

In-Class Team Project 3

ECO391

Name Your Team: _____

Team Members:

Due April 24, Wednesday.

Worth 50 Points

Part I

The distribution center for an online retailer has been experiencing quite a few “pick errors” (i.e., retrieving the wrong item). Although the warehouse manager thinks most errors are due to inexperienced workers, she believes that a training program also may help to reduce them. Before sending all employees to train, she examines data from a pilot study of 30 employees (half attended the training and the other half did not) to decide whether the training program will be effective. The file **Chapter 17 Pick_Errors** contains the data.

1. Specify Regression Model and Indicate the Expected Sign of the Coefficients

Specify (write out) your regression model (use all the variables):

Indicate the expected sign for each regression coefficient -- positive, negative, or ambiguous -- and provide a brief explanation supporting your expectation.

How does including a dummy variable in regression affect the regression line?
Demonstrate using a graph of the above regression.

Does the inclusion of the dummy variable affect both the slope and the intercept?

2. Specify another regression model that includes an interaction variable.

In your model, include an interaction of years of experience and training and specify the model in the space below.

3. Estimate both models on Excel and note down the goodness-of-fit measures for both the models.

4. Which model is a better fit? How did you come to this conclusion?

5. Regression Results

Using your model of choice, indicate which variables are statistically significant at the 5% significance level. Show working.

Using your model of choice, indicate which variables are statistically significant at the 10% significance level. Show working.

Are the signs on all the coefficients as you expected?

Interpret the coefficients on Experience and Training.

Interpret the coefficient on the interaction variable.

How does inclusion of a dummy variable in regression along with an interaction variable affect the regression line? Demonstrate using a graph of the above regression.

Does the inclusion of the dummy variable affect both the slope and the intercept?

6. **Use the chosen model to predict the number of pick errors for an employee with 10 years of experience who attended the training program, and for an employee with 20 years of experience who did not attend the training program. Show working.**

What does this tell you of the importance of experience for employees working for this online retailer?

7. **Consider (hypothetically) that, in order to train their employees, the online retailer is considering four different types of training programs each conducted by a different training center. Each employee can attend only one of these programs. The online retailer, therefore, gathers attendance data on these four types of training programs and the corresponding numbers of pick errors.**

Since there are four categories of training programs, we can have four dummy variables. However, in this case, how many dummy variables can we include in our regression?

The category that we exclude from our regression is called a _____ category.

We exclude one of categories from our regression in order to avoid a _____.

What common violation is this type of problem related to? How?

Part II

Baseball has always been a favorite pastime in America and is rife with statistics and theories. In a recent paper, researchers showed that major league players who have nicknames live 2.5 years longer than those without them (*The Wall Street Journal*, July 16, 2009). You do not believe in this result and decide to collect data on the lifespan of 30 baseball players along with a nickname variable that equals 1 if the player had a nickname, and 0 otherwise. Use dataset Chapter 10 Nicknames.

1. Create two subsamples consisting of players with and without nicknames. Calculate the average longevity for each subsample.
2. Specify the hypotheses to contradict the claim made by the researchers.
3. Use Excel to conduct a hypothesis test of the difference in population means between baseball players who have nicknames versus those who do not. State the conclusion of the test using a 5% level of significance. Assume the population variances are unknown but equal.