Name:

SM339 · Applied Statistics Spring 2023 · Uhan

Exam 1 – Part 1 – 2/15/2023

Instructions

- This part is worth 60 points total. The exam (both parts) is worth 100 points total.
- You have until the end of the class period to complete this exam.
- You may use your plebe-issue TI-36X Pro calculator.
- You may refer to notes that you have handwritten, not to exceed one side of an 8.5" × 11" piece of paper.
- You may not use any other materials.
- **No collaboration allowed.** All work must be your own.
- **Show all your work.** To receive full credit, your solutions must be completely correct, sufficiently justified, and easy to follow.
- Keep this booklet intact.
- **Do not discuss the contents of this exam with any midshipmen until it is returned to you.**
- Copy and sign the honor statement below. This exam will not be graded without a signed honor statement.

> The Naval Service I am a part of is bound by honor and integrity. I will not compromise our values by giving or receiving unauthorized help on this exam.

Problem	Weight	Score
1a	0.4	
1b	0.4	
2a	0.4	
2b	0.4	
2c	0.4	
2d	0.4	
2e	0.4	
3a	0.4	
3b	0.4	
3c	0.4	
3d	0.4	
4a	0.4	
4b	0.4	
4c	0.4	
4d	0.4	
Total		/ 60

Name:

Problem 1. You are studying customer behavior on YouView, a new streaming video service. You want to estimate the average length of a video on YouView. You take a random sample of 17 videos and find that the sample mean video length is 12.74 minutes with a sample variance of 24.1.

a. Construct a 95% confidence interval for the mean length of videos on YouView. Provide your answer rounded to three decimal places.

   You may find some of the following R output helpful:

Code	Output
`qnorm(1 - 0.05/2, mean = 0, sd = 1)`	1.959964
`qnorm(1 - 0.95/2, mean = 0, sd = 1)`	0.062706
`qt(1 - 0.05/2, df = 16)`	0.063697
`qt(1 - 0.95/2, df = 16)`	2.119905

b. You write in your report that you are "95% confident" that the interval you found in part a contains the true mean video length. Briefly explain what this means.

Problem 2. The Blackwell Brewery has recently hired you as an analyst. In the past, the brewery produced an average of 725 thousand barrels of beer per day. You've been asked to test whether the average daily production has dropped significantly over the past year.

Using the records for the past year, based on 260 operating days, the sample mean of daily production is 712 thousand barrels per day, and the sample standard deviation is 118 thousand barrels per day.

Let μ be the average number of barrels of beer produced per day, in thousands.

Perform a hypothesis test by answering the following prompts.

a. State the null and alternative hypotheses.

b. Calculate the test statistic. Provide your answer to three decimal places.

c. Suppose the p-value is 0.038. Using a significance level of 0.05, do you reject or fail to reject the null hypothesis? Briefly explain.

d. Based on your answer to part c, state your conclusion about the average number of barrels of beer produced per day.

e. Suppose we fail to reject the null hypothesis when in fact it is false. What type of error have we committed?

Name:

Problem 3. Back to studying YouView. Suppose you are interested in predicting the number of *Likes* (in 1000s) a video on YouView has based on its *Length* (in minutes). You have collected this data for 43 videos. With this data, you fit a simple linear regression model. You obtain the following least squares line:

$$\widehat{Likes} = 17.626 + 1.24Length$$

a. Interpret the slope in the context of the problem.

b. Predict the number of likes for a video that is 5 minutes long.

c. Calculate the residual for a video that is 5 minutes long and has 20,194 likes.

d. Suppose the model results in a sum of squared error of 3,456. Calculate the size of a "typical" error.

Problem 4. Back to Blackwell Brewery. You have been asked to evaluate your competition. You have collected data on the *ABV* (alcohol by volume) and *Rating* from the Beer Advocate website for 44 beers. Using R, you fit a simple linear regression model with *Rating* as the response variable and *ABV* as the explanatory variable. The output is below. Note that some parts are missing.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  85.4707     1.6279   52.50   <2e-16 ***
ABV           0.2937     0.2623    1.12    0.269
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.72 on 42 degrees of freedom
Multiple R-squared:  ???????, Adjusted R-squared:  ????????
F-statistic: 1.254 on 1 and 42 DF,  p-value: 0.2692
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
	<int>	<dbl>	<dbl>	<dbl>	<dbl>
ABV	1	17.34882	17.34882	1.253707	0.2692081
Residuals	42	581.19664	13.83802	NA	NA

a. Write the fitted model.

b. Suppose you perform a t-test for simple linear regression slope. Using a significance level of 0.05, do you reject or fail to reject the null hypothesis? Briefly explain. Circle the R output you used to make your decision.

c. Based on your decision in part b, state your conclusion about the relationship between *Rating* and *ABV*.

d. What percent of the variability in *Rating* is explained by the model? Provide your answer rounded to three decimal places.