Welcome to

Linear Statistical Models 1

CHAPTER 0

What is a Statistical Model?

Introduction:

Statistical models are useful for answering all kinds of questions.

Example 1:

Can we use the number of miles that a used car has been driven to predict the price that is being asked for the car?

- How much less can we expect to pay for each additional 1000 miles that the car has been driven?
- Would it be better to base our price predictions on the age of the car in years, rather than its mileage?
- Is it helpful to consider both age and mileage, or do we learn roughly as much about price by considering only one of these?
- Would the impact of mileage on the predicted price be different for a Honda as opposed to a Porsche?

These questions reveal several purpose of statistical modeling:

a. Making Predictions.

b. Understanding relationship.

c. Assessing differences.

Fundamental Terminology:

Data: any collection of numbers, characters, images, or other items that provide information about something.

Population: is the entire collection of objects or outcomes about which information is sought. (N)

- Parameter: a number, usually unknown, that applies to the whole population.

Sample: is a subset of a population, containing the objects or outcomes that are actually observed.

- Statistic: a number computed from the sample.

Observations: are the people, objective, or cases which data are recorded.

Example 2:

Suppose that there exists 7 million college students in the United States today. And, let's assume the average GPA of all these college students is 3.0. If we select 150 college students to compute the average GPA.

1. What is the population?

2. What is the sample?

3. What is the observation?

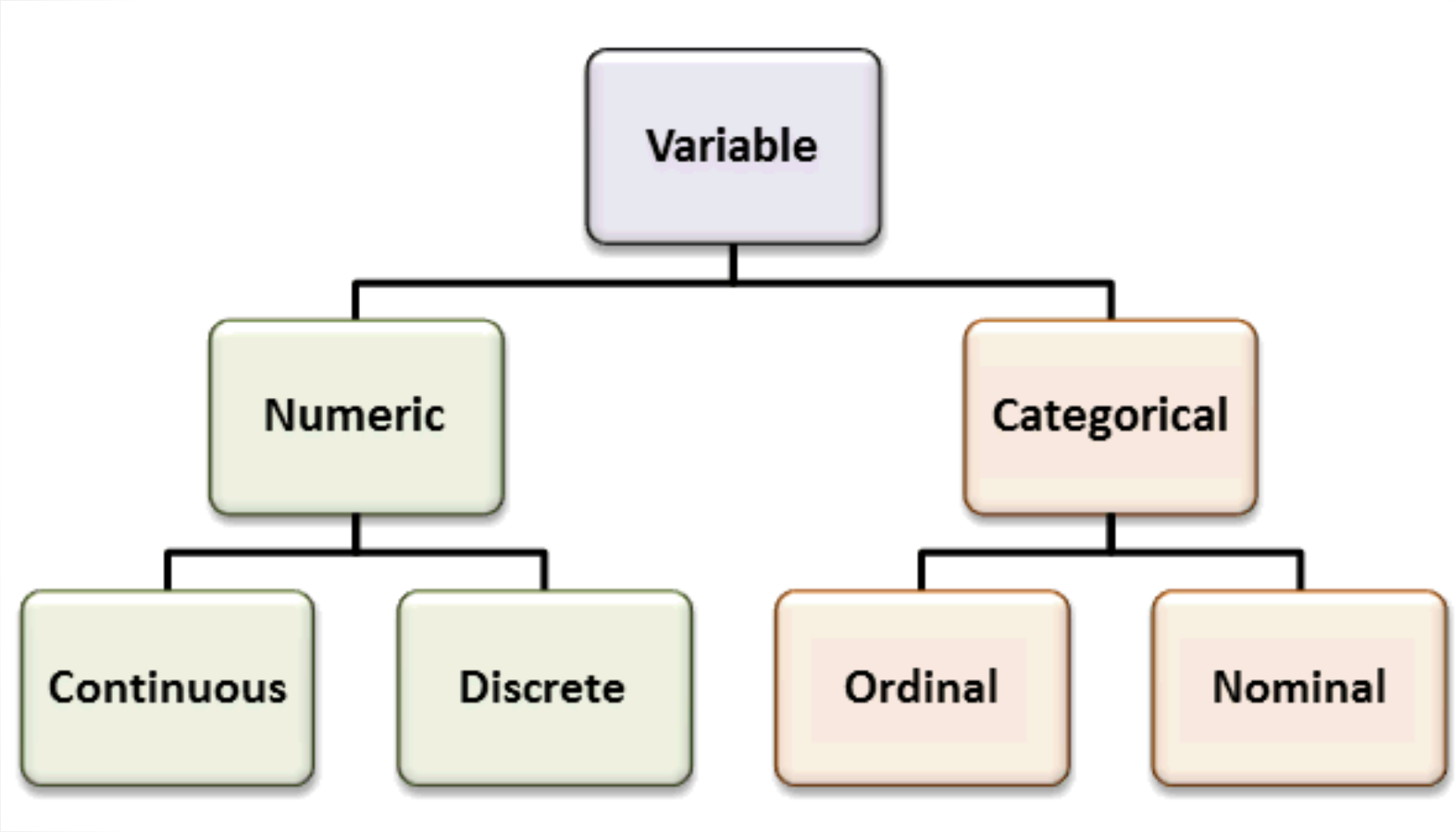
Variables: are the characteristic that are measured or recorded about each observational unit.

Variables can be classified into two types:

1. **Numerical (Quantitative) variable:** records numbers about the observational units.

2. **Categorical (Qualitative) variable:** records a category designation about the observational units.

   - If there are only two possible categories, the variable is also said to be **binary**.


Variable
Numeric
Categorical
Continuous
Discrete
Ordinal
Nominal

Another important consideration is the role played by each variable in the study.

- Response (dependent, outcome) variable: measures the outcome of interest.

- Explanatory (independent, predictor) variable: is one may explain or may cause differences in the response variable.

Response or Explanatory?

Example 4:

In example 1:

- The price of the car is . . . . . Variable.
- The mileage is . . . . . Variable.
- The age is . . . . . Variable.

The reason that these classifications are important is that the choice of the appropriate analysis procedure depends on the type of the variables in the study and their roles.

Response	Explanatory	Procedure	Chapter
Quantitative	Single quantitative	Simple linear regression	1, 2
Quantitative	Multiple quantitative	Multiple linear regression	3, 4
Categorical	Single quantitative	Simple logistic regression	9
Categorical	Multiple quantitative	Multiple logistic regression	10, 11

Types of Studies:

1. Controlled experiment: the researchers manipulates the explanatory variable by assigning the explanatory group or value to observational units.
2. Observational study: the researchers do not assign the explanatory variable but rather passive observe and record its information.

Statistical Model:

We will employ a four-step process for statistical modeling throughout this class.

1. Choose a form for the model.
2. Fit that model to the data.
3. Assess how well the model describes the data.
4. Use the model to address the question that motivated collecting the data in the first place.

Four-Step Process:

A statistical model can be written as:

$$DATA = MODEL + ERROR$$

or as

$$Y = f(X) + \epsilon$$

where Y is the variable being modeled (**response** variable), X is the variable used to do modeling (**explanatory** variable), and f is a function, ϵ is called the **error**, meaning the part of the response variable Y that remains unexplained after considering the predictor X.

Example 6: In a study reported in the *Journal of Preventative Medicine*, 85 nutrition experts were asked to scoop themselves as much ice cream as they wanted. Some of them were **randomly** given a large bowl (34 ounces) , and the others were given a smaller bowl (17 ounces). Similarly, some were **randomly** given a large spoon (2 ounces) and the others were given a small spoon (2 ounces). Researchers then recorded how much ice cream each subject scooped for him- or herself. Their conjecture was that those given a large bowl would tend to scoop more ice cream, as would those given a large spoon.

a. Identify the observational units in this study.

b. Is this an observational study or a controlled experiment? Explain how you know.

c. Identify the response variable in this study, and classify it as quantitative or categorical.

d. Identify the explanatory variable(s) in this study, and classify it (them) as quantitative or categorical.

Reading Assignment

Read chapter 0.