Part II: Producing Data

Producing Data accounts for 10%-15% of the material covered on the AP Exam. "*Data must be collected according to a well-developed plan if valid information on a conjecture is to be obtained. This plan includes clarifying the question and deciding upon a method of data collection and analysis.*"

Check your understanding of the following topics from | Chapters *10-12* in our Stats textbook |

II. Sampling and Experimentation: Planning and Conducting a Study

A. Overview of methods of data collection

- ☐ 1. Census
- ☐ 2. Sample survey
- ☐ 3. Experiment
- ☐ 4. Observational study

B. Planning and conducting surveys

- ☐ 1. Characteristics of a well-designed and well-conducted survey
- ☐ 2. Populations, samples, and random selection
- ☐ 3. Sources of bias in sampling and surveys
- ☐ 4. Sampling methods, including simple random sampling, stratified random sampling, and cluster sampling

C. Planning and conducting experiments

- ☐ 1. Characteristics of a well-designed and well-conducted experiment
- ☐ 2. Treatments, control groups, experimental units, random assignments, and replication
- ☐ 3. Sources of bias and confounding, including placebo effect and blinding
- ☐ 4. Completely randomized design
- ☐ 5. Randomized block design, including matched pairs design

D. Generalizability of results and types of conclusions that can be drawn from observational studies, experiments, and surveys

Review questions from Previous AP Exams:

2012 #6a

6. Two students at a large high school, Peter and Rania, wanted to estimate μ, the mean number of soft drinks that a student at their school consumes in a week. A complete roster of the names and genders for the 2,000 students at their school was available. Peter selected a simple random sample of 100 students. Rania, knowing that 60 percent of the students at the school are female, selected a simple random sample of 60 females and an independent simple random sample of 40 males. Both asked all of the students in their samples how many soft drinks they typically consume in a week.

   (a) Describe a method Peter could have used to select a simple random sample of 100 students from the school.

2016 #2

3. Alzheimer's disease results in a loss of cognitive ability beyond what is expected with typical aging. A local newspaper published an article with the following headline.

> Study Finds Strong Association Between Smoking and Alzheimer's

The article reported that a study tracked the medical histories of 21,123 men and women for 23 years. The article stated that, for those who smoked at least two packs of cigarettes a day, the risk of developing Alzheimer's disease was 2.57 times the risk for those who did not smoke.

(a) Identify the explanatory and response variables in the study.

Explanatory variable:

Response variable:

(b) Is the study described in the article an observational study or an experiment? Explain.

(c) Exercise status (regular weekly exercise versus no regular weekly exercise) was mentioned in the article as a possible confounding variable. Explain how exercise status could be a confounding variable in the study.

2015 Multiple Choice Practice Exam

2. A researcher wanted to estimate the average amount of money spent on extracurricular activities per school in a certain region. The researcher randomly selected 20 public schools and 20 private schools in the region to use for a sample. Which of the following best describes the type of sample that was taken?

   (A) A census

   (B) A cluster sample

   (C) A convenience sample

   (D) A simple random sample

   (E) A stratified random sample

4. A researcher conducting a telephone survey is concerned about possible sources of bias. Of the following, which is the best example of nonresponse bias?

   (A) The wording of the questions in the survey leads people to respond in a certain way.

   (B) The behavior of the interviewer leads people to respond in a certain way.

   (C) People might be uncomfortable with the survey questions and, as a result, might not always respond to those questions truthfully.

   (D) Many of the people selected to participate in the survey who do not respond might have opinions different from those who do respond.

   (E) People without telephones are overlooked in the sampling procedure used to determine who is surveyed.

5. At a large conference of teachers from a variety of subjects, a random sample of 50 mathematics teachers attending the conference was selected. Among the selected mathematics teachers, 28 percent had taken one or more courses in statistics. For which of the following populations is 28 percent a reasonable estimate of the percentage of those who have taken one or more courses in statistics?

   (A) All mathematics teachers

   (B) All mathematics teachers who attended the conference

   (C) All mathematics teachers who have taken one or more courses in statistics

   (D) All teachers who attended the conference

   (E) All teachers

39. In a certain school, students can choose whether to eat in the school's cafeteria. A reporter working for the school's newspaper polled students on their reactions to changes in the menu at the cafeteria. For each student leaving the cafeteria in one 30-minute time period, the reporter used a coin to determine whether to stop the student and ask how he or she felt about the new menu. In the reporter's article it was stated that a random sample of the students showed that 89 percent of the school's student population was happy with the new menu. Which of the following statements is true?

   (A) Because each student leaving the cafeteria was randomly selected and could choose to answer or not, this is a random sample of the student population, and the 89% is an accurate measurement of the school population's view of the new menu.

   (B) Because students self-selected whether to eat in the cafeteria, the sampling method might be biased and the sample might not be representative of all students in the school.

   (C) The survey would have been more effective if the reporter had collected the data in one 15-minute time period rather than in one 30-minute time period.

   (D) The survey would have been more effective if students who cared about the food could have called the reporter to tell how they felt about the new menu, so that only students with opinions on the subject would have been surveyed.

   (E) Because no treatment was imposed on the students eating in the cafeteria, one cannot make any conclusions about the new menu.

2014 Multiple Choice Practice Exam

3. A well-designed experiment should have which of the following characteristics?

I. Subjects assigned randomly to treatments

II. A control group or at least two treatment groups

III. Replication

(A) I only

(B) I and II only

(C) I and III only

(D) II and III only

(E) I, II, and III

6. Approximately 52 percent of all recent births were boys. In a simple random sample of 100 recent births, 49 were boys and 51 were girls. The most likely explanation for the difference between the observed results and the expected results in this case is

(A) bias
(B) variability due to sampling
(C) nonsampling error
(D) a sampling frame that is incomplete
(E) confounding

9. A survey was administered to parents of high school students in a certain state to see if the parents thought the students' academic needs were being met. To select the sample, the parents were divided into two groups—one group of parents who live in cities with populations of more than 100,000 and the other group of parents who live in cities with populations less than or equal to 100,000. A random sample of 100 parents from each group was taken. Which of the following statements about the sample of 200 parents is true?

(A) It is a convenience sample because the sample of parents was easily obtained.

(B) It is a stratified random sample because parents were randomly selected from each group.

(C) It is a random cluster sample because parents were randomly selected from each group.

(D) It is a random cluster sample because groups of high schools were randomly selected.

(E) It is a systematic sample because the parents were systematically divided into two groups.

21. In a recent poll of 1,500 randomly selected eligible voters, only 525 (35 percent) said that they did not vote in the last election. However, a vote count showed that 80 percent of eligible voters actually did not vote in the last election. Which of the following types of bias is most likely to have occurred in the poll?

(A) Nonresponse bias

(B) Sampling bias

(C) Selection bias

(D) Response bias

(E) Undercoverage bias

33. A large clinical trial was designed to determine whether a certain vitamin improves the general health of adults. The investigators first identified 85 variables that measure various aspects of the general health of adults. Because each adult in the clinical trial was to serve as his or her own control, the 85 variables were measured for each adult, both before taking the vitamin and after taking the vitamin for three months. The investigators then performed 85 matched-pair t-tests, one for each variable. They found statistically significant results at the 0.05 level in 2 of the variables, both in the direction of improved general health. Which of the following should the investigators conclude?

(A) There is evidence that the vitamin improves the health of adults because 2 of the 85 tests were statistically significant.

(B) There is evidence that the vitamin improves the health of adults because in clinical trials investigators typically underestimate the proportion of the population that is helped by a vitamin.

(C) There is insufficient evidence that the vitamin improves the health of adults because at the 0.05 significance level, one could easily get statistically significant results in 2 out of 85 tests just due to chance variability.

(D) There is insufficient evidence that the vitamin improves the health of adults because the sample size of 85 is not large enough to draw a conclusion.

(E) No conclusion can be drawn because an even number of variables is needed for a matched-pairs design.

38. Which of the following distinguishes an observational study from a randomized experiment?

(A) In an observational study volunteers are always used, whereas in a randomized experiment a random sample is always taken from the population.

(B) In an observational study a random sample is always taken from the population, whereas in a randomized experiment volunteers are always used.

(C) In an observational study treatments are not randomly assigned, whereas in a randomized experiment treatments are randomly assigned.

(D) In an observational study a control group is never used, whereas in a randomized experiment a control group is always used.

(E) An observational study can be double-blind, whereas a randomized experiment can only be single-blind because the experimenter determines who is randomly assigned to each treatment.

2014 #4

4. As part of its twenty-fifth reunion celebration, the class of 1988 (students who graduated in 1988) at a state university held a reception on campus. In an informal survey, the director of alumni development asked 50 of the attendees about their incomes. The director computed the mean income of the 50 attendees to be $189,952. In a news release, the director announced, "The members of our class of 1988 enjoyed resounding success. Last year's mean income of its members was $189,952!"

   (a) What would be a statistical advantage of using the median of the reported incomes, rather than the mean, as the estimate of the typical income?

   (b) The director felt the members who attended the reception may be different from the class as a whole. A more detailed survey of the class was planned to find a better estimate of the income as well as other facts about the alumni. The staff developed two methods based on the available funds to carry out the survey.

   Method 1: Send out an e-mail to all 6,826 members of the class asking them to complete an online form. The staff estimates that at least 600 members will respond.

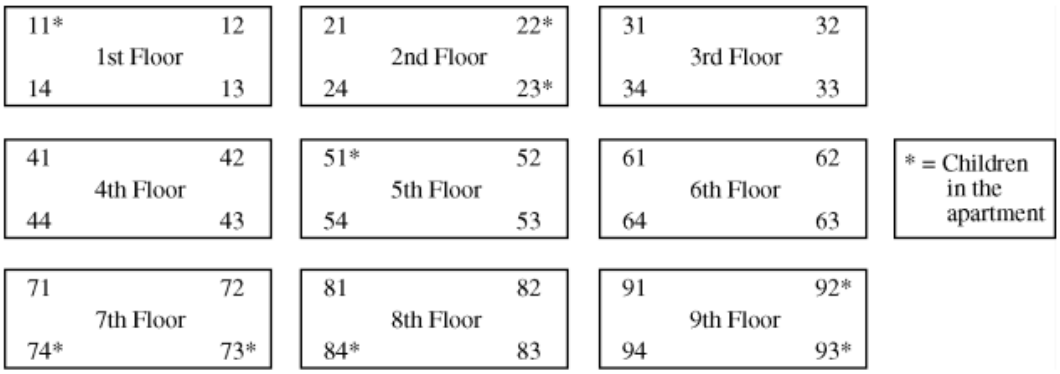
   Method 2: Select a simple random sample of members of the class and contact the selected members directly by phone. Follow up to ensure that all responses are obtained. Because method 2 will require more time than method 1, the staff estimates that only 100 members of the class could be contacted using method 2.

   Which of the two methods would you select for estimating the average yearly income of all 6,826 members of the class of 1988 ? Explain your reasoning by comparing the two methods and the effect of each method on the estimate.

2011 #3

3. An apartment building has nine floors and each floor has four apartments. The building owner wants to install new carpeting in eight apartments to see how well it wears before she decides whether to replace the carpet in the entire building.

   The figure below shows the floors of apartments in the building with their apartment numbers. Only the nine apartments indicated with an asterisk (*) have children in the apartment.



   (a) For convenience, the apartment building owner wants to use a cluster sampling method, in which the floors are clusters, to select the eight apartments. Describe a process for randomly selecting eight different apartments using this method.

   (b) An alternative sampling method would be to select a stratified random sample of eight apartments, where the strata are apartments with children and apartments with no children. A stratified random sample of size eight might include two randomly selected apartments with children and six randomly selected apartments with no children. In the context of this situation, give one statistical advantage of selecting such a stratified sample as opposed to a cluster sample of eight apartments using the floors as clusters.