MgtOp 215 **Chapter 1** **Dr. Ahn**

Statistics:
1) commonly means a collection of numerical facts or data.
2) a science of collecting, presenting, analyzing, and interpreting numerical data.

The word *statistics* originally meant the collection of facts useful to the person involved with the affairs of state.

Why do we learn statistics?
1) to have a better feel for statistics: what it is; how and when to apply statistical techniques to decision making situations; how to interpret the results.
2) to develop analytical skills and critical thinking.

The field of statistics
1) **descriptive statistics**: organizing and summarizing numerical data in order to describe the various features of the data (Chaps. 2 and 3).
2) **inferential statistics**: procedures and methods that can be used to make inferences (or generalization), predictions, and decisions about the population using the information contained in a sample (Chaps. 8 and on).

Probability:
1) commonly means a chance.
2) a mathematical means of studying uncertainty, and used as a tool for inferential statistics (Chaps. 4, 5, 6, and 7).

Variable: a characteristic of items or individuals.

Population: a collection of all the elements under study (about which we are trying to draw a conclusion) or a collection of all the measurements under study.

Sample: a part of the population.

Parameter: a numerical descriptive measure that represents a characteristic of the population.

Statistic: a numerical descriptive measure from a sample.

"Statistics is a science of getting cheaper but good-quality information, although not perfect, about the population, especially about the parameter, instead of getting perfect information which tends to be infeasible and very expensive."

"It is better to be roughly right than to be precisely wrong." *Maynard Keynes*

Census: evaluation of each and every unit (element) in the population under study; the **census data** contain complete (and perfect) information about the population.

Difficulties of collecting the census data:
1) expensive (especially when the population is large),
2) time consuming,
3) could be imperfect as they are more prone to human errors and due to lies,
4) may not be feasible (especially when measurement requires destruction of unit).

A "representative" sample contains the relevant characteristic of the population, even though the information contained in the sample is imperfect.

Simple random sample: a sample drawn such a way that each element of the population has the equal chance of being selected. Three ways of taking a simple random sample: 1) draw tickets out of a hat; 2) use random number tables (an example is on Table E.1 on pp. 742-743 and discussion pp.22-23); 3) use MS Excel: **Formulas > Insert Function (f_x) >Math & Trig >RANDBETWEEN**.

Hard Copy (HC) Homework: Visit the companion Web-site (www.pearsonhighered.com/levine) as described in Appendix C.1 on p. 729 and down load the Excel data file "Retirement Funds." Then randomly select 15 retirement funds and list their fund numbers along with Type and 1 Year Return %. (Since the names are between rows 2 and 317, it is convenient to draw numbers between 2 and 317, inclusive.)

Data are the different values associated with variables, and can be thought of the information needed to help us make informed decision.

Sources of data:

Primary: the data collector using the data for analysis
Secondary: the person analyzing the data is not the data collector

Types of data:

qualitative: nonnumeric and categorical
quantitative: numeric
discrete: assumes only specific points on a scale
continuous: assumes all values within an interval

Levels of measurements:

The **nominal scale** applies to data that are divided into different categories, and these categories are used for identification purposes, e.g., female=0, male=1; ISP: Verizon=1, AT&T=2, Sprint=3, T-Mobile=4, Other=5.

The **ordinal scale** applies to data that are divided into different categories that can be ranked, e.g., bad=0, fair=1, good=2; bond ratings: AAA, AA, A, BBB, BB; Student grades: A, B, C, D, F, etc.

The **interval scale** applies to data that can be ranked and for which the difference between two values can be calculated and interpreted and do not involve a true zero point, e.g., temperature, calendar time, SAT score.

The **ratio scale** applies to data that can be ranked and for which the ratio between two values can be calculated and interpreted, e.g., income, weight.

MyStatLab (MSL) Homework: 1.1, 1.3, 1.5, 1.7, 1.8 on p. 17 (These are labeled in MyStatLab as 1.1-2.1,1.1-2.3, 1.1-2.5, 1.1-2.7, 1.1-2.8)

Cross sectional data are collected on different elements at the same point in time or for the same period of time.

Time series data are collected on the same element for the same variable at different points in time or different periods of time. Usually line graphs or bar graphs are used.