Name:

SM339 · Applied Statistics Spring 2023 · Uhan

Quiz 6 – 3/22/2023

Instructions. You have 15 minutes to complete this quiz. You may use your plebe-issue TI-36X Pro calculator. You may not use any other materials.

Show all your work. To receive full credit, your solutions must be completely correct, sufficiently justified, and easy to follow.

Problem	Weight	Score
1a	0.5	
1b	0.5	
1c	1	
2a	1	
2b	1	
Total		/ 40

Problem 1. To study the fertility of fish stocked in Lake Ontario, researchers collected samples of female lake trout from Lake Ontario in September and November of 2002 through 2004. The data contains three variables for each fish: *Age*, the age of the fish (in years); *PctDM*, the percentage of total egg material that is solid (a measure of egg viability); and *Sept*, which is equal to 1 for fish collected in September, and 0 for those collected in November. The researchers fit the following model:

$$PctDM = \beta_0 + \beta_1 Age + \beta_2 Sept + \beta_3 (Age \times Sept) + \varepsilon \qquad \varepsilon \sim N(0, \sigma_\varepsilon^2).$$

The R summary output is below.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 39.39733    1.07376  36.691   <2e-16 ***
Age         -0.21821    0.08942  -2.440   0.0206 *
Sept        -1.27623    1.51190  -0.844   0.4051
Age:Sept    -0.02144    0.12782  -0.168   0.8679
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.242 on 31 degrees of freedom
Multiple R-squared:  0.4303,  Adjusted R-squared:  0.3752
F-statistic: 7.806 on 3 and 31 DF,  p-value: 0.000505
```

a. What is the estimated slope of *Age* for fish collected in September?

b. What is the estimated slope of *Age* for fish collected in November?

c. Is there a statistically significant difference in the slopes? Briefly explain. Circle the R output you used to make your decision. Assume a signficance level of 0.05.

Problem 2. A model was fit to a dataset containing measurements of the girth, height, and volume of timber in 31 felled black cherry trees. For this dataset, *Girth* is the diameter of the tree (in inches) measured at 4 ft 6 in above the ground. The *Height* of the tree is measured in feet and the *Volume* is measured in cubic feet. The R summary output is below.

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  69.39632   23.83575   2.911  0.00713 **
Girth        -5.85585    1.92134  -3.048  0.00511 **
Height       -1.29708    0.30984  -4.186  0.00027 ***
Girth:Height  0.13465    0.02438   5.524 7.48e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.709 on 27 degrees of freedom
Multiple R-squared:  0.9756,    Adjusted R-squared:  0.9728
F-statistic: 359.3 on  3 and 27 DF,  p-value: < 2.2e-16
```

a. Use the model output to estimate the volume of a tree with *Girth* = 12 inches and *Height* = 70 feet. Provide the answer rounded to three decimal places.

b. Why might it be surprising that $\hat{\beta}_1$ and $\hat{\beta}_2$ are negative? Give a reasonable explanation for why the signs are negative.