**HW 3**

**STAT 4315**

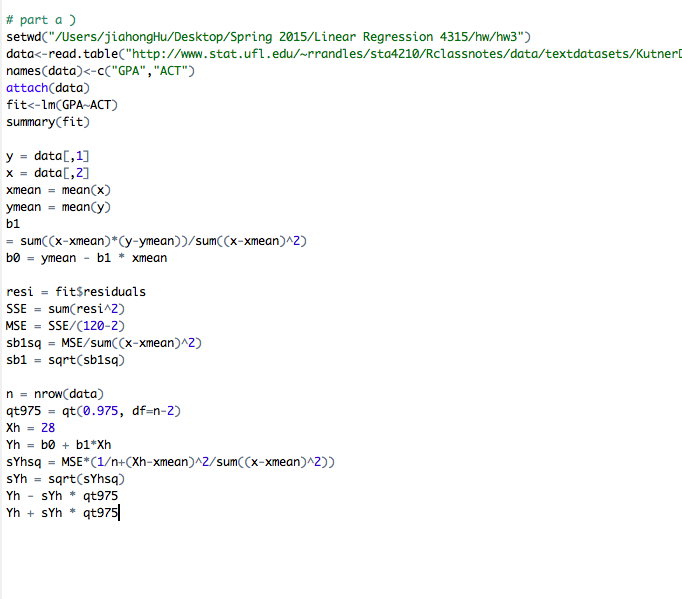
**Jiahong Hu**

**Jh3561**

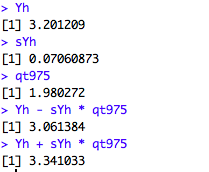
**Problem 2.13**

**Part a**

**Code:**

****

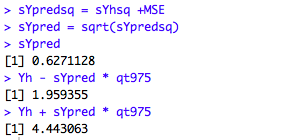
**Result:**

****

**Interpretation:**

The 95% confident interval of the mean freshman GPA for students whose ACT score is 28 is (3.061384,3.41033). We conclude with confidence coefficient 0.95 that the mean freshman GPA for students with ACT score of 28 is somewhere between 3.061384 and 3.41033.

**Part b**

****

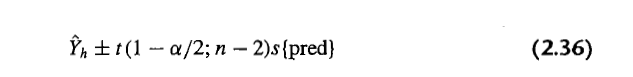
The 95% predict level is [1.959355, 4.443063]. With confidence coefficient 0.95, we predict Mary’s freshman GPA for with ACT score 28 will be somewhere between 1.959355 and 4.443063.

**Part c**

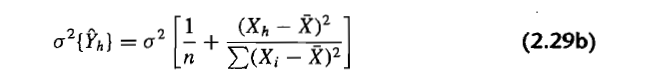
Yes, the prediction level is wider than the confidence level;

Yes, it should be the case.

Formula used for prediction level:

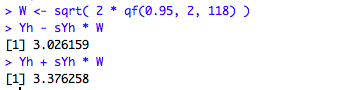


Macintosh HD:Users:jiahongHu:Desktop:Screen Shot 2015-02-17 at 9.57.08 AM.png

Formula used from confidence level:Macintosh HD:Users:jiahongHu:Desktop:Screen Shot 2015-02-17 at 9.59.36 AM.png

As the formula indicated, the variance of prediction error is always larger than the variance of sampling distribution of mean Y as long as there exists > 0. With other components same, the prediction level is always wider than the confidence level .

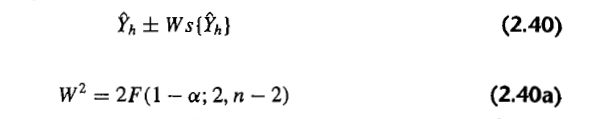
**Part d:**

****

The 95% confidence band of freshmen’s GPA with ACT score 28 is between 3.026159 and 3.376358.

Yes, the 95% confidence band at this point is wider than the 95% confidence interval.

Yes, it should be the case.

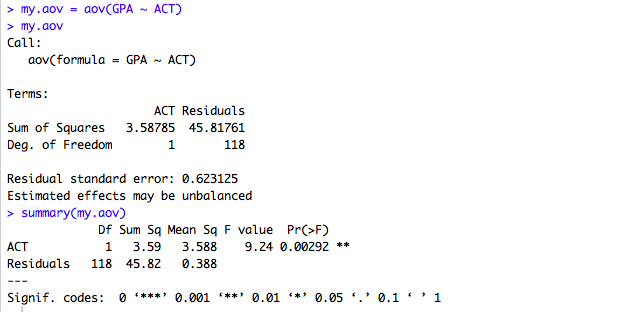
****

As indicated above, the formula for boundary value is of exactly same form as formula for the confidence limits for the mean response at Xh, except that the t multiple has been replaced by the W multiple. F distribution usually results a larger number than t distribution. The confidence band simultaneously convers mean response at every possible predictor level. Consequently, the boundary point of the confidence band for the regression line is wider apart from Xh than the confidence interval.

**Problem 2.23**

**Part a)**

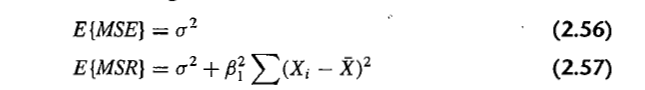
**Code:**

****

**ANOVA Table**

|  |  |  |  |
| --- | --- | --- | --- |
| **Source of Variation** | **SS** | **df** | **MS** |
| **Regression** | **3.59** | **1** | **3.588** |
| **Error** | **45.82** | **118** | **0.388** |
| **Total** | **49. 41** | **119** |  |

**Part b:**

****

Given the formula above, the MSE is an estimator of  and MSR is an estimator of Macintosh HD:Users:jiahongHu:Desktop:Screen Shot 2015-02-17 at 11.10.39 AM.png.

MSE = MSR= when = 0

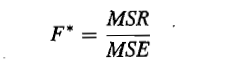
**Part c:**

Hypothesis Test:

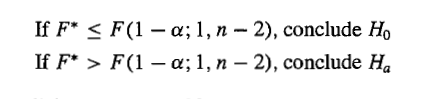
Macintosh HD:Users:jiahongHu:Desktop:Screen Shot 2015-02-17 at 11.14.40 AM.png

The risk of a Type I error is to be controlled at 0.01

Test Statistics



Construction of decision rule:



n =120,

**Result**:

Macintosh HD:Users:jiahongHu:Desktop:Screen Shot 2015-02-17 at 11.21.54 AM.png

F \* = 3.588/0.388 = 9.247423

F (0.99; 1, 118) = 6.855

F\* > 6.855 , Conclude Ha.

(We could see the direct result in the summary of ANOVA table too)

**Part d:**

Regression sum of square may be considered as absolute magnitude of the reduction in the variation of Y when X is introduced into the regression. In this case, SSR = 3.59

The relative reduction = SSR/ SSTO = 3.59/49.41=0.07265736 0.07%, which is called coefficient of determination.

**Part e:**

R^2 = SSR/SSTO = 0.07265736; r = sqrt(R^2) = 0.2695503

**Part f:**

R^2 gives us the fraction of the total variability in Y, SSTO that is explained by the regression model or X, calculated us SSR. On the other hand, r provides the strength of linear association between X and Y. Here, R^2 provides the more clear-cut interpretation.