Name: Feedback

SM339 · Applied Statistics Spring 2024 · Uhan

Exam 2 – Part 1 – 4/2/2024

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	0.50	
1b	0.50	
1c	0.50	
1d	0.50	
1e	0.75	
1f	0.75	
2a	0.50	
2b	0.50	
2c	0.75	
3	0.75	
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. The Simplexville Snack Company would like to understand how various factors influence the sales of Primal Pretzels, their new line of artisanal soft pretzels. They have collected data on the sales of the product (*Sales*, in thousands of dollars), as well as the price of the product (*Price*, in dollars), the amount spent on advertising (*Advertising*, in thousands of dollars), and the number of competitors (*Competitors*) in a variety of different markets.

They used multiple linear regression to model the relationship between these variables. Using R to fit their model, they obtained the following summary output:

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

Residuals:
    Min      1Q  Median      3Q     Max
-343.01 -141.45   18.61   99.37  367.22

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1125.396    608.370   1.850   0.0721 .
Price       -116.992    121.421  -0.964   0.3414
Advertising   11.205      5.408   2.072   0.0451 *
Competitors   -5.592     17.060  -0.328   0.7449
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 179.3 on 38 degrees of freedom
Multiple R-squared:  0.1164,  Adjusted R-squared:  0.04666
F-statistic: 1.669 on 3 and 38 DF,  p-value: 0.1899
```

a. What is the fitted model? Report all coefficients to 3 decimal places.

See Example 1 in Lesson 16 Part 1 for a similar example.

b. Predict the sales for a new market where the price is 6 dollars, the advertising budget is 10 thousand dollars, and there are 5 competitors. Provide your answer in dollars.

See Example 1 in Lesson 15 Part 1 for a similar example. Be careful with units!

Name:

c. Determine the total number of markets used in this regression.

See Example 2 in Lesson 16 Part 1 for a similar example.

d. Interpret the coefficient for *Price* in terms of its impact on the sales of the product.

See Example 1 in Lesson 16 Part 1 for a similar example. Be careful with units!

Also see page 1 of Lesson 15 Part 1 for guidance on how to correctly interpret estimated coefficients of a multiple linear regression model; in particular, for the conditions under which the interpretation holds.

e. Perform an appropriate hypothesis test to determine whether the sales of the product is significantly associated with the amount spent on advertising. Use a significance level 0.05. In particular,

   i. State the full name of the hypothesis test you chose.
   ii. Perform all four steps of the hypothesis test.

See Example 1 in Lesson 16 Part 1 for a similar example.

f. Perform an appropriate hypothesis test to determine whether the model as a whole is effective. Use a significance level of 0.05. In particular,

   i. State the full name of the hypothesis test you chose.
   ii. Perform all four steps of the hypothesis test.

See Example 4 in Lesson 16 Part 1 for a similar example.

Problem 2. Markov Motors, a used car dealership, is interested in the relationship between the type of fuel a car uses, its mileage, and its sale price. Their data analyst has retrieved data on a number of 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 sale *Price* (in dollars). The analyst used R to fit the following model:

$$Price = \beta_0 + \beta_1 Hybrid + \beta_2 Mileage + \beta_3 (Hybrid \times Mileage) + \varepsilon \qquad \varepsilon \sim \text{iid } N(0, \sigma_\varepsilon^2)$$

Here is the summary output:

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

Residuals:
    Min      1Q  Median      3Q     Max
-9320.2 -2244.2   -23.7  2223.5  8467.5

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)     13054.5491  9.610e+02  13.584  < 2e-16 ***
Hybrid           2132.8014  1.342e+03   1.590  0.11313
Mileage            -0.07443  1.251e-02  -5.951 8.68e-09 ***
Hybrid:Mileage     0.04948  1.705e-02   2.902  0.00403 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3302 on 258 degrees of freedom
Multiple R-squared:  0.4727,  Adjusted R-squared:  0.4665
F-statistic: 77.08 on 3 and 258 DF,  p-value: < 2.2e-16
```

Name:

a. What is the estimated slope for the relationship between mileage and sale price for gasoline cars? Provide your answer to 3 decimal places.

See Section 3 of Lesson 18 Part 1 for the general idea. Also, see the feedback for Quiz 7 for hints on a similar problem.

b. What is the estimated slope for the relationship between mileage and sale price for hybrid cars? Provide your answer to 3 decimal places.

See Section 3 of Lesson 18 Part 1 for the general idea. Also, see the feedback for Quiz 7 for hints on a similar problem.

c. Perform an appropriate hypothesis test to determine whether the slope for the relationship between mileage and sale price is different for gasoline versus hybrid cars. Use a significance level of 0.05. In particular,

   i. State the full name of the hypothesis test you chose.
   ii. Perform all four steps of the hypothesis test.

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

Problem 3. You have been given data on a sample of 1200 births over the past year. For each birth, you have the birth *Weight* in ounces, the mother's *Age* at birth, and the mother's *Race*. The variable *Race* is categorical, and codes the mother's race as either Black, Hispanic, White, or Other. The first few rows of the data look like this:

Weight	*Age*	*Race*
111	32	White
116	32	Black
138	27	Black
136	27	Hispanic
121	25	Other
117	25	Other

Write a multiple linear regression model that predicts birth weight based on the mother's age and race. In particular, write the population-level model.

Make sure to define any additional variables you use in your model. Your model should be written using mathematical expressions; do not write your model using R code.

See Lesson 19 Part 1 for guidance on formulating multiple linear regression models with categorical predictors. Note that the problem asks for the population-level model.