Chapter 19 – From Data Production to Inference: Part III Review

Test Yourself Exercise Answers are answers or sketches. All of these problems are similar to ones found in Chapters 12–18, for which the solutions in this manual provide more detail.

19.1 (a) S = {male, female}. **(b)** S = {6, 7, 8, …, 19, 20}. **(c)** S = {All values $2.5 \leq VO_2 \leq 6.1$ liters per minute}. **(d)** S = {All heart rates such that heart rate > 0 bpm} (students may have other choices for the minimum in this sample space).

19.2 (a) 0.33. These are disjoint events, so P(Microsoft or Yahoo) = 0.21 + 0.12 = 0.33.

19.3 (a) 0.01. The calculation is $1 - 0.64 - 0.21 - 0.12 - 0.02 = 0.01$.

19.4 (c) 0.36. The calculation is $1 - 0.64 = 0.36$.

19.5 $\{Y > 1\}$ or $\{Y \geq 2\}$; $P(Y > 1) = 1 - 0.28 = 0.72$.

19.6 $P(2 < Y \leq 4) = P(3 \leq Y \leq 4) = P(Y = 3) + P(Y = 4) = 0.16 + 0.13 = 0.29$.

19.7 (d) 0.66. The calculation is $1 - 0.34 = 0.66$.

19.8 All of the probabilities are between 0 and 1 (inclusive), and they sum to 1. This is a legitimate discrete probability model.

19.9 This is the event that a woman between the ages of 15 and 44 has given birth to two or fewer children. $P(X \leq 2) = 0.476 + 0.167 + 0.199 = 0.842$ (84.2% of women between the ages of 15 and 44 have given birth to two or fewer children).

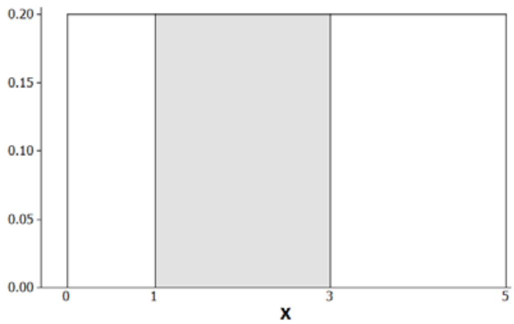
19.10 (c) 0.643. The calculation is $P(X < 2) = P(X \leq 1) = 0.476 + 0.167 = 0.643$.

19.11 $\{X \geq 3\}$; $P(X \geq 3) = 0.101 + 0.038 + 0.019 = 0.158$.

19.12 (b) continuous, but not Normal.

19.13 (a) 0.2. The height of the density curve is $1/5 = 0.2$, because the area under the density function must be 1. See the graph for Exercise 19.14.

19.14 The graph is provided below. $P(1 \leq Y \leq 3) = 2/5 = 0.4$.



19.15 There is no area above Y = 5 (the rectangle stops at 5), so $P(4 < Y < 7) = P(4 < Y \leq 5) = 1/5 = 0.2$.

19.16 (b) 0.3707. This is $P(Z \geq 1/3)$, or $P(Z \geq 0.33)$.

19.17 (c) Mean = 100, standard deviation = 1.94. Standard deviation is calculated by $15/\sqrt{60}$ = 1.94 (rounded).

19.18 (a) 0.0049. This is $P(Z \geq 2.58)$.

19.19 The answer in Exercise 19.16 would change, because this refers to the population distribution, which is now non-Normal (we most likely could not determine this probability). The answer in Exercise 19.17 would not change—the mean of $\bar{x}$ is 100, and the standard deviation of $\bar{x}$ is 1.94, regardless of the population distribution. The answer in Exercise 19.18 would, essentially, not change. The central limit theorem tells us that the sampling distribution of $\bar{x}$ is approximately Normal when n is large enough (and 60 should be large enough), no matter what the population distribution.

19.20 Whether n = 15 or n = 150, the mean of $\bar{x}$ is 445 ms. If n = 15, the standard deviation of $\bar{x}$ is $82/\sqrt{15}$ = 21.17 ms. If n = 150, the standard deviation of $\bar{x}$ is $82/\sqrt{150}$= 6.70 ms.

19.21 If the population from which we're sampling is heavily skewed, then a larger sample is required for the central limit theorem to apply. If n = 15, the sampling distribution of $\bar{x}$ may not be approximately Normal, but if n = 150, it will surely be approximately Normal.

19.22 $P(\bar{x} > 450) = P(Z > 0.75)$ = 0.2266.

19.23 (c) 322.35 to 391.65. This calculation is $357 \pm 1.96 \frac{50}{\sqrt{8}}$ = 322.35 to 391.65.

19.24 (b) 327.92 to 386.08. This calculation is $357 \pm 1.645 \frac{50}{\sqrt{8}}$ = 327.92 to 386.08.

19.25 $357 \pm 1.282 \frac{50}{\sqrt{8}}$ = 334.34 to 379.66.

19.26 As the confidence level decreases, the margin of error decreases, resulting in a narrower confidence interval.

19.27 (b) 197 ± 18.46 mg/dl. This calculation is $197 \pm 1.645\frac{42}{\sqrt{14}} = 197 \pm 1.645(11.22) = 197 \pm 18.46$ mg/dl. Alternatively, if students omit the intermediate rounding step, they may select option (d) none of the above. This calculation results in $197 \pm 1.645\frac{42}{\sqrt{14}} = 197 \pm 18.47$ mg/dl.

19.28 (d) 56. To cut the margin of error in half, we need to quadruple the sample size from 14 to 56.

19.29 (c) 191. $n = (1.645 \times 42/5)^2 = 190.94$, so $n = 191$.

19.30 (a) $H_0\colon \mu = 50, H_a\colon \mu < 50$.

19.31 (c) $H_0\colon \mu = 14.25, H_a\colon \mu \neq 14.25$. We want to know if your college differed, so the alternative is two-sided. Also, hypotheses are in terms of population parameters, not sample statistics.

19.32 (b) −2.41. This calculation is $z = \frac{197 - 224}{\frac{42}{\sqrt{14}}} = -2.41$.

19.33 (c) $\alpha = 0.01$ but not at $\alpha = 0.005$. The P-value is 0.0080.

19.34 (a) $\alpha = 0.001$. Now $z = \frac{197 - 224}{42/\sqrt{56}} = -4.81$, so the P-value is essentially 0.

19.35 (d) no more than 0.01. $z = \frac{357 - 100}{50/\sqrt{8}} = 14.538$. The corresponding P-value is essentially 0.

19.36 We are 95% confident the mean IQ at age 20 for men who had very low birth weight is in the interval $87.6 \pm 1.96\frac{15}{\sqrt{113}} = 87.6 \pm 2.77 = 84.83$ to 90.37.

19.37 We test $H_0\colon \mu = 100$ vs. $H_a\colon \mu < 100$; $z = \frac{87.6 - 100}{15/\sqrt{113}} = -8.79$; P-value is essentially 0. This is overwhelming evidence that the mean IQ for the very-low-birth-weight population is less than 100.

19.38 (c) The statement "no differences were seen" means that the observed differences were not statistically significant at the significance level used by the researchers.

19.39 $P = 0.74$ means that the observed difference is easily explained by random chance (if there is actually no difference, we have a 74% chance of seeing the observed or a larger difference in cholesterol). P = 0.013 means that the observed difference was unlikely to have occurred by chance alone; such a result (or something more extreme) would be expected only 13 times in 1000 repetitions of this study.

19.40 Here, r^2 = 0.61 means that 61% of the total variability in number of wildfires is explained by our model (by knowing the year). If there is really no relationship between number of fires and year (a surrogate for population here), then an observed linear relationship in our data as strong as that observed (r^2 = 0.61) would have been very unlikely to occur by chance alone. It seems reasonable to conclude that year and wildfires are positively associated—fires have increased over time, suggesting that population (or changing weather) increases wildfires. However, a cause-and-effect conclusion is not possible.

19.41 (b) Byron's personal probability that Ohio State and Alabama will play in the championship football game this year. This is a personal probability. It is Byron's opinion, and not something based on many repetitions of a football season (which would be impossible).

19.42 **(a)** 11,479/14,099 = 0.8142. **(b)** 6457/(6457 + 1818) = 0.7803. **(c)** These were not independent. If they were, the probabilities in part (a) and part (b) would be the same; that is, we would have P(male) = P(male | accidental).

19.43 (c) 0.30. Let W be the event that you see a whale and D be the event you see a dolphin. Then $P(W \text{ or } D) = P(W) + P(D) - P(W \text{ and } D)$, so rearranging we have $P(W) = P(W \text{ or } D) - P(D) + P(W \text{ and } D) = 0.85 - 0.65 + 0.1 = 0.3$.

19.44 (b) 0.20. Let W be the event that you see a whale and D be the event that you see a dolphin. Then $P(W \text{ and not } D) = P(W) - P(W \text{ and } D) = 0.3 - 0.1 = 0.2$.

19.45 (d) 0.9775. For a single day, $P(\text{not } D \text{ and not } W) = 1 - 0.85 = 0.15$. So, for two independent days, the probability of not seeing a dolphin or a whale is $(0.15)^2$ = 0.0225. Thus, the probability of seeing a dolphin or a whale on at least one of the two days is 1 − 0.0225 = 0.9775.

19.46 (c) 0.023. Because of independence, the probabilities can be multiplied, resulting in (0.6)(0.6)(0.4)(0.4)(0.4) = 0.023.

19.47 (a) 0.259. This is binomial, with n = 5 and p = 0.4.

19.48 (c) $N(1838.5, 26.78)$. $\mu = np = 3014(0.61) = 1838.5$ and $n(1-p) = 3014(0.39) = 1175.46$; both are more than 10, so the Normal approximation holds. Finally, $\sigma = \sqrt{np(1-p)} = 26.78$.
19.49 $P(X \geq 1900) = P(Z \geq 2.30) = 0.0107$.

Supplementary Exercises

19.50 There are many possible answers; the key is that the events A and B must be able to occur together. One possibility: consider undergraduate students; let A = {the student is female} and B = {the student is a freshman}.

19.51 (a) All probabilities are between 0 and 1, and their sum is 1. **(b)** Let R_1 be Taster 1's rating and R_2 be Taster 2's rating. Add the probabilities on the diagonal (upper left to lower right): $P(R_1 = R_2) = 0.05 + 0.08 + 0.25 + 0.18 + 0.08 = 0.64$. **(c)** $P(R_1 > R_2) = 0.18$. This is the sum of the ten numbers in the lower left part of the table: the bottom four numbers from the first column, the bottom three from the second column, the bottom two from the third column, and the last number in the fourth column. These entries correspond to, for example, "Taster 2 gives a rating of 1, and Taster 1 gives a rating more than 1." $P(R_2 > R_1) = 0.18$; this is the sum of the ten numbers in the upper right part of the table. We could also find this by noting that this probability and the other two in this exercise must add to 1 (because they account for all of the entries in the table).

19.52 $P(\text{A}) = P(\text{B}) = \cdots = P(\text{F}) = (1 - 0.28)/6 = 0.12$ and $P(1) = P(2) = \cdots = P(8) = (0.28)/8 = 0.035$.

19.53 (a) Out of 100 seniors, nearly all should be in the range $\mu \pm 3\sigma = 3.3 \pm 3(0.8) =$ 0.9 to 5.7. **(b)** The sample mean $\bar{x}$ has an $N(\mu, \sigma/\sqrt{100}) = N(3.3, 0.08)$ distribution, so nearly all such means should be in the range $3.3 \pm 3(0.08) = 3.3 \pm 0.24$, or 3.06 to 3.54.

19.54 To cut the range of values of $\bar{x}$ in half, we need to halve the standard deviation of the distribution of $\bar{x}$, which requires increasing the sample size by a factor of 4, to $n = 400$. Those 400 individual NSSE scores will be as variable as the 100 individual scores. Because of the limited scale, it would be almost impossible to have outliers.

19.55 (a) The provided stemplot confirms the description given in the text. (Arguably, there are two "mild outliers" visible in the stemplot, although the 1.5 × *IQR* criterion only flags the highest as an outlier.)

```
 96 | 8
 97 | 344
 97 | 888889
 98 | 0133
 98 | 5789
 99 |
 99 | 6
100 | 2
```

(b) STATE: Is there evidence that the mean body temperature for all healthy adults is not equal to 98.6°F? PLAN: Let μ be the mean body temperature. We test H_0: μ = 98.6°F vs. H_a: $\mu \neq$ 98.6°F; the alternative is two-sided because we had no suspicion (before looking at the data) that μ might be higher or lower than 98.6°F. SOLVE: Assume we have a Normal distribution and an SRS. The average body temperature in our sample is $\bar{x}$ = 98.203°F, so the test statistic is $z = \frac{98.203 - 98.6}{0.7/\sqrt{20}} = -2.54$. The two-sided P-value is $P = 2P(Z < -2.54) = 0.011$. CONCLUDE: We have fairly strong evidence—significant at $\alpha = 0.05$, but not at $\alpha = 0.01$—that mean body temperature is not equal to 98.6°F. (Specifically, the data suggest that mean body temperature is lower.)

19.56 (a) The provided stemplot confirms the description given in the text.

```
 7 | 6
 8 |
 8 | 89
 9 | 234
 9 | 578
10 | 1234
10 | 5566788999
11 | 4
11 | 6
12 | 14
12 | 69
13 |
13 | 7
```

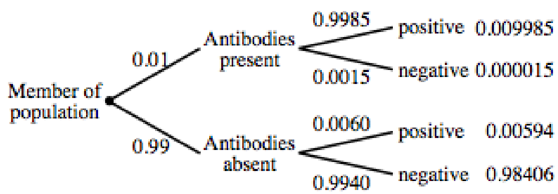
(b) STATE: Does the presence of a lavender odor increase the mean time spent in the restaurant? PLAN: Let μ be the mean time spent in the restaurant with the lavender odor. We test H_0: μ = 90 minutes vs. H_a: $\mu > 90$ minutes; the alternative is one-sided because we are looking specifically for an increased time spent in the restaurant. SOLVE: The provided stemplot in part (a) suggests that the distribution of customer times is reasonably Normal; we also assume we have an SRS. We find that $\bar{x}$ = 105.7, so the test statistic is $z = \frac{105.7 - 90}{15/\sqrt{30}} = 5.73$, and the P-value is extremely small ($P = P(Z > 5.73) \approx 0$). CONCLUDE: This is overwhelming evidence that the mean time spent in the restaurant increases when the lavender odor is present.

19.57 STATE: What is the mean body temperature μ for healthy adults? PLAN: We will estimate μ by giving a 90% confidence interval. SOLVE: Assume we have a Normal distribution and an SRS. With $\bar{x}$ = 98.203, our 90% confidence interval for μ is $98.203 \pm 1.645(0.7/\sqrt{20}) = 98.203 \pm 0.257$, or 97.95°F to 98.46°F. CONCLUDE: We are 90% confident that the mean body temperature for healthy adults is between 97.95°F and 98.46°F.

19.58 STATE: What is μ, the mean time spent in the restaurant on Saturday nights when a lavender odor is present? PLAN: We will estimate μ by giving a 95% confidence interval. SOLVE: We assume that we have an SRS of the population and that the distribution is roughly Normal with standard deviation 15 minutes. With $\bar{x}$ = 105.7, our 95% confidence interval for μ is $105.7 \pm 1.96(15/\sqrt{30}) = 105.7 \pm 5.37$, or 100.33 to 111.07 minutes. CONCLUDE: We are 95% confident that the mean time spent in the restaurant is between 100.33 and 111.07 minutes.

19.59 For the two-sided test H_0: M = \$50,000 vs. H_a: $M \neq$ \$50,000 with significance level α = 0.10, we can reject H_0 because \$50,000 falls outside the 90% confidence interval.

19.60 (a) The tree diagram is provided.



(b) P(positive) = 0.009985 + 0.00594 = 0.015925.

19.61 Let H be the event the student was home schooled. Let R be the event the student attended a regular public school. We want $P(H \mid \text{not } R)$. Note that the event H and not R equals H. Then $P(H \mid \text{not } R) = \frac{P(H)}{P(not\ R)} = \frac{0.006}{1 - 0.758} = 0.025$.

19.62 P(has antibody | positive) = 0.009985/0.015925 = 0.627.

19.63 (a) For n = 300 people beginning the program, $\mu = np = 300(0.18) = 54$ people and $\sigma = \sqrt{np(1-p)} = \sqrt{300(0.18)(1-0.18)} = 6.65$ people. **(b)** $np = 54$ and $n(1-p) = 246$; both are more than 10 so the Normal approximation holds. If at least 235 people of 300 remain in the program, then no more than 65 people drop out. Let X be the number of people that dropped out of the program. Then, using the Normal approximation: $P(X \leq 65) = P\left(Z \leq \frac{65-54}{6.65}\right) = P(Z \leq 1.65) = 0.9505$. Using software to find the exact binomial probability, $P(X \leq 65) = 0.9554$. The Normal approximation 0.9505 misses the true probability by about 0.0049.

19.64 A low-power test has a small probability of rejecting the null hypothesis, at least for some alternatives. That is, we run a fairly high risk of making a Type II error (failing to reject H_0 when it is false) for such alternatives. Knowing that this can happen, we should not conclude that H_0 is true simply because we failed to reject it.

19.65 A Type I error means that we conclude the mean IQ is less than 100 when it really is 100 (or more). A Type II error means that we conclude the mean IQ is 100 (or more) when it really less than 100.