Assignment 2

your name goes here

Due: Wednesday, January 31, 2018, noon, to Sakai

Summary

The tasks in this homework assignment include running inference on the SLR model, including confidence and prediction intervals at a particular value of the explanatory variable. Additionally, you will gain practice breaking down the sums of squares of a model and interpreting the values.

Assignment

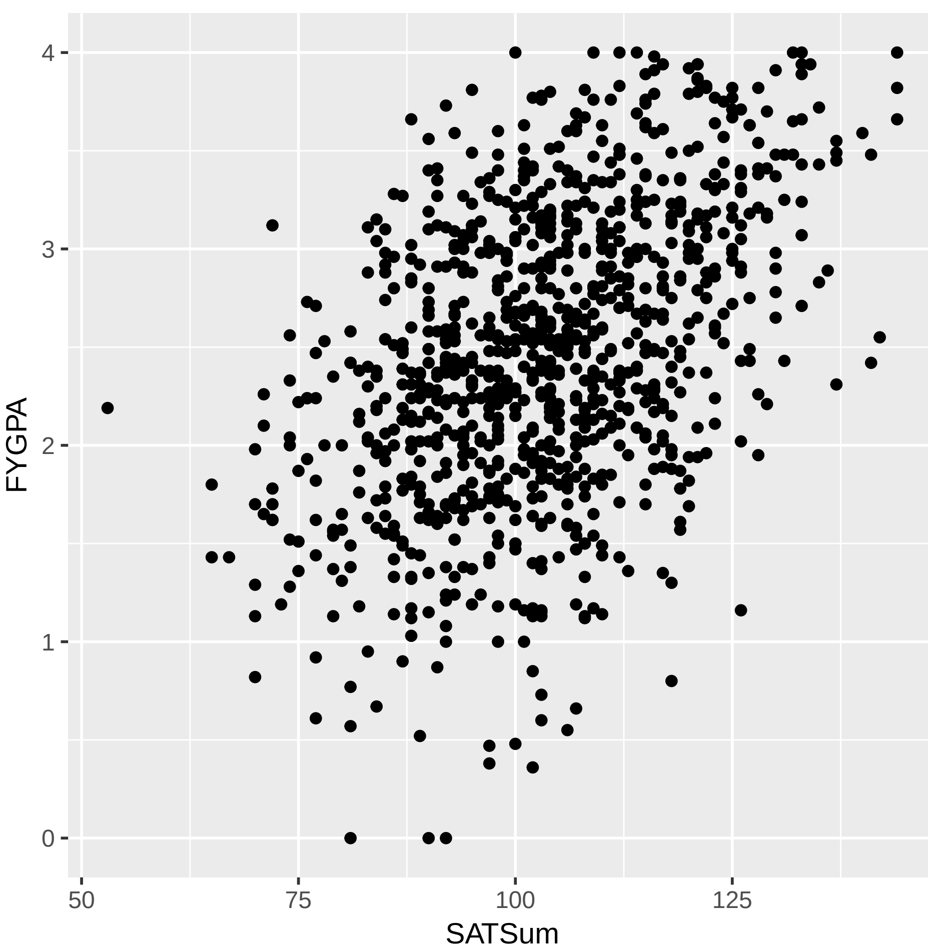
1. For each of the following questions, explain whether a confidence interval for a mean response or a prediction interval for a new observation is appropriate.
   (a) What will be the humidity level in this greenhouse tomorrow when we set the temperature level at 31°C?
   (b) How much do families whose disposable income is $23,500 spend, on the average, for meals away from home?
   (c) How many kilowatt-hours of electricity will be consumed next month by commercial and industrial users in the Twin Cities service area, given that the index of business activity for the area remains at its present level?
2. Can Var(individual predicted value) = $\sigma^2\{\text{ pred }\}$ (eq 2.37 in text) be brought increasingly close to 0 as n becomes large? Is this also true for Var(mean value) = $\sigma^2\{\hat{Y}_h\}$ (eq 2.29b)? What is the implication of the difference?
3. For testing the parameter β_1, why is the t-test more versatile than the F-test?
4. In a small-scale regression study, five observations on Y were obtained corresponding to $X = 1, 4, 10, 11, 14$. Assume that $\sigma = 0.6$, $\beta_0 = 5$, and $\beta_1 = 3$.
   (a) What are the expected values of MSR and MSE here?
   (b) For determining whether or not a regression relation exists, would it have been better or worse to have made the five observations ad $X = 6, 7, 8, 9, 10$? Why? Would the same answer apply if the principal purpose were to estimate the mean response for $X = 8$? Discuss.
5. The normal error regression model is assumed to be applicable. (See section 2.8, pages 72-73, for F test structure / algorithm.)
   (a) When testing $H_0 : \beta_1 = 5$ versus $H_a : \beta_1 \neq 5$ by means of a general linear test, what is the reduced model? What are the degrees of freedom for SSE(R)?
   (b) When testing $H_0 : \beta_0 = 2, \beta_1 = 5$ versus $H_a :$ not both $\beta_0 = 2, \beta_1 = 5$ by means of a general linear test, what is the reduced model? What are the degrees of freedom for SSE(R)?

6. Consider the dataset collected by ETS on 1000 randomly selected students (at an unnamed college). The variables of interest are SAT and GPA (note that SAT is measured in terms of percentiles).

```
require(openintro)
glimpse(satGPA)

## Observations: 1,000
## Variables: 6
## $ sex    <int> 1, 2, 2, 1, 1, 2, 1, 1, 2, 1, 1, 2, 2, 2, 1, 2, 2, 1, 2...
## $ SATV   <int> 65, 58, 56, 42, 55, 55, 57, 53, 67, 41, 58, 45, 43, 50,...
## $ SATM   <int> 62, 64, 60, 53, 52, 56, 65, 62, 77, 44, 70, 57, 45, 58,...
## $ SATSum <int> 127, 122, 116, 95, 107, 111, 122, 115, 144, 85, 128, 10...
## $ HSGPA  <dbl> 3.40, 4.00, 3.75, 3.75, 4.00, 4.00, 2.80, 3.80, 4.00, 2...
## $ FYGPA  <dbl> 3.18, 3.33, 3.25, 2.42, 2.63, 2.91, 2.83, 2.51, 3.82, 2...

ggplot(satGPA, aes(x=SATSum, y=FYGPA)) + geom_point()
```



(a) Consider predicting first year GPA from the sum of the verbal and math SAT percentiles. Obtain a 99% CI for β_1. Interpret the CI. Does it contain zero? Why might it be interesting to know whether or not the CI contains zero?

(b) Using the test statistic t^*, test whether linear association exists between a student's sum of SAT (X) and first year GPA (Y). Test at a 0.01 significance level. State the hypotheses, p-value, and conclusion in terms of the problem.

(c) Obtain a 95% interval for the average first year GPA for students whose SAT percentiles sum to 120. Interpret your interval

(d) Obtain a 95% interval for the first year GPA of Veronica Davis who who scored an SAT sum of 120. Interpret your interval.

(e) Would you expect the two intervals above to be wider or more narrow for considering an SAT sum of 100? Explain using both the mathematical formula for creating the intervals and using the intuition given by what we've seen in class of the variability of the line.

(f) Set up the ANOVA table for the regression model above.

(g) What is estimated by MSR? What about MSE? When do they estimate the same thing?

(h) Conduct an F test of whether or not $\beta_1 = 0$. State the hypotheses, p-value, and conclusion in context of the problem.

(i) What is the absolute magnitude in the reduction of variation in first year GPA when the sum of the SAT is introduced into the regression model? What is the relative reduction? What is the name of the latter measure?

(j) Obtain r (with the appropriate sign).