MA684 Homework from Class 2

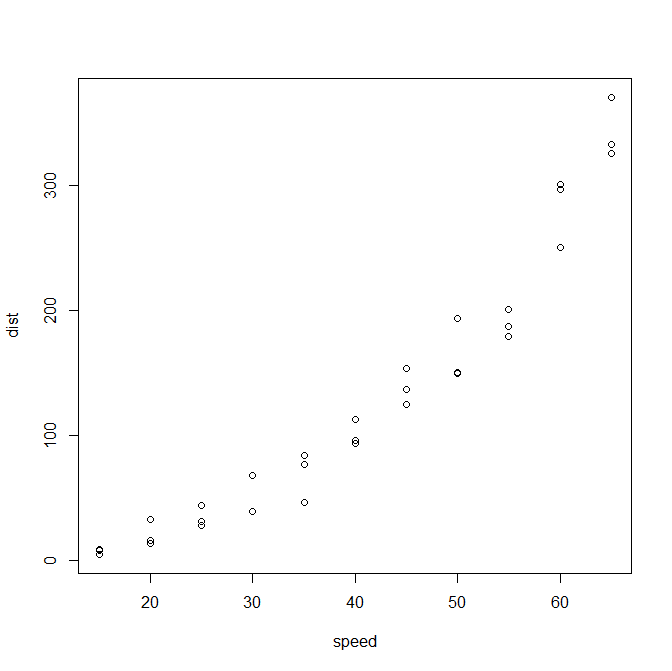
Simple Linear Regression

Hongbo Xue

U72135949

HW2

1. (Based on an example from the Chapter 5 exercises in KKM and N) A study was conducted to evaluate the relationship between driving speed (mph) and the distance a vehicle travels once brakes are applied (in feet). The scatter plot (drawn in R) from a sample of 33 trials is presented below:



What are the assumptions of the linear regression model, and are these assumptions met for these data? Explain.

**Answer:**

i) Independent, random sample from underlying population

ii) Linearity, the means of Y|X fall on a straight line

iii) Homoscedasticity, the variance of Y|X is the same for all X (or equivalently, the variance of

Ei is the same for all X)

iv) Normality, the distribution of Y|X follows a normal distribution for all X (or equivalently, Ei

follows a normal distribution).

v) Existence, the model holds for valid values of X

By looking at this scatter plot, it can be seen that variables **X** and **Y** have a close relationship that may be reasonably represented by a straight line. So that means speed and distance have a close relationship-----linear regression model. This would be represented mathematically as Y = a+bX+e where a describes where the line crosses the y-axis, b describes the slope of the line, and e is an error term that describes the variation of the real data above and below the line.

2. Is there an association between maternal age and a child’s cognitive ability? The following hypothetical results are from a linear regression predicting child IQ scores at age 10 (IQ scores are scaled to have a mean of 100 and standard deviation of 15 in the general population) from maternal age (in years, at the birth of the child) for a sample of n=30 mother-child pairs:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Parameter  Estimate | Standard  Error | t-value  (df) | p-value | 95% CI |
| Intercept  Maternal Age | 115.44  -0.49 | 14.73  0.56 | ---  0.8750 | ---  <0.001 | ---  -1.6369 , 0.6569 |

(n=30, sy|x = 13.5, mean +/- sd for maternal age 27.0 +/- 4.5)

2a. Complete the above table. What null hypothesis is being tested by the p-value for maternal age in this table? Summarize your conclusions based on the p-value. Give an interpretation for the 95% confidence interval for the slope for maternal age. How does this confidence interval relate to the p-value for maternal ages?

**Answer:**

**Null hypothesis:** There is not an association between maternal age and a child’s cognitive ability.

**P-value and conclusions:** Since this p-value is below the conventional cut-off value of 0.001, we can conclude that we have significant evidence to show an association between maternal age and a child’s cognitive ability.

**Interpretation:** 95% confident that, in the population of children, mean child’s cognitive ability increases or decreases somewhere between -1.6369 to 0.6569 scores for each year of age.

**CI with P-value:** If the null hypothesis of no an association between maternal age and a child’s cognitive ability was true, then the difference in means would be 0. So, if the confidence interval for the difference in means contains 0, there is no strong evidence against the null, and the p-value from the two-sample t-test will be greater than 0.05 (assuming a 95% CI). If the confidence interval for the difference in means does not contain 0, then we are 95% confident that the means are not equal, and the p-value from the two-sample t-test will be less than 0.05.

2b. What is the predicted IQ for a child born to a 20 year old mother? For a child born to a 30 year old mother?

Answer:

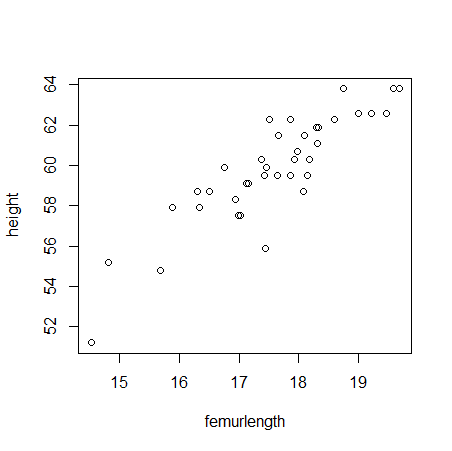
IQ(20)=115.44-0.49(20)=105.64

IQ(30)=115.44-0.49(30)=100.74

3. Data from a hypothetical sample of n=40 subjects was used to examine the association between femur (thigh bone) length and height, both measured in inches. Data are saved in the attached ‘CSI femur stature inches’ files.

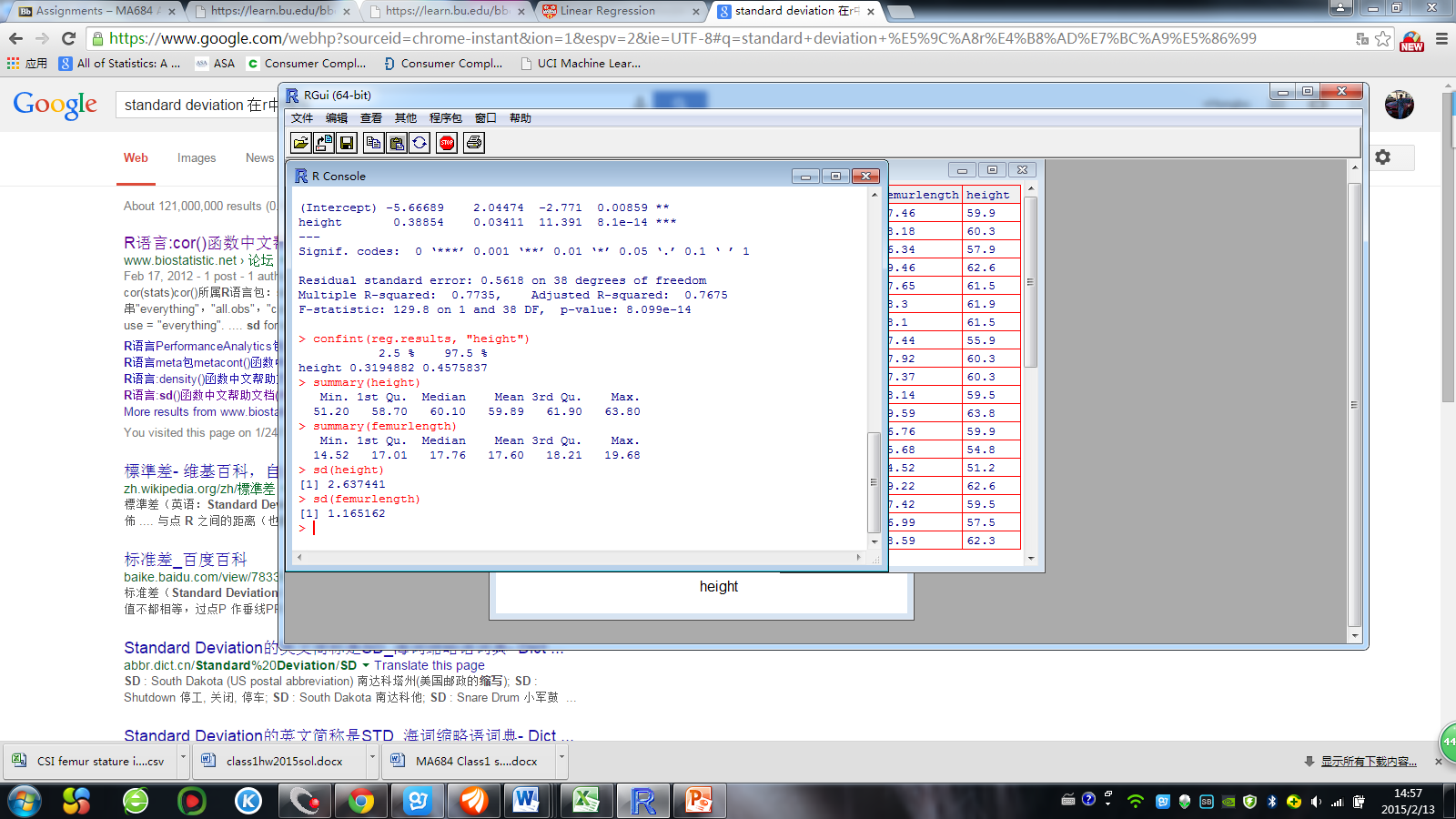
3a. Find the scatter plot between height and femur length.

**Answer:**



3b. Give the mean +/- sd for femur length and for height in this sample.

**Answer:**

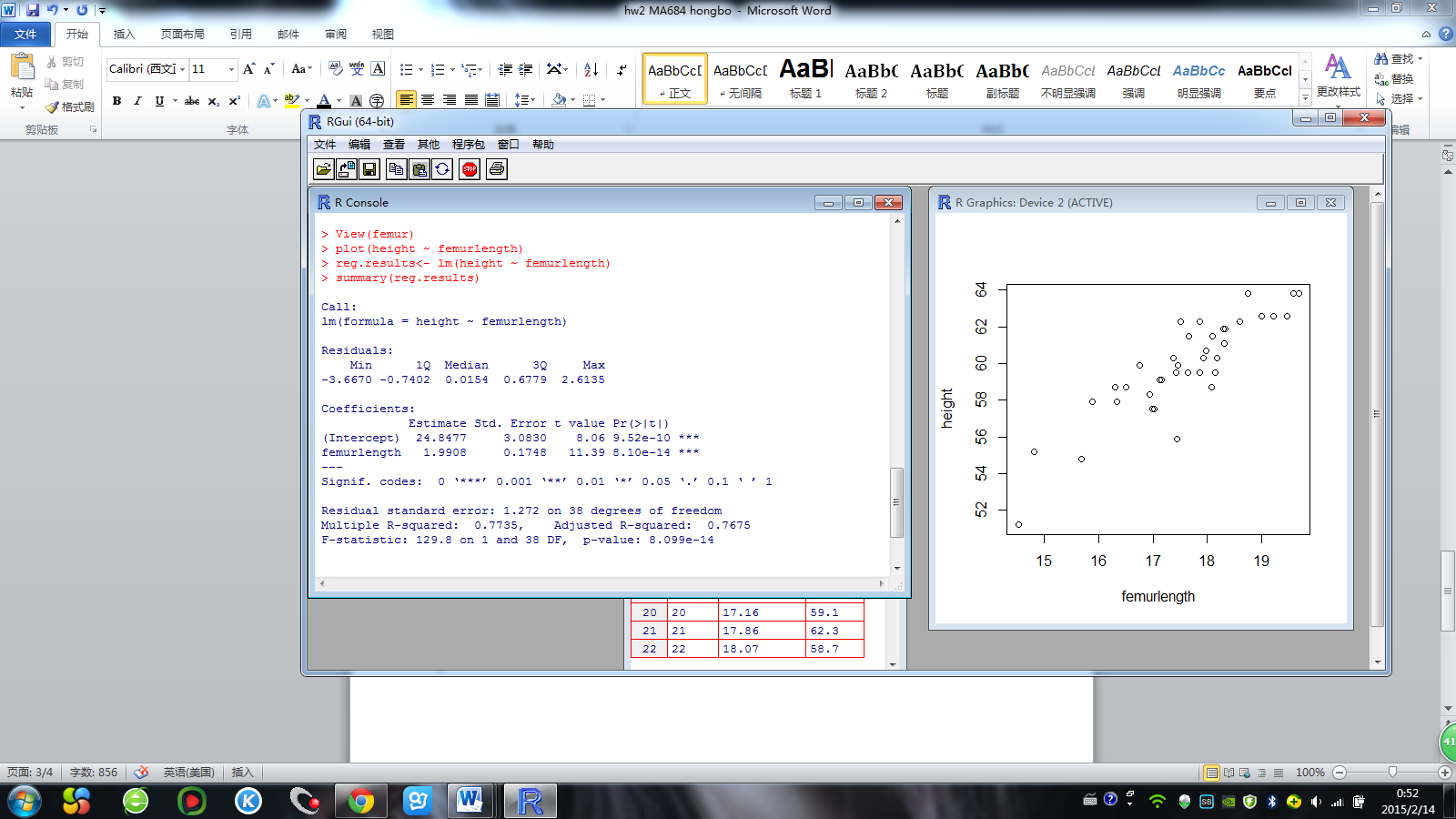


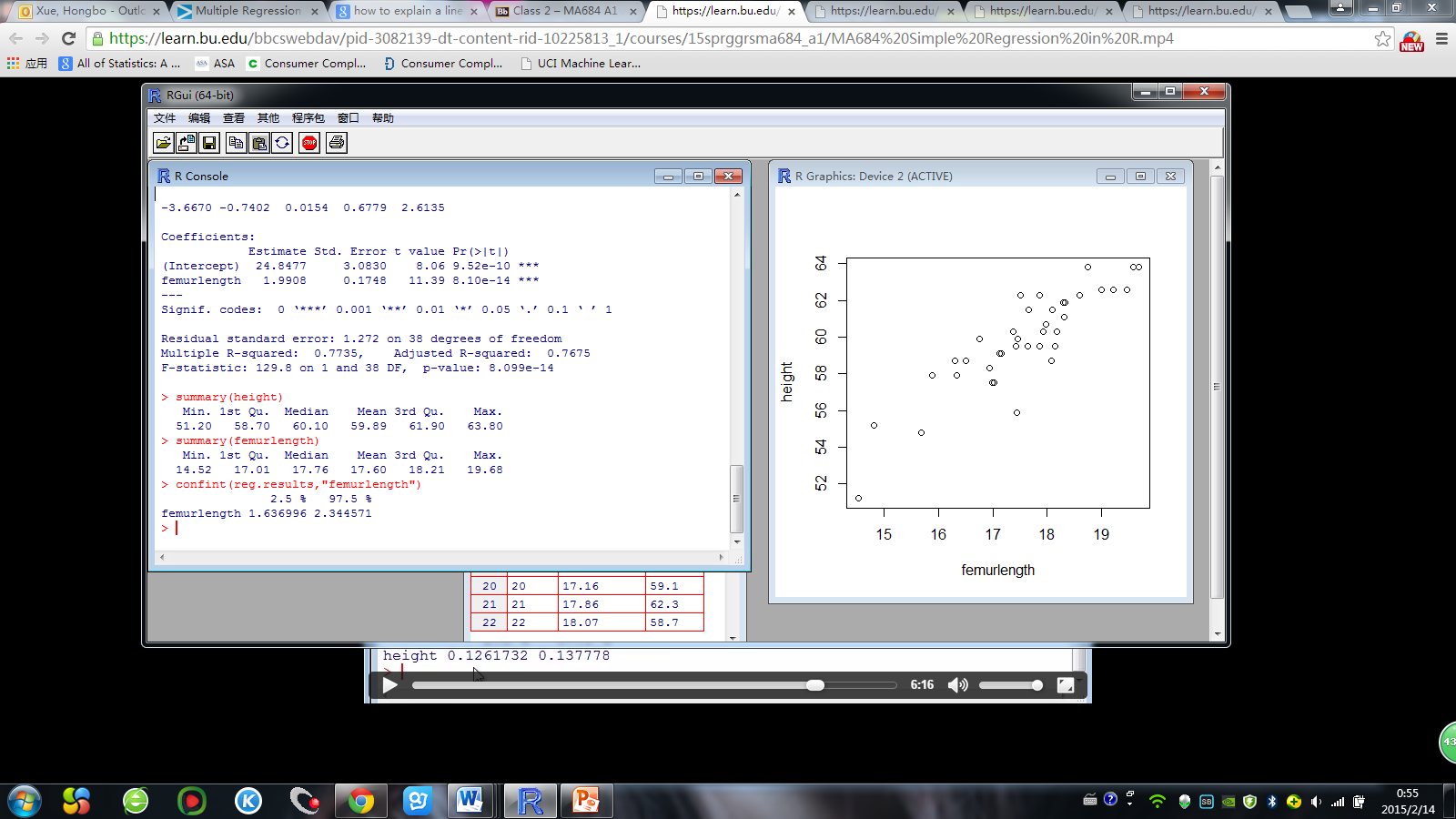
3c. Complete the following table:

Linear regression predicting height (inches) from femur length (inches)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Parameter  Estimate | Standard  Error | t-value  (df) | p-value | 95% CI |
| Intercept  Femur Length | 24.8477  1.9908 | 3.0830  0.1748 | ---  11.39 | ---  <0.001 | ---  1.636996 , 2.344571 |

Using information from this table,





3d. Is there a significant association between femur length and height? Explain.

**Answer:**

Yes, By looking at this scatter plot, it can be seen that variables **femurlength** and **height** have a close relationship that may be reasonably represented by a straight line.

3e. Describe the association between femur length and height by giving an interpretation of the slope for femur length from this regression.

**Answer:**

From the form, we could find the slope of linear function is 1.9908, and the function is

Height = 1.9908\*Femurlength+24.8477. That means the femur length will increase 1.9908 as the Femurlength increase 1.

3f. What is the value of s(y|x) from this regression (this value is not included in the above table, but is given in the computer output for the regression)? Give an interpretation of s(y|x) in the context of this analysis (in terms of femur lengths and heights).

**Answer:**

s(y|x)=1.272

Residual standard deviation is referred to as the standard deviation of points around a fitted line. In here, that means there is 1.272 error when we use the linear function of lengths and heights.