Introduction to Data Analytics 1

Enter Your Name

2019-08-17

# Part 1: Variables, Hypothesis, Designs

*Title:* Offshore outsourcing: Its advantages, disadvantages, and effect on the American economy

*Abstract*: The United States has trained some of the world’s best computer programmers and technology experts. Despite all of this training, many businesses do not have a full understanding of information technology. As the importance of technology in the business world grows, many companies are wasting money on extensive technology projects. When problems arise, they expect that further investment will solve these issues. To prevent such problems, many companies have begun to outsource these functions in an effort to reduce costs and improve performance. The majority of these outsourced information technology and call center jobs are going to low-wage countries, such as India and China where English-speaking college graduates are being hired at substantially lower wages. The purpose of this study is to evaluate the positive and negative aspects of offshore outsourcing with a focus on the outsourcing markets in India and China, arguably the two most popular destinations for outsourcers. The cost savings associated with offshore outsourcing will be evaluated in relation to the security risks and other weakness of offshore outsourcing. In addition, an analysis of the number of jobs sent overseas versus the number of jobs created in the United States will be used to assess the effects that outsourcing is having on the American economy and job market. Finally, the value of jobs lost from the American economy will be compared to the value of jobs created. The goal of these analyses is to create a clear picture of this increasingly popular business strategy.

Answer the following questions about the abstract above:

1. What is a potential hypothesis of the researchers?
2. What is one of the independent variables?
   1. What type of variable is the independent variable?
3. What is one of the dependent variables?
   1. What type of variable is the dependent variable?
4. What might cause some measurement error in this experiment?
5. What type of research design is the experiment?
   1. Why?
6. How might you measure the reliability of your dependent variable?
7. Is this study ecologically valid?
8. Can this study claim cause and effect?
   1. Why/why not?
9. What type of data collection did the researchers use (please note that #5 is a different question)?

# Part 2: Use the assessment scores dataset (03\_lab.csv) to answer these questions.

The provided dataset includes the following information created to match the abstract:

* Jobs: the percent of outsourced jobs for a call center.
* Cost: one calculation of the cost savings for the business.
* Cost2: a separate way to calculate cost savings for the business.
* Business: an ID number for each business.
* Where: where the jobs were outsourced to.

Calculate the following information:

1. Create a frequency table of the percent of outsourced jobs.
2. Create histograms of the two types of cost savings. You will want to add the breaks argument to the hist() function. This argument adds more bars to the histogram, which makes it easier to answer the following questions:

hist(dataset$column, breaks = 15)

15 is a great number to pick, but it can be any number. For this assignment, try 15 to see a medium number of bars.

1. Examine these histograms to answer the following questions:
   1. Which cost savings appears the most normal?
   2. Which cost savings data is multimodal?
   3. Which cost savings data looks the most skewed (and in which direction positive or negative)?
   4. Which cost savings data looks the most kurtotic?
2. Calculate the z-scores for each cost savings, so they are all on the same scale.
3. How many of the cost saving scores were more extreme than 95% of the data (i.e., number of z-scores at a *p* < .05)?
   1. Cost Savings 1:
   2. Cost Savings 2:
4. Which business had:
   1. the highest cost savings?
   2. the the lowest cost savings?
   3. Use both cost savings columns and find the ID number of the business with the lowest and highest z-score.