Name:

SM339 · Applied Statistics Spring 2024 · Uhan

Exam 1 – Part 1 – 2/13/2024

Instructions

- This part is worth 60 points total. The exam (both parts) is worth 100 points total.
- You have 50 minutes to complete this part of the 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.**

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	
3e	0.4	
4a	0.4	
4b	0.4	
5	0.4	
Total		/ 60

Problem 0. 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.

Signature:

Problem 1. You have just been hired as an operations research analyst at MedTech Diagnostics, a medical testing company. They are interested in the average patient waiting time at their Simplexville location over the past year. You take a random sample of 23 patients at the Simplexville location over the past year and find that the sample mean patient waiting time is 16.74 minutes with a sample variance of 23.9.

a. Construct a 95% confidence interval for the mean patient waiting time at the Simplexville location. Provide your answer to 3 decimal places.

   You may find some of the following R output helpful:

Code	Output
`qnorm(1 - 0.05/2, mean = 0, sd = 1)`	1.9600
`qnorm(1 - 0.95/2, mean = 0, sd = 1)`	0.0627
`qt(1 - 0.05/2, df = 22)`	2.0739
`qt(1 - 0.95/2, df = 22)`	0.0634

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

Name:

Problem 2. Continuing with the setting from Problem 1...

You learned from your predecessor at MedTech Diagnostics that two years ago, the average patient waiting time at the Simplexville location was 16 minutes. You're interested in whether the average patient waiting time at the Simplexville location over the past year has increased in a statistically significant way.

You use the same data as before: based on a random sample of 23 patients at the Simplexville location over the past year, the sample mean patient waiting time is 16.74 minutes with a sample variance of 23.9.

Perform a t-test for one population mean by answering the following prompts.

a. Let μ be the mean patient waiting time at the Simplexville location over the past year. State the null and alternative hypotheses.

b. Calculate the test statistic. Provide your answer to 3 decimal places.

c. Suppose the p-value is 0.2378. Using a significance level of 0.10, 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 patient waiting time at the Simplexville location.

e. Suppose we fail to reject the null hypothesis when in fact it is false. What type of error have we committed? No explanation necessary.

Problem 3. You are working with data for 36 breakfast cereals. Your data consists of two variables: *Calories* per serving and grams of *Fiber* per serving. You are interested in predicting *Calories* based on *Fiber*. With this data, you fit a simple linear regression model.

Below is output from `summary()` for your model:

```
Call:
lm(formula = Calories ~ Fiber, data = Cereal)

Residuals:
    Min      1Q  Median      3Q     Max
-17.363  -7.363  -4.005   1.413  55.801

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 117.3635     3.7216  31.536  < 2e-16 ***
Fiber        -4.3881     0.7358  -5.964  9.6e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.72 on 34 degrees of freedom
Multiple R-squared:  0.5112,  Adjusted R-squared:  0.4969
F-statistic: 35.56 on 1 and 34 DF,  p-value: 9.603e-07
```

a. Write the fitted model. Report all coefficients to 3 decimal places.

b. Interpret the slope in the context of the problem. Make sure to include units in your interpretation.

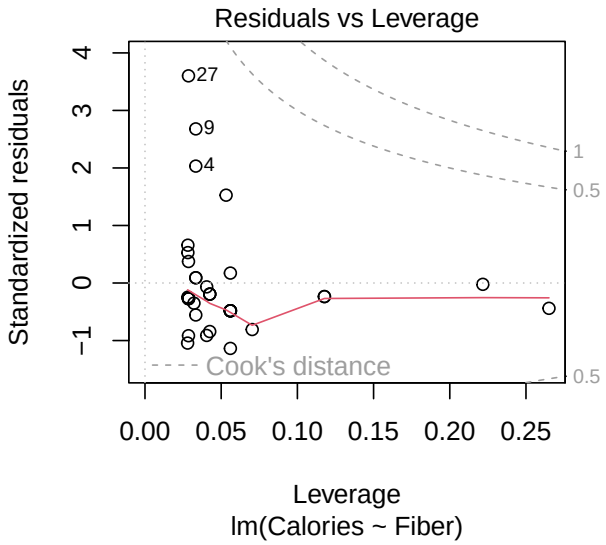
c. Predict the number of calories per serving for cereal that has 5 grams of fiber per serving. Provide your answer to 3 decimal places.

d. Compute the residual for a cereal that has 3 grams of fiber and 100 calories per serving. Provide your answer to 3 decimal places.

e. What is the size of a typical error when predicting carlories from fiber content?

Problem 4. Continuing with the setting from Problem 3...

Using R, you generate the diagnostic plot on the right for the model you fit in Problem 3.



a. Based on the rules of thumb we covered in class, circle and label the points that are classified as "very unusual" leverage points. Briefly explain your reasoning below.

b. Based on the rules of thumb we covered in class, circle and label the points that are classified as "very unusual" outliers. Briefly explain your reasoning below.

Problem 5. A capacitor was charged with a 9-volt battery and then a voltmeter recorded the voltage as the capacitor was discharged. Measurements were taken every 0.02 seconds. Your data consists of two variables: *Voltage* (in volts) and *Time* (in seconds). You are interested in predicting *Voltage* based on *Time*.

After exploring the data, you decide that applying a log transformation to *Voltage* is appropriate. You fit a simple linear regression model with $log(Voltage)$ as the response variable, and *Time* as the explanatory variable (assume *log* is the natural logarithm). Your fitted model is

$$\widehat{log(Voltage)} = 2.19 - 2.06 Time$$

Use your fitted model to predict *Voltage* when *Time* = 0.05 seconds.