STAT 350

Assignment 6

1. Consider a design with 5 data points $x_1 = -2$, $x_2 = -1$, $x_3 = 0$, $x_4 = 1$ and $x_5 = 2$ so that the design matrix is

$$\begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} .$$

If we fit a simple linear regression of the form

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

for which the design matrix is as above, evaluate the non-centrality parameter of the t test of $\beta_1 = 0$ when in fact $\beta_1 = 2$ and $\sigma^2 = 2.5$. What would the power of a two sided 1% level t test of this null hypothesis be? How many times would we have to replicate this design to get a power of 0.9 for a 1% level test?

Solution: We have

$$X^T X = \begin{bmatrix} 5 & 0 \\ 0 & 10 \end{bmatrix}$$

and

$$(X^T X)^{-1} = \begin{bmatrix} \frac{1}{5} & 0 \\ 0 & \frac{1}{10} \end{bmatrix} .$$

The variance of $\hat{\beta}_1$ is $\sigma^2/10$ and the non-centrality parameter is $\beta_1/\sqrt{\sigma^2/10} = 4$. From Table B 5 p 1347 I get a power of 0.31 (3 df for error). If we replicate the design m times we would have a non-centrality parameter of $4\sqrt{m}$. The infinity line in the table would suggest an m of 1 but because the degrees of freedom is so low we have to do some trial and error. An m of 2 would give a non-centrality parameter of 5.66. There would be 8 degrees of freedom and the power would be around 0.95 so $m = 2$ would be adequate.

2. For the design points x_i as in the previous question evaluate the non-centrality parameter of the F test of the hypothesis $\beta_1 = \beta_2 = 0$ in the model

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2(x_i^2 - 2) + \epsilon_i \, .$$

Assume that in fact $\beta_1 = 1$ and $\beta_2 = 2$.

Solution: The non-centrality parameter is

$$E(RegressionSS)/\sigma^2 - 2 = [\beta_1, \beta_2] X_2^T (I - H_1) X_2 [\beta_1, \beta_2]^T / \sigma^2$$

Here $X_1 = [1,1,1,1,1]^T$, $H_1 = X_1(X_1^T X_1)X_1^T$ is a 5×5 matrix with all entries equal to 1/5 and

$$X_2 = \begin{bmatrix} -2 & 2 \\ -1 & -1 \\ 0 & -2 \\ 1 & -1 \\ 2 & 2 \end{bmatrix}.$$

We find $[\beta_1, \beta_2]X_2^T = [2, -3, -4, -1, 6]$ and finally that the non-centrality parameter is 66/2.5=26.4.

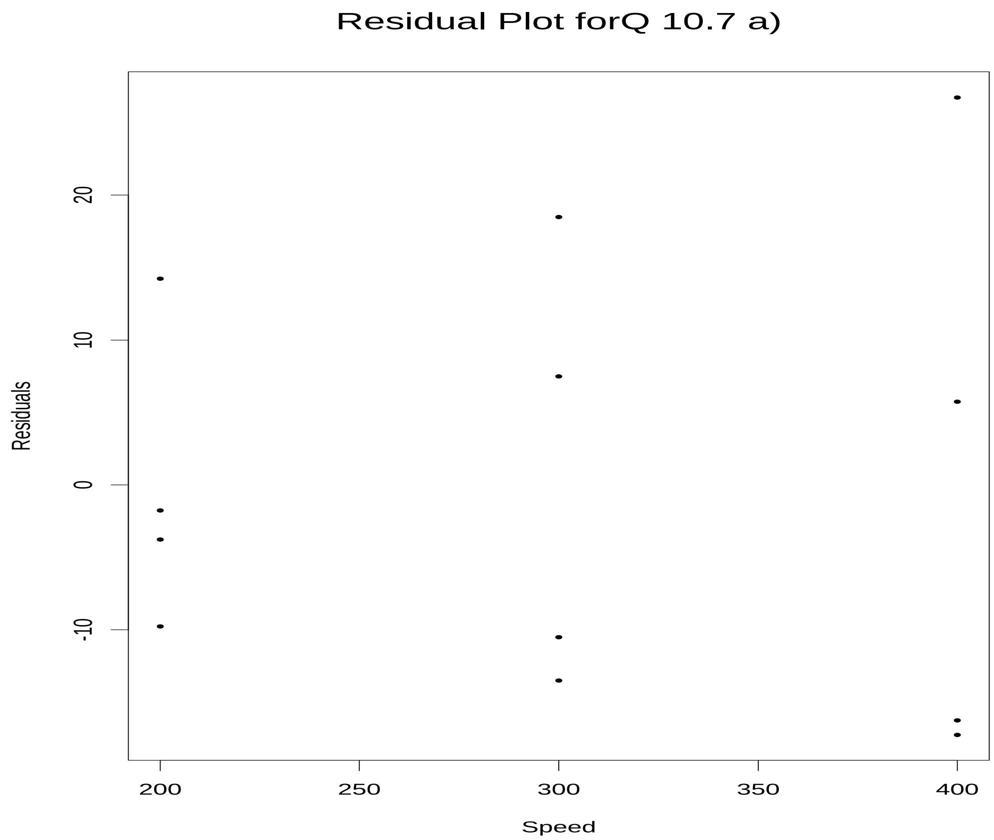
3. Question 10.7 parts a, c, d, e and f.

The data set in the text has one number different than the data set on the computer disk. The results below are for the data off the disk. Students who took the data from the book got somewhat different results but I marked them right, since they were right for the data they used.

Solution: The fitted line is

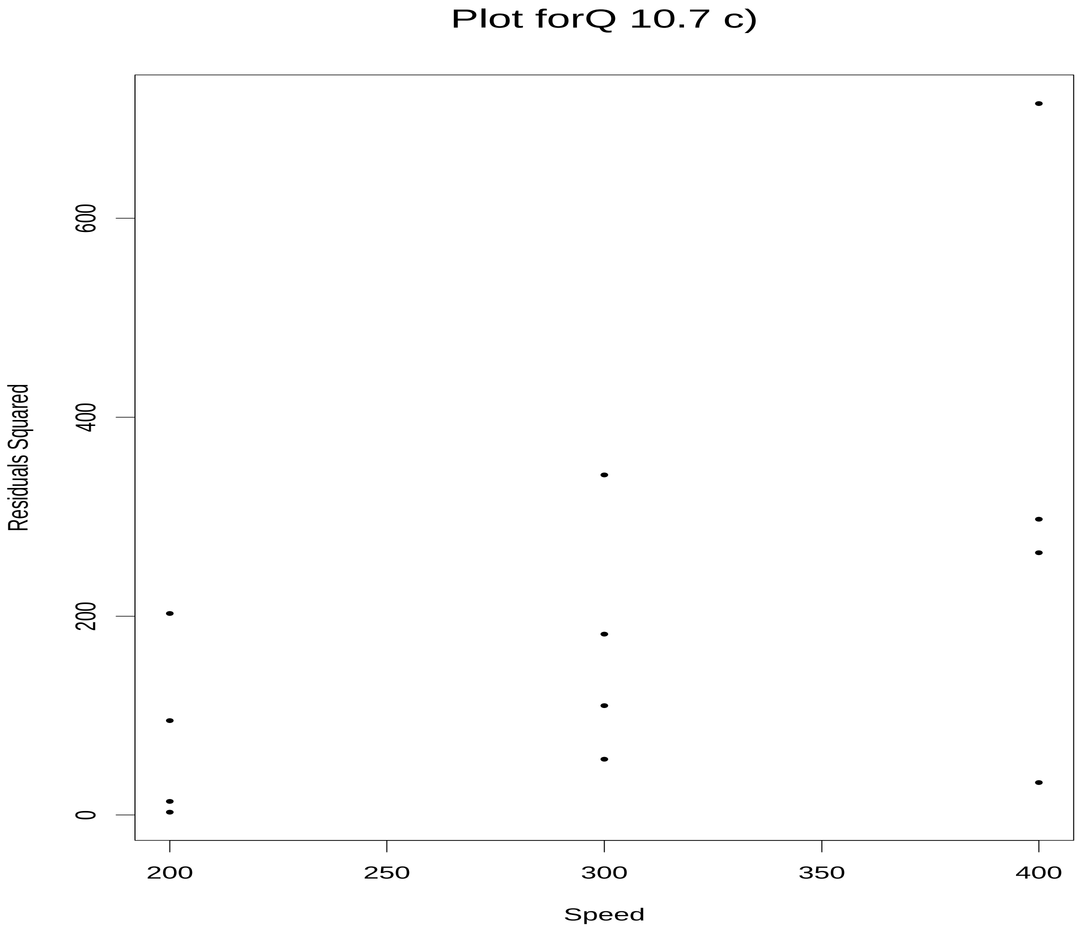
$$y = -5.75 + 0.1875x\,.$$

a) The residual plot is



This plot suggests that the residuals are bigger when Speed is bigger; this is evidence of heteroscedasticity.

c) The desired plot is



This plot suggests a linear fit of $\hat{\epsilon}_i^2$ against Speed might work.

d) The regression equation is

$$\sigma_i^2 = -180.0833 + 1.24375 X_i$$

The required weights are

```
0.0146 0.0032 0.0052 0.0032 0.0146 0.0052
0.0052 0.0032 0.0146 0.0032 0.0146 0.0052
```

e) The weighted least squares regression line is

$$y = -6.2332 + 0.1891x\,.$$

These estimates are little changed compared to the standard errors.

Parameter	OLS Estimate	WLS Estimate	OLS SE	WLS SE
Intercept	-5.75	-6.23	16.73	13.17
Slope	0.1875	0.1891	0.0538	0.0506

f) The standard errors from weighted least squares are rather smaller though more so for the less important parameter, the intercept.

4. Suppose

$$Y_i = \beta_0 + \epsilon_i$$

where the errors ϵ_i have variances σ^2/w_i and the w_i are known quantities. Find an explicit algebraic formula for the weighted least squares estimate of β_0.

Solution: The matrix X is a single column of n ones. The matrix W is diagonal with w_i down the diagonal. The vector WX has entries $w_1, \ldots, w_n$ and finally $X^T W X = \sum w_i$. Next $X^T W Y$ is the inner product between the vector WX with ith entry w_i and Y with ith entry Y_i so $X^T W Y = \sum w_i Y_i$. Finally $(X^T W X)^{-1} X^T W Y = \sum w_i Y_i / \sum w_i$ which is what is called a weighted average.