

# ISyE 6414 Regression Analysis

## Homework 3

### Due September 15, 2022

Enterprise Inc. produces Clean, a brand of liquid laundry detergent. In order to study the relationship between demand and price, the company gathered the data for 30 sales periods (each sales period is a month) where:

- $y$ : demand for the large bottle of Clean (in hundreds of thousands of bottles) in the sales periods.
- $x$ : The price difference per bottle (in dollars) between the Clean price and the average price of competitors' similar detergents in the sales period.

By using the data set (6414\_HW3\_Clean.csv), answer Questions 1–9. Submit the relevant outputs and R codes.

1. Construct and submit a scatter plot of  $y$  versus  $x$ . Does a simple linear regression model seem appropriate here?
2. Construct and run a simple linear regression model using  $x$  as the predictor variable and  $y$  as the response variable. By using the output, identify and report  $\hat{\beta}_0$ ,  $\hat{\beta}_1$ ,  $\hat{\sigma}$ ,  $se(\hat{\beta}_0)$ , and  $se(\hat{\beta}_1)$ .
3. Calculate a 95% confidence interval for the slope  $\beta_1$  and give a statement in plain English interpreting this interval.
4. Conduct a hypothesis test using  $\alpha = 0.05$  to see whether or not the predictor  $x$  is statistically significant. State the null and alternative hypotheses, the test statistic, critical  $t$ -value, and your conclusion. Give a statement in plain English interpreting this result.
5. Identify the  $p$ -value for testing the intercept  $\beta_0$ . By using the  $p$ -value, state whether or not the intercept is statistically significant.
6. Identify the  $p$ -value for testing  $\beta_1$ . Using the  $p$ -value, determine whether we can reject  $H_0$  by setting  $\alpha$  to 0.10, 0.05, and 0.005. What do you conclude about the strength of the relationship between  $y$  and  $x$ ?

7. By using R, find a point estimate of and a 95% confidence interval for the *mean demand* when the price difference is 0.10.
8. By using R, find a point prediction of and a 95% prediction interval for the *actual demand* for Clean in an individual sales period when the price difference is 0.10. Is this interval wider than the confidence interval for the mean demand found in Question 7? Answer by comparing the half-lengths numerically.
9. Find a point estimate of and a 95% confidence interval for the mean demand when the price difference is 0.25. Is the half-length of this confidence interval identical to the one found in Question 7? If not, why? Answer by comparing the half-lengths numerically.

### Derivation Question

10. Suppose you know that the true  $\beta_0$  (intercept) is zero in a simple linear regression model, and you do not want to include the intercept in the model, i.e., the no-intercept model is  $y = \beta_1 x + \epsilon$ . Show that the least squares estimate of  $\beta_1$  for such a model that minimizes  $SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$  is

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$$