Interactive Statistics 3rd Edition: Chapter 2 Full Solutions

2.1
35% is a parameter – it is a numerical summary of the population.
28% is a statistic – it is a numerical summary of a sample from the population.

2.2
(a) The value of 9% is a parameter – it is a numerical summary of the population.
(b) The value of 12.5% is a statistics – it is a numerical summary of a sample from the population.

2.3
The proportion of register voters in Ann Arbor who would vote "Yes" on Proposal A is an example of (c) a parameter.

2.4
(a) The population consists of (the planned vote for) the 100 U.S. Senators.
(b) $N = 100$.
(c) The sample consists of (the planned vote for) the ten selected U.S. Senators.
(d) $n = 10$.

2.5
For a simple random sample of $n = 200$ and the number of items that were defective = 5. All we can say is that (c) the percent of defective items in the sample is 5/200 = 2.5%. $\hat{p}$ = 5/200 = 0.025 or 2.5%.

2.6
(a) 22.4.
(b) 19. No.
(c) 22.5. No. No.
(d) 18 and 21. The sample mean age is 19.5.
18 and 26. The sample mean age is 22.
20 and 27. The sample mean age is 23.5.
20 and 21. The sample mean age is 20.5.
20 and 26. The sample mean age is 23.
27 and 21. The sample mean age is 24.
27 and 26. The sample mean age is 26.5.

2.7
(a) The value of 34 is a statistic.
(b) Response bias.

2.8
(a) Statistic since the 54% is a sample percentage, not of the population.
(b) Nonresponse bias.

2.9
Response bias.

2.10
Nonresponse bias.

2.11
Nonresponse bias.

2.12
False. When only 272 out of 1000 people respond, for a 27.2% response rate, this results in nonresponse bias.

2.13
The best answer is (c).

2.14
Results will vary.
(a) Using the calculator with a seed value of 291, the selected persons are 39 (79 years old), person 3 (75 years old) and person 24 (70 years old). The average age is 74.67 years.
(b) 22 (36 years old), 34 (89 years old) and 29 (89 years old). The average age is 71.33 years, different from the mean calculated in part (a).
(c) 69.19 years.
(d) 74.67 and 71.33 are statistics, while the mean in part (c) is a parameter.

2.15
(a) Yes, each sample of size 20 has the same chance as any other sample of size 20 to be selected (assuming all tags are exactly the same and the box is thoroughly mixed).
(b) It is drawn without replacement.

2.16
(a) 12/100.
(b) With the graphing calculator the selected employees are: 77, 51, 72, 40, 71, 42, 17, 34, 62, 23, 35, and 12.
From the random table the selected employees are: 7, 5, 69, 76, 28, 33, 78, 70, 99, 98, 42, and 80.

2.17
(a) H_0: The population proportion of dissatisfied customers equals 0.10.
H_1: The population proportion of dissatisfied customers is less than 0.10.
(b) With the calculator the selected customers are: 34318, 15553, 8461, and 614. With the random table the selected customers are: 15409, 23336, 29490, and 30414.
(c) Type II error.
(d) 0.21.
(e) No, the p-value is > 0.05.
(f) (ii) Statistic.

2.18
Using the TI: Label the sites from 1 to 80. With a seed value of 29 the five labels, and thus sites, selected at random are as follows: Sites #50, #66, #43, #49, and #74. Using the Random Number Table: Since there are more than 10, but less than 100, sites, we can use double-digit labels, such as 01, 02, 03, through 80. With Row 10 of the table, starting at Column 1, the simple random sample of sites consists of sites #47, #53, #68, #57, and #34.

2.19
Stratified random sampling. You are dividing up your population by class rank then selecting 100 students from within each stratum at random.

2.20
(a) Stratifying by declared major (field of study) would be a good stratification variable. The cost of textbooks is likely to vary from one field of study to the next, but not vary as much within a particular field of study (many students have to take the same basic classes for their field of study and hence may be paying similar textbook costs.) You would also learn about the costs for each field of major as well as costs overall.
(b) Stratifying by gender may not be as useful as a stratification variable. The cost of textbooks may vary a lot for females and vary a lot for males, and there may not be much variation between males and females. However, if you wanted to learn about the costs for each gender as well as overall, you might wish to stratify by gender.
(c) Stratifying by class rank may be useful as a stratification variable. The cost of textbooks may vary from one class-rank to another. However, there may be quite a lot of variation within each class-rank as well. If you wanted to learn about the costs for each class-rank as well as overall, you might wish to stratify by class-rank.

2.21

(a) 45 8th-grade students; 25 10th-grade students; 48 12th-grade students. For a total in the sample of 118 students.
(b) With the calculator the selected students are: 193, 127, 430, 100, and 427. With the random number table the selected students are: 241, 304, 22, 364, and 151.

2.22
(a) 100/200 = 1/2 = 0.50 or 50%.
(b) 100/1000 = 1/10 = 0.10 or 10%.
(c) The chance of being chosen is 0.10. This is NOT a simple random sample. For this stratified random sampling plan each possible sample would contain exactly 100 males and 20 females. All samples of size 120 are not equally likely (as it should be for simple random sample). Some samples of size 120 are not even possible, for example, having 120 males.

2.23
We should use a weighted average: $\left(\frac{40}{100}\right)(70)+\left(\frac{60}{100}\right)(63)=65.8$ inches

2.24
(a) 20/200 = 0.10.
(b) With the calculator the first five females selected in the sample are: 132, 195, 147, 171, and 85. With the random number table: To each woman we assigned 5 numbers, for example 1, 201, 401, 601 and 801. The numbers from the table were 521, 625, 391, 646, and 369. So the first five selected females are: 121, 25, 191, 46, and 169.
(c) The weighted average is: (0.75)8 + (0.25)5 = 7.25.

2.25
(a) Stratified random sampling.
(b) With the calculator the first five selected homes are: 386, 81, 379, 211, and 156. With the random number table the first five selected homes are: 94, 299, 396, 378, and 363.
(c) (0.60)(2100) + (0.40)(2600) = 2300 square feet.

2.26
(0.20)(16) + (0.50)(43) + (0.30)(71) = 46 years.

2.27
(a) Stratified random sampling.
(b) With the calculator the first six homes selected from the 1000 homes in County I are: 918, 193, 902, 502, 370, and 5. With the random number table the first six homes selected from the 1000 homes in County I are: 963, 19, 197, 705, 463, and 79 where the homes are labeled 0 to 999.
(c) (0.50)(175) + (0.30)(200) + (0.20)(195) = 186.5 thousands of dollars or $186,500.

2.28
(a) With the calculator, the label of the first student selected is 3. With the random number table, the label of the first student selected is 3.
(b) The students in the sample are those with ID numbers 3, (3 + 4 =) 7, (7 + 4 =) 11, and (11 + 4=) 15, for a total of 4 students.

2.29
(a) With the calculator with the first 100 addresses labeled 1 through 100, the sample is: Addresses = #79, #179, #279, #379, #479. Using the random number table with the first 100 addresses labeled 01, 02, ... , 98, 99, 00, the sample is: Addresses = #30, #130, #230, #330, #430.
(b) 1/100 = 0.01 or 1% There are 100 possible systematic samples of size 5, each equally likely.

2.30
(a) Stratified random sampling.
(b) No, if there are more people whose family name begins with say A and fewer people whose family name begins with Z, then those with Z will have a higher chance of being selected.
(c) Selection bias, a systematic tendency to exclude those with unlisted phone numbers.

2.31
Yes, the chance is 1/5 or 0.20 or 20%. There are 5 possible systematic samples, each equally likely.

2.32
(a) 1/14 = 0.0714.
(b) Note that there are 555 members so 555/14 → 39.64 or 39 groups of 14 and one last group of 9. For the 17th member to be selected, the starting point must have been a 3, that is, the 3rd member in each group of 14 is selected. This will result in a sample of 40 members.

2.33
(a) 1/8 = 0.125.
(b) Note that there are 350 members so 350/8 → 43.75 or 43 groups of 8 and one last group of 6. For the 300th member to be selected, the starting point must have been a 4, that is, the 4th member in each group of 8 is selected. This will result in a sample of 44 members.

2.34
(a) 1/40 = 0.025.
(b) Using the TI calculator with seed of 19, and with the first 40 members labeled 1 through 40, the first five members are 7, 47, 87, 127, and 167. Using row 15, column 1, and labeling the first 40 members as 01, 02, ... , 39, 40, the first 5 members are 7, 47, 87, 127, and 167.
(c) Note that there are 2220 members so 2220/40 → 55.5 or 55 groups of 40 and one last group of 20. Since the starting point was a 7, that is, the 7th member in each group of 40 was selected. This would result in a complete sample of 56 members.

2.35
False, the chance depends on the number of clusters.

2.36
(a) With the calculator, the two departments selected were 2 and 4; Chemistry and Mathematics. Using the random number table, the two department selected were 2 and 3; Chemistry and Biology.
(b) With calculator, the sample size is 80. With random number table, the sample size is 90.

2.37
(a) Cluster sampling. The faculty are grouped into departments which serve as clusters. Six of the clusters are selected at random. All of the units in the cluster are in the sample.
(b) Label the list of the 60 departments 1, 2, 3, ..., 60. The numbers generated and thus the departments selected using the TI calculator with seed = 79 are: 54, 37, 49, 5, 15, and 43. If you are using the random number table, you might label the list of the 60 departments 01, 02, 03, ... , 60. The numbers generated and thus the departments selected starting at row 60, column 1 are: 23, 22, 47, 40, 25, and 37.
(c) Yes we can determine the chance that any specific professor will be selected. The chance of selecting a professor is the same chance that his/her department or cluster will be selected. The reason being if a cluster is selected, every element in the cluster is selected. Therefore the probability that a cluster is selected is 6/60 = 0.10, the chance that any specific professor will be selected.

2.38

(a) This is cluster sampling since the students are first divided into clusters (undergraduate classes). A class/cluster is then selected using a simple random sample and all students from that class/cluster are sampled.

(b) No, since not all students take the same number of classes. Students who attend more classes have a greater chance of being selected.

(c) No it is not biased since the cluster was selected at random. It is a case of poor design together with bad luck. When clustering the variability between clusters should not be more important than the variability within clusters. Here we have poor design because there might be more variability between a class with many students on athletic scholarship and a class without any, than variability within each of those classes.

2.39

(a) The type of sampling performed in each dorm is cluster sampling, with the rooms forming the clusters and 3 clusters were selected at random from each of the four dorms.

(b) If each cluster selected has one student, we would have 3 students from each of the four dorms for a minimum sample size of 12 students.

(c) If each cluster selected has three students, we would have 9 students from each of the four dorms for a maximum sample size of 36 students.

(d) We do not know, it will depend on how many clusters selected are rooms with women. The number of women could be as low as 0 and as high as 36.

(e) No, there will be anywhere from 3 to 9 freshmen students sampled from the freshmen dorm, as well as from 3 to 9 sophomores, from 3 and 9 juniors, and from 3 to 9 seniors.

2.40

(a) Cluster sampling.

(b) No, if the number of adults per city block is unknown. Yes, if you have the seed value and you know how many adults are in each block.

2.41

(a) With the calculator or the random number table, the selected region is 3 = Southwest.

(b) Stratified random sampling.

(c) (i) 1-in-10 systematic sampling.
(ii) 0.10.
(iii) With the calculator, the first five selected cans are 7, 17, 27, 37, and 47. With the random number table we might label the first can 1, the second can 2, …, and the 10th can 0. Then the first five selected cans are 7, 17, 27, 37, and 47.
(iv) Note that 125/10 = 12.5 → 12 or 13 cans. However, there will not be a 7th can to select in last group. Thus the total number of cans in the sample will be 12.

(d) (i) No.
(ii) Two possible values are 0.12 and 0.15.
(iii) Yes, a Type II error.

2.42

(a) (i) Convenience sampling.
(ii) Yes, a selection bias.
(iii) The calculated average is expected to be higher than the true average, as all of the books in the sample have already been checked out at least once, and may include some of the more popular books.

(b) Cluster sampling.

(c) (i) Stratified random sampling.
(ii) Overall estimate: (40/1200)(20) + (200/1200)(15) + (600/1200)(10) = 14.2 times checked out.

(d) For each of the three categories of books the following stages are followed.
Stage 1: Divide the books into clusters according to the last digit of the call number (0 through 9). Take a simple random sample of 3 digits from the list of 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9.
The clusters of books (in that category) with call numbers ending with those selected digits are selected.
Stage 2: Within each of the selected clusters of books from Stage 2, select a simple random sample of 7 books.
Note that with this multistage sampling plan, we will have a total of 3 categories x 3 clusters x 7 books = 63 books.

2.43

(a) A cluster sampling of blocks.
(b) Yes, the chance is 5/50 = 0.10.
(c) No, we do not know how many households are in each block. Additionally, if the number of households in each block was not the same across all blocks, then we would also need to know which of the 5 blocks were selected.
(d) With the calculator, the five selected blocks are 8, 44, 33, 6, and 38. With the random number table we might label the blocks 01 to 50, then the five selected blocks are 12, 18, 11, 43, and 05.
(e) Response bias because the interviewers are college students. People may not feel comfortable telling these students they want to forbid loud music at parties in the college dorms.

2.44
Answer is (c) statistic is to a sample.

2.45
The proportion 153/200 is a parameter since the instructor polled her entire class and that was the group that she was interested in learning about.

2.46
(a) Population: Adults U.S. residents. The sample size $n = 1500$.
(b) Population: Today's shipment of 1-gallon milk cartons. The sample size $n = 5$.
(c) Population: The 740 members of the local women's business association. The sample size $n = 100$.

2.47
Nonresponse bias.

2.48
(a) Selection bias, where the sampling frame is either incomplete (such as sampling only one dealership) or incorrect.
(b) Nonresponse bias, where persons who do not respond to a survey (such as the lazy car owners) may have different opinions to those who do.
(c) Response bias, where respondents may have a tendency to lie (such as car dealerships that believe in lower miles per gallon) or refuse to answer.

2.49
This survey may be subject to nonresponse bias. Only those alumni who respond and report their income will be included. Alumni who perhaps are currently unemployed or in a low paying position may elect not to respond. Therefore, the reported average income based on such a survey may be biased upwards -- the average may be larger than the actual average for all alumni.

2.50
(a) The sampling design is a simple random sample of size 227 taken from the cocaine users who called.
(b) The population this sample is drawn from is cocaine users who called the National help line between February and March.
(c) The sample is not from the population of workers, but only from those who called the hotline, so the survey results do not generalize to the population of workers. Also, "more people" implies a comparison for which no data were given.

2.51
(a) With the calculator the selected ID numbers are: 179, 2274, and 3327. With the random number table the selected ID numbers are: 2398, 2258, and 3540.
(b) A statistic.

2.52
(a) True. If we fail to reject the null hypothesis, them the population consists of all males.
(b) False.
(c) True.

2.53

(a) Based on her decision rule, Jane rejects H_0.
(b) Since Jane rejected the null hypothesis, the data are statistically significant.
(c) No, based on a sample of two \$1 Jane can not be certain of which purse she has since both purses contain at least two \$1.
(d) Type I error: reject H_0 when H_0 is true.
(e) Simple random samples of size 2 from null purse:
$\$1_1$ and $\$1_2$, $\$1_1$ and $\$5_1$, $\$1_1$ and $\$5_2$
$\$1_2$ and $\$5_1$, $\$1_2$ and $\$5_2$, $\$5_1$ and $\$5_2$
(f) Simple random samples of size 2 from alternative purse:
$\$1_1$ and $\$1_2$, $\$1_1$ and $\$1_3$, $\$1_1$ and $\$1_4$
$\$1_2$ and $\$1_3$, $\$1_2$ and $\$1_4$, $\$1_3$ and $\$1_4$
(g) The p-value is the chance of getting two \$1 bills or more extreme (in the direction of H_0, but in this case, there is no "more extreme") if H_0 is true. The p-value is the chance of getting two \$1 bills if H_0 is true, i.e., 1/6.

2.54
You first need to label the 4000 signatures. Hopefully no one signed the petition more than one time. If you use a calculator or computer, the labels can simply be 1 to 4000. If you use the random number table, you could label the signatures from 0001 to 4000. Using a seed value of say 29 with the calculator or row 120, column 11 of the random number table, you can proceed to take a simple random sample of 400 signatures.

2.55
Stratified random sampling. You are dividing up your population by gender then selecting your sample within the strata at random.

2.56
(a) Stratified random sampling.
(b) 75/416.
(c) (0.20)(15/25) + (0.80)(60/75) = 0.76.
(d) Statistic.

2.57
(a) All former university graduate students.
(b) Stratified random sampling.
(c) False.

2.58
(a) Stratified random sampling.
(b) Selection bias.

2.59
(a) Stratified random sampling
(b) The chance is 0, only 1 of the two fiction books (A, B) will be selected.
(c) Two books that could be selected are Book A and Book C.
(d) Another pair of books that could be selected is Book A and Book D.
(e) The chance that the total number of pages exceeds 800 is the same as the chance that the two selected books are Book B and Book D. There are 4 possible pairs of books that could be selected of which the (B, D) pair is 1, so the chance is 1/4 = 0.25.

2.60
(a) Stratified random sampling.
(b) High: 20 clients, Moderate: 125 clients, Low: 45 clients.
(c) With the calculator the selected clients are: 163, 2196, 214, 2462, 740.
With the random number table the selected clients are: 1887, 1209, 2294, 954, and 1869.

2.61
(a) With the calculator the first student selected is 1. With the random number table the first student selected is 1.
(b) The sample size is 6.

2.62
(a) Systematic sampling (1-in-30). Once you select the sample you can determine the number of freshmen you selected. Also knowing the population list you can calculate the number of freshmen you would select for each of the 30 possible samples (depending on the starting point.)
(b) Stratified random sampling. Yes, you would have 25 freshmen.

2.63
(a) (i) 0.02(500)+0.03(1200)+0.05(18000) = 10+36+900 = 946.
(ii) Stratified Random Sampling.
(iii) With the calculator the selected labels are 432, 232, 304, 412, and 372. With the random number table, we might assign the first High category driver the labels 001 and 501. We would assign the second High category driver 002 and 502. This assignment pattern would continue until the 500th High category driver who would be assigned the labels 000 and 500. Reading off labels from row 60, column 1, we have: 789, 191, 947, 423, and 632. This would correspond to selecting the 289th, 191st, 447th, 423rd, and 132nd High category drivers in the list of 500 High category drivers.
(b) (i) With the calculator or the random number table, the first selected label is 15. Thus the selected labels will be 15, 35, 55, 75, 95, and so on.
(ii) Since the 500 High category drivers divide evenly into groups of 20 (500/20 = 25), there will be a total of 25 High category drivers in the systematic 1-in-20 sample.

2.64
(a) 907.
(b) Cluster sample.
(c) (iii) Selection bias.

2.65
These results are based on a study of 125 aerobic classes in five health clubs, not all aerobic classes in all health clubs. Thus, the 60% figure is a statistic and the sample size is $n = 125$.

2.66
(a) It is a systematic 1-in-45 sampling resulting in 50 students (1 from each of the 50 sections, the 33rd in each of the 50 lists of 45 students).
(b) It is a cluster sample and you cannot know the sample size (number of students selected) because we do not know how many students are in the various major clusters.

2.67
(a) Cluster sampling.
(b) 1/5.
(c) Response bias.

2.68
(a) False.
(b) False.
(c) True.

2.69
(a) A 1-in-40 systematic sample.
(b) 1/40 is the chance that any specific address is chosen since one of the first 40 addresses is picked at random. The other addresses are directly linked to that first random pick.

2.70

(a) H_0: The population mean increase in the GMAT score is 40 points.
H_1: The population mean increase in the GMAT score is less than 40 points.

(d) With the calculator the first four selected students are: 209, 218, 7, and 750.
With the random number table the first four selected students are: 070, 569, 762, and 833.

(c) (i) A possible p-value is 0.03.
(ii) Reject H_0.
(iii) Type I error.
(iv) Yes, since the p-value was ≤ 0.05, it would also be ≤ 0.10.

(d) (ii) Statistic.

2.71

(a) Multistage, with the first stage being a cluster sample of 3 lab sections and the second stage being a simple random sample of 25% of the students in the 3 selected labs.

(b) 0.0625.

(c) The 78% is a statistic since it was based on the sample of students surveyed.

2.72

Nonresponse bias is the distortion that can arise because a large number of units selected for the sample do not respond or refuse to respond, and these nonresponders have a tendency to be different from the responders. So nonresponse bias has to do with *who* responds. Response bias is the distortion that can arise because the wording of a question and the behavior of the interviewer can affect the responses received. So response bias has to do with *how* the responders answer.

2.73

(a) Since 15 patients were selected with a 1-in-9 systematic sample, there were at least 15 groups of 9 patients each or 135 patients in all.

(b) (ii) Statistic.

2.74

(a) Since there is no prior information, we will *sample in proportion to the size of the stratum relative to the size of the population.* So the sample size from Stratum I is 3 (since there are 10/40 large facilities in all, 25% of the sample could be large facilities, for 25% of 12 = 3), and the sample size from Stratum II is 12 – 3 or 9.

(b) Using the calculator, the stratum I selected facilities and responses are: 7 (Yes), 10 (Yes), 8 (No), and the stratum II selected facilities and responses are: 7 (No), 17 (No), 9 (Yes), 24 (No), 3 (No), 21 (No), 8 (Yes), 12 (No), and 1 (Yes),
Using the random number table we will label the stratum I facilities as: 1 has label 1, Facility 2 has label 2, ... , Facility 10 has label 0. Then the selected facilities and responses are: 2 (Yes), 6 (Yes), 4 (Yes). We will label the stratum II facilities as: 1 has label 01; Facility 2 has label 02; up to Facility 30 has label 30. Then the selected facilities and responses are: 05 (No), 36 (skip), 60 (skip), 42 (skip), 13 (Yes), 25 (Yes), 66 (skip), 92 (skip), 64 (skip), 22 (Yes), ..., 04 (Yes), 06 (No), 20 (No), 12 (No), 18 (Yes).

(c) Using the calculator we have: $\hat{p} = \text{estimate} = \left(\frac{10}{40}\right)\left(\frac{2}{3}\right) + \left(\frac{30}{40}\right)\left(\frac{3}{9}\right) = 0.416$.

Using the table we have: $\hat{p} = \text{estimate} = \left(\frac{10}{40}\right)\left(\frac{3}{3}\right) + \left(\frac{30}{40}\right)\left(\frac{5}{9}\right) = 0.667$.

(d) The true population proportion is $p = 20/40 = 0.50$.

(e) In general, estimates may not be exactly equal to p. The calculator estimate was slightly too small, while the random number table estimate was too large.

2.75

(a) Answers will vary. See the web site.

(b) Summaries will vary.