STAT 420 Spring 2015
Dalpiaz

Exam #2, Part #1

(due Wednesday, May 6, by 3:00 p.m.)

1. In order to compare the average GPA for the members of three fraternities at Anytown State University, six students were randomly chosen from each of the three fraternities (Deltas, Omegas, Sigmas), the students' GPA (y) and the average time spent studying per week (x) in hours was recorded.

Deltas		Omegas		Sigmas	
x	y	x	y	x	y
0	1.2	11	2.5	4	2.6
1	0.7	10	3.2	7	1.7
5	0.9	17	3.9	12	2.8
7	1.1	14	3.2	13	2.7
11	2.1	18	3.4	15	3.5
15	2.7	20	3.6	18	3.2

The data can be found in **`gpa.csv`**. Be sure to tell RStudio the file contains a header.

a) Fit the model

$$Y = \beta_0 + \beta_1 v_1 + \beta_2 v_2 + \beta_3 x + \varepsilon, \qquad \text{(a)}$$

where

$v_1 = 1$ if a student is from Deltas fraternity, 0 otherwise,

$v_2 = 1$ if a student is from Omegas fraternity, 0 otherwise.

State the fitted regression line.

b) Test whether the relationship between GPA and time spent studying is the same for all three fraternities. Use $\alpha = 0.05$. State the null hypothesis in terms of the model parameters, report the value of the test statistic, report the p-value and state your decision.

c) Test whether the time spent studying has no effect on the GPA. Use $\alpha = 0.05$. State the null hypothesis in terms of the model parameters, report the value of the test statistic, report the p-value and state your decision.

d) Let $v_3 = 1$ if a student is from Sigmas fraternity, 0 otherwise.

(i) Find and compare the predicted values of model (A) and the model

$$Y = \beta_1 v_1 + \beta_2 v_2 + \beta_3 v_3 + \beta_4 x + \varepsilon. \qquad (D)$$

Output the fitted values of both models.

(ii) Find and compare the residuals of model (A) and model (D). Output the residuals of both rounded to the first decimal place.

"Hint": To fit a model without the *y*-intercept, use:

```
lm(y ~ ... + 0)
```

e) The null model from part (b) and the null model from part (c) are NOT nested. (Two models are nested if both contain the same terms and one has at least one additional term.) So they cannot be compared with an F test.

It may be helpful to use the `broom` package and the `glance()` function to complete the following parts:

(i) Find the AIC values for the null model from part (b) and the null model from part (c). Which model is preferred?

(ii) Find the BIC values for the null model from part (b) and the null model from part (c). Which model is preferred?

(iii) Find the Adjusted *R*-squared values for the null model from part (b) and the null model from part (c). Which model is preferred?

f) Fit the model

$$Y = \beta_0 + \beta_1 v_1 + \beta_2 v_2 + \beta_3 x + \beta_4 v_1 x + \beta_5 v_2 x + \varepsilon,$$

Test for interaction between fraternities and time spent studying. Use $\alpha = 0.05$. State the null hypothesis in terms of the model parameters, report the value of the test statistic, report the p-value and state your decision.