

## Chapter 1: The Nature of Probability and Statistics

*Statistics* is the science of conducting studies to collect, organize, summarize, and present data with the possible goal to analyze and draw conclusions relevant to a population

*Descriptive statistics* consists of the collection, organization, summarization, and presentation of data

*Inferential statistics* consists of analyzing the data collected in a sample to draw conclusions relevant to a population

*Data* is a collection of values, measurements, or observations

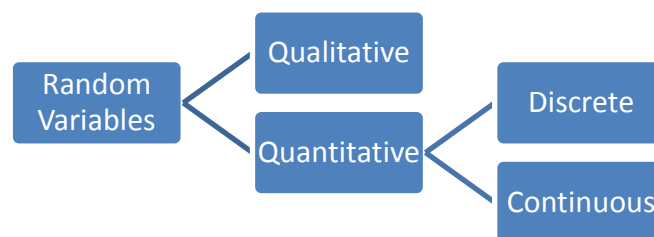
*Note: the singular form for data is datum*

A *random variable* is any characteristic of a value, measurement, or observation; the value of a random variable is determined by chance

A *population* consists of all members under study

A *sample* is a group of members selected from a population

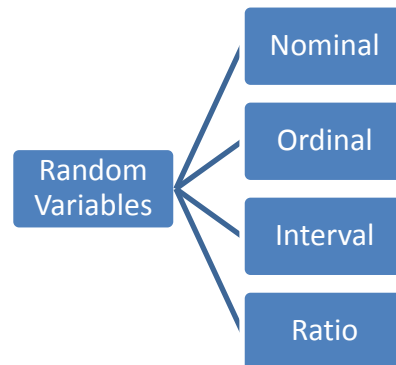
Random variables can be *classified* in two ways:



A *qualitative* random variable assumes values that are categorical

A *discrete* quantitative random variable assumes values that can be counted

A *continuous* quantitative random variable assumes values that cannot be counted



A *nominal* random variable assumes values that are mutually exclusive and categorical without an order or ranking

An *ordinal* random variable assumes values that are mutually exclusive and categorical but can be ranked

An *interval* random variable assumes values that are mutually exclusive, that can be ranked, and precise differences between units of measure exist; however, a meaningful zero does not exist

A *ratio* random variable assumes values that are mutually exclusive, that can be ranked, and precise differences between units of measure exist; a true zero exists

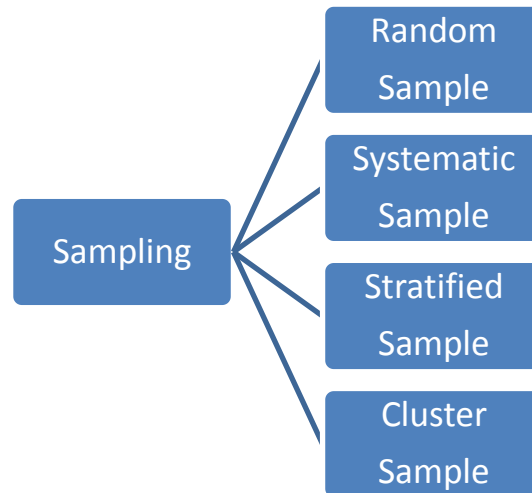
*Population* is the entire group of members that we want information about

*Sample* is a part of the population that is examined to gather information about the population

*Note: generally one can say that a larger sample-size yields more accurate results*

A *bias* favors certain outcomes

*Note: a sampling bias occurs when some members of the population are more likely to be included than others (over-coverage), some members of the population are left out of the sample (under-coverage), or when a member of the sample does not respond (non-response)*



*Note: voluntary response samples and convenience samples are potentially biased and should be avoided in elementary statistics. Voluntary response samples consist of persons who choose themselves by responding to a general appeal; most likely to respond are people with strong opinions. In convenience samples the members are chosen because they are conveniently reached*

The *independent* or *explanatory* variable is the variable that is manipulated by the researcher in an experimental study

The *dependent* variable is the variable that is influenced by the change in the independent variable

A *confounding* variable is a variable that that influences the dependent variable but is not recognized as such when the experimental study was designed

In an *observational* study the researcher merely observes and draws conclusions based on these observations

In an *experimental* study the researcher manipulates the independent variables and tries to determine how this manipulation influences the dependent variable

Be aware of misuses in statistics:

*Small samples:* "Three out of four college students scored an "A" in this test" is a meaningless statement if only one sample of size four was studied.

*Biased samples:* often the members of a college class are used in samples in the classroom. College students, however, are only an unbiased sample for the population of college students, not for the entire population of kids between 18 and 26 year. For sure, the members of one college class are not randomly selected college students from the entire population of college students.

*Using the wrong quantity:* mean and median are different quantities and may have different values in the same problem.

*Using different units:* "Expenditures increased by only 4% during my administration" and "Expenditures increased by a whopping \$6,000,000 during her administration" sound very different but may actually mean the same thing.

*Detached statistics:* "This yogurt has 33 percent more nutrients" is a meaningless statistic since no comparison is made.

*Implied connections between variables:* "Eating fish may increase your health," "Studies suggest you will lose weight," and "This medicine will decrease cholesterol in some people" do not state facts that can be applied to a population.

*Faulty and leading survey questions:* survey questions should be clear and concise and should not make you feel bad answering them like "Do you vote for the animals shelter's "no kill policy" since animal shelters in surrounding counties euthanize already 30,000 animals per year?"

*Flawed interpretations:* "Only 13% of Americans think the country spends too much on helping the poor" but "44% of Americans think the country spends too much on welfare."