**3. Ch.2 Problem 2.10**

a)

One should use prediction interval, since it is trying to infer a new observation, which is the humidity level.

b)

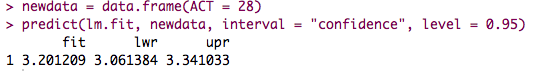
One should use the confidence interval, since it is trying to infer the average of families.

c)

One should use the prediction interval, since it is trying to make inference on a single observation, which is the amount of electricity measured by kilowatt-hours.

**4. Ch. 2 Problem 2.13**

a)



The confidence interval:

The lower bound is 3.061384 and the upper bound is 3.341033. It means that there are 95% probability that the true mean GPA for freshman with 28 on ACT is between 3.061384 and 3.341033.

b)

Macintosh HD:Users:Qihong:Desktop:Screen Shot 2015-02-12 at 9.23.44 PM Feb 12.png

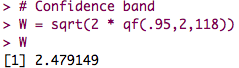
If there is a new student with 28 on the entrance test, this student has 95% chance to obtain a score above 3.201209 on first year GPA.

c)

Yes, the prediction interval is wider and it should be. Since there is additional variability of 1 added in the formula that computes the prediction interval.

d)

Confidence band



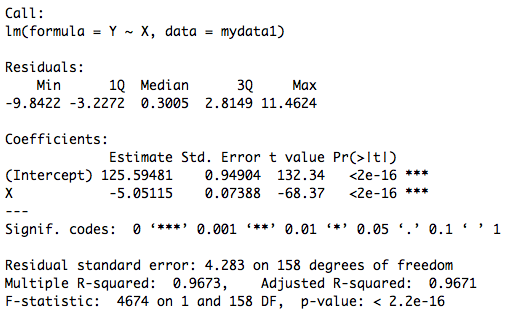
Yes, it is wider than the confidence interval, and it should be wider.

**WHY CONFIDENCE BAND SHOULD BE WIDER?**

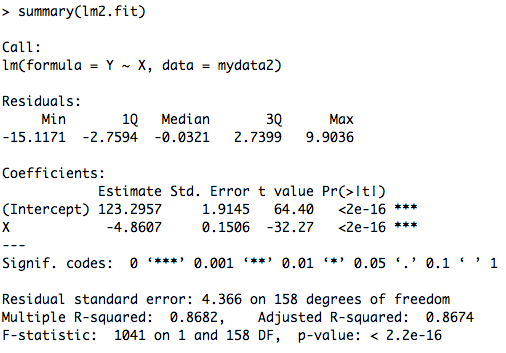
**6. Simulation**

**Part A**

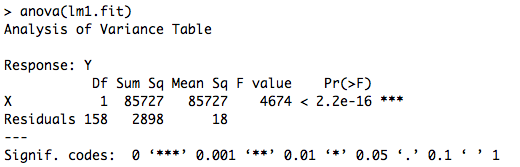
The summary for the first data set



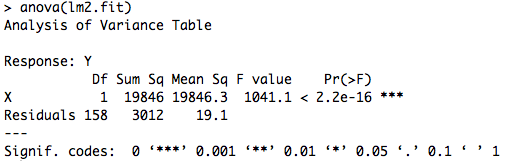
The summary for the second data set



The ANOVA table for the first data set

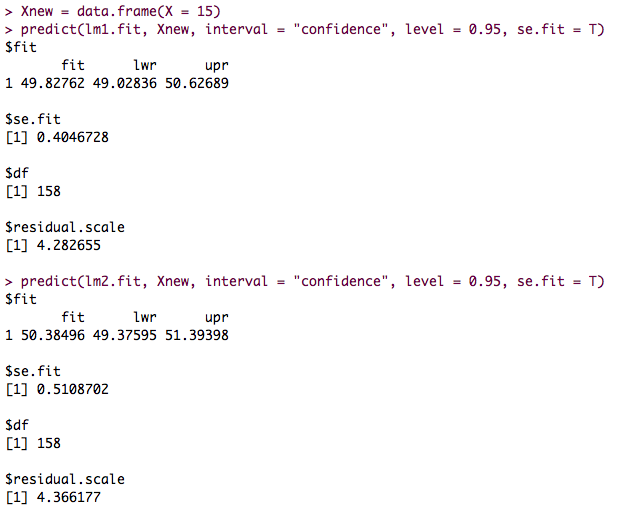


The ANOVA table for the second data set

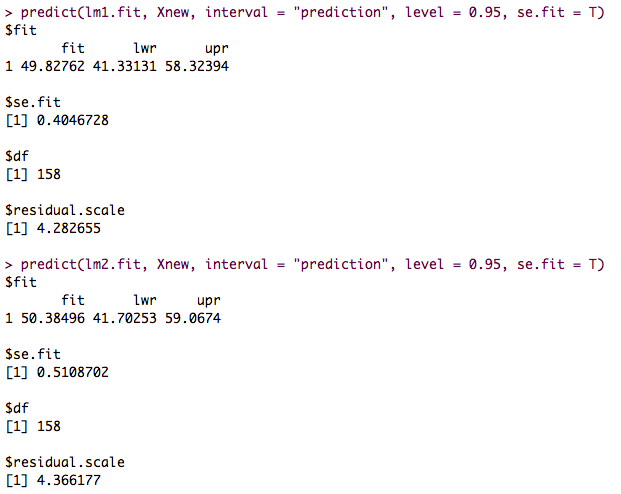


The first simulated data, which is the data with higher variance in X, has a better approximation to the true regression function. Namely, the estimated ß0 and ß1 for the first data set is closer to the truth. Moreover, the standard errors for both ß0 and ß1 are smaller in the first data compared to the second data.

Compare the confidence intervals

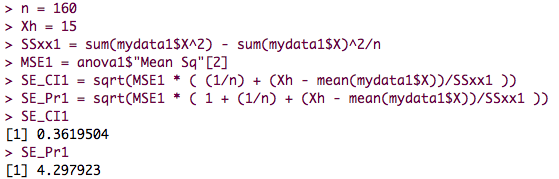


Compare the prediction intervals

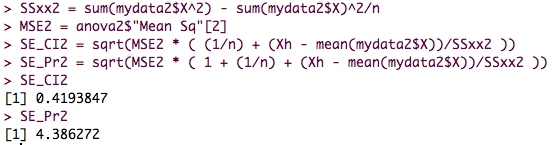


The data set with larger variability had smaller confidence interval, as well as the prediction intervals.

The standard errors for the prediction interval and confidence interval for data1



The standard errors for the prediction interval and confidence interval for data2



The standard error for the confidence interval for the first data set: 0.3619504

The standard error for the confidence interval for the second data set: 0.4193847

The standard error for the prediction interval for the first data set: 4.297923

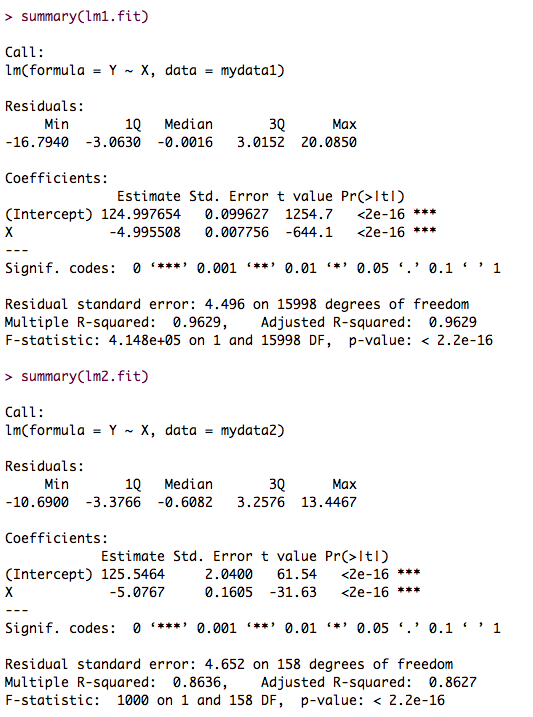
The standard error for the prediction interval for the second data set: 4.386272

In conclusion, the standard error for the first data set that has larger variability also has smaller standard error for both confidence interval and prediction interval.

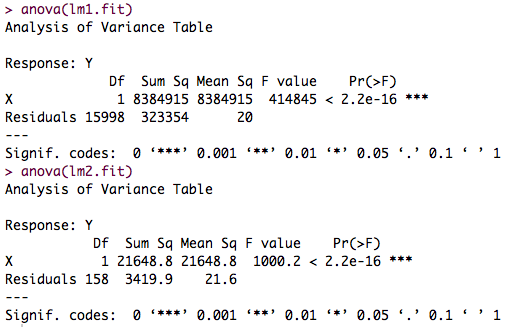
**6. Part B**

Simulation with larger sample sizes

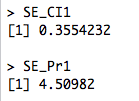
The summary statistics for both models:



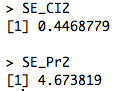
The ANOVA table for both data sets



The standard error for the confidence interval and prediction interval for the first data set:



The standard error for the confidence interval and prediction interval for the second data set:



Compare to previous simulation, which had smaller sample sizes, the estimated ß0, ß1 and sigma squared become even closer to the truth.

When simulating with larger sample size, the standard error for the prediction interval become larger. For confidence, there is also a similar trend to become larger, but it is unclear based on current results.