Lesson 10.1.2

10-13. a. 4

b. See bold values in table below.

Sample	1, 2	1, 3	1, 4	**1, 5**	**2, 3**	**2, 4**	**2, 5**	**3, 5**	**3, 4**	**4, 5**
Range	1	2	3	**4**	**1**	**2**	**3**	**2**	**1**	**1**

c. 2

d. No. The mean of the sampling distribution 2 is not equal to the true range 4.

10-14. a. These two distributions are both extremely right skewed. The mean of the population variances is about 3.13 while that of the sample variances is about 3.91 and the standard deviations are about 2.80 and 3.50 respectively.

b. With samples of size 5, the mean of this distribution is 3.13. The true population variance is 3.91. This is pretty different!

c. The mean of this statistic distribution should be near 3.91, though it varies somewhat. However, it is certainly much closer to the true variance!

d. You want the sample variance to be an unbiased estimator for the population variance, which the (n – 1) formula seems to provide.

10-15. a. Both are biased, though the n – 1 is better! The sample standard deviation, even when dividing by n – 1, has a mean slightly lower than the true population standard deviation.

b. The mean of 4, 64, and 100 is 56.

c. $\sqrt{56} \approx 7.483$

d. The three standard deviations are 2, 8, 10, with mean of 6.67, undershooting the true standard deviation.

10-16. Select "IQR" from one of the drop-down menus and notice that the mean value of the IQR distribution is about 1.84, while the true IQR is 2. Like the median, this seems to underestimate the true value a bit, though the bias certainly is not bad.

10-17. The sample standard deviation has slightly less variability. You can tell by comparing the standard deviations of the IQR distribution is about 1.2, while the standard deviation of the sample st dev distribution is about 0.80.

10-18. Yes, both distributions narrow, IQR to a new standard deviation of about 0.67, sample standard deviations to a new standard deviation of about 0.33.

 © 2018 CPM Educational Program. All rights reserved.

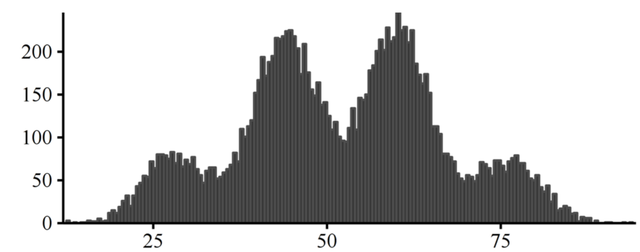
10-19. a. Two peaks, one at 25 and 75. Very symmetric. Mean about 50.

b. two highs = $0.5 \cdot 05 = 0.25$, two lows = 0.25, one of each = 0.5 (because it could be high-then-low or low-then-high)

c. 75, 25, 50

d. About half of the values make the peak centered at 50, one fourth make the smaller peak at 25, and one fourth make the smaller peak at 75. Those match the probabilities and centers calculated in part (c).

e. See graph at right—key features should be four peaks instead of three. Short ones at 25 and 75 (that correspond to three highs and three lows) and taller ones around 37.5 and 62.5 that correspond to two lows/one high and two highs/one low.



10-20. a. Meg used a stratified random sample.

b. A side-by-side bar graph or stacked bar graph may be appropriate.

c. This graph reveals that students a Three Puddles and Riverland are more likely to choose basketball while Eagle Lakes is strongly for hockey.

d. The appropriate test would be a chi-square test for homogeneity. H_0: There is no significant difference of sport preference among the schools. H_A: There is an association between school and sport preference.

10-21. a. $\hat{y} = 8.32 + -0.0138(x)$, where: x is plant height and $\hat{y}$ is predicted soil acidity.

b. For every 1 cm increase in plant height the predicted soil acidity will decrease by –0.0138 pH.

c. $\hat{y} = 8.32 + -0.0138(120.72) = 6.65$ pH

d. The predicted value at $x = 136.7$ cm is $\hat{y} = 8.32 + -0.0138(136.7) = 6.43$ pH. The residual is $Y_{\text{observed}} - Y_{\text{predicted}}$. Residual = 6.14 – 6.43 = –0.29 pH

e. There is no discernable curved pattern in the scatterplot which is evidence that a linear model is appropriate.

© 2018 CPM Educational Program. All rights reserved.