HOMEWORK #7: SOLUTIONS TO OPTIONAL PROBLEMS

Chapter 3.

3.14 (a) A diagram is shown below. (b) Label the subjects from 01 through 20. From line 131, we choose

05, 19, 04, 20, 16, 18, 07, 13, 02, and 08;
that is, Decker, Travers, Chen, Ullmann, Quinones, Thompson, Fluharty, Lucero, Afifi, and Gerson for one group, and the rest for the other. See note on page 50 about using Table B.

3.43. First, create a column containing the numbers 0 to 439, representing the 440 districts. Just follow the steps listed below. It is self explanatory.

You can use Make Patterned Data to create a variety of data patterns. The Make Patterned Data command is not as quick and interactive as Autofill, but it allows you to more easily create large data sets with repeated values.

Calc > Make Patterned Data >

Simple Set of Numbers - fill a column with a pattern of equally spaced numbers

More details follow:

Calc > Make Patterned Data > Simple Set of Numbers

Provides an easy way to fill a column with numbers that follow a pattern, such as the numbers 1 through 100, or five sets of 1, 2, and 3. This is very useful for entering factor levels for analysis of variance designs.

With this command, you can specify a pattern of equally spaced numbers, such as 10 20 30. To specify an unequally spaced pattern of numbers, such as the numbers 10 20 50, use the Arbitrary Set of Numbers command.

Dialog box items

Store patterned data in: Specify a column to store the new data.
From first value: Enter the starting point of the sequence. You may enter a number or a stored constant (K).

To last value: Enter the end point of the sequence. You may enter a number or a stored constant (K).

In steps of: Enter a number if you want to increment the starting value by a specified amount until you reach the end value.

List each value: Enter the number of times you want each value listed.

List the whole sequence: Enter the number of times you want the entire group of numbers listed

Next, ask Minitab to choose a random sample from the column you just made. Again, just follow the steps shown below. It is fast and easy.

Calc > Random Data > Sample From Columns

Randomly samples the same rows from one or more columns. You can sample with replacement (select the same row more than once), or without replacement (select each row only once). To generate the same random sample on multiple occasions, see Calc > Set Base.

Dialog box items

Sample ___ rows from column(s): Specify the number of rows to randomly select, then enter the column(s) you want to sample from. If you sample from several columns at once, they must all have the same length.

Store samples in: Specify the column(s) where you want to store the sampled values. The number of storage columns must be the same as the number of columns sampled from.

Sample with replacement: Check to sample with replacement. Leave unchecked to sample without replacement (sample size must be less than or equal to the length of the columns).

3.55(a) This design would omit households without telephones or with unlisted numbers. Such households would likely be made up of poor individuals (who cannot afford a phone), those who choose not to have phones, and those who do not wish to have their phone numbers published. (b) Those with unlisted numbers would be included in the sampling frame when a random-digit dialer is used.
3.56 (a) There were 14,484 responses. (Note that we have no guarantee that these came from 14,484 distinct people; some may have voted more than once.) (b) This voluntary response sample collects only the opinions of those who visit this site and feel strongly enough to respond.

3.67 (a) The sample size for Hispanics was smaller. Smaller sample sizes give less information about the population, and therefore lead to larger margins of error (with the same confidence level). (b) The sample size was so small, and the margin of error so large, that the results could not be viewed as an accurate reflection of the population of Cubans.

3.75 (a) The scores will vary depending on the starting row. Note that the smallest possible mean is 61.75 (from the sample 58, 62, 62, 65) and the largest is 77.25 (from 73, 74, 80, 82). (b) Answers will vary; shown below are two views of the (exact) sampling distribution. The first shows all possible values of the experiment (so the first rectangle is for 61.75, the next is for 62.00, etc.); the other shows values grouped from 61 to 61.75, 62 to 62.75, etc. (which makes the histogram less bumpy). The tallest rectangle in the first picture is 8 units; in the second, the tallest is 28 units.

   Note: These histograms were found by considering all \( \binom{104}{4} = 210 \) of the possible samples. It happens that half (105) of those samples yield a mean smaller than 69.4 and half yield a greater mean.

3.89. (a) One possible population: all full-time undergraduate students in the fall term on a list provided by the registrar. (b) A stratified sample with 125 students from each year is one possibility. (c) Mailed (or e-mailed) questionnaires might have high nonresponse rates. Telephone interviews exclude those without phones and may mean repeated calling for those who are not home. Face-to-face interviews might be more costly than your funding will allow. There might also be some response bias: Some students might be hesitant about criticizing the faculty (while others might be far too eager to do so).
Chapter 5

5.1.  (a) This could be reasonably viewed as binomial with \( n = 500 \) and \( p = 1/12 \), because there is a fixed number (500) of independent trials with the same chance of success (1/12) on each try. (b) Not binomial: There is no fixed number of attempts \( (n) \). (c) Not binomial: There are no separate “trials” or “attempts” being observed here.

5.14.  For \( \hat{p} \), \( \mu = 0.49 \) and \( \sigma = \sqrt{p(1-p)/n} \approx 0.01576 \). As \( \hat{p} \) is approximately normally distributed with this mean and standard deviation, we find

\[
P(0.46 < \hat{p} < 0.52) \approx P(-1.90 < Z < 1.90) = 0.9426
\]

(Exact calculation gives 0.94565.)

5.16.  When \( n = 250 \), the distribution of \( \hat{p} \) is approximately normal with mean 0.49 and standard deviation 0.03162 (about twice that in Exercise 5.14). When \( n = 4000 \), the standard deviation drops to 0.00790 (half as big as in Exercise 5.14). Therefore,

\[
\begin{align*}
\text{\( n = 250 \):} & \quad P(0.46 < \hat{p} < 0.52) \approx P(-0.95 < Z < 0.95) = 0.6578 \\
\text{\( n = 4000 \):} & \quad P(0.46 < \hat{p} < 0.52) \approx P(-3.80 < Z < 3.80) = 0.9998
\end{align*}
\]

Larger samples give a better probability that \( \hat{p} \) will be close to the true proportion \( p \). (Exact calculation of the first probability gives 0.68853, but this more accurate answer does not change our conclusion.)

5.20.  (a) \( \mu = (1500)(0.7) = 1050 \) and \( \sigma = \sqrt{15} \approx 17.7482 \) students. (b) \( P(X = 1000) \approx P(Z = -2.82) = 0.9976 \) (0.9978 with continuity correction; see the first line of the table below). (c) \( P(X = 1201) \approx P(Z = 8.51) < 0.00005 \) (it’s very small). (d) With \( n = 1700 \), \( P(X = 1201) \approx P(Z = 0.58) = 0.2810 \). Other answers are shown in the second line of the table below.

<table>
<thead>
<tr>
<th>Continuity correction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table normal</td>
</tr>
<tr>
<td>0.9976</td>
</tr>
<tr>
<td>0.2810</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Continuity correction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table normal</td>
</tr>
<tr>
<td>0.9978</td>
</tr>
<tr>
<td>0.2877</td>
</tr>
</tbody>
</table>