Problems Before Experimental Design Assessment #1 Answers

1. Below is a list of cities. Assign numbers to each city from left to right starting at 01. Use the line of random digits shown below to choose a sample of 6 cities. Circle the cities that the table selects.

71950  22494  00369  51269  87073  73694  97751  17857  52352  21392  22930  43776

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>Haworth, NJ</td>
<td>02 Glen Rock, NJ</td>
<td>03 Closter, NJ</td>
</tr>
<tr>
<td>04</td>
<td>Red Bank, NJ</td>
<td>05 Glassboro, NJ</td>
<td>06 Rockville, MD</td>
</tr>
<tr>
<td>07</td>
<td>Silver Springs, MD</td>
<td>08 Portland, OR</td>
<td>09 Clemmons, NC</td>
</tr>
<tr>
<td>10</td>
<td>Winston-Salem, NC</td>
<td>11 Wilmington, NC</td>
<td>12 Wayne, NJ</td>
</tr>
<tr>
<td>13</td>
<td>Maywood, NJ</td>
<td>14 Montvale, NJ</td>
<td>15 Woodcliff Lake, NJ</td>
</tr>
</tbody>
</table>

2. A headline in a local newspaper announced “Video game playing can lead to better spatial reasoning abilities.” The article reported that a study found “statistically significant differences” between teens who play video games and teens who do not, with teens who play video games testing better in spatial reasoning. Do you think the headline was appropriate? Explain.

No. We are not certain if the video games help with the spatial reasoning or if kids with good spatial reasoning like to play video games.

3. A college group is investigating student opinions about funding of the military. They phone a random sample of students at the college, asking each person one of these questions (randomly chosen):

A: “Do you think that funding of the military should be increased so that the United States can better protect its citizens?”

B: “Do you think that funding of the military should be increased?”

Which question do you expect will elicit greater support for increased military funding? Explain. What kind of bias is this?

Choice A. This question is attempting to tie into peoples fears of being harmed and that the only way to prevent ourselves from harm is more military spending. Response Bias
4. A psychology professor wants to study the behavior of the 8500 college students at
the State University. She decides to obtaining sample of 50 students and asking,
“What is the average number of hours you sleep on a weekday?” For each of the
following, identify the type of sample obtained. (*convenience, voluntary, systematic
random, simple random, or stratified random*)

a. Each student is assigned a number 0001 to 8500. A number from
001 to 170 is randomly selected and every 170th students from the
list from that point on is then included in the sample.

**Systematic Random**

b. The psychology professor asks 50 students in her classes.

**Convenience**

c. Students are listed by their school residence location (dormitory or
apartment building). Four residence locations are randomly
chosen; then one floor for each is randomly selected. Students
from these locations are then randomly chosen and combined to be
the sample of the population.

**Stratified Random**

d. The psychology professor posts a listing in the school newspaper
asking for students to come in to be interviewed.

**Voluntary**

e. Students are listed by number 0001-8500 and a computer is used to
generate a list of 50 numbers representing the students to be used
in the sample.

**Simple Random**
5. A member of the City Council has proposed a resolution opposing construction of a new state prison there. The council members decide they want to assess public opinion before they vote on this resolution.

   a. Below are some of the methods that are proposed to sample local residents to determine the level of public support for the resolution. Match each with one of the listed sampling techniques.

   **Voluntary**  Place an announcement in the newspaper asking people to call their council representatives to register their opinions. Council members will tally the calls they receive.

   **Convenience**  Have each council member survey 50 friends, neighbors, or co-workers.

   **Simple Random**  Have the Board of Elections assign each voter a number, then select 400 of them using a random number table.

   **Judgment**  Go to a downtown street corner, a grocery store, and a shopping mall; interview 100 typical shoppers at each location.

   **Stratified Random**  Randomly pick 50 voters from each election district.

   **Systematic Random**  Call every 500th person in the phone book.

   **Multistage**  Randomly pick several city blocks, then randomly pick 10 residents from each block.

   **Cluster**  Randomly select several city blocks; interview all the adults living on each block.

   b. The City Council decides to conduct a telephone poll. Pollsters ask a carefully chosen random sample of adults this question: “Do you favor the construction of a new prison to deal with the high level of violent crime in our State?” In what way might the proportion of “Yes” answers fail to accurately reflect true public opinion? Explain briefly. What kind of bias is this?

   **People who answer “Yes” could possible be answering “Yes” that they want something to be done about violent crimes, not that they want to build new prison. The wording creates a **Response Bias**.**
6. Administrators at a hospital are concerned about the possibility of drug abuse by people who work there. They decide to check on the extent of the problem by having a random sample of the employees undergo a drug test.

   a. Several plans for choosing the sample are proposed. Name the sampling strategy in each.

   **Stratified Random** There are four employee classifications: doctors, medical staff (nurses, technicians, etc.) office staff, and support staff (custodians, maintenance, etc.). Randomly select ten people from each category.

   **Simple Random** Each employee has a 4-digit ID number. Randomly choose 40 numbers.

   **Systematic Random** At the start of each shift, choose every tenth person who arrives for work.

   **Cluster** Randomly select a department (say, radiology) and test all the people who work in that department – doctors, nurses, technicians, clerks, custodians, etc.

   b. Explain why the last plan suggested above, sampling an entire department, might be biased. Be sure to name the kind(s) of bias you describe.

   By not allowing all groups to be part of the study we will only get one general point of view from a specific group. This would under-coverage.

   c. Listed are the names of the 20 pharmacists on the hospital staff. Use the random numbers listed below to select three of them to be in the sample. Clearly explain your method.

      Spiridinov    Lavine    Grubb

   d. Name and describe the kind of bias that might be present if the administration decides that instead of subjecting people to random testing they’ll just…

      i. interview employees about possible drug abuse.

      People might lie. **Response Bias**

      ii. ask people to volunteer to be tested.

      Only those not taking drugs would Volunteer. **Under-coverage**
7. Julie is an 80% free throw shooter. After practice she attempts 10 free throws.

a. Assign digits as shown in the following way.
   Make: 1-8  Miss: 9-0
   Look at groups of 10 single digit numbers. If at least 7 of the numbers represent a make than that simulation is a success. There were 16 successes so the experimental probability is $\frac{16}{20} = 0.8$

b. Assign digits as shown in the following way.
   Make: 1-8  Miss: 9-0
   Look at groups of 4 single digit numbers. If at least 1 of the numbers represent a make than that simulation is a success. There were 20 successes so the experimental probability is $\frac{20}{20} = 1$

8. In a putting contest at a local golf course, Mr. Postman bragged that he sinks putts under 6 feet 80% of the time.

a. Assign digits as shown in the following way.
   Make: 1-8  Miss: 9-0
   Look at groups of 10 single digit numbers. If exactly 8 of the numbers represent a make than that simulation is a success. There were 2 successes so the experimental probability is $\frac{7}{20} = 0.35$

b. Assign digits as shown in the following way.
   Make: 1-8  Miss: 9-0
   Look at groups of 10 single digit numbers. If at least 7 of the numbers represent a make than that simulation is a success. There were 18 successes so the experimental probability is $\frac{18}{20} = 0.9$

c. Assign digits as shown in the following way.
   Make: 1-8  Miss: 9-0
   Look at groups of 2 single digit numbers. If the first is a miss(0-9) and the second make(1-8) than the simulation is a success. There were 2 successes so the experimental probability is $\frac{2}{20} = 0.1$
9. Mr. Postman was the MVP of his college intramural softball league, where he got a hit 60% of the time.

a. Assign digits as shown in the following way.
   Hit: 1-6 No Hit: 7-0
   Look at groups of 20 single digit numbers. If exactly 12 of the numbers represent a Hit than that simulation is a success. There were 2 successes so the experimental probability is 2/10 = 0.2

b. Assign digits as shown in the following way.
   Hit: 1-6 No Hit: 7-0
   Look at groups of 20 single digit numbers. If there are less than 11 of the numbers represent a hit than that simulation is a success. There were 3 successes so the experimental probability is 3/10 = 0.3

c. Assign digits as shown in the following way.
   Hit: 1-6 No Hit: 7-0
   Look at 40 single digit numbers, if the number represents a hit than the simulation is a success. There were 21 successes so the experimental probability is 21/40 = 0.525

10. a. Describe how you will use a random number table to conduct this simulation.

   Assign digits as shown in the following way.
   Good CD: 01-95 Bad CD: 95-00
   Choose groups of 2 digits numbers until you get 5 Good CDs. Record the number of CDs chosen.

b. Show three trials by clearly labeling the random number table given below. Specify the outcome for each trial.

   Trial Simulation Outcome
   #1 03242 50692 18977 28370(5 CDs required)
   #2 78695 21402 85525 81183(5 CDs required)
   #3 60809 06765 39996 81915(5 CDs required)
11. Assign digits as shown in the following way.

“M”: 01-25  “&”: 26-35  “Try Again”: 36-00
Choose groups of 2 digits numbers until you get 2 “M” and one “&”. Record the number of CDs chosen.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th>(6 Bags)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>M</td>
<td>&amp;</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trail 1:</td>
<td>6 9 0 7 4</td>
<td>9 1 9 7 6</td>
<td>3 3 5 8 4</td>
</tr>
<tr>
<td>Trail 2:</td>
<td>4 8 3 2 4</td>
<td>7 7 9 2 8</td>
<td>3 1 2 4 9</td>
</tr>
</tbody>
</table>

12. A large manufacturer of batteries knows that, historically, 10% of its batteries come off the production line defective, and the remaining 90% of batteries come off the production line in working condition. Conduct a simulation to estimate how many batteries the company needs to pull off the production line in order to be sure of ending up with 10 working batteries.

a. Describe how you will use a random number table to conduct this simulation.

Assign digits as shown in the following way.

Defective: 1  Working: 2-0
Choose single digits numbers until you get 10 Working batteries. Record the number of batteries chosen.

b. Show three trials by clearly labeling the random number table given below. Specify the outcome of each trial.

Trial 1:
10242 50692 18977 28370 82669 83236 77479 90618 43707 78695
12 batteries chosen to get 10 working batteries

Trial 2:
81183 48554 60809 39996 81915 25404 33366 92082 04822 79866
12 batteries chosen to get 10 working batteries

Trial 3:
06765 67041 20479 54612 13411 36837 69983 53082 43589 27865
11 batteries chosen to get 10 working batteries
Every Monday a local radio station gives coupons away to 50 people who correctly answer a question about a news fact from the previous day’s newspaper. The coupons given away are numbered from 1 to 50, with the first person receiving coupon 1, the second person receiving coupon 2, and so on, until all 50 coupons are given away. On the following Saturday, the radio station randomly draws numbers from 1 to 50 and awards cash prizes to the holders of the coupons with these numbers. Numbers continue to be drawn without replacement until the total amount awarded first equals or exceeds $300. If selected, coupons 1 through 5 each have a cash value of $200, coupons 6 through 20 each have a cash value of $100, and coupons 21 through 50 each have a cash value of $50.

(a) Explain how you would conduct a simulation using the random number table provided below to estimate the distribution of the number of prize winners each week.

(b) Perform your simulation 3 times. (That is, run 3 trials of your simulation.) Start at the leftmost digit in the first row of the table and move across. Make your procedure clear so that someone can follow what you did. You must do this by marking directly on or above the table. Report the number of winners in each of your 3 trials.

```
72749 13347 65030 26128 49067 02904 49953 74674 94617 13317
81638 36566 42709 33717 59943 12027 46547 61303 46699 76423
38449 46438 91579 01907 72146 05764 22400 94490 49833 09258
```

Part (a):

1. **Scheme:** Obtain a two-digit random number from the random number table. If it is between 01 and 50, use it to represent the selected ticket. Ignore numbers 00 and 51 – 99.

2. **Stopping Rule:** Determine the amount of the prize associated with the chosen ticket, and add this amount to the total amount awarded so far. If the total amount awarded so far is less than $300, repeat this process.

3. **Count:** Note the total number of winners.

4. **Non-Replacement:** Ignore any ticket number that has already been awarded a prize in this trial.

Part (b):

Solution will depend on answer to part (a).

For example, using scheme above:

<table>
<thead>
<tr>
<th>Trial 1</th>
<th>Trial 2</th>
<th>Trial 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>72</td>
<td>02</td>
<td>06</td>
</tr>
<tr>
<td>74</td>
<td>200</td>
<td>100</td>
</tr>
<tr>
<td>91</td>
<td>50</td>
<td>28</td>
</tr>
<tr>
<td>33</td>
<td>100</td>
<td>50</td>
</tr>
<tr>
<td>47</td>
<td>ignore</td>
<td>48</td>
</tr>
<tr>
<td>65</td>
<td>ignore</td>
<td>50</td>
</tr>
<tr>
<td>03</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>Total number of winners: 3</td>
<td>Total number of winners: 3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trial 2</th>
<th>Total so far</th>
<th>Trial 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>02</td>
<td>200</td>
<td>06</td>
</tr>
<tr>
<td>200</td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>28</td>
<td>50</td>
<td>29</td>
</tr>
<tr>
<td>250</td>
<td></td>
<td>50</td>
</tr>
<tr>
<td>48</td>
<td>50</td>
<td>60</td>
</tr>
<tr>
<td>300</td>
<td></td>
<td>200</td>
</tr>
</tbody>
</table>

Total number of winners: 3
The dentists in a dental clinic would like to determine if there is a difference between the number of new cavities in people who eat an apple a day and in people who eat less than one apple a week. They are going to conduct a study with 50 people in each group.

Fifty clinic patients who report that they routinely eat an apple a day and 50 clinic patients who report that they eat less than one apple a week will be identified. The dentists will examine the patients and their records to determine the number of new cavities the patients have had over the past two years. They will then compare the number of new cavities in the two groups.

a. Why is this an observational study and not an experiment?

b. Explain the concept of confounding in the context of this study. Include an example of a possible confounding variable.

c. If the mean number of new cavities for those who ate an apple a day was statistically significantly smaller than the mean number of new cavities for those who ate less than one apple a week, could one conclude that the lower number of new cavities can be attributed to eating an apple a day? Explain.

   a. The student can appeal to any of three reasons in judging this study not an experiment:

      1. there is no random assignment of subjects to treatments;
      2. there are no treatments imposed;
      3. existing data is being used.

   b. Two variables are confounded if their effect on the number of new cavities cannot be distinguished from one another. The student must mention not only that the confounding variables may affect the outcome but that they have differential effects within the two groups. For instance: confounding would occur if patients who eat an apple a day differ from those who eat less than one apple a week on some variable that is related to dental health. In this example, diet or general level of health are examples of what might be confounding variables. For example, it is possible that people who eat an apple a day are more nutrition conscious and have a more healthy diet in general than those who eat one or fewer apples per week, and this might explain the observed difference in dental health.

c. No, because it is not an experiment, and cause-and-effect conclusions cannot be drawn from an observational study.

   OR

No, because there are possible confounding variables.
15. A polling organization is investigating public opinion about cloning. They phone a random sample of 1200 adults, asking each person one of these questions (randomly chosen):

A: “Do you favor allowing doctors to use cloned cells in attempts to find cures for such terrible diseases as Alzheimer’s, diabetes, and Parkinson’s?”

B: “Should research scientists be allowed to use cloned human embryos in their experiments?”

Which question do you expect will elicit greater support for cloning? Explain. What kind of bias is this?

A: “Do you favor allowing doctors to use cloned cells in attempts to find cures for such terrible diseases as Alzheimer’s, diabetes, and Parkinson’s?”

This question presents only the positive aspects of cloning therefore when a person reads the question it is likely to cause them to respond in favor of cloning. Since the wording of the questions is meant to elicit a certain response this is an example of Response Bias
Solution

Part (a):

The administrators could number an alphabetical list of students from 1 to 2,500. They could then use a random number generator from a calculator or computer to generate 200 unique random integers from 1 to 2,500. The students corresponding to those 200 numbers would be asked to participate in the survey.

Part (b):

One possible stratification variable might be the school level of the student (elementary, middle, high school). The students’ perceptions of the importance of good nutrition in food served may differ depending on the students’ ages and therefore on school levels. For example, there may be a difference between what elementary students value in food served as opposed to middle school and high school students.

Part (c):

One statistical advantage of using stratified random sampling as opposed to simple random sampling is, for example, if the elementary, middle and high school strata create groups that differ with respect to what they value — and are therefore more homogeneous with respect to opinion on this issue — then for the same overall sample size a more accurate estimate of the overall proportion of students who are satisfied with the food under this contract may result. Another advantage is that stratified random sampling guarantees that each of the school-level strata will have some representation, because it is possible that a simple random sample would miss one or more of the strata completely.
Part (a):

of new car buyers who bought model E, \( p \), which has a value of \( p = \frac{2,323}{297,354} \approx 0.0078 \). The expected value of the number of model E buyers in a simple random sample of 2,000 is therefore \( n \times p = 2,000 \times 0.0078 \approx 15.62 \). The variance is \( n \times p \times (1 - p) = 2,000 \times 0.0078 \times (1 - 0.0078) = 15.50 \), so the standard deviation is \( \sqrt{15.50} \approx 3.94 \).

Part (b):

For the reason given in part (a), the binomial distribution with \( n = 2,000 \) and \( p = 0.0078 \) can be used here. The probability that the sample would contain fewer than 12 owners of model E is calculated from the binomial distribution to be

\[
\sum_{x=0}^{11} \binom{2,000}{x} (0.0078)^x (0.9922)^{2,000-x} = 0.147.
\]

This probability is small enough that the result (fewer than 12 owners of model E in the sample) is not likely, but this probability is also not small enough to consider the result very unlikely.

Part (c):

Stratified random sampling addresses the concern about the number of owners for models D and E. By stratifying on car model and then taking a simple random sample of at least 12 owners from the population of owners for each model, the company can ensure that at least 12 owners are included in the sample for each model while maintaining a total sample size of 2,000. For example, the company could select simple random samples of sizes 755, 647, 560, 22 and 16 for models A, B, C, D and E, respectively, to make the sample size approximately proportional to the size of the owner population for each model.