Solutions for Review Problems for Hypothesis Testing with Bivariate Regressions

1(a) To respond to the supposition made by the TTC's management, we want to specify it as a hypothesis, and then use our statistical results to test that hypothesis. If the TTC's management believes that there is no impact of the average number of delays on customer service ratings, then we can use the regression we estimated to test the following hypothesis:

$$H_0: \beta_1 = 0$$

$$H_a: \beta_1 \neq 0$$

In this case, the null hypothesis asserts that the population regression coefficient on "Average Number of Delays Per Ride" is zero, which means that there is no correlation between delays and customer service reviews. To test this null hypothesis, we could use one of two approaches: a t-test, or a confidence interval. Let's start with a t-test:

$$t = \frac{\hat{b}_1 - \beta_1}{se(\hat{b}_1)} = \frac{-1.5 - 0}{0.25} = -6$$

In this case, since the absolute value of the t-statistic is greater than 2, then we reject the null hypothesis at the 5% level of significance. This means that there is less than a 5% chance that the population coefficient is zero, given the sample regression results that we've generated. In practical terms, we would tell the TTC management that it is highly unlikely that delays have no impact on customer service ratings.

The other way to test this hypothesis is with a confidence interval. Here, the confidence interval around our sample regression coefficient is:

$$\hat{b}_1 \pm 2 * se\left(\hat{b}_1\right) = -1.5 \pm 2 * (0.25) = (-2, -1)$$

In this case, zero lies outside of our confidence interval, so we reject the null hypothesis at the 5% level of significance.

(b) If a rider experiences an average of one delay per ride, then we could use a conditional mean calculation from our regression to predict that their customer service rating would be:

$$Customer\ Service\ Rating = 8 - 1.5(1) = 6.5$$

So we would expect that this type of customer would provide a service rating of 6.5, but we also need to recognize that this is a prediction, so there's a degree of error that's associated with this guess. To represent this uncertainty, we need to create a confidence interval around this estimate:

$$6.5 \pm 2 * (Standard\ Error\ of\ the\ Regression) = 6.5 \pm 2 * (0.75) = (5{,}8)$$

So the 95% confidence interval ranges from 5 to 8.

With this confidence interval in hand, we can now address the question at hand: how would we react to the claim that a rider experiencing one delay (on average) would have a customer service rating of 6? We would treat this as a hypothesis:

$$H_0: Conditional\ mean = 6$$

$$H_a: Conditional\ mean \neq 6$$

In this case, we recognize that the null hypothesis (a conditional mean of 6) lies within the 95% confidence interval, so we would fail to reject the null hypothesis at the 5% level of significance. In plain language, we would say that our sample regression results can't dismiss the idea that a rider experiencing one delay (on average) will have a service rating of 6.

2(a) To assess the environmental group's claim, we will again try to conceive of the claim as a hypothesis. In particular, if they think that prices and sales are unrelated, then the group is predicting that there's a zero correlation between the two factors, which would imply that the coefficient on price is zero. In this case, the relevant hypotheses would be:

$$H_0: \beta_1 = 0$$

$$H_a: \beta_1 \neq 0$$

To test the null hypothesis, we can (again) use either a t-test or a confidence interval. Here's the t-test:

$$t = \frac{\hat{b}_1 - \beta_1}{se(\hat{b}_1)} = \frac{-0.02 - 0}{0.002} = -10$$

Since the absolute value of the t-statistic is greater than 2, then we reject the null hypothesis at the 5% level of significance. This means that there is less than a 5% chance that the population coefficient is zero, given the sample regression results that we've generated. In practical terms, we would tell the environmental group that it is highly unlikely that the sales of gas are unrelated to the price of gas.

The other way to test this hypothesis is with a confidence interval. Here, the confidence interval around our sample regression coefficient is:

$$\hat{b}_1 \pm 2 * se\left(\hat{b}_1\right) = -0.02 \pm 2 * (0.002) = (-0.024, -0.016)$$

In this case, zero lies outside of our confidence interval, so we reject the null hypothesis at the 5% level of significance.

(b) If the environmental group is predicting that the demand for gas would be 4500 Litres per gas station if the price of gas were 50 cents per Litre, then we can test this prediction with our regression. Since we would need to use both coefficients from our regression to predict gas sales, we know that we need to create a hypothesis test with the conditional mean from the regression:

$$H_0: Conditional\ mean = 4.5$$

$$H_a: Conditional\ mean \neq 4.5$$

To test this hypothesis, we need to determine the conditional mean that our sample regression would predict given a price of gas equal to 50 cents per Litre. This can be accomplished with a simple calculation:

$$Gas\ sales\ per\ day = \hat{b}_0 + \hat{b}_1(Price\ of\ gas\ on\ a\ given\ day)$$

$$= 5 - 0.02(50) = 4$$

Therefore, our regression equation predicts sales of 4000 Litres when the price is 50 cents per Litre. Furthermore, the confidence interval around this conditional mean is:

$$4 \pm 2 * (Standard\ Error\ of\ the\ Regression) = 4 \pm 2 * (0.1) = (3.8{,}4.2)$$

In this case, we can formally say that since the conditional mean of the null hypothesis (that a 50-cent price would induce sales of 4500 Litres) lies outside of our confidence interval, then we can reject the null hypothesis at the 5% level of significance. In plain language, we would say that our sample regression results can dismiss the idea that a 50-cent per Litre price of gas will induce sales of 4500 Litres.

3(a) To assess our view about the impact of commission rates on sales, we need to conceptualize this view as a hypothesis. Referring back to our sample regression, we can see that the appropriate hypothesis would be:

$$H_0: \beta_1 = 0$$

$$H_a: \beta_1 \neq 0$$

The reason that this null hypothesis is appropriate is because a coefficient of zero on "Commission rate of the salesperson" would be evident if there was no relationship between the commission and sales. Our ability (or inability) to reject this idea will address the key question we face: does a higher commission rate alter sales? We can test this hypothesis with either a t-test or a confidence interval. Here's the t-test:

$$t = \frac{\hat{b}_1 - \beta_1}{se(\hat{b}_1)} = \frac{200 - 0}{150} = 1.33$$

Since the absolute value of the t-statistic is smaller than 2, then we **<u>can't</u>** reject the null hypothesis at the 5% level of significance. This means that there is **<u>more</u>** than a 5% chance that the population coefficient is zero, given the sample regression results that we've generated. In practical terms, we would recognize that our sample regression results are consistent with the notion that commission rates don't alter sales.

The other way to test this hypothesis is with a confidence interval. Here, the confidence interval around our sample regression coefficient is:

$$\hat{b}_1 \pm 2 * se(\hat{b}_1) = 200 \pm 2 * (150) = (-100, 500)$$

In this case, zero lies **<u>inside</u>** of our confidence interval, so we fail to reject the null hypothesis at the 5% level of significance.

(b) To assess the view that a 4% commission rate would induce daily sales of $1500, we would begin by recognizing that this is a conditional mean prediction: based upon a particular value of the variable on the right-hand-side of our equation, we are making a guess about the variable on the left-hand-side of the equation. To examine this more formally, we can use our sample regression to see what we would predict about sales under a 4% commission rate:

$$Daily\ sales\ by\ the\ salesperson = 1000 + 200(4) = 1800$$

Therefore, our regression equation predicts sales of $1800 when the commission is 4%. Furthermore, the confidence interval around this conditional mean is:

$$1800 \pm 2 * (Standard\ Error\ of\ the\ Regression) = 1800 \pm 2 * (50) = (1700{,}1900)$$

In this case, we can formally say that since a sales level of $1500 lies outside of our confidence interval, then we can reject the null hypothesis at the 5% level of significance. In plain language, we would say that our sample regression results can dismiss the idea that a 4% commission rate would induce sales of $1800.