Mini-Lecture 1.1
Introduction to the Practice of Statistics

Objectives
1. Define statistics and statistical thinking
2. Explain the process of statistics
3. Distinguish between qualitative and quantitative variables
4. Distinguish between discrete and continuous variables
5. Determine the level of measurement of a variable

Examples
1. Researcher Elisabeth Kvaavik and others studied factors that affect the eating habits of adults in their mid-thirties. (Source: Kvaavik E, et. Al. Psychological explanatorys of eating habits among adults in their mid-30's (2005) International Journal of Behavioral Nutrition and Physical Activity (2)9.) Classify each of the following variables considered in the study as qualitative or quantitative. Determine whether the quantitative variables are discrete or continuous. In addition, specify the level of measurement for each variable.
   a. Nationality (Qualitative; Nominal)
   b. Class rank (i.e. freshman, sophomore, etc.) (Qualitative; Ordinal)
   c. Number of children (Quantitative, Discrete; Ratio)
   d. Household income in the previous year (Quantitative, Discrete but frequently treated as continuous; Ratio)
   e. Education level (Qualitative; Ordinal)
   f. Today's high temperature (Quantitative, Continuous; Interval)
   g. Daily intake of whole grains (measured in grams per day) (Quantitative, Continuous; Ratio)

2. A study was conducted to assess school eating patterns in high schools in the United States. The study analyzed the impact of vending machines and school policies on student food consumption. A total of 1088 students in 20 schools were surveyed. (Source: Neumark-Sztainer D, French SA, Hannan PJ, Story M and Fulkerson JA (2005) School lunch and snacking patterns among high school students: associations with school food environment and policies. International Journal of Behavioral Nutrition and Physical Activity 2005, (2)14.) Classify each of the following variables considered in the study as qualitative or quantitative. Determine whether the quantitative variables are discrete or continuous. In addition, specify the level of measurement for each variable.
   a. Number of vending machines in the school (Quantitative, Discrete; Ratio)
   b. Whether or not the school has a closed campus policy during lunch (Qualitative; Nominal)
   c. Whether the vending machines are off during lunch (Qualitative; Nominal)
   d. Number of days per week a student eats school lunch (Quantitative, Discrete; Ratio)
   e. Name of the school (Qualitative; Nominal)

Mini-Lecture 1.2

Observational Studies, Experiments, and Simple Random Sampling

Objectives
1. Distinguish between an observational study and an experiment
2. Explain the various types of observational studies

Examples
1. Determine whether each of the following studies depict an observational study or an experiment. If the researchers conducted an observational study, determine the type of the observational study.
    a. Researchers wanted to assess the long-term psychological effects on children evacuated during World War II. They obtained a sample of 169 former evacuees and a control group of 43 people who were children during the war but were not evacuated. The subjects' mental states were evaluated using questionnaires. It was determined that the psychological well-being of the individuals was adversely affected by evacuation. (Source: Foster D, Davies S, and Steele H (2003) The evacuation of British children during World War II: a preliminary investigation into the long-term psychological effects. Aging & Mental Health (7)5.) (Observational study; Case-control study)
    b. Xylitol has proven effective in preventing dental caries (cavities) when included in food or gum. A total of 75 Peruvian children were given milk with and without xylitol and were asked to evaluate the taste of each. Overall, the children preferred the milk flavored with xylitol. (Source: Castillo JL, et al (2005) Children's acceptance of milk with xylitol or sorbitol for dental caries prevention. BMC Oral Health (5)6.) (Experiment)
    c. A total of 974 homeless women in the Los Angeles area were surveyed to determine their level of satisfaction with the healthcare provided by shelter clinics versus the healthcare provided by government clinics. The women reported greater quality satisfaction with the shelter and outreach clinics compared to the government clinics. (Source: Swanson KA, Andersen R, Gelberg L (2003) Patient satisfaction for homeless women. Journal of Women's Health (12)7.) (Observational study; Cross-sectional study)

2. Some researchers have claimed that cell phones may cause benign brain tumors. To address this question, a group of pre-teens is identified for participation in a study. Their cell phone use is recorded over the next twenty years. At the end of the study, the researchers compare the development of brain tumors to the time each of the subjects spent using a cell phone. What type of observational study does this describe? (Cohort study)

Mini-Lecture 1.3
Simple Random Sampling

Objective
1. Obtain a simple random sample

Examples
1. A statistics instructor wants to obtain a simple random sample (SRS) of three students from her class of 27 students. She assigns each student a number from 1 to 27 based on her roll, which lists the students alphabetically, by last name. Now, which (if any) of the following scenarios could she use to obtain a SRS?
      a. The instructor asks a student in the class to randomly choose three numbers from 1 - 27. The student chooses 3, 18, and 24. The students with those assigned numbers represent the SRS of size three from this class. (Not a SRS – maybe this student only "likes" numbers that are divisible by three)
      b. The instructor puts pieces of paper numbered 1 – 27 in a pile and closes her eyes and selects 6,15, and 22. The students with those assigned numbers represent the SRS of size three from this class. (A SRS)
      c. The instructor of a class of 27 students uses the TI-83 command randInt(1,27,3), which reports {8,11,19}. The students with those assigned numbers represent the SRS of size three from this class. (A SRS)

2. Main Street in Rexburg is a historic and economically important corridor in the small Idaho city. The mayor of Rexburg is considering the implementation of reverse angle parking along this road. Reverse angle parking is a way of painting parking stalls so that motorists drive past the stall, and then back in at an angle. Studies have shown that this parking method results in a 30% fewer accidents, compared to traditional angle parking. Business owners are concerned that some residents would avoid the new parking stalls, and that it would negatively impact their sales. In 2007, a group of Introductory Statistics students was invited to survey a simple random sample of city utility customers to determine how their desire to shop on Main Street would be impacted if reverse angle parking was implemented. The mayor has a list of all 1,100 of the city utility customers. Describe how a simple random sample of 50 utility customers could be obtained. (Assign each utility customer a number between 1 and 1,100. Repeatedly using the TI-83 command randInt(1,1100,1), the students could get a list of numbers corresponding to the customers to be surveyed. Duplicate values were ignored. _Results of the survey_: About 25% said they would be more likely to park on Main Street, about 25% said they would be less likely to park on Main Street, and about 50% said the likelihood that they would park on Main Street would be unchanged.)

Mini-Lecture 1.4
Other Effective Types of Sampling

Objectives
1. Obtain a stratified sample
2. Obtain a systematic sample
3. Obtain a cluster sample

Examples
1. Identify the type of sampling used.
   a. A study was conducted at a Swedish University with 5000 students in which every tenth person was selected, after starting with a randomly chosen number. The students were surveyed on their personality traits and aspects of their health. (Source: von Bothmer MIK, Fridlund B (2003) Self-rated health among university students in relation to sense of coherence and other personality traits. Scandinavian Journal of Caring Sciences (17) pp. 347–357.) (Systematic)
   b. A sample of British secondary-school students was obtained in the following manner. The schools were divided into three groups, based on the abilities of the students in each school. A simple random sample of students from schools in each of the groups was selected. What sampling scheme was employed? (Source: Ireson J and Hallam S (2005) Pupil's liking for school. British Journal of Educational Psychology (75).) (Stratified)

2. Michael Siegel and his research team wanted to assess the impact of smoking restrictions in restaurants on youth tobacco exposure. (Source: Siegel M, Albers AB, Cheng DM, Bierner L, Rigotti NA (2004) Effect of local restaurant smoking regulations on environmental tobacco smoke exposure among youths. American Journal of Public Health (94)2.) To obtain their sample, the researchers randomly dialed phone numbers of Massachusetts households. They then requested parental permission to interview all resident youths between 12 and 17 years of age. In their article, the researchers state that "Because a simple random sample was obtained, with no stratification or clustering, there were no design effects." It was determined that in towns with stronger restaurant smoking restrictions, youths were exposed to less second-hand smoke. Answer the following questions.
   a. The researchers stated that they took a simple random sample. What were the experimental units in their simple random sample? (Massachusetts phone numbers)
   b. What is the experimental unit of interest? (Youths ages 12-17)
   c. What design was actually used to select a sample of youths? (Cluster sample of Massachusetts households with telephones)

Mini-Lecture 1.5
Bias in Sampling

Objective
  1. Explain the sources of bias in sampling

Examples
  1. A cluster sample was conducted to assess the mortality rates in Iraq before and after Operation Iraqi Freedom in 2003. A total of 33 locations were selected and from each location, the closest households were surveyed to determine the total number of household members who died in the 15 months prior to the invasion compared to the 18 months after the invasion. (Source: Roberts L, Lafta R, Garfield R, Khudhairi J, Burnham G (2004) Mortality before and after the 2003 invasion of Iraq: cluster sample survey. The Lancet (364) 9448.) The authors were later sharply criticized. The cluster design they employed is appropriate for estimating natural mortality rates. However, given the violent and sporadic nature of many of these deaths, one cluster per 739,000 individuals was deemed insufficient. In addition, due to the nature of the causes of death, households may be inclined to include deceased extended family members in their counts, even if they would not otherwise have resided together. (Source: Apfelroth S, Roberts L, Burnham G, Garfield R, Abad-Franch F (2005) Mortality in Iraq/Authors' reply. The Lancet (365)9465.) In what ways do these criticisms weaken the results of the study? (Answers will vary) Are the flaws in this study due to the sampling method or the survey itself? (Flawed sampling method)

  2. To investigate students' opinions on a proposed tuition increase, a reporter from the school newspaper polled her roommates and friends. 67% of those surveyed opposed the proposed tuition increase. Discuss potential problems with this study. (Answers will vary) Are the flaws in this study due to the sampling method or the survey itself? (Flawed sampling method)

  3. After receiving criticism for polling her friends, the same reported decided it would be better to get a sample from other students. She plans to choose a location on campus and poll everyone who is willing to respond to her survey. Discuss the pros and cons of using the following locations. (Answers will vary)
       a. The main entrance of the school library.
       b. The hallway outside of a senior-level biology class.
       c. The hallway outside the auditorium in which the freshman psychology classes are taught.
       d. A large parking lot on campus.

  4. A student wants to estimate the prevalence of illicit drug use on campus. Using the student directory, he obtained a SRS of students. He telephoned each randomly selected student and as part of the survey asked, "In the last 12 months, have you ever used any illegal drugs?" What are some potential problems with

this study? (Answers will vary) Are the flaws in this study due to the sampling method or the survey itself? (Sampling method: Interviewer error)

Mini-Lecture 1.6
The Design of Experiments

Objectives
1. Describe characteristics of an experiment
2. Explain the steps in designing an experiment
3. Explain the completely randomized design
4. Explain the matched-pairs design
5. Explain the randomized block design

Examples
1. Xylitol has proven effective in preventing dental caries (cavities) when included in food or gum. A total of 75 Peruvian children were given milk with and without xylitol and were asked to evaluate the taste of each. The researchers measured the children's ratings of the two types of milk. (Source: Castillo JL, et al (2005) Children's acceptance of milk with xylitol or sorbitol for dental caries prevention. BMC Oral Health (5)6.)
   a. What is the response variable in this experiment? (The children's rating of each type of milk)
   b. Think of some of the factors in the study. Which are controlled? Which factor is manipulated? (Answers will vary but could include age and gender of the children; Milk with and without xylitol is the factor that was manipulated.)
   c. What are the treatments? How many treatments are there? (Milk with xylitol and milk without xylitol; 2)
   d. How are factors that are not controlled or manipulated dealt with? (Random assignment)
   e. What type of experimental design is this? (Matched-pairs design)
   f. Identify the subjects (75 Peruvian children)
   g. Draw a diagram similar to Figure 7, 8, or 10 to illustrate the design (Diagram should be similar to Figure 8)

2. The English Department is considering adopting an online version of the freshman English course. To compare the new online course to the traditional course, the English Department faculty randomly split a large section of this course. Half of the students take the traditional course and the other half take an online version. At the end of the semester, both groups will be given a test to determine which group has a better mastery of the concepts in the course.
   a. What is the response variable in this experiment? (The final test scores)
   b. Think of some of the factors in the study. Which are controlled? Which factor is manipulated? (Answers will vary but could include gender, GPA, SAT scores, IQ, etc.; Controlled: none; Manipulated: method of teaching.)
   c. What are the treatments? How many treatments are there? (Traditional course and online course; 2)
   d. How are factors that are not controlled or manipulated dealt with? (Random assignment)

e. What type of experimental design is this? (Completely randomized design)
f. Identify the subjects (the students in the large section of the English course)
g. Draw a diagram similar to Figure 7, 8, or 10 to illustrate the design (Diagram should be similar to Figure 7)

3. After some deliberation, the English Department thinks that there may be a difference in the performance of the men and women in the traditional and online courses described in the previous problem. To accommodate any potential differences, they randomly assign half the men to each of the two courses and they do the same for the women.
a. What is the response variable in this experiment? (The final test scores)
b. Think of some of the factors in the study. Which are controlled? Which factor is manipulated? (Answers will vary but could include gender, GPA, SAT scores, IQ, etc.; Controlled: gender; Manipulated: method of teaching.)
c. What are the treatments? How many treatments are there? (Traditional course and online course; 2)
d. How are factors that are not controlled or manipulated dealt with? (Random assignment)
e   What type of experimental design is this? (Randomized block design)
f. Identify the subjects (the students in the large section of the English course)
g.   Draw a diagram similar to Figure 7, 8, or 10 to illustrate the design (Diagram should be similar to Figure 10)