

Name _____

Spring 2021

STAT 587-3

Final exam
(53 points)

Instructions:

1. Full credit will be given only if you show your work.
2. The questions are not necessarily ordered from easiest to hardest.
3. You are allowed to use any resource except aid from another individual.
4. Aid from another individual will automatically earn you a 0.

1. Short calculations

- (a) If the correlation is -0.8, what is the coefficient of determination?
- (b) If the sample average of the response variable is 1, the sample average of the explanatory variable is -1, and the slope of the simple linear regression line is 2, what is the intercept?
- (c) If there are 20 observations, 3 explanatory variables (not including the intercept), and the residual sum of squares is 48, what is the MLE for the residual variance?

- (d) If a regression equation is

$$E[Y] = -3 + 2X_1 + 4X_2$$

what is the expected response when $X_1 = -2$ and $X_2 = 1$?

- (e) In simple linear regression with 10 observations, what is the t -critical value associated with a 95% confidence interval?

3. An experiment was performed to understand the effect of building material and span length on bridge strength. Answer the following questions based on the output below.

```
## Analysis of Variance Table
##
## Response: strength
##              Df Sum Sq Mean Sq F value Pr(>F)
## material      2  5.0792   2.5396   2.4708 0.12994
## factor(length) 3  6.8473   2.2824   2.2206 0.14304
## material:factor(length) 5 17.3723   3.4745   3.3804 0.04294 *
## Residuals    11 11.3062   1.0278
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(a) How many different materials were used?

(b) How many different lengths were used?

(c) Is the design complete?

(d) What is the estimate for σ^2 in this model?

(e) For the `factor(length)` line,

i. what is the reduced model?

ii. what is the full model?

iii. what model is used for the estimate of σ^2 ?

iv. using a significance level of 0.05, what decision would you make?

v. using a significance level of 0.05, what conclusion would you draw?

4. The file `imports.csv` contains the total value of imports (in billions) into the United States from Canada and Mexico by Air and Ship for each month from 2018 to the current. The variable `covid` indicates the months from March 2020 until now. Answer the following questions based on the model fit in this code.

```
months = c("January", "February", "March", "April", "May", "June", "July", "August",  
           "September", "October", "November", "December")  
d = read.csv("imports.csv")  
d$month = factor(d$month, levels = months)  
m = lm(log(value) ~ I(year-2021) + month + covid, data = d)
```

- (a) Provide the point estimate, 95% confidence interval, and interpretation for the following quantities:

i. Intercept

ii. Coefficient for 'I(year-2021)'

iii. Coefficient for the indicator of a COVID month

- (b) What proportion (not percentage) of the variability in the logarithm of value is explained by this model? (1 pt)
- (c) Provide an estimate and 95% **prediction** interval for the value (in billions of dollars) of imports in March 2021 (assuming we are still in COVID times).
- (d) Conduct an F-test for comparing the additive model to the model that does not include **Month**. Provide the F-statistic, numerator and denominator degrees of freedom, and p -value.
- (e) Conduct an F-test for comparing the additive model to the model that includes an interaction between month and covid. Provide the F-statistic, numerator and denominator degrees of freedom, and p -value. In addition, describe the support in the data for including this interaction.