Bonus assignment 4

Ludwig Tranheden
ludtra@student.chalmers.se
9406238791
MVE155

2016/03/10

Contents

1 Problem 14.52

We are given a data set containing normal body temperature and heart rates for 65 males and 65 females.

1.1 a)

For both females and males we are to make scatterplots of the heart rate versus the body temperature. The result in shown i figure 1.

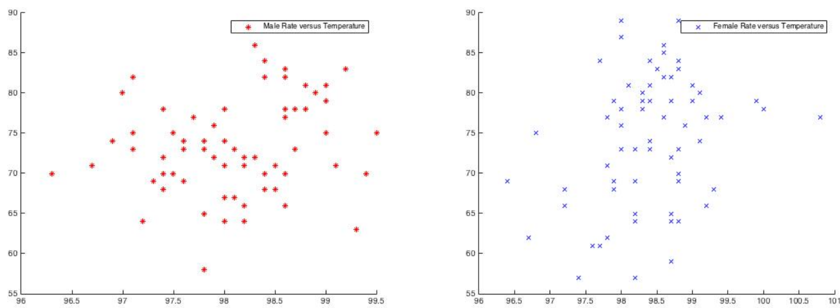

Figure 1: Scatter plots, rate vs temperature

The male scatter plot show some kind of relationship, even if we have a wide spread. As the rate increases the temperature increases as well. The female plot show a very weak relationship, we basically just have more data around the temperature interval (97.5,99.5) where the rate ranges from max-value to min-value.

1.2 b)

The question now is if there is any difference in relationship in the male and female case. To examine this we plot both scatterplots in the same graph.

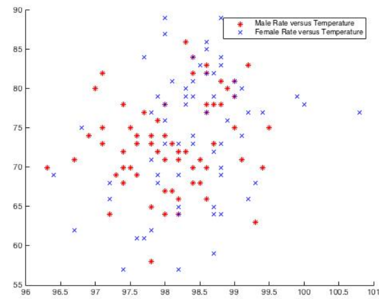

Figure 2: Male and female scatter in same grapgh

The relationship do look similar for the female and male case. Though the data are concentrated at different intervals so we should expect some difference in the linear regression model for the two cases.

1.3 c)

Now we are to fit a linear regression line for the males. We use the heart rate as the random response y and the temperature as the explanatory variable x and the simple linear regression model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \tag{1}$$

where $\epsilon_i \sim N(0, \sigma^2)$.
The least square estimates is.

$$b_0 = \hat{y} - b_1 \hat{x} \tag{2}$$

$$b_1 = \frac{r s_y}{s_x} \tag{3}$$

where r is the sample correlation coefficient. The regression line and the residuals given by $e_i = y_i - \hat{y}_i$ are shown in figure 3.

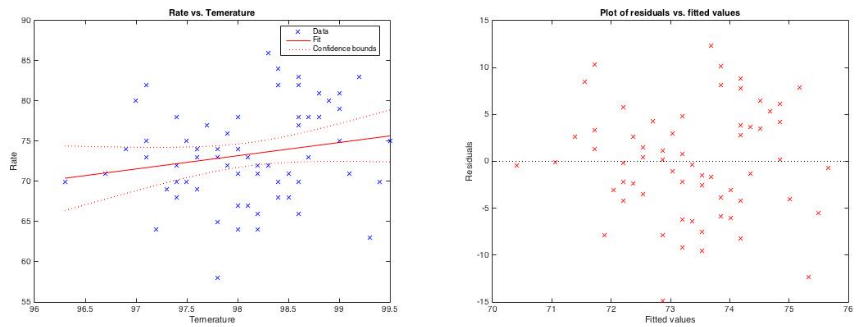

Figure 3: The regression line and residuals

The relationship does not seem linear at all. Further the coefficient of determination, r^2, is 0.0383. Which indicates that our model is not very good. The value of the estimated slope b_1 is 1.6445. Using the standard error formula

$$s_{b_1} = \frac{s}{s_x\sqrt{n-1}} \tag{4}$$

We get $s_{b_1} = 1.0389$

Listing 1: Code for linear regression

```
load Data
tbl = table([Male.Rate]',[Male.Temperature]','VariableNames',{'Rate','Temerature'});
lm = fitlm(tbl,'Rate~Temerature')
figure(1)
plot(lm)
figure(2)
plotResiduals(lm,'fitted','Color','r')
```

1.4 d)

Now we are to fit a linear regression line for the females. We use the heart rate as the random response y and the temperature as the explanatory variable x. The method to produce the result is exactly the same as in section c. But now we acquire the regression line and residuals shown in figure 4.

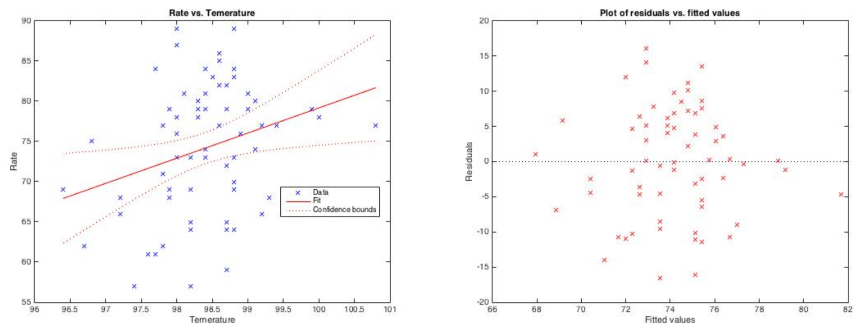

Figure 4: The regression line and residuals

Again the relationship does not look very linear. The coefficient of determination,r^2 , is 0.0823. Which again indicates that our model is not very good, the temperature does not explain the rate. We have $b_1 = 3.128$ and $s_{b_1} = 1.3157$. So the slope is three times larger then in the male case and the models are different as we predicted in subsection b. The code is the same as for the previous subsection.

1.5 e)

Now we are to test if the slopes of the two different models are the same. In other words we are to test

$$H_0 : \beta_{1,m} - \beta_{1,f} = 0 \qquad \text{The slopes are equal} \tag{5}$$

where m denotes Male and f Female, against the two sided alternative. We have the normality assumption that

$$b_{j,i} \sim N(\beta_{j,i}, \sigma^2_{i,j})$$

Using the large sample test we get that

$$b_{1,m} - b_{1,f} \sim^a N(\beta_{1,m} - \beta_{1,f}, s^2_{b_{1,m}} + s^2_{b_{1,f}}) \tag{6}$$

And we get the test statistic

$$Z = \frac{b_{1,m} - b_{1,f}}{\sqrt{s^2_{b_{1,m}} + s^2_{b_{1,f}}}}$$

which is approximately $N(0,1)$ under the null hypothesis. Inserting our values we get $Z = -0.8849$. This means we get a p value of 0.3762 and we do not reject the null hypothesis.

1.6 f)

Now we are to test if the intercepts of the two different models are the same.

$$H_0 : \beta_{0,m} - \beta_{0,f} = 0 \qquad \text{The intercepts are equal} \tag{7}$$

The formulas used is the same as in subsection e but with (1,m) and (1,f) replaced with (0,m) and (0,f). However standard error is different.

$$s_{b_{0,i}} = \frac{s\sqrt{\sum_1^n x_i^2}}{s_x\sqrt{n(n-1)}} \tag{8}$$

Using this instead of the standard error of b_1 in the formulas we get a test statistic $Z = 0.8840$ and a p value of 0.3767. Again we do not reject the null hypothesis.