Statistics Name _____

Sampling and Experiments

Sampling Video: Follow along with the video and answer the following questions.

1. Are recent U.S. Censuses more or less accurate than early Censuses?

2. Why is the U.S. Census undercount, which is quite small as a percent of the population, so important?

3. What is a simple random sample?

4. How many uses of sampling can you spot in the account of Frito-Lay potato chips?

SAMPLING:

Key definitions:

- A _____ is a method of sampling where the entire population is surveyed

- _____: The entire group of individuals about whom we hope to learn
- _____: A representative subset of a population, examined in hope of learning about the population as a whole
- _____: Any systematic failure of a sampling method to represent its population.

- _____: A sample where people choose themselves by responding to a general appeal.
- _____: A sample where individuals who are easiest to reach are chosen.
- _____: The best defense against bias, each individual is given a fair, random chance of selection
- _____: Each individual has an equal chance of being selected, and each combination of people is also equally likely

How to take a Simple Random Sample (SRS)

1. _____

2. _____

Examples

1. A local television station takes quick polls of public opinion by announcing a question on the 6 o'clock news and asking viewers to call-in or text their opinion of "Yes" or "No" to the station. The results are announced on the 11 o'clock news. One such poll finds that 73% of those who called in or texted are opposed to a proposed local gun control ordinance.

a. What do you think the population is in this situation?

b. Explain why this sampling method is biased. Is the percent of the population who oppose gun control probably higher or lower than the 73% of the sample who are opposed?

2. The students named below are enrolled in a new statistics course. Use a random digits table or a calculator/computer (this is what we will use) number generator to choose five of these students at random to be interviewed in detail about the quality of the course. Explain how you chose your sample.

Agarwal	Dewald	Hixson	Puri
Anderson	Fernandez	Klassen	Rodriguez
Baxter	Frank	Mihalko	Rubin
Bowman	Fuhrmann	Moser	Santiago
Bruvold	Goel	Naber	Shen
Casella	Gupta	Petrucelli	Shyr
Cordero	Hicks	Pliego	Sundheim

3. On the Hudson Valley, NY Patch Facebook page, readers were asked to send in stories of awful Valentine's Day gifts. The following were selected:

- Leftover chocolate (and he had eaten one!)
- Flowers purchased the day BEFORE Valentine's because it was cheaper to buy them the day before.
- A recycled card from an ex-boyfriend with an open box of chocolates.

Readers were then asked to vote on the best "worst Valentine's Day gift ever" story.

a. Describe the population

b. Describe the sample

c. Do you think the response to this poll is representative of the views of the residents of Hudson Valley, New York? Explain.

Sampling #1

For questions 1-4, identify the population and the sample.

1. A high school's student newspaper plans to survey local businesses about the importance of students as customers. From telephone book listings, the newspaper staff chooses 150 businesses at random. Of these, 73 return the questionnaire mailed by the staff.

2. An archaeological dig turns up large numbers of pottery shards, broken stone tools and other artifacts. Students working on the project classify each artifact and assign it a number. The counts in different categories are important for understanding the site, so the project director chooses 2% of the artifacts at random and checks the students' work.

3. A large retailer prepares its customers' monthly credit card bills using an automatic machine that folds the bills, stuffs them into envelopes, and seals the envelopes for mailing. Are the envelopes completely sealed? Inspectors choose 40 envelopes from the 1000 stuffed each hour for visual inspection.

4. A department store mails a customer satisfaction survey to people who make credit card purchases at the store. This month, 45,000 people made credit card purchases. Surveys are mailed to 137 of these people, chosen at random, and 100 people return the survey from.

5. A newspaper advertisement for an upcoming TV show said: "Should handgun control be tougher? You call the shots in a special call-in poll tonight. If yes, call 1-900-720-6181. If no, call 1-900-720-6182. Charge is 50 cents for the first minute." Explain why this opinion poll is almost certainly biased.

6. You are in the staff of a member of Congress who is considering a bill that would provide government sponsored insurance for nursing home care. You report that 1128 letters have been received on the issue, of which 871 oppose the legislation. "I'm surprised that most of my constituents oppose the bill. I thought it would be quite popular," says the congresswoman. Are you convinced that a majority of voter oppose the bill? How would you explain that statistical issue to the congresswoman.

7. A recent online poll posed the question "Should female athletes be paid the same as men for the work they do?" In all, 13,146 (44%) said "Yes," 15,182 (50%) said "No'" and the remaining 1448 said "Don't know." In spite of the large sample size for this survey, we can't trust the result. Why not?

8. How much sleep do high school students get on a typical school night? An interested student designed a survey to find out. To make data collection easier, the student surveyed the first 100 students to arrive at school on a particular morning. These students reported an average of 7.2 hours of sleep on the previous night.

a. What type of sample did the student obtain?

b. Explain why this sampling method is biased. Is 7.2 hours probably higher or lower than the true average amount of sleep last night for all students at the school? Why?

9. To gather data on a 1200-acre pine forest in Louisiana, the US Forest Service laid a grid of 1420 equally space circular plots over a map of the forest. A ground survey visited a sample of 10% of these plots.

a. Explain how you would use your to choose an SRS of 141 plots.

b. Use your method in (a) to choose the first 3 plots.

Follow along with the video and answer the following questions

1. Why was the *Literary Digest* poll so far wrong in predicting the outcome of the 1936 presidential election?

2. Why would a simple random sample of counties in a state give results that might not represent the entire state?

3. In sampling, what are strata?

4. You are an interviewer for an opinion poll. How should you react to answers that seem anti-social or immoral?

Types of Sampling Methods

Stratified Random Sampling

Cluster Random Sampling

Systematic Sampling

What can go Wrong

Undercoverage

Nonresponse

Response Bias

Wording of Questions

1) State whether the sampling procedure is convenience sampling, voluntary response sampling, stratified random sampling, simple random sampling, cluster sampling, or systematic.

   a) Redbook magazine asks its readers to complete a survey and mail it in. _____

   b) The 1650 students at RHS are labeled 0001 to 1650 and a *random* number generator is used to create a sample of size 100. _____

   c) The sales manager of a company stands outside his office and asks the first 40 staff members that walk-by their average gross sales per month. _____

d) The 1650 students at RHS are divided into grade levels and we *randomly* select 25 from each grade. _____

e) We *randomly* select 5 of the 66 English classes at RHS and ask each student in those 5 classes to complete a short survey. _____

f) A researcher randomly selects and interviews fifty male and fifty female teachers._____

g) A researcher for an airline interviews all of the passengers on five randomly selected flights. _____

h) Based on 12,500 responses from 42,000 surveys sent to its alumni, a major university estimated that the annual salary of its alumni was 92,500. _____

i) A community college student interviews everyone in a biology class to determine the percentage of students that own a car. _____

j) A market researcher randomly selects 200 drivers under 35 years of age and 100 drivers over 35 years of age. _____

k) All of the teachers from 85 randomly selected nation's middle schools were interviewed. _____

l) To avoid working late, the quality control manager inspects the last 10 items produced that day. _____

Sampling #2

Classify each of the following with the correct type of sampling method. (SRS, stratified, systematic, convenience, voluntary response, or cluster).

1. In a large school district, all teachers from two randomly selected buildings are interviewed to determine whether they believe the students have less homework to do now than in previous years.

2. Every seventh customer entering the shopping mall is asked to select her or his favorite store.

3. Nursing supervisors are selected using random numbers in order to determine annual salaries.

4. Every 100^{th} hamburger manufactured is checked to determine its fat content.

5. Mail carriers of a large city are divided into four groups according to gender and according to whether they walk or ride on their routes. Then 10 are selected from each group and interviewed to determine whether a dog has bitten them in the last year.

6. A group of hospitals wants to survey patients discharged this past year to obtain patient satisfaction profiles. They wish to use a sample of such patients.

a. Obtain a list of patients discharged. Divide the patients according to length of hospital stay (2 days or less, 3 – 7 days, 8 – 14 days, more than 14 days). Randomly select patients from each group.

b. Randomly select some of the hospitals, and then include all the patients on the discharge lists of the selected hospitals.

c. Obtain lists of patients discharged from all the hospitals. Number the patients, and then use a random number table to obtain the sample.

d. At the beginning of the year, instruct each hospital to survey every 500^{th} patient discharged.

e. Instruct each hospital to survey 10 discharged patients this week and send in the results.

7. An important part of employee compensation is a benefits package. Suppose you want to conduct a survey of benefit packages available in private businesses in Hawaii. You want a sample size of 100.

a. Assign each business in the Island Business Directory a number, and then use a random number table to select the businesses to be included in the sample.

b. Use the postal Zip Codes to divide the state into regions. Pick a random sample of 10 Zip Code areas and then include all the businesses in each selected Zip Code area.

c. Send a team of five research assistants to Bishop Street in downtown Honolulu. Let each assistant select a block or building and interview an employee from each business found. Each researcher can have the rest of the day off after getting responses from 20 different businesses.

d. Use the Island Business Directory, number all the businesses. Select a starting point at random and then use every 50^{th} business listed until you have 100 businesses.

e. Group the businesses according to type: medical, shipping, retail, manufacturing, financial, construction, restaurant, hotel, tourism, other. Then select a random sample of 10 businesses from each business type.

8. A supply of military ammunition consists of 1,000 boxes, with each box containing 50 rounds. To check if military specifications are being met, 10 boxes of ammunition are randomly selected, and each round from the 10 boxes is tested.

9. A large corporation wants to survey its employees to estimate the average amount that they would be willing to contribute to a company-assisted savings plan. The company will randomly select 50 employees from each of 3 groups consisting of management, skilled, and non-skilled workers.

10. A soft drink manufacturer has adopted a new bottling process. To check on the amount of carbonation, every 20^{th} bottle is removed from the bottling line, and the amount of carbon dioxide is measured.

11. A taxation board for a city wants to estimate the proportion of homes for which major improvements have been made without a building permit. To conduct a survey, officials will randomly select a sample of 100 city blocks, and then inspect each house in each selected block.

12. Label each of the following situations with the type of bias indicated. Question wording, under coverage, response bias, non-response bias.

a. The population of interest is all teachers at RHS but the sample is selected from those teachers who listed their phone numbers in the staff school directory. _____

b. In a survey about addictive habits, an embarrassed teacher deliberately gives an incorrect answer. _____

c. Selected teachers cannot be contacted or refuse to respond. _____

d. A survey asks the question "Do you agree that a national system of health insurance should be favored because it would provide needed health insurance for everyone and would reduce administrative costs." _____

Experiment vs. Observational Study

- An _____ observes individuals and measures the variables of interest but does not attempt to influence the responses.

  Goal: _____

  _____

- An _____ deliberately imposes some treatment on individuals to measure their responses.

  Goal: _____

  _____

Which is Which? Describe the following as an experiment or observational study.

A study took random sample of adults and asked them about their bedtime habits. The data showed that people who drank a cup of tea before bedtime were more likely to go to sleep earlier than those who didn't drink tea.

Another study took a group of adults and randomly divided them into two groups. One group was told to drink tea every night for a week, while the other group was told not to drink tea that week. Researchers then compared when each group fell asleep.

<u>Response Variable vs. Explanatory Variable</u>

- A _____ measures the outcome of a study.

- An _____ may help explain or influence changes in a response variable.

Language of Experiments

- _____ are the individuals to which the treatment is applied.
- When the units are human being, they are often called _____.
- A specific condition applied to the individuals in an experiment is called a _____ .
- The _____ are the explanatory variables in the experiment.

Does adding fertilizer affect the productivity of tomato plants? How about the amount of water given to the plants? To answer these questions, a gardener plants 24 similar tomato plants in identical pots in his greenhouse. He will add fertilizer to the soil in half of the pots. Also, he will water 8 of the plants with 0.5 gallons of water per day, 8 of the plants with 1 gallon of water per day and the remaining 8 plants with 1.5 gallons of water per day. At the end of three months he will record the total weight of tomatoes produced on each plant.

State the experimental units, explanatory variable, response variable and treatments.

Randomized Comparative Experiment

- _____ produces groups that are similar in all respects before we apply the treatments.
- _____ ensures that influences other than the experimental treatments operate equally on all groups.

This way, differences in the response variable must be due to the effects of the treatment.

Control Group

- In an experiment, the Control Group does not receive the treatment, gets a placebo or gets a known treatment. The primary purpose of the control group is to provide a _____ the effects of the other treatments.

Principles of Experimental Design

- Random Assignment: Use _____ to assign experimental units to treatments. This helps to create roughly equivalent groups by balancing the effects of _____ _____ .

- Control: Lurking variables might effect the response variable. Use a comparative design and ensure that the only difference between groups is the _____.

- Replication: Use enough experimental units in each group so that any differences in the effects of the treatments _____ .

Placebo

- The Placebo is the _____ in an experiment.

- The _____ is when subjects respond to the dummy treatment.

Blindness:

- In a _____ experiment, the subjects do not know what treatment they received.

- In a _____ experiment, neither the subjects nor the people administering the treatment know who received what treatment.

Statistically Significant:

- An observed effect so large that it would rarely _____ is called statistically significant.

Experiments 3

1. A University of Helsinki study wanted to determine if chocolate consumption during pregnancy had an effect on infant temperament at age 6 months. Researchers began by asking 305 healthy pregnant women to report their chocolate consumption. Six months after birth, the researchers asked mother to rate their infants' temperament, including smiling, laughter, and fear. The babies born to women who had been eating chocolate daily during pregnancy were found to be more active and "positively reactive" – a measure that the investigators said encompasses traits like smiling and laughter.

a. Was this an observational study or an experiment? Justify your answer.

b. What are the explanatory and response variables?

c. Does this study show that eating chocolate regularly during pregnancy helps produce infants with good temperament? Explain

2. An educator wants to compare the effectiveness of computer software for teaching biology with that of a textbook presentation. She gives a biology pretest to each of a group of high school juniors, then randomly divides them into two groups. One group uses the computer, and the other studies the text. At the end of the year, she tests all students again and compares the increase in biology test scores in the two groups.

a. Is this an observational study and an experiment? Justify your answer.

b. If the group using the computer has a much higher average increase in test scores than the group using the textbook, what conclusions, if any, could the educator draw?

3. One study of cell phones and the risk of brain cancer looked at a group of 469 people who have brain cancer. The investigators matched each cancer patient with a person of the same age, gender, and race who did not have brain cancer, then asked about the use of cell phones. Result: "Our data suggest that the use of handheld cellular phones is not associated with risk of brain cancer."

a. Is this an observational study or an experiment? Justify your answer.

b. Based on this study, would you conclude that cell phones do not increase the risk of brain cancer? Why or why not?

For the experiments described in 4 - 5, identify the experimental units or subjects, the explanatory variables (factors), the treatments, and the response variables.

4. Ability to grow in shade may help pines found in the dry forests of Arizona to resist drought. How well do these pines grow in shade? Investigators planted pine seedlings in a green house in either full light, light reduced to 25% of normal by shade cloth, or light reduced to 5% of normal. At the end of the study , they dried the young trees and weighed them.

5. A maker of fabric for clothing is setting up a new line to "finish" the raw fabric. The line will use either metal rollers or natural-bristle rollers to raise the surface of the fabric; a dyeing-cycle time of either 30 or 40 minutes; and a temperature of either 150° of 175° Celsius. An experiment will compare all combinations of these choices. Three specimens of fabric will be subjected to each treatment and scored for quality.

Experimental Design

1. An ad for OptiGro plant fertilizer claims that with this product you will grow "juicier, tastier" tomatoes. You'd like to test this claim, and wonder whether you might be able to get by with half the specified dose. You have 24 tomato plants to works with. Describe or outline a completely randomized experiment to test the ad's claim.

2. Most people know that drunk driving is dangerous. Even so, some drivers are cited multiple times for DWI. Is there anything that can be done to change their behavior?

Design an experiment for 300 people who have been convicted of DWI three times in one year. The treatments are a fine plus suspended jail sentence plus one of (1) no treatment, (2) attend an alcoholism clinic, (3) participate in AA.

3. Researchers would like to see if taking ginkgo biloba tree extract will improve memory. 230 elderly community members volunteer to participate. Diagram a completely randomized experiment to see if taking ginkgo biloba tree supplement will improve memory.

Experiments 4

1. Doctors identify "chronic tension-type headaches" as headaches that occur almost daily for at least six months. Can antidepressant medications or stress management training reduce the number and severity of theses headaches? Are both together more effective than either alone? Investigator compared four treatments: antidepressant alone, placebo alone, antidepressant and stress management, and placebo plus stress management. There are 36 headache sufferers that have agreed to participate in the study.
Outline the design of the experiment.

2. A large study about how to treat prostate disease used records form Canada's national health care system to compare the effectiveness of two ways to treat prostate disease. The two treatments are traditional surgery and a new method that does not require surgery. The records described many patients whose doctors had chosen each method. The study found that patients treated by the new method were significantly more likely to die within 8 years.

You have 300 prostate patients who are willing to serve as subjects in an experiment to compare two methods. Outline the design of the experiment.

3. A biologist would like to determine which two brands of weed killer is less likely to harm the plants in a garden at the university. Before spraying near the plants, the biologist decides to conduct an experiment using 24 individual plants. Which of the following two plans for randomly assigning the treatments should the biologist use? Why?

Plan A: Choose the 12 healthiest looking plants. Apply Brand X weed killer to all 12 those plants. Apply Brand Y week killer to the remaining 12 plants.

Plan B: Choose 12 of the 24 plants at random. Apply Brand X weed killer to those 12 plants and Brand Y weed killer to the remaining 12 plants.

4. A study sought to determine whether the ancient Chinese art of acupuncture could help infertile women become pregnant. 160 healthy women undergoing treatment for infertility were recruited for the study. Half of the subjects were randomly assigned to receive acupuncture treatment 25 minutes before embryo transfer and again 25 minutes after the transfer. The remaining 80 subjects were instructed to lie still for 25 minutes after the embryo transfer. *Results:* In the acupuncture group, 34 women became pregnant. In the control group, 21 women became pregnant.

a. The difference in the percent of women who became pregnant in the two groups is statistically significant. Explain what this means to someone who know little statistics.

b. Explain why the placebo effect prevents us from concluding that acupuncture caused the difference in pregnancy rates.