Interpretation of Regression Coefficients, Fitted Values and Residuals

Reading the data

```
library(faraway)
data("uswages")
head(uswages)
```

```
        wage educ exper race smsa ne mw so we pt
6085  771.60   18    18    0    1  1  0  0  0  0
23701 617.28   15    20    0    1  0  0  0  1  0
16208 957.83   16     9    0    1  0  0  1  0  0
2720  617.28   12    24    0    1  1  0  0  0  0
9723  902.18   14    12    0    1  0  1  0  0  0
22239 299.15   12    33    0    1  0  0  0  1  0
```

Task 1: Fitting the model

```
attach(uswages)
fit = lm(wage ~ educ + exper, data = uswages)
S = summary(fit)
S
```

```
Call:
lm(formula = wage ~ educ + exper, data = uswages)
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-1018.2  -237.9   -50.9   149.9  7228.6

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -242.7994    50.6816  -4.791 1.78e-06 ***
educ          51.1753     3.3419  15.313  < 2e-16 ***
exper          9.7748     0.7506  13.023  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 427.9 on 1997 degrees of freedom
Multiple R-squared:  0.1351,    Adjusted R-squared:  0.1343
F-statistic:   156 on 2 and 1997 DF,  p-value: < 2.2e-16
```

Task 2: Regression coefficient of education

```
S$coeff["educ", "Estimate"]
```

```
[1] 51.17527
```

The simple interpretation is that each one year of education corresponds to increasing the weekly pay by about 51.18 dollars.

Task 3: Fitted values and residuals

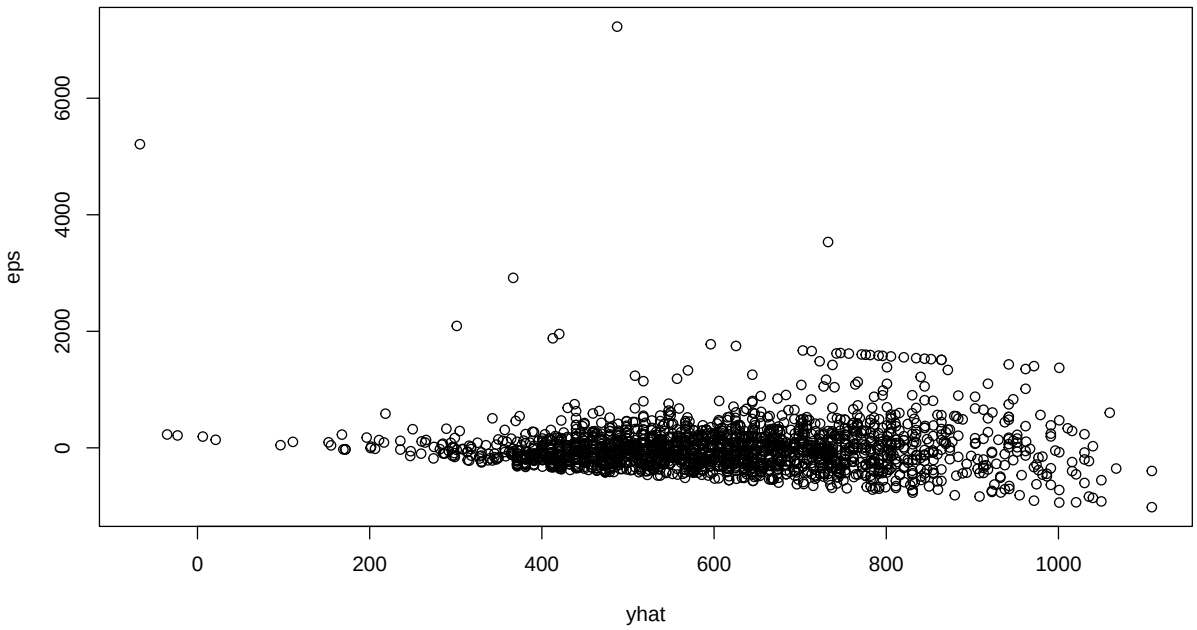
```
yhat = fitted(fit)
eps = resid(fit)

crossprod(yhat, eps)
```

```
             [,1]
[1,] 1.092849e-08
```

As we can see - the dot product of the fitted values and residuals is very close to 0. It will rarely be exactly 0 taking into consideration the constraints of computer arithmetics. It is enough though to infer that these vectors are ortogonal - as expected.

Task 4: Plotting the residuals vs fitted values

```
plot(yhat, eps)
```



It seems like the errors are centered, but their variance might be dependant on something else in our data.

Task 5: Consequences for the LS estimator

If the errors were not center we might find that the LS estimator is biased.