1 Chapter 17 review exercises

1. (a) The smallest sum will be obtained if only 1's are drawn, yielding $100 \times 1 = 100$. Similarly, we obtain the largest sum when only 10's are drawn, namely $100 \times 10 = 1000$.

   (b) To figure out what the chance is that the sum will be between 650 and 750, we use the normal curve.

   First, we compute the expected value per draw; it is equal to the average of the values in the box, i.e.,

$$\frac{1+6+7+9+9+10}{6} = \frac{42}{6} = 7.$$

   We thus know that the expected sum of 100 draws is $100 \times 7 = 700$.

   Second, we calculate the standard deviation (SD). We already know that the average value of the tickets in the box is 7, and the deviations from the average are thus

$$-6 \quad -1 \quad 0 \quad 2 \quad 2 \quad 3$$

   Using these values, we find that the SD of the box is

$$\begin{aligned} \sqrt{\frac{(-6)^2+(-1)^2+0^2+2^2+2^2+3^2}{6}} &= \sqrt{\frac{36+1+0+4+4+9}{6}} \\ &= \sqrt{\frac{54}{6}} = \sqrt{9} = 3 \end{aligned}$$

   Next, according to the "square root law," we have to compute the standard error (SE).

$$\begin{aligned} SE &= \sqrt{number\ of\ draws} \times (SD\ of\ box) \\ &= \sqrt{100} \times 3 = 10 \times 3 = 30 \end{aligned}$$

   Given that the SE is 30, 650 is $\frac{650-700}{30} = \frac{-50}{30} \approx -1.67$ SEs away from the expected value for the sum. Likewise, 750 is ≈ 1.67 SEs above 700. Using the normal table, we find that for a z-score of 1.67, the area under the curve is 90.5%.

   Based on this calculation we finally conclude that the sum of 100 draws is between 650 and 750 with a chance of about <u>90%</u>.

3. For a variable that takes only two values (e.g., 0 and 1, or -1 and 2), the SD is

$$(\text{high value} - \text{low value}) \times \sqrt{\frac{\#\ \text{high}}{\text{total}} \times \frac{\#\ \text{low}}{\text{total}}}$$

   Using this formula, we can make the matches:

   (a) SD is (iii)
   (b) SD is (i)
   (c) SD is (v)
   (d) SD is (iv)
   (e) SD is (ii)

7. (a) The average is $321/100 = 3.21$.

   (b) The sum is $3.78 \times 100 = 378$.

   (c) To answer this question, we use the normal curve. Note that saying that "the average of the draws is between 3 and 4" is equivalent to stating that the sum of the 100 draws is between 300 and 400.

   First, compute the expected value for one draw:

$$\frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3.5$$

   Thus, the expected sum of 100 draws is $3.5 \times 100 = 350$.

   Next, find the SD:

$$\text{deviations from the expected value: } -2.5 \quad -1.5 \quad -0.5 \quad +0.5 \quad +1.5 \quad +2.5$$

$$\begin{aligned} \Rightarrow SD &= \sqrt{\frac{(-2.5)^2 + (-1.5)^2 + (-0.5)^2 + 0.5^2 + 1.5^2 + 2.5^2}{6}} \\ &= \sqrt{\frac{6.25 + 2.25 + 0.25 + 0.25 + 2.25 + 6.25}{6}} \\ &= \sqrt{\frac{17.5}{6}} \approx 1.708 \end{aligned}$$

   The SE is given by:

$$SE = \sqrt{100} \times 1.708 = 10 \times 1.708 = 17.08$$

   As before, given that the SE is 17.08, 300 is $\frac{300-350}{17.08} = \frac{-50}{17.08} \approx -2.927$ SEs away from the expected value of the sum of 100 draws. Likewise, 400 is ≈ 2.927 SEs above 350. Using the normal table, we find that for a z-score of 2.93 the area under the curve is 99.66%.

   We thus conclude that the chance that the average of the draws is between 3 and 4 is 99.66%.

12. Note that in this question two of the four blanks have to be filled in with the same answer in all three cases, (a), (b), and (c). The only values that change from case to case are the "observed value" and the "chance error."

    The expected value for one draw is:

$$\frac{1+2+3+4+5+6+7}{7} = \frac{28}{7} = 4$$

    The expected value for the sum of the 100 draws is:

$$(\text{number of draws}) \times (\text{average of box}) = 100 \times 4 = 400.$$

    The SD can be found as follows:

$$\text{deviations: } -3 \quad -2 \quad -1 \quad 0 \quad +1 \quad +2 \quad +3$$

$$\begin{aligned}\Rightarrow SD &= \sqrt{\frac{(-3)^2+(-2)^2+(-1)^2+0^2+1^2+2^2+3^2}{7}}\\ &= \sqrt{\frac{9+4+1+0+1+4+9}{7}}\\ &= \sqrt{\frac{28}{7}} = \sqrt{4} = 2\end{aligned}$$

The SE is therefore given by:

$$SE = \sqrt{100}\times 2 = 10\times 2 = 20$$

We are now ready to fill in all the blanks:

(a) The sum of the draws is 431. The expected value for the sum of the draws is 400, the observed value is 431, the chance error is 31 ($431-400$), and the standard error is 20.

(b) The sum of the draws is 386. The expected value for the sum of the draws is 400, the observed value is 386, the chance error is -14 ($386-400$), and the standard error is 20.

(c) The sum of the draws is 417. The expected value for the sum of the draws is 400, the observed value is 417, the chance error is 17 ($417-400$), and the standard error is 20.

2 Chapter 18 review exercises

2. (a) As before, we have to find the expected value for the sum, the SD, and the SE in order to answer the question. The average is 4, so the expected sum is $400\times 4 = 1600$. The SD is $\sqrt{\frac{9+1+1+9}{4}} = \sqrt{5} \approx 2.236$. The SE is $20\times 2.236 = 44.72$. For the z-score, we obtain $\frac{1500-1600}{44.72} \approx -2.236$. Using the Normal table again, we find that with a z-score of 2.24, 1.25% of the area under the curve is in one tail. This means that we expect that in 100% - 1.25% = <u>98.75%</u> of all cases the sum of the draws will be more than 1,500.

   (b) In this case, it is easiest to "transform" the box that we are given. We will code obtaining a "3" as a success and assign this event a value of 1, while all other outcomes will be assigned a zero. The transformed box thus still contains four tickets, but now three of those have a zero written on them, and one has a 1 on it (the one that previously was the "3").

   The expected value for one draw from our transformed box is $\frac{0+1+0+0}{4} = \frac{1}{4}$, so our expected sum for 400 draws is $\frac{1}{4}\times 400 = 100$. The SD is $\sqrt{1/4\times 3/4} \approx 0.433$, using the shortcut method. The SE for the sum of draws is given by $\sqrt{400}\times 0.433 = 20\times 0.433 = 8.66$.

   If there are 90 3's, then the observed sum of draws from the transformed box would be 90, which corresponds to the following z-score: $\frac{90-100}{8.66} \approx -1.1547$. The Normal table now shows that this (1.15) translates to an area of 0.1251 in the left tail, meaning that our estimate for the chance that there will be fewer than 90 3's is <u>12.51%</u>.

5. (iii) is a histogram for the numbers drawn, (i) is the probability histogram for the sum, and (ii) is the probability histogram for the product. The histogram for the numbers drawn can have only three blocks (one each for 1, 2, and 3), ruling out (i) and (ii). (i) has the familiar "bell" shape of a probability histogram for a sum, and (iii) that of a probability histogram for the product—which has the most possible outcomes and thus the most blocks.

9. (a) True. To check that the assertions are true, we first have to find the average: $\frac{(99\times 0)+(1\times 1)}{100} = 0.01$. It follows that the expected sum is predicted to be $(100) \times 0.01 = 1$, confirming the first part of the sentence. Now, the SD is given by $\sqrt{\frac{99\times(-0.01)^2+0.99^2}{100}} = \sqrt{0.0099} \approx 0.0995$. For the SE, we thus obtain $\sqrt{100}\times 0.0995 = 10\times 0.0995 = 0.995 \approx 1$. Therefore, the second part of the sentence is also correct.

   (b) Using the empirical rule, this statement would appear to be true. However, because the contents of the box are lopsided, the empirical rule is not appropriate. For example, the empirical rule would tell us that the sum will be between 0 and 1 about 34% (= 68%/2) of the time, and that it will be between −1 and 1 about 47.5% (= 95%/2) of the time. We would thus conclude that the sum would be between −1 and 0 about 47.5% − 34% = 13.5% of the time — but this is impossible, since the sum cannot be negative! Since the empirical rule does not apply in this scenario, the statement is false.

11. Before the math begins, we can rule out one of the options right away, namely "number of 1's." Why? Both (a) and (b) state that the observed value is *above* the one we would expect, yet only 17 1's were drawn (as opposed to 25 expected 1's). We are thus left faced a choice between "sum of the draws" and "number of 2's." But, which one goes where?

    First, we'll compute the SE for the 2's. To do so quickly, once more we "convert" our box to *1 = success* (drawing a 2) and *0 = failure* (drawing a 1 or 5). We expect 50 successes out of 100 draws, and the SD is now 1/2 (found using the shortcut formula: $\sqrt{1/2 \times 1/2}$). The SE is then $\sqrt{100} \times 1/2 = 5$. Thus, the number of 2's will be around 50, give or take 5 or so. The observed value of 54 2's is $(54 - 50)/5 = 0.8$ SEs above the expected value.

    We can then identify the number of 2's as the answer to (a), and we can identify the number of 1's as the answer to (b) by the process of elimination.

    (a) For the number of 2's, the observed value is 0.8 SEs above the expected value.

    (b) For the sum of the draws, the observed value is 1.33 SEs above the expected value.

3 Chapter 20 review exercises

1. Your table should have the following entries:

	# of heads		% of heads	
# of tosses	Expected value	SE	Expected value	SE
100	50	5	50%	5%
2,500	1,250	25	50%	1%
10,000	5,000	50	50%	0.5%
1,000,000	500,000	500	50%	0.05%

   The values in the first column are given. The ones in the second column are found via 1/2×(# of tosses). The values in column three can be calculated via $\sqrt{\#\ of\ tosses} \times 1/2$. For the fourth column, we always expect 50% of the tosses to be heads, so this value remains the same no matter how many tosses we have. For column five, however, we have to compute the size of the absolute SE relative to the # of tosses, i.e., divide the values from column three by those from column one.

6. (i) is the best choice. Refer back to FPP (p. 367, emphasis added):

> When estimating percentages, *it is the absolute size of the sample which determines accuracy*, not the size relative to the population. This is true if the sample is only a small part of the population, which is the usual case.

In this problem the calculated SE would be very small for both states, but California's would be smaller. Because the SEs for both states would be so small, there would be little difference in the accuracy between the states, so (i) is the best answer.

7. (a) This is <u>true</u>. There are a total of $60,000 + 20,000 = 80,000$ tickets in the box, and a quarter of them are 1's. When drawing at random from a 0-1 box, the expected value for the percentage of 1's among the draws is equal to the percentage of 1's in the box. In other words, we expect *exactly* 25% of the draws to be 1's (note that we make this statement *before* we begin drawing).

   (b) This statement is <u>false</u>. We expect the value to be *exactly* 25%, not "around 25%, give or take 2% or so."

   (c) <u>True</u>. Due to chance error, the actual outcome of 500 draws will in most cases be different from our *ex ante* expectation of exactly 25%. What we have to do here is compute the likely range of values that might be observed, rather than a point value.

   The expected value for the sum is $500 \times 1/4 = 125$. The SD is $\sqrt{3/4 \times 1/4} \approx 0.433$. The SE is thus $\sqrt{500} \times 0.433 \approx 9.6822$.

   We are not done yet, though, as the statement refers to the sample percentage, which means that we have to convert to percent relative to the size of the sample. 125 out of 500 is 25%, and 9.6822 out of 500 is $\approx 1.936\%$. Therefore, the percentage of 1's among the draws will be around 25%, give or take 2% or so.

   (d) This statement is <u>false</u>. Given that chance error will influence the outcome of the actual draw, it is extremely unlikely that we will observe exactly 25% of the draws being 1's.

   (e) This statement is <u>true</u>. 20,000 out of 80,000 tickets in the box are 1's, so the percentage of 1's in the box is exactly 25% (note that there's no drawing going on here).

   (f) This is <u>false</u>. There is no uncertainty about the percentage of 1's in the box. It is *exactly* 25%, not somewhere around 25%.

10. <u>Yes</u>, the person has computed the right SE. With a coin, we know that we can use the 0-1 box model. We also know that the SD is 1/2 ($\sqrt{1/2 \times 1/2} = \sqrt{1/4} = 1/2$), and that the shortcut formula $\sqrt{\# \ of \ draws} \times SD$ applies. Therefore, $SE = \sqrt{2000} \times 0.5 \approx 22$.

    As a cross-check, you can look at the table from question 20.1. There we found that for 2,500 tosses of a fair coin, we expect the SE to be 25. As 2,000 is somewhat smaller than 2,500 and 22 is also somewhat smaller than 25, even without the computation we can verify that an SE of 22 will not be too far off the mark.