AP Exam Practice: Sampling

1. At a certain university, students who live in the dormitories eat at a common dining hall. Recently, some students have been complaining about the quality of the food served there. The dining hall manager decided to do a survey to estimate the proportion of students living in the dormitories who think that the quality of the food should be improved. One evening, the manager asked the first 100 students entering the dining hall to answer the following question.

<table>
<thead>
<tr>
<th>Many students believe that the food in the dining hall needs improvement. Do you think that the quality of food served here needs improvement, even though that would increase the cost of the meal plan?</th>
</tr>
</thead>
<tbody>
<tr>
<td>_____ Yes _____ No _____ No opinion</td>
</tr>
</tbody>
</table>

(a) In this setting, explain how bias may have been introduced based on the way this convenience sample was selected and suggest how the sample could have been selected differently to avoid that bias.

(b) In this setting, explain how bias may have been introduced based on the way the question was worded and suggest how it could have been worded differently to avoid that bias.
2. An administrator at a large university wants to conduct a survey to estimate the proportion of students who are satisfied with the appearance of the university buildings and grounds. The administrator is considering three methods of obtaining a sample of 500 students from the 70,000 students at the university.

(a) Because of financial constraints, the first method the administrator is considering consists of taking a convenience sample to keep the expenses low. A very large number of students will attend the first football game of the season, and the first 500 students who enter the football stadium could be used as a sample. Why might such a sampling method be biased in producing an estimate of the proportion of students who are satisfied with the appearance of the buildings and grounds?

(b) Because of the large number of students at the university, the second method the administrator is considering consists of using a computer with a random number generator to select a simple random sample of 500 students from a list of 70,000 student names. Describe how to implement such a method.

(c) Because stratification can often provide a more precise estimate than a simple random sample, the third method the administrator is considering consists of selecting a stratified random sample of 500 students. The university has two campuses with male and female students at each campus. Under what circumstance(s) would stratification by campus provide a more precise estimate of the proportion of students who are satisfied with the appearance of the university buildings and grounds than stratification by gender?
3. Corn tortillas are made at a large facility that produces 100,000 tortillas per day on each of its two production lines. The distribution of the diameters of the tortillas produced on production line A is approximately normal with mean 5.9 inches, and the distribution of the diameters of the tortillas produced on production line B is approximately normal with mean 6.1 inches. The figure below shows the distributions of diameters for the two production lines.

![Graph showing distributions of diameters for Production Line A and Production Line B.]

The tortillas produced at the factory are advertised as having a diameter of 6 inches. For the purpose of quality control, a sample of 200 tortillas is selected and the diameters are measured. From the sample of 200 tortillas, the manager of the facility wants to estimate the mean diameter, in inches, of the 200,000 tortillas produced on a given day. Two sampling methods have been proposed.

**Method 1:** Take a random sample of 200 tortillas from the 200,000 tortillas produced on a given day. Measure the diameter of each selected tortilla.

**Method 2:** Randomly select one of the two production lines on a given day. Take a random sample of 200 tortillas from the 100,000 tortillas produced by the selected production line. Measure the diameter of each selected tortilla.

(a) Will a sample obtained using Method 2 be representative of the population of all tortillas made that day, with respect to the diameters of the tortillas? Explain why or why not.
(b) The figure below is a histogram of 200 diameters obtained by using one of the two sampling methods described. Considering the shape of the histogram, explain which method, Method 1 or Method 2, was most likely used to obtain such a sample.

(c) Which of the two sampling methods, Method 1 or Method 2, will result in less variability in the diameters of the 200 tortillas in the sample on a given day? Explain.

(Parts (d)-(f) covered sampling distribution topics)
4. 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.

![Floor Plan](image)

(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.
ANSWERS to AP Exam Practice: Sampling

1. 2004B #2--Dining Hall
   a) Students who arrive early may have different opinions about food quality than other students (late-diners), thus producing a biased sample. This bias can be avoided by taking a simple random sample of all dorm residents.
   b) The phrase about the dining hall food needing improvement may lead more students to respond that the food really does need improvement. Also, the phrase indicating increased cost may also influence students to answer insincerely. Better wording: “Do you feel the quality of food needs improvement?”

2. 2013 #2—Buildings and Grounds Survey
   a) The first 500 students who enter the football stadium were not likely to be representative of the population of all students at the university. In other words, these 500 students were likely to differ systematically from the population with regard to many variables. For example, these 500 students might have more school pride than the population of students as a whole, which might be related to their opinions about the appearance of university buildings and grounds. Perhaps their school pride is related to having more positive opinions about the appearance of university buildings and grounds, in which case the sample proportion of students who were satisfied would be biased toward overestimating the population proportion of students who were satisfied.
   b) Obtain a list of all 70,000 students at the university. Assign an identification number from 1 to 70,000 to each student. Then use a computer to generate 500 random integers between 1 and 70,000 without replacement. The students whose ID numbers correspond to those numbers were then selected for the sample.
   c) Stratifying by campus would be more advantageous than stratifying by gender provided that opinions about appearance of university buildings and grounds between the two campuses differ more than the opinions about appearance of university buildings and grounds between the two genders.

3. 2015 #6 (abc only)
   In part (a) the response says no AND either argues that the sample will only be selected from one production line and not the entire population
   In part (b) the response chooses Method 1 AND refers to a relevant characteristic of the histogram (shape, center, or variability).
   In part (c) the response chooses Method 2 AND either justifies by stating that the sample comes from only one production line.

4. 2011 #3--Sampling Apartments
   a) Use a random number generator to select two different random integers from 1 to 9. The two numbers correspond to the two floors that will be sampled, for a total of eight apartments.
   b) Using the cluster sampling method described above, it would be possible to select NO apartments with children if, say, floors 3 and 6 were picked. Since it’s reasonable to believe that carpets of apartments with children will have more wear than apartments without children, a cluster sample may not provide the owner with a good representative sample of carpet wear in this building.
   A stratified sample as described will allow the owner to see how well the carpet wears in both apartments with and without children.
Which student’s answer is better? Explain.

(a) In this setting, explain how bias may have been introduced based on the way this convenience sample was selected and suggest how the sample could have been selected differently to avoid that bias:

This survey can be biased because not everybody who live in the dormitories eat at the dining hall. Also, the first 100 students entering the dining hall might not be the people who live in the dormitories. To make better survey, they should give a number to all the people who live in dormitories, then by using random table, choose first 100 people, then ask them true questions.

(a) In this setting, explain how bias may have been introduced based on the way this convenience sample was selected and suggest how the sample could have been selected differently to avoid that bias.

The first students to enter the dining hall may enjoy the food served more than the student population because they may enjoy it more, they arrive earlier. To prevent this bias, each student should be allocated a number and 100 students should be chosen using a random digit generator. This way, every sample of 100 students have an equal chance of being selected.
2013 #2: Buildings and Grounds Survey (Student Samples)

Student #1:

(a) Because of the large number of students at the university, the second method the administrator is considering consists of using a computer with a random number generator to select a simple random sample of 500 students from a list of 70,000 student names. Describe how to implement such a method.

Get an alphabetical list of all the students and assign each individual a number, 1-70,000. Then use a random number generator to select 500 numbers, disregarding repeats, and use the people whose names correspond with the 500 numbers selected to participate in the study.

(b) No mention of how/why this scenario is biased.

Student #2:

(a) Because of financial constraints, the first method the administrator is considering consists of taking a convenience sample to keep the expenses low. A very large number of students will attend the first football game of the season, and the first 500 students who enter the football stadium could be used as a sample. Why might such a sampling method be biased in producing an estimate of the proportion of students who are satisfied with the appearance of the buildings and grounds?

This could be biased because the students who are attending the football game are obviously very proud of their school and support their team and school. These students most likely will not have a problem with their school’s appearance because they’re so proud of the school itself and what it stands for.

1: Did not select random numbers “out of 70,000.” (P)
2: No mention of HOW/WHY this scenario is biased. (P)
(b) Because of the large number of students at the university, the second method the administrator is considering consists of using a computer with a random number generator to select a simple random sample of 500 students from a list of 70,000 student names. Describe how to implement such a method.

First each student should be assigned a number that falls between 00000 and 69,999. Now all 70,000 students have a number assigned to them. Then use a computer to generate a random digits table and pick out the first 500 numbers that fall between 00000 and 69,999. When you have your 500 numbers, correspond the correct number to the correct student and you have a sample of 500 random students.

(c) Because stratification can often provide a more precise estimate than a simple random sample, the third method the administrator is considering consists of selecting a stratified random sample of 500 students. The university has two campuses with male and female students at each campus. Under what circumstance(s) would stratification by campus provide a more precise estimate of the proportion of students who are satisfied with the appearance of the university buildings and grounds than stratification by gender?

Stratification by campus would give a more precise estimate than gender because if one campus is nicer than another, that proportion will obviously be more satisfied with the appearance of the university, compared to the less appealing campus.

b) No mention of “without replacement” (P)

c) No comparison to gender option. (P)
2011 #3: Apartments (Student Samples)

Student #1, Part (a):

(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.

The way this method is used is a floor is first randomly selected, then an apartment in the floor is randomly selected. And you keep on doing it until the person selected eight different apartments.

Part (b):

The advantage of this that you get to see how the carpet wears with an apartment with children and in an apartment without children. If you use the cluster sampling method is you run the risk of not selecting an apartment with children for example, or maybe not maybe not enough to render it statistically significant. With the stratified random sample it’s a guarantee that you are going to get enough of both.
Student #2:
Part (a):

You could randomly select 2 floors to carpet instead of randomly selecting individual rooms to carpet.

Part (b):

An advantage to this would be to see if kids play a big role in the carpets wear because if you use a cluster sample on the floors you might not get any kids located on that floor.

1a: No cluster, no method (I); b: (E)
2a: No "how" (I); b: weak, but (E)