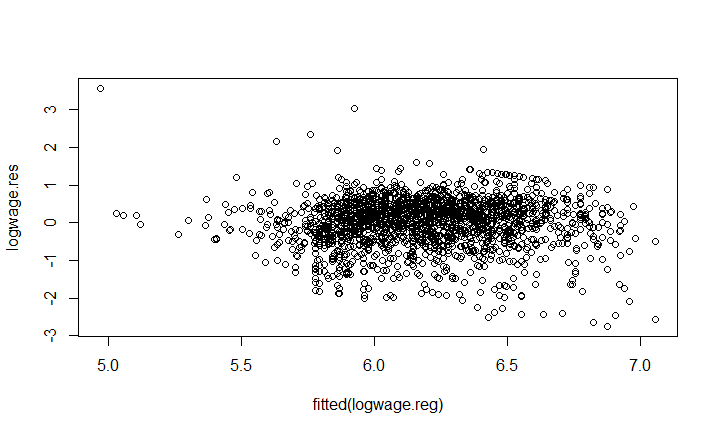
6.

The difference in predicted weekly wages would be 9.3287 dollars per week.

7.

If fit the same model but with log(weekly wages) as the response, the regression coefficient for experience is 0.016516.

Plot residuals against fitted values



The second model with log(weekly wlingeages) as response has a more natural interpretation because log(weekly wages) has a distribution similar to normal distribution,the scale is smaller and can compare with the scale of experience, and the residuals look more random as supposed

R-code used:

library(faraway)

data(uswages)

attach(uswages)

help(uswages)

summary(uswages)

uswages$exper[uswages$exper<0] <- NA

wage.reg<-lm(uswages$wage~uswages$educ+uswages$exper)

summary(wage.reg)

wage.res=residuals(wage.reg)

wage.largestno=which.max(wage.res)

wage.resmean=mean(wage.res)

wage.resmedian=median(wage.res)

summary(fitted(wage.reg))

wage.cor=cor(wage.res,fitted(wage.reg))

plot(fitted(wage.reg),wage.res)

logwage=log(uswages$wage)

logwage.reg<-lm(logwage~uswages$educ+uswages$exper)

summary(logwage.reg)

logwage.res=residuals(logwage.reg)

plot(fitted(logwage.reg),logwage.res)

plot(factor(uswages$exper),logwage)

abline(logwage.reg,col=2)