Statistics: Chapter 10 Solutions

Lesson 10.1.1

10-1. a. The distribution has a single peak and skewed right, with a center (both mean and median) near 4. Most of the data is clustered near the mean (the IQR and SD are both just 2) but the maximum value of 14 shows how far the distribution is skewed.

b. No! The distribution is clearly not symmetric.

10-2. a. Answers will vary.

b. Answers will vary, but should have been already found in previous question! Whether or not the sample is a good estimate also varies—samples of size 5 are not very big, so there is no guarantee!

c. Since only one mean has been drawn, there is no spread or variability in the distribution.

d. It should hopefully be near the center.

10-3. The histogram slowly adds bars and height as you add a mean for each sample.

10-4. a. The entire sampling distribution would require that you get every single possible sample of 5 from the original population, but the simulation does not do that—it just randomly selects samples.

b. The sampling distribution is single peaked, slightly skewed right, with a mean of 4.08 letters and a standard deviation of about 0.89 letters, while the original population has more skew, a standard deviation of 1.98 letters but also has a center of 4.08 letters. So the centers are the same, but the sampling distribution has a similar but less extremely skewed shape, and significantly lower spread.

c. The spread of the sampling distribution describes how much you can expect sample means to vary from the overall mean value of the sampling distribution—this is variability!

d. The mean of the sampling distribution is generally 4.08, which is identical to the population mean. Thus, the sample mean can be seen as an unbiased estimator of the population mean.

e. No. It is still noticeably skewed.

10-5. a. The median distribution is generally single peaked and right skewed, just like the other two distribution. Its mean is often a tad lower than the mean of other two (3.8 or so) while the median is consistently 4. The standard deviation is usually about 1—higher than the sampling distribution of the means but lower than the original population. Because the median in this case is always an integer, the distribution is much more discrete in its values or "blockier" than that of the means.

b. The mean of the sampling distribution of medians is about 3.70, which is quite a bit lower than the true median of 4. Technically, this means the sample median is a slightly biased estimator for the population median in this case.

c. A correct answer will compare the standard deviations or variances of the sampling distributions. The standard deviation of the median distribution is significantly wider with this population than the standard deviation of the mean distribution, so the median is more variable in this case.

d. Definitely not. Skew and too discrete or "blocky."

10-6. a. About 90% of all samples of size 5 have a mean between 2.7 and 5.5.

b. This does not feel as accurate because there are so few bars, but more than 90% of the data is contained in bars 3 through 5.

10-7. a. It should decrease, as larger samples will be more like each other and more like the true population.

b. The centers of both distributions are unchanged. The spreads of both distributions decrease significantly, with the standard deviation of the medians decreasing to about 0.47 while the standard deviation of the means decreases to about 0.36. The shape of the distributions change too, with both of them becoming significantly more symmetric, though the discreteness of the median distribution makes it "blocky" looking.

c. Yes, the sampling distribution of the means now appears to be approximately normal.

10-8. a. bias = high, variability = low

b. bias = high, variability = high

c. bias = low, variability = low

d. bias = low, variability = high

10-9. a. For these, the mean of the sampling distribution must be on or near 172 to be unbiased. A appears to be biased a little high, but could be unbiased. B looks pretty close to unbiased but also maybe a bit high. C is REALLY hard to tell because of the skew, but could be unbiased or could be too high. D is clearly biased to overestimate the value.

b. C, A, D, B

c. Stat C is the closest to the true value the most often, and seems like the best option. This is the actual statistic that is used most frequently for situations like this one!

10-10. a. See bold answers in table below.

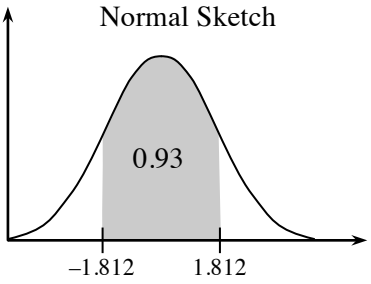
b. See bold answers in table below.

	Teenager	Young Adult	Older Adult	
Used phone	62 **44.67**	**110 (109.81)**	54 **71.52**	226
Did not use phone	141 **158.33**	**389 (389.19)**	**271 (253.48)**	**801**
	203	**499**	325	1027

c. The $\chi^2 = 14.12$, since with 2 degrees of freedom there is less than a 0.05 chance that $\chi^2 > 5.991$ could occur as a result of pure chance, Anna should reject the hypothesis that phone use is independent of age group.

d. Examining the expected values compared to the actual values, it appears that more teens use their phone than expected and less older adults use their phone than expected. It appears about 22% of the observed people use their phone at the light.

10-11. a. Identify: One proportion confidence interval. Given: $n = 891, x = 157, \hat{p} = 0.176$. See sketch at right.



Check: Random selection, to avoid bias—"randomly" was mentioned in the problem description with regards to sampling technique. Independent trials, for an accurate σ—it must be assumed that the population of interest is larger than ten times the sample size of 891. Large counts, so the sampling distribution is ≈ normal— $n\hat{p} = 157$, and $n(1 - \hat{p}) = 734$, which are both at least 10.

Calculate: SE = $\sqrt{\frac{0.1762(1-0.1762)}{891}} = 0.0128$, z* = invnorm(0.965, 0, 1) = 1.812, CI = $\hat{p} \pm$ (z*)(SE), CI = $0.176 \pm (1.812)(0.0128)$, CI = $(0.1531 < p < 0.1993)$.

Conclude: Students are 93% confident the interval from 0.1531 to 0.1993 contains the population proportion of people who have a middle birth order among one's siblings.

b. Solve $1.812\sqrt{\frac{0.176(0.824)}{n}} \le 0.015$ and find $n \ge 2116.3$, so Davis should use at least 2117 people.

10-12. The median for both types of gum was about 18 minutes of flavor time. The times for 10 were skewed, while the times for Strident were symmetric. The lower half of the distributions for both gums was the same. But there was much more variability in the upper half of people chewing 10 than in the upper half of Strident. Indeed, more than 25% of 10 chewers reported flavor lasting longer than any of the Strident chewers. Neither gum had outliers in flavor time. There was more variability in flavor time for 10—the IQR was about 9 minutes (25 – 16 = 9). The IQR of 4 minutes (20 – 16 = 4) for Strident was less than half that of 10. That variability is an advantage. If you chew 10, you will probably be no worse off than chewing Strident, and you could have much longer flavor.