5. A survey will be conducted to examine the educational level of adult heads of households in the United States. Each respondent in the survey will be placed into one of the following two categories:

- Does not have a high school diploma
- Has a high school diploma

The survey will be conducted using a telephone interview. Random-digit dialing will be used to select the sample.

(a) For this survey, state one potential source of bias and describe how it might affect the estimate of the proportion of adult heads of households in the United States who do not have a high school diploma.

(b) A pilot survey indicated that about 22 percent of the population of adult heads of households do not have a high school diploma. Using this information, how many respondents should be obtained if the goal of the survey is to estimate the proportion of the population who do not have a high school diploma to within 0.03 with 95 percent confidence? Justify your answer.

(c) Since education is largely the responsibility of each state, the agency wants to be sure that estimates are available for each state as well as for the nation. Identify a sampling method that will achieve this additional goal and briefly describe a way to select the survey sample using this method.
Solution

Part (a):
Since random digit dialing will be used, individuals without phones will not be included in the sample. People without a high school diploma are more likely to have lower paying jobs and therefore may not be able to afford a telephone. Thus, the estimated proportion of adult heads of households in the United States without a high school diploma may be less than the true population proportion.

Part (b):
The sample size necessary to estimate the proportion of the population that does not have a high school diploma, \( p \), within .03 with 95% confidence is:

\[
.03 = z^* \left( \sqrt{\frac{p^* (1-p^*)}{n}} \right), \quad \text{or}
\]

\[
.03 = 1.96 \left( \sqrt{\frac{.22 (1-.22)}{n}} \right), \quad \text{so}
\]

\[
n = 0.22(.78) \left[ \frac{1.96^2}{.03} \right] = 732.4651.
\]

Thus, we would need 733 respondents.

Part (c):
To achieve this additional goal, the agency should use stratified random sampling by taking samples within each state. Each state would be a stratum. Within each state, a random sample of adult heads of households would be selected and surveyed. The sample size within each state will be based on the desired precision. Data from the individual states should be combined to obtain the national estimate.
Scoring

Part (a) will be scored as E, P, or I:

Essentially correct (E) if both
1. a possible source of sampling or nonsampling bias is provided and linked to whether or not a person has a high school diploma.
   AND
2. the impact on the survey is correctly described.

Other potential sources of bias:
- Wording of the questions or the tone of the interviewers may make callers less likely to reveal that they do not have a high school diploma (nonsampling bias), forcing the estimate to be too low.
- Calls may be made during the day when individuals with diplomas will be at work, leading to an over reporting of individuals without a high school diploma (sampling bias). Thus, the estimate will be too high.
- More educated people may be more likely to have Caller ID and not answer the phone when they see the call is from a survey firm. Thus, the estimate will be too high.
- Since people do not like responding to cold calling (consider the size of the national do not call list), the response rates might be so low that the survey results are not useful at all. The impact of this bias on the estimate is not clear.

Partially correct (P) if
1 is correct but not 2
   OR
Bias described in 1 is improperly named but 2 is consistent with description.

Incorrect (I) if the source of bias is not reasonable for this survey.
Part (b) will be scored as E, P, or I:

Essentially correct (E) if
- the appropriate critical value, margin of error, and a standard deviation based on a value of $p \ (0.22 \ or \ 0.5)$ are used to calculate the number of necessary respondents
  AND
- work is shown
  AND
- the numeric response is rounded up.

Other possible solutions for Part (b):
- The sample size necessary to estimate the proportion of the population that does not have a high school diploma, $p$, within $0.03$ with 95% confidence is:
  \[
  n = p^* (1 - p^*) \left[ \frac{z^*}{m} \right]^2 = 0.22 \left( \frac{1.96}{0.03} \right)^2 = 732.4651. \ 
  \text{So, we would need 733 respondents.}
  \]
- Conservative approximation ($p^* = 0.5$):
  \[
  n = \left[ \frac{z^*}{2m} \right]^2 = 0.22 \left( \frac{1.96}{2 \times 0.03} \right)^2 = 1067.1111, \text{ so we need 1068 respondents.}
  \]
- Wilson estimate:
  \[
  n + 4 = p^* (1 - p^*) \left[ \frac{z^*}{m} \right]^2 = 0.22 \left( \frac{1.96}{0.03} \right)^2 = 732.4651, \text{ so we need 729 respondents.}
  \]
- Wilson estimate with conservative approach ($p^* = 0.5$):
  \[
  n + 4 = \left[ \frac{z^*}{2m} \right]^2 = 0.22 \left( \frac{1.96}{2 \times 0.03} \right)^2 = 1067.1111, \text{ so we need 1064 respondents.}
  \]

Partially correct (P) if work is shown and no more than one of the following occur:
- an incorrect critical value is used in the calculation
- an incorrect margin of error is used in the calculation
- an incorrect standard deviation is used in the calculation
- numeric response is rounded down or is not an integer.

Incorrect (I) if a solution is provided with no justification or an incorrect formula is used to justify the calculation.
Part (c) is scored E, P, or I:

Essentially correct (E) if stratified random sampling is used to select a random sample from each state. The student must indicate that:

1. the states are the strata. The student must use the phrase “strata” or “stratified.”
   AND
2. a random sample is taken in each state.

Partially correct (P) if only one of the two items necessary for an essentially correct score is provided.

Incorrect (I) if the student suggests that random digit dialing is continued until large enough samples are obtained for all 50 states.

4 Complete Response (3E)
   All three parts essentially correct

3 Substantial Response (2E 1P)
   Two parts essentially correct and 1 part partially correct

2 Developing Response (2E 0P or 1E 2P or 3P)
   2 parts essentially correct and no parts partially correct
   OR
   One part essentially correct and 2 parts partially correct
   OR
   3 parts partially correct

1 Minimal Response (1E 1P or 1E 0P or 0E 2P)
   One part essentially correct and either 0 or 1 parts partially correct
   OR
   No parts essentially correct and 2 parts partially
**Undercoverage/Selection Bias:** Tendency for samples to differ from the corresponding population as a result of systematic exclusion of some part of the population

**Example:** Those without high school diplomas are less likely to have telephones.

**Measurement or Response Bias:** Tendency for samples to differ from the corresponding population because the method of observation tends to produce values that differ from the true value

**Example:** Those without high school diplomas may be embarrassed and lie.

**Nonresponse Bias:** Tendency for samples to differ from the corresponding population because data are not obtained from all individuals selected for inclusion in the sample

**Example:** Those without high school diplomas are more likely to not have a job. Thus, they would be more likely to be home and answer the phone than those with a high school diploma.

**Voluntary Response Bias:** Bias introduced to a sample when individuals can choose on their own whether to participate in the sample

**Not Applicable to This Survey**

**Convenience Sample:** A sample that is selected by convenience

**Not Applicable to This Survey**

**Stratified Random Sample:** A sampling design in which the population is divided into several subpopulations, or strata, and random samples are then drawn from each stratum

**Simple Random Sample:** A sample that is selected from a population in a way that ensures that every different possible sample of the desired size has the same chance of being selected

**Cluster Sample:** A sampling design in which entire groups are chosen at random and every member in the group surveyed

**Multistage sample:** Sampling schemes that combine several sampling methods

**Systematic sample:** A sample drawn by selecting individuals systematically from a sampling frame
### Workshop Exam Materials
#### Score Legend
**2005 AP® Statistics**

<table>
<thead>
<tr>
<th>Question #/Prompt</th>
<th>Sample Identifier</th>
<th>Score Point</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1A</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>1B</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1C</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2A</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>2B</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>2C</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3A</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>3B</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>3C</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>4A</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4B</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>4C</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>5A</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>5B</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>5C</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>6A</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>6B</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>6C</td>
<td>2</td>
</tr>
</tbody>
</table>
Workshop Exam Materials
Scoring Commentary
2005 AP® Statistics

Question 5

Overview

In this question, students are given information on an upcoming survey. The goal is to estimate the proportion of heads of households in the United States with (or without) a high school diploma. Random digit dialing is to be used to select heads of households for inclusion in the sample. In Part (a), students had to identify a potential source of bias for this survey, explain how that source of bias would be related to whether or not the head of household had a high school diploma, and describe the impact of the bias on the estimated proportion. Part (b) tested whether students could determine the sample size that would be needed to obtain an estimate of the proportion with a desired level of precision. In Part (c), students had to recognize that stratified random sampling should be used with states as strata and random sampling should be done within each state. This process would yield both state and national estimates of the proportion of heads of households without a high school diploma.

Sample: 5A
Score: 4

In Part (a), one source of nonresponse bias is correctly identified as not being at home and therefore not answering the phone. In addition, not being at home is linked to not having a high school diploma. The direction of the impact on the survey is correctly described.

Parts (b) and (c) were nicely done. In Part (b), the z* value was reported as 1.959 (incorrectly rounded from the calculator value) rather than 1.96, resulting in a value of 731.718. The student then rounds appropriately.

The student correctly specified the need for stratified random sampling in Part (c). States were clearly used as strata. It would have been better if the student had used a method other than random digit dialing because of the problems with this method identified in Part (a). The student indicated that both state and national estimates could be made using the survey. It could have been noted that proper weighting of the state estimates is needed to obtain the national estimate.

Sample: 5B
Score: 3

The student correctly identifies households not having a telephone as a source of bias (selection or undercoverage) in Part (a). The lack of telephone service is linked to not having a high school diploma. The response should indicate the effect on the estimate derived from this survey due to the source of bias. This student’s answer does not address this requirement.

Parts (b) and (c) were both well done. The answer in (b) provides the correct formula, substitution and computation of the sample size. The decimal answer is not required since 733 is the correct solution if .22 is used for the proportion.

In (c), the student identifies “a stratified random sample” as the desired method. A random sample is to be selected in each state.
Sample: 5C
Score: 2

The student’s response in Part (a) correctly explains the source of bias (*nonresponse*), links having a high school diploma to not being home to respond to the phone survey and indicates the impact of the source of bias on the estimate produced by the survey. This student refers to the bias as “convenience bias,” which is not correct.

The student incorrectly substitutes the variance for the standard deviation in Part (b). The computed decimal answer is indicated and is correctly rounded up.

In Part (c), the student used the word “strata,” identified the states as the strata and provided a method of random selection.
5. A survey will be conducted to examine the educational level of adult heads of households in the United States. Each respondent in the survey will be placed into one of the following two categories:

- Does not have a high school diploma
- Has a high school diploma

The survey will be conducted using a telephone interview. Random-digit dialing will be used to select the sample.

(a) For this survey, state one potential source of bias and describe how it might affect the estimate of the proportion of adult heads of households in the United States who do not have a high school diploma.

One potential source of bias might be nonresponse bias. Those who don't have a high school diploma must work double shifts or longer hours to support their families. If the interviewer calls, and no one is at home to answer, the estimate of those who don't have a high school diploma might be underestimated.

(b) A pilot survey indicated that about 22 percent of the population of adult heads of households do not have a high school diploma. Using this information, how many respondents should be obtained if the goal of the survey is to estimate the proportion of the population who do not have a high school diploma to within 0.03 with 95 percent confidence? Justify your answer.

\[
\begin{align*}
\text{ME} &= \frac{z^* \sqrt{p(1-p)}}{n} \\
0.03 &= 1.959 \frac{0.22(1-0.22)}{n} \\
n &\approx 731.768 \\
n &\approx 732
\end{align*}
\]

You should have 732 respondents.
(c) Since education is largely the responsibility of each state, the agency wants to be sure that estimates are available for each state as well as for the nation. Identify a sampling method that will achieve this additional goal and briefly describe a way to select the survey sample using this method.

I would use stratified random sampling. For each state, I would have 100 or more respondents give the interviewers their information, and then I would pool all the respondents to form the sample for the entire nation. In order to implement this method, I will use a random number generator and find random starting points using each state. Starting from left to right I will find groups of 7 digits. These 7 digits will be a phone number that the interviewer will call. I will then repeat this pick many more times.
5. A survey will be conducted to examine the educational level of adult heads of households in the United States. Each respondent in the survey will be placed into one of the following two categories:

- Does not have a high school diploma
- Has a high school diploma

The survey will be conducted using a telephone interview. Random-digit dialing will be used to select the sample.

(a) For this survey, state one potential source of bias and describe how it might affect the estimate of the proportion of adult heads of households in the United States who do not have a high school diploma.

Because this sample is done by telephone, those without phone access cannot be reached. This may be a source of bias because those who do not have a diploma may not be able to afford telephone service and thus be excluded from this survey.

(b) A pilot survey indicated that about 22 percent of the population of adult heads of households do not have a high school diploma. Using this information, how many respondents should be obtained if the goal of the survey is to estimate the proportion of the population who do not have a high school diploma to within 0.03 with 95 percent confidence? Justify your answer.

For sample size:

\[ n \geq \left( \frac{Z^* \sqrt{p^*(1-p^*)}}{m} \right)^2 \]

\[ 1.96 \sqrt{\frac{0.22(0.78)}{n}} \leq 0.03 \]

\[ n \geq 733 \text{ people in order to produce a 95\% CI with a margin of error less than or equal to 0.03.} \]
(c) Since education is largely the responsibility of each state, the agency wants to be sure that estimates are available for each state as well as for the nation. Identify a sampling method that will achieve this additional goal and briefly describe a way to select the survey sample using this method.

In each of the 50 states, we can also randomly select 1000 households to find proportions within each state (this is also a stratified random sample). For example, we can assign numbers to each household in the state, and randomly choose 1000 with a table of random digits.
5. A survey will be conducted to examine the educational level of adult heads of households in the United States. Each respondent in the survey will be placed into one of the following two categories:

- Does not have a high school diploma
- Has a high school diploma

The survey will be conducted using a telephone interview. Random-digit dialing will be used to select the sample.

(a) For this survey, state one potential source of bias and describe how it might affect the estimate of the proportion of adult heads of households in the United States who do not have a high school diploma.

convenience bias - some adults won’t be able to be reached (may not answer phone, be out of house, etc. and therefore the people who are reached and respond might not represent true population

This bias may lead to overestimating prop. of adult who have a diploma. Olm adults with a high school diploma typically have jobs and won’t be home to answer phone.

(b) A pilot survey indicated that about 22 percent of the population of adult heads of households do not have a high school diploma. Using this information, how many respondents should be obtained if the goal of the survey is to estimate the proportion of the population who do not have a high school diploma to within 0.03 with 95 percent confidence? Justify your answer.

\[
1.96 \left( \frac{.22(.78)}{\sqrt{n}} \right) \leq .03
\]

\[
\sqrt{1.96} \left( \frac{.1716}{\sqrt{n}} \right) \leq .03 \sqrt{n}
\]

\[
1.96 \leq \frac{.01716}{.0153} \sqrt{n}
\]

\[
.01716 \leq .0153 \sqrt{n}
\]

\[
11.216 \leq \sqrt{n}
\]

\[
125.99 \leq n
\]

\[
126 = n
\]

if we contact 1260 respondents, n is big and our margin of error will decrease
(c) Since education is largely the responsibility of each state, the agency wants to be sure that estimates are available for each state as well as for the nation. Identify a sampling method that will achieve this additional goal and briefly describe a way to select the survey sample using this method.

Simple Random Sampling

random digit dialing of phone numbers (and each state draws a certain number of phone numbers; proportion choose these)

strata (states)
in each state, use random digit dialing of residents of state assign each telephone number a random digit number; choose a sample size of random digits with replacement and call those numbers

GO ON TO THE NEXT PAGE.