Test 2 - part 1

STAT 021

Swarthmore College

Do not flip this page until instructed to do so.

Test organization: There are X questions in total on this test and they are organized into two subsections: the first **X** questions are select all that apply multiple choice questions and the last **X** questions are free response. If you need additional paper you may come to the front of the class and pick some up. There are a total of XX points possible on this test.

Instructions: The first part of this test are multiple choice questions that do not require any additional explanation or work. No extra work will be considered in the grading of these questions but *you can get partial credit* for many of these questions. The last part of this test involves short answer questions. For these questions, you must show all your work and/or provide enough justification and explain your reasoning in order to get full credit or be considered for partial credit. You do not need a calculator to evaluate any expressions. For any calculation problems, simply writing out the formula to find the answer will suffice.

First and Last Name: _____

Swarthmore Username: _____

Take a deep breath.

You have prepared for this test and with a clear and well-rested mind, you are ready to show me what you have learned this semester. As with the other tests, the purpose of this test is to measure your understanding of the material we have covered. This is nothing more than a metric for me to evaluate your preparedness to think statistically at this particular moment in time and in this particular setting. This is not a perfect measure of your knowledge and does not predict your future statistical skills.

Multiple choice problems (2 points each)

Suppose we are trying to understand how the aerial biomass (response variable) production of a certain type of marsh grass is related to the three predictor variables

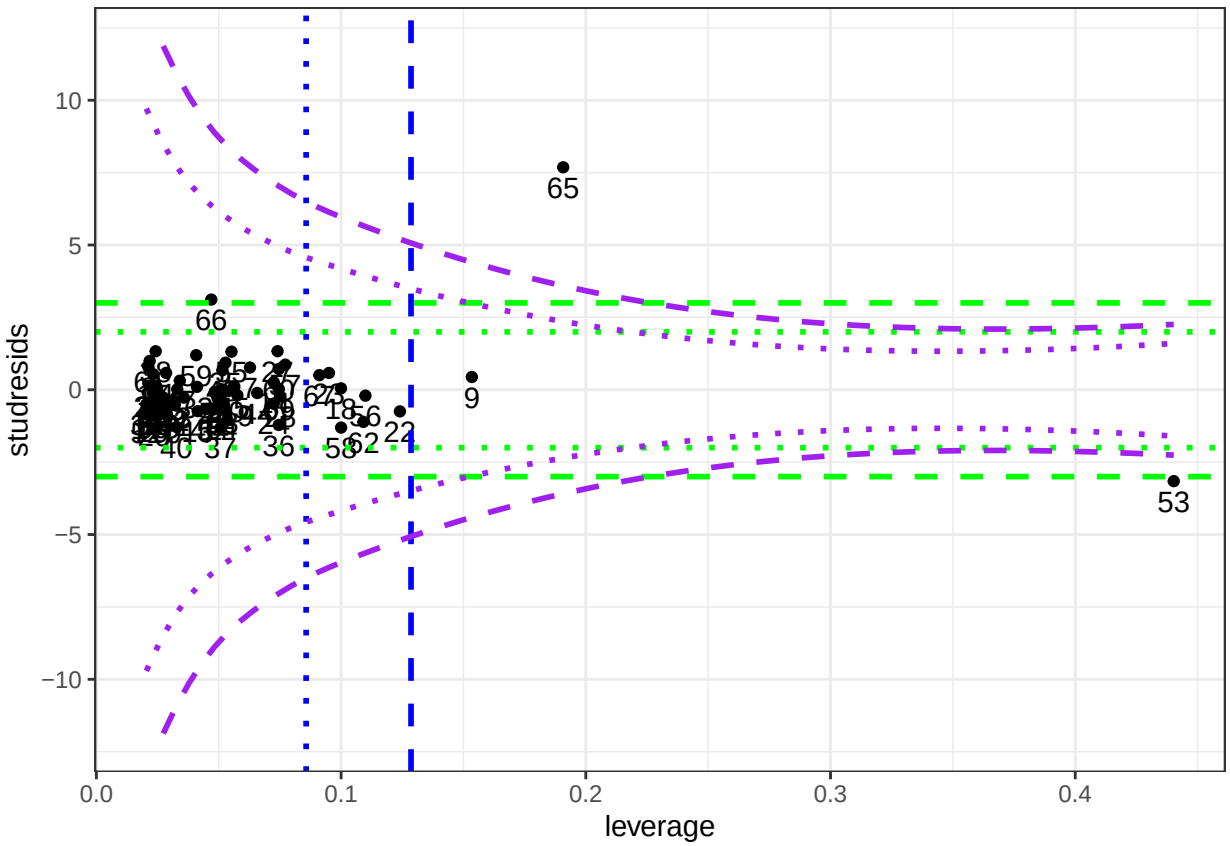
- pH, which measures the pH level of the soil (from 0-14),
- K, which measures the potassium level of the soil (in ppm), and
- the categorical variable *location* which can be one of three different spots ("OI" is short for Oak Island, "SI" is short for Smith Island, and "SM" is short for Snows Marsh).

Aerial biomass is called `BIO` and is measured with units gm^{-2}.

Consider the main effects model whose summary is shown on the next page. Below the model summary is a plot of the studentized residuals and leverage values for each data point based on this model. The heavy dashed lines represent "extreme" cutoffs and the short dashed lines represent "moderate" cutoffs for leverage (blue), studentized residuals (green), and Cook's distance (purple), respectively.

Use this information to answer questions 1-4.

```
##
## Call:
## lm(formula = BIO ~ Location + pH + K, data = biomass)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -811.15 -190.99  -37.70   96.78 1056.67
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   43.0122   299.1682   0.144   0.8864
## LocationSI  -497.6490   163.4252  -3.045   0.0041 **
## LocationSM    58.1814   131.6870   0.442   0.6610
## pH           414.9021    43.3381   9.574 6.68e-12 ***
## K             -1.0095     0.2324  -4.344 9.32e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 354.2 on 40 degrees of freedom
## Multiple R-squared:  0.7382, Adjusted R-squared:  0.712
## F-statistic: 28.19 on 4 and 40 DF,  p-value: 3.613e-11
```



Problem 1

Based on the output above, which of the following statements are supported (circle all that apply)

(a) the effect of changing pH level of the soil has a greater impact on the biomass than changing the potassium levels?

(b)

(c)

(d)

Problem 2

Which of the following represents the reduced model if we want to determine whether or not to include the categorical predictor variable? (circle all that apply)

(a)
(b)
(c)
(d)

Problem 3

Which of the following data points XXXXXX (circle all that apply)

(a) Observation 66

(b) Observation 65

(c) Observation 9

(d) Observation 53

Problem 4

Which of the following data points XXXXXX (circle all that apply)

(a) Observation 66

(b) Observation 65

(c) Observation 9

(d) Observation 53