Name: Feedback

SM339 · Applied Statistics Spring 2023 · Uhan

# Exam 2 – Part 1 – 4/5/2023

## 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 | 0.5 | |
| 1b | 0.5 | |
| 1c | 0.5 | |
| 1d | 0.5 | |
| 1e | 1.0 | |
| 2a | 0.5 | |
| 2b | 1.0 | |
| 3a | 0.5 | |
| 3b | 0.5 | |
| 3c | 0.5 | |
| Total | | / 60 |

Name:

**Problem 1.** Jellyfish Insights, a marketing research firm, is interested in understanding how various factors influence the sales of EcoBrew, a new line of environmentally-friendly coffee makers. They have collected data on the sales of the product (*Sales*, in thousands of dollars), as well as information on the price of the product (*Price*, in dollars), the amount of advertising spent (*Advertising*, in thousands of dollars), and the number of competitors (*Competitors*) in a number of different markets.

They use multiple linear regression to model the relationship between these variables, and using R to fit their model, they obtain the following output:

```
Call:
lm(formula = Sales ~ Price + Advertising + Competitors, data = product_data)

Residuals:
Min      1Q  Median      3Q     Max
-64.536 -20.872   0.740  19.670  67.115

Coefficients:
             Estimate Std. Error t-value  Pr(>|t|)
(Intercept) 400.72904   15.04895  26.625   < 2e-16 ***
Price        -0.05329    0.00654  -8.146  4.92e-14 ***
Advertising  22.05848    4.14992   5.313  2.33e-07 ***
Competitors -11.26764    3.79331  -2.967   0.00384 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 28.04 on 96 degrees of freedom
Multiple R-squared:  0.8371,  Adjusted R-squared:  0.8312
F-statistic: 144.5 on 3 and 96 DF,  p-value: < 2.2e-16
```

a. What is the fitted model?

See Example 1c from Lesson 13 Part 1 for a similar example.

Don't forget to use a $\widehat{\text{hat}}$ to signal that you are estimating the response variable in the fitted model.

b. Interpret the coefficient for *Advertising* in terms of its impact on the sales of the product.

See Example 1g from Lesson 13 Part 1 for a similar example.

Remember that the estimated coefficient of a predictor in a multiple linear regression model describes the relationship bewteen that predictor and the response variable on average and when all the other explanatory variables are held constant.

c. Predict the sales for a new market where the price is 100 dollars, the advertising budget is 50 thousand dollars, and there are 5 competitors. Provide your answer to 3 decimal places.

See Example 1e in Lesson 13 Part 1 for a similar problem. Also Problem 3ab in the Review Problems for Exam 2.

Be careful with units! Note that *Sales* and *Advertising* are given in thousands of dollars, while *Price* is given in dollars.

d. Determine the total number of observations.

See Example 2a in Lesson 14 Part 1 for a similar problem.

e. Perform a hypothesis test to determine whether the coefficient for *Competitors* is statistically significant at the 0.05 level. State all four steps.

See Example 2d in Lesson 14 Part 1 for a similar problem.

Name:

**Problem 2.** Rocket Motors, a used car dealership, is interested in the relationship between the type of fuel a car uses, its mileage, and its sale price. The dealership collected data on 100 used car sales transactions and recorded the type of fuel used in the variable *Hybrid* (0 = gasoline, 1 = hybrid), the *Mileage* of the car (in miles), and the *SalePrice* (in dollars). The analyst used R to fit the following model:

$$SalePrice = \beta_0 + \beta_1 Mileage + \beta_2 Hybrid + \beta_3(Mileage \times Hybrid) + \varepsilon \qquad \varepsilon \sim N(0, \sigma_\varepsilon^2)$$

Here is the output:

```
Call:
lm(formula = SalePrice ~ Mileage + Hybrid + Mileage:Hybrid, data = cars)

Residuals:
    Min      1Q  Median      3Q     Max
-5163.8 -1638.8  -204.9  1619.2  5488.2

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)     10129.4     1251.3   8.102 4.94e-13 ***
Mileage             1.5        0.3   5.260 2.17e-07 ***
Hybrid            505.4     1769.9   0.285    0.776
Mileage:Hybrid      0.4        0.4   1.087    0.280
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2321 on 96 degrees of freedom
Multiple R-squared:  0.4329,  Adjusted R-squared:  0.412
F-statistic: 20.87 on 3 and 96 DF,  p-value: 1.073e-10
```

a. What is the estimated slope for the relationship between mileage and sale price for hybrid cars?

See Section 4 of Lesson 17 Part 1 for the general idea.

b. Perform a hypothesis test to determine whether the slope for the relationship between mileage and sale price is different for gasoline and hybrid cars. Use a significance level of 0.05. State all 4 steps.

See Example 2b in Lesson 17 Part 2 and Problem 1e in the Lesson 17 Part 2 Exercises for similar examples.

**Problem 3.** You've been tasked with investigating the effectiveness of three different shooting training methods on the accuracy of soldiers when using a particular weapon. Your randomly assign 90 soldiers to one of three training groups – A, B, C – with 30 soldiers in each group. The soldiers complete the assigned training and then participate in a shooting accuracy test. You record the accuracy score for each soldier.

You fit a one-way ANOVA model to your data. The parameter estimates are:

$$\hat{\mu} = 74.718 \qquad \hat{\alpha}_A = -1.614 \quad \hat{\alpha}_B = -3.690 \quad \hat{\alpha}_C = 5.305$$

The ANOVA table is below:

```
                Df Sum Sq Mean Sq F value Pr(>F)
TrainingMethod   2   5012  2506.2   13.08 1.1e-05 ***
Residuals       87  18808   216.3
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

a. Is this an observational study or an experiment? Briefly explain.

See Problem 2a from the Review Problems for Exam 2 for a similar problem. Also see Section 3 of Lesson 24 Part 1 for the relevant details.

b. Compute $\hat{\mu}_A$, $\hat{\mu}_B$, and $\hat{\mu}_C$. Provide your answers to 3 decimal places.

See Example 1 in Lesson 25 for a similar problem.

c. Compute the effect size of training group C. Provide your answer to 3 decimal places.

See Example 2a in Lesson 25 for a simlar problem.