Lesson 7.1.2

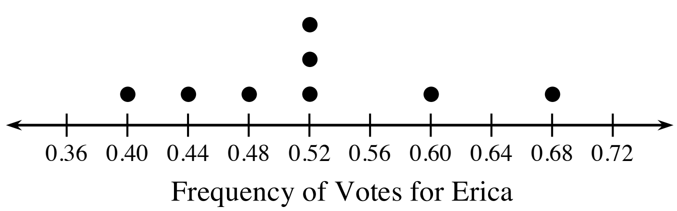
7-8. a. A vote is a census, and the population is all students at Jefferson High who vote. Students may want to call it a sample since not everybody votes, but since the non-voters have no effect on anything and it does not actually matter how they *would have* voted, they are simply excluded from the population instead. The individuals are the voters, the only variable you know about is "choice for president," though there may be other variables, such as secretary or VP. The domain for the president variable are the names of the nominees (plus perhaps "other"), and it is categorical data.

b. The proportion of students who voted for each nominee.

7-9. a. To simulate this, the most efficient thing is to go two digits at a time and let 00 through 55 represent students who vote for Erica and 56 through 99 those who do not. Another option should be to let 1 through 56 represent Erica and 57 through 99 and 00 represent those who do not. Other options are possible too: as long as 56% of checked values vote for Erica (and no values are skipped) then the system is fine. Counting by twos, one row of fifty digits is exactly the correct length to get 25 students.

Using the 00 through 55 = Erica method, row 1 has 11, or 0.44, for Erica. Row 2 has 13 (0.52) for Erica. Row 3 has 13 (0.52), row 4 has 10 (0.4), row 5 has 15 (0.6), row 6 has 17 (0.68), row 7 has 13 (0.52), and row 8 has 12 (0.48). Numbers will vary if different strategies are used.

b. The distribution will vary, but should include descriptions of shape, center, spread, and any outliers, as well as some sort of graphical display—dot plots usually work well. The dot plot below comes from using the 8 pieces of data provided in the answer to part (a). Even with a full 8 samples, this is not a good representation of the sampling distribution; it is far too small.

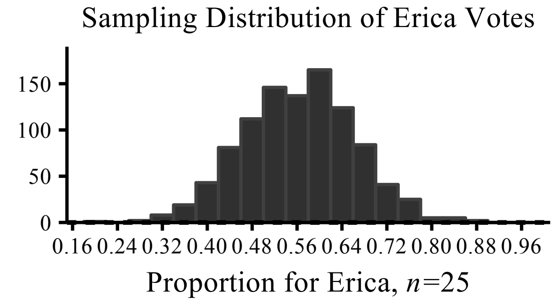


c. No. There are $\binom{320}{25} \approx 1.05 \times 10^{37}$ possible samples in this situation. That is significantly more than the number of stars in the known universe, and about as many water molecules as in the Atlantic Ocean.

Solution continues on next page →

7-9. *Solution continued from previous page.*

d. One possible histogram is shown at right: it should be a unimodal, symmetric distribution with a center around 0.56 and a standard deviation of around 0.10, so most values are between 0.36 and 0.76.



e. It looks like about $\frac{732}{1000}$, or 70%, of the values in the provided simulation voted for Erica. So she has about a 70% chance of getting a sample where more than half vote for her.

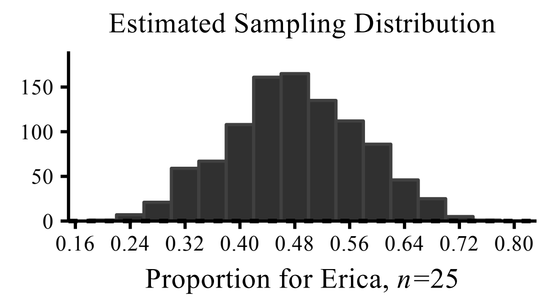
f. Using the histogram provided, 0.44 to 0.68, or 0.56 ± 0.12, would be reasonable middle 90% interval.

7-10. a. Yes. It is solidly in the middle 90% interval.

b. Her best guess is her sample proportion, 0.48 or 48%.

7-11. Since Erica does not know the true $p = 0.56$, she is going to have to use her best estimate, 0.48, instead to do her sample.

7-12. See the histogram at right for one example. The center is different from the original sample, centered at 0.48 rather than 0.56. The spread and shape, however, are essentially identical. If standard deviations are calculated, they are usually quite close.



7-13. a. Using the histogram above, a reasonable estimate would be 0.36 to 0.64.

b. Yes. The true parameter was 0.56, which is within the interval given in part (a).

7-14. $\frac{43}{100} = 0.43$ so by entering that proportion into a sampling program or calculator command with a sample size of 5, Caitlin could create a model of a sampling distribution, as in problem 7-8. Given the histogram, she could estimate the probability of a sample containing 3 or more Statistics students simply by adding up the bars above $\hat{p} = 0.60$ and dividing by the number of samples.

7-15. The middle 95% of the histogram is about 0.74 to 0.86, or 0.80 ± 0.06. These values were found by dropping about 2.5% of the data from each tail of the distribution.

7-16. This is a binomial setting. One way to find the value is to use a binomial cdf function with trials = 5, $p = 0.43$, and $k = 2$, then subtract the answer from one. Using that technique, the probability is 0.37.

7-17. a. It should be clear that a meeting either does or does not start more than 5 minutes late. The situation would need to be posed in terms of a fixed number of meetings. It would have to be assumed that the 65% is the true probability that any given meeting will start at least 5 minutes late. It would also have to be assumed that the meetings starting late are independent.

b. Binomial distribution, $n = 6, p = 0.65$. $P(X = 5) = 0.2437$

7-18. Binomial distribution, $n = 5, p = 0.90$. $P(X \geq 3) = 1 - P(X \leq 2) = 0.9914$

7-19. See possible diagram below. Home alarm systems will call police in events (A) or (D) from the diagram. That would be 199 + 1595 = 1794 calls per year. Of those alarms only the 199 from event (A) are actual break-ins. When the police dispatcher receives a call from a home alarm system there is only a $\frac{199}{1794} = 0.111$ chance it is due to criminal activity.

	200 are broken into	319,000 are *not* broken into
Alarm is correct 0.995	(A) 199	(B) 317,405
Alarm is *not* correct 0.005	(C) 1	(D) 1595 *false alarms*

