Real Econometrics - Chapter 1

True and False Questions:

1. **True** or False: The dependent variable is the outcome of interest.
2. True or **False**: X1 is the slope coefficient in the following model $Y_i = B_0 + B_1X_{1i} + e_i$.
3. True or **False**: $Y_i=B_0+10X_{1i}+e_i$, a 10 unit increase in X1 is expected to lead to a 1 unit increase in Y.
4. **True** or False: An independent variable is endogenous if changes in it are related to factors in the error term.
5. True or **False**: In randomized experiments we determine the value of the dependent variable via a random process.

Multiple Choice Questions:

1. What is the goal of a randomized experiment?
   a. **To create exogeneity.**
   b. To create endogeneity.
   c. To deal with problems of external validity.
   d. To increase sample size.

2. Which of the following are fundamental challenges when trying to infer relationships from data?
   a. Exogeneity
   b. **Endogeneity**
   c. Correlation
   d. Math

3. An independent variable is endogenous if:
   a. It is correlated with another independent variable in the model.
   b. **It is correlated with the error term.**
   c. It is not correlated with the error term.
   d. If it is correlated with Y.

4. Assume that we are looking at the effect of education on wages, which of the following factors could most likely lead to endogeneity?
   a. Height
   b. Religion
   c. Number of siblings
   d. **Intelligence**

5. Assume that we trying to explain infant mortality as a function of GDP per capita in a sample of countries from around the world. Which of the following factors could most likely lead to endogeneity?
   a. Whether the country has been at war
   b. Population size
   c. (Geographic) size of country

d. Use of proportional representation system

6. Given $Income_i = 10 + 1.5Age_i + e_i$, interpret the model
   a. A one year increase in age is expected to lead to a 1.5 increase in income
   b. A 1.5 increase in age is expected to lead to a 1 increase in income
   c. A 1.5 increase in income is expected to lead to a 1 unit increase in age
   d. Income is 10

7. Given $GPA_i=1.0+0.001SAT_i+e_i$ what is the expected GPA of someone with an SAT score of 2000
   a. 1.0
   b. 1.001
   c. 2000
   d. 3.0

8. When an independent variable is exogenous
   a. It is correlated with the error term
   b. It is not correlated with the error term
   c. It is correlated with the slope
   d. It is not correlated with the slope

9. Suppose we are trying to understand the effect of maternal smoking on the birthweight of newborn babies. Which of the following is not a likely source of endogeneity?
   a. Mother's health consciousness
   b. Tax rates on cigarettes
   c. Mother's education
   d. Father's smoking habits

10. The error term
    a. Contains all other factors that affect the dependent variable.
    b. Contains all other factors that affect the independent variable.
    c. Contains all other factors that affect the intercept.
    d. Is a random variable with no relation to other variables.

Short Answer:

1. List and explain the three fundamental challenges in statistics.
   Answer: Randomness, endogeneity and specification. Randomness means that we might observe a relationship between two variables that is due to chance rather than causation. Endogeneity means that what looks like a relationship between two variables may actually be due to a third (or fourth or fifth…) variable that is associated with both variables. Specification refers to identifying the variables and functional form of an appropriate model.

2. Describe the basic process behind an experiment.
   Answer: Individuals are divided into treatment and control groups. After treatment is administered, compare the results between the two groups in order to determine if

the treatment had an effect. Because the treatment is random, it will not be correlated with anything in the error term.

3. What is internal and external validity?

   Answer: Internal validity refers to whether an inference is biased while external validity refers to whether an inference can be applied more generally.

4. Imagine that you are looking at the relationship between income and level of education, list some of the factors that could lead to endogeneity.

   Answer: Parents income, type of school (private/public), intelligence, height, age.

5. Imagine that you are looking at the relationship between political affiliation and religion, list some of the factors that could lead to endogeneity.

   Answer: parents political affiliation, city/state, income, marital status, age, level of education.

Real Stats – Chapter 2

True or False Questions:

1. **True** or False: The standard deviation measures how widely dispersed the values of the observations are.
2. **True** or False: If an analysis cannot be replicated, it cannot be trusted.
3. **True** or False: Replication files are used by researches to verify another researcher's projects results.
4. **True** or False: Statistical projects should document both the data and the methods used to arrive at the conclusion.
5. True or **False**: The statistical package R is easier to use than Stata.

Short Answer Question:

1. Please list the descriptive statistics that are used to better understand the data.
   Answer: Mean, number of observations, standard deviation, minimum, maximum.

2. Explain the importance of keeping replication files and properly documenting the data/research.
   Answer: Important because it allows others to check the robustness of the results and therefore helps verify if a study was done in an accurate manner.

3. Describe/list some of the information that generally goes into a codebook.
   Answer: Type of data, source of data, variable information, and individual/organization that collected the data.

4. What is the purpose of plotting the data?
   Answer: A useful first step in order to help identify any patterns or anomalies in the data prior to moving forward with further analysis.

5. Compare and contrasts the basic benefits/drawbacks of Stata and R.
   Answer: Stata is more user friendly in the beginning, however once you get a hang of it, R could be easier to use because the coding is more direct and isn't hidden behind a backend/GUI.