Stat 500 – Homework 2 (Solutions)

1. The R output (only relevant part) for the fitted model is given below.

```
> library(faraway)
> data(uswages)
> g<-lm(wage~educ+exper,uswages)
> summary(g)
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -239.1146    50.7111  -4.715 2.58e-06 ***
educ          51.8654     3.3423  15.518  < 2e-16 ***
exper          9.3287     0.7602  12.271  < 2e-16 ***

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. The percentage of variation in the response is given by the Multiple R-squared, which is 13.48%.

3. The observation (case number) with the largest residual is given by 1550 if you go with index or 15387 if you use the label given in the data – these come from the following:

```
> which.max(residuals(g))
15387
 1550
```

4. The mean and median are given together as a vector, rounded to 8 places for ease of output (rounding and vector representation are of course optional).

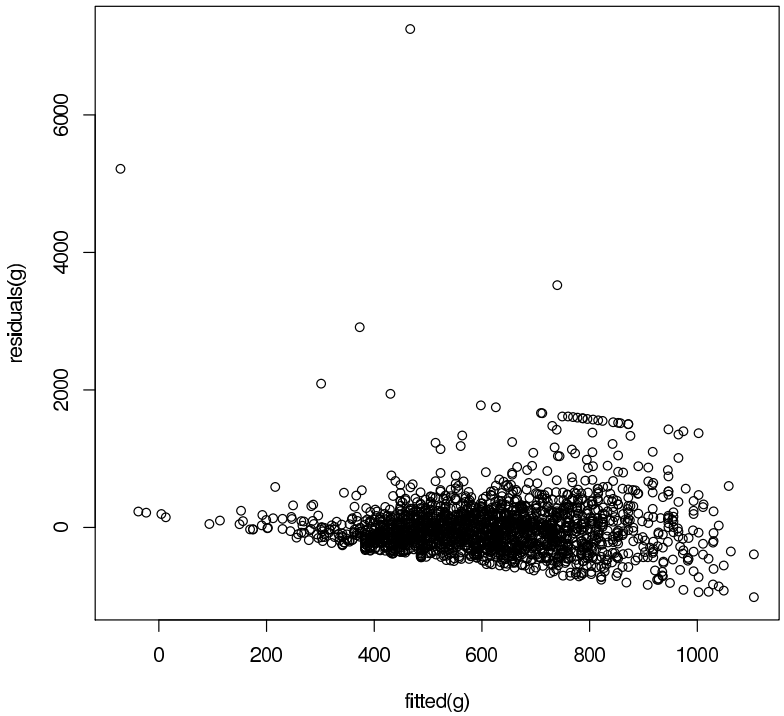
```
> round(c(mean(residuals(g)),median(residuals(g))),8)
[1]   0.00000 -52.14337
```

   The difference between the mean and the median indicates that the distribution of residuals is not perfectly symmetric. Since mean is greater than median, the distribution is slightly right-skewed.

5. The correlation between the residuals and fitted values is essentially zero.

```
> cor(residuals(g),fitted(g))
[1] 5.079906e-17
> round(cor(fitted(g),residuals(g)),8)
[1] 0
```

   A plot of residuals vs. fitted values is given below.



6. For two people with the same education and one year difference in experience, the more experienced person's predicted weekly wage will be \$9.3287 (see the coefficient of **exper**) higher.

7. A model with log(weekly wages) as the response and education and experience as predictors is fitted.

```
> g1<-lm(log(wage)~educ+exper,uswages)
> summary(g1)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 4.675518   0.077085   60.65   <2e-16 ***
educ        0.091940   0.005080   18.10   <2e-16 ***
exper       0.016516   0.001156   14.29   <2e-16 ***

Residual standard error: 0.6488 on 1964 degrees of freedom
  (33 observations deleted due to missingness)
Multiple R-Squared: 0.1747,     Adjusted R-squared: 0.1739
F-statistic: 207.9 on 2 and 1964 DF,  p-value: < 2.2e-16
```

The regression coefficient for experience is interpreted as follows. When education is held constant, one unit increase in experience will increase the `log(wage)` by 0.016516. That means, one unit increase in experience will multiply the weekly wage amount by a factor of $exp(0.016516) = 1.016653$.

Obviously, the first model (without logarithm) has a more natural interpretation.