

Unit 18

The Central Limit Theorem

Objectives:

- To distinguish between the sampling distribution of a sample mean \bar{x} and the sampling distribution of a sample proportion \bar{p}
- To understand that the sampling distribution of \bar{x} and the sampling distribution of \bar{p} exhibit the same properties with simple random samples of size n as n increases

The Central Limit Theorem tells us that when selecting a simple random sample from a parent population which can be treated as an infinite population with mean μ and standard deviation σ , then the following are true:

- (1) The mean of the sampling distribution of \bar{x} is $\mu_{\bar{x}} = \mu$.
- (2) The standard deviation of the sampling distribution of \bar{x} is $\sigma_{\bar{x}} = \sigma / \sqrt{n}$.
- (3) If the parent population is normally distributed, then the sampling distribution of \bar{x} with samples of size n will be normally distributed for any n ; if the parent population is not normally distributed, then the sampling distribution of \bar{x} can be accurately approximated by a normal distribution for sufficiently large n .

We have previously applied these facts to find probabilities concerning the sampling distribution of \bar{x} by using the standard normal distribution (Table A.2). Instead of a mathematical proof of The Central Limit Theorem, we provided an intuitive justification using Tables 14-1, 14-2a, and 14-2b together with Figures 14-3a, 14-3b, and 14-4. These tables and figures were created by considering a cube with the values 0, 6, 9, 10, 11, and 12 each painted on one of its six sides; these values comprised our parent population, and our interest was in how well we might estimate the population mean μ using the sample mean \bar{x} from a simple random sample.

Let us now change this situation somewhat. Suppose there are no values on the six sides of the cube, but instead four of the sides are each painted red, and the other two sides are each painted blue. Instead of estimating a mean μ , our interest will be in estimating the proportion of sides which are red. Of course, you can easily see that the proportion of red sides $4/6 = 2/3 = 0.667$. However, as we did previously for the sake of illustration, we want you to imagine that you have a friend who does not know this proportion and whose job it is to estimate what the proportion of red sides is.

The reason why we are imagining that your friend does not know the proportion is to study how accurately your friend would be able to estimate the value of the proportion using simple random sampling. This will hopefully provide us with further insight into the use of inferential statistics to draw conclusions about the value of a parameter corresponding to a

Table 18-1
Samples of Size $n = 3$ from Rolling a Cube with Four Sides Painted Red and Two Sides Painted Blue

<i>Outcomes on Each Roll</i>			<i>Number of Samples</i>	<i>Sample Proportion of Red</i>
<i>1st</i>	<i>2nd</i>	<i>3rd</i>		
Red	Red	Red	64	$3/3 = 1.000$
Blue	Red	Red	32	$2/3 = 0.667$
Red	Blue	Red	32	$2/3 = 0.667$
Red	Red	Blue	32	$2/3 = 0.667$
Red	Blue	Blue	16	$1/3 = 0.333$
Blue	Red	Blue	16	$1/3 = 0.333$
Blue	Blue	Red	16	$1/3 = 0.333$
Blue	Blue	Blue	8	$0/3 = 0.000$

population, based on the value of a statistic from a simple random sample. In our illustration, the population proportion of red sides on the cube is the parameter of interest, and the proportion of red sides from a simple random sample of rolls is the statistic used to estimate the population proportion.

Let us now imagine that your friend is permitted to roll the cube three times and record the color facing upward each time. Essentially, your friend is selecting a simple random sample of size $n = 3$; your friend is going to use the sample proportion to estimate the population proportion. Before proceeding any further, we must define some notation to distinguish between the two different proportions that we are talking about here. Previously, we have defined μ to represent the mean in a population, and we have defined \bar{x} to represent the mean in a sample; in other words, \bar{x} is a statistic and μ is a parameter.

We shall now define the Greek letter λ

(lambda) to represent the proportion in a population, and define \bar{p} to represent the proportion in a sample. In our present illustration with the cube, $\lambda = 2/3 = 0.667$ is the proportion of red sides in the population consisting of the colors painted on each of the six sides of the cube, and \bar{p} is the sample proportion of red sides resulting from $n = 3$ rolls of the cube. As we have emphasized previously, a parameter such as λ is always a fixed, single value that we are trying to estimate, because we do not know its value, and a statistic such as \bar{p} is a random value which depends on the sample we happen to have selected.

Let us now return to the illustration where your friend does not know that $\lambda = 2/3 = 0.667$, but is going to estimate λ with the value of \bar{p} from a random sample of $n = 3$ rolls of the cube. In order to evaluate your friend's chances of obtaining an estimate of the population proportion which is reasonably close to the correct value $= 2/3 = 0.667$, we shall first need to consider a list of all the samples of size $n = 3$ which are possible. In Table 14-1, where we needed to construct a list of all possible samples of size $n = 2$, we individually listed each possible sample. In this present illustration, it would be a very tedious, long process for us to list each possible sample of size $n = 3$, but we can avoid this by using what is called the *multiplication principle*.

The multiplication principle gives us the number of results which are possible with an ordered sequence of actions. Suppose you are about to order a quick lunch at a cafeteria, and you must first choose one of 5 sandwiches (hamburger, hot dog, grilled cheese, hoagie, tuna fish) after which you must choose one of 3 drinks (soda, juice, milk). How many different combinations are there for lunch? If you match each one of the 5 sandwich choices with each one of the 3 drink choices, you find that there are $(5)(3) = 15$ lunch combinations. You could easily write down all the different possible combinations to see that this is true. In other words, the number of possible combinations can be obtained by multiplying the number of choices for a sandwich and the number of choices for a drink.

Now look at Table 18-1. The first column displays the different types of samples possible with $n = 3$ rolls of the cube, and the second column displays the number of possible samples of each type. The second column was obtained by multiplying the number of choices for the first roll, the number of choices for the second roll, and the number of choices for the third roll. For instance, look at the second row of Table 18-1 where the first roll results in Blue, the second roll results in Red, and the third roll results in Red. Since there are 2 sides (choices) for Blue and 4 sides (choices) for Red, then the number of possible samples must be $(2)(4)(4) = 32$, which is displayed in the second column of the second row. For a more detailed explanation of

<i>Outcomes on Each Roll</i>			<i>Number of Samples</i>	<i>Sample Mean</i>
<i>1st</i>	<i>2nd</i>	<i>3rd</i>		
1	1	1	64	$3/3 = 1.000$
0	1	1	32	$2/3 = 0.667$
1	0	1	32	$2/3 = 0.667$
1	1	0	32	$2/3 = 0.667$
1	0	0	16	$1/3 = 0.333$
0	1	0	16	$1/3 = 0.333$
0	0	1	16	$1/3 = 0.333$
0	0	0	8	$0/3 = 0.000$

how the multiplication principle was applied to obtain the second column of Table 18-1, the interested reader may consult Section B.1 of Appendix B.

In the third column of Table 18-1, we calculated the proportion of red sides for each sample by dividing the number of red sides in the sample by the sample size $n = 3$. This, of course, is the obvious way to calculate the proportions. However, if we were to use the code value 1 (one) to represent a red side and use the code value 0 (zero) to represent a blue side, we would find that calculating a proportion is really the same as calculating a mean! Table 18-2 is an exact copy of Table 18-1 except that "Red" has been replaced by "1" and "Blue" has been replaced by "0" in the first column, and that the heading for the third column has been changed from "Sample Proportion of Red" to "Sample Mean." We see then that each sample proportion of red sides is exactly the same as the mean of the 0s and 1s in the sample. In other words, when we use the codes 0 and 1, \bar{p} and \bar{x} are really the same thing!

This concept can be applied whether we are talking about samples or populations. Our population consists of 4 red sides and 2 blue sides, implying that the population proportion of red sides is $\lambda = 4/6 = 2/3 = 0.667$. However, if we once again code each red side with the value 1, and code each blue side with the value 0, we find that the population mean is $\mu = (1 + 1 + 1 + 1 + 0 + 0) / 6 = 4/6 = 2/3 = 0.667$. In other words, when we use the codes 0 and 1, λ and μ are really the same thing.

From either Table 18-1 or 18-2, we see that the total number of possible samples of size $n = 3$ is $64 + 32 + 32 + 32 + 16 + 16 + 16 + 8 = 216$. With simple random sampling, each of these 216 samples has an equal chance (1

out of 216) of being the one which your friend will obtain with $n = 3$ rolls of the cube and therefore the one from which your friend's estimate of λ will come. In order to evaluate your friend's chances of obtaining an estimate of the population proportion which is reasonably close to the correct value of $\lambda = 2/3 = 0.667$, we construct the sampling distribution of \bar{p} which is a frequency distribution displaying the possible values of \bar{p} together with the corresponding relative frequency of times each value occurs with simple random sampling.

Table 18-3 lists each of the possible sample proportions together with the frequency of times it occurs. This frequency distribution represents the *sampling distribution of \bar{p}* based on simple random sampling with $n = 3$, and Figure 18-1 is a histogram displaying the relative frequencies of this sampling distribution. Looking at either Table 18-3 or Figure 18-1, you should notice that there are only four possible sample proportions that we can observe with $n = 3$ rolls: 0.000, 0.333, 0.667, or 1.000. The sample proportion with the highest probability is 0.667, which is exactly equal to the population proportion λ that your friend is trying to estimate. Your friend has a 44.44% chance of observing a sample proportion equal to the population proportion that you are trying to estimate. Let us find out how this would change with a larger sample size. (In Self-Test Problem 18-1, we have you consider the sampling distribution of \bar{p} with $n = 4$.)

<i>Sample Proportion of Red</i>	<i>Relative Frequency</i>
0/3 = 0.000	8/216 = 3.70%
1/3 = 0.333	48/216 = 22.22%
2/3 = 0.667	96/216 = 44.44%
3/3 = 1.000	64/216 = 29.63%

<i>Sample Proportion of Red</i>	<i>Relative Frequency</i>
0/10 = 0.0	0.002%
1/10 = 0.1	0.034%
2/10 = 0.2	0.305%
3/10 = 0.3	1.626%
4/10 = 0.4	5.690%
5/10 = 0.5	13.656%
6/10 = 0.6	22.761%
7/10 = 0.7	26.012%
8/10 = 0.8	19.509%
9/10 = 0.9	8.671%
10/10 = 1.0	1.734%

It would be a time consuming task to construct by hand the sampling distribution of \bar{p} with $n = 10$ rolls, but with the aid of a computer, we did construct this sampling distribution, which is displayed in Table 18-4. Figure 18-2 is a histogram displaying this sampling distribution. Looking at either Table 18-4 or Figure 18-2, you should notice that there are only 11 possible sample proportions that we can observe with $n = 10$ rolls: 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0. The sample proportions with the highest probabilities are 0.6 and 0.7, each of which is close to the population proportion λ that your friend is trying to estimate.

As you compare Figures 18-1 and 18-2, you should notice two things. First, you should realize that although the values of the sample proportion resulting from $n = 3$ rolls and the values of the sample proportion resulting from $n = 10$ rolls must both be between 0 and 1, there are a larger number of possible values for the sample proportion resulting from $n = 10$ rolls than there are for the sample proportion resulting from $n = 3$ rolls. In a sense, then, we might say that the values of the sample proportion resulting $n = 10$ rolls is closer to being on a continuum than the values of the sample proportion resulting from $n = 3$ rolls. The second thing to notice is that each of the two extreme values of 0 and 1 are considerably less likely to occur with the sample proportion resulting from $n = 10$ rolls than with the sample proportion resulting from $n = 3$ rolls; with $n = 10$ rolls, it is more likely to observe a sample proportion in between closer to $\lambda = 2/3 = 0.667$, than it is with $n = 3$ rolls.

It is important to keep in mind that we are actually considering many different populations simultaneously: the parent population and each of the populations consisting of the possible values of \bar{p} corresponding to a given n . We use λ to represent the proportion of red sides in the parent population, but we know that if we label the red sides with 1 and label the blue sides with 0, then the mean corresponding to the parent population is λ , that is, $\mu = \lambda$. We shall use $\mu_{\bar{p}}$ to represent the mean corresponding to the population consisting of the possible values of \bar{p} corresponding to a given sample size n ; that is, $\mu_{\bar{p}}$ is the mean of the sampling distribution of \bar{p} with samples of size n .

By thinking of the sample proportions as sample means (calculated from 0s and 1s), we should expect to observe the same phenomenon that we have observed previously with sample means. Specifically, we should expect that when simple random samples of size n are selected from a parent population with mean $\mu = \lambda$, then the following three things will be true: first, $\mu_{\bar{p}} = \mu = \lambda$; second, the variation in the sampling distribution of \bar{p}

Figure 18-1

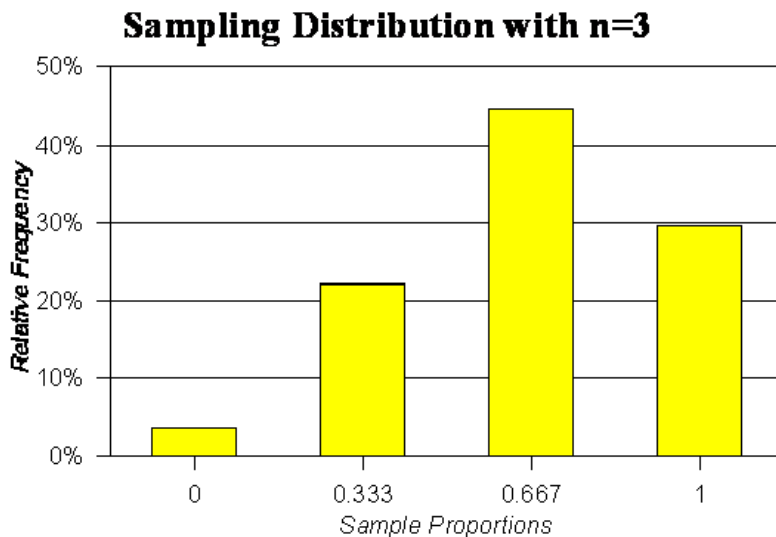
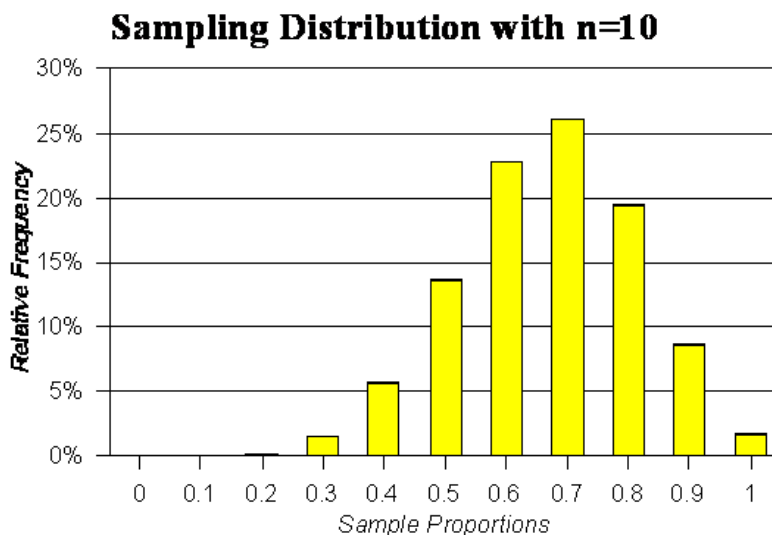


Figure 18-2



is less than in the parent population and decreases as n increases; third, the sampling distribution of \bar{p} becomes more bell-shaped as n increases. Returning to Table 18-3, we find that 8 sample proportions are equal to 0, 48 sample proportions are equal to 0.333, 96 sample proportions are equal to 0.667, and 64 sample proportions are equal to 1.000. Using this information, we find

$$\mu_{\bar{p}} = \frac{(8)(0) + (48)(0.333) + (96)(0.667) + (64)(1)}{216} = \frac{144}{216} = \frac{2}{3} = 0.667 = \lambda,$$

which demonstrates that the population mean of all sample proportions is equal to the mean of the parent population; that is, $\mu_{\bar{p}} = \mu = \lambda$. By comparing Figures 18-1 and 18-2, we see how the sampling distribution of \bar{p} shows less variation and becomes more bell-shaped as n increases. Figure 18-1 has somewhat of a bell-shape but is negatively skewed, whereas Figure 18-2 has a very bell-shaped distribution and shows considerably less variation than Figure 18-1.

Table 18-5
Samples of Size $n = 4$ from Rolling a Cube with Four Sides Painted Red and Two Sides Painted Blue

<i>Outcomes on Each Roll</i>				<i>Number of Samples</i>	<i>Sample Proportion of Red</i>
<i>1st</i>	<i>2nd</i>	<i>3rd</i>	<i>4th</i>		
Red	Red	Red	Red	256	$3/3 = 1.000$
Blue	Red	Red	Red	128	$2/3 = 0.667$
Red	Blue	Red	Red	128	$2/3 = 0.667$
Red	Red	Blue	Red	128	$2/3 = 0.667$
Red	Red	Red	Blue	128	
Blue	Blue	Red	Red	64	
Blue	Red	Blue	Red	64	
Blue	Red	Red	Blue	64	
Red	Blue	Blue	Red	64	
Red	Blue	Red	Blue	64	
Red	Red	Blue	Blue	64	
Blue	Blue	Blue	Red	32	
Blue	Blue	Red	Blue	32	
Blue	Red	Blue	Blue	32	
Red	Blue	Blue	Blue	32	
Blue	Blue	Blue	Blue	16	

Self-Test Problem 18-1. Suppose that 4 sides of a cube are each painted red, and the other 2 sides are each painted blue. The first column of Table 18-5 lists the different types of samples that are possible if the cube is rolled $n = 4$ times. The second column of Table 18-5 lists the number of ways each corresponding type of sample can occur, and the third column lists the corresponding sample proportions of red.

- (a) Optional: Use the multiplication principle to verify that the second column of Table 18-5 is correct.
- (b) Complete the third column of Table 18-5.
- (c) What is the total number of possible samples of size $n = 4$?
- (d) Construct a relative frequency table and a corresponding histogram to display the sampling distribution of \bar{p} with $n = 4$.
- (e) Find the mean ($\mu_{\bar{p}}$) of the sampling distribution of \bar{p} with $n = 4$, and verify that $\mu_{\bar{p}} = \lambda$.
- (f) How does the dispersion in the sampling distribution of \bar{p} with $n = 4$ compare with the dispersion in the sampling distribution of \bar{p} with $n = 3$ (displayed in Figure 18-1)?
- (g) How does the dispersion in the sampling distribution of \bar{p} with $n = 4$ compare with the dispersion in the sampling distribution of \bar{p} with $n = 10$ (displayed in Figure 18-2)?
- (h) How does the shape of the sampling distribution of \bar{p} with $n = 4$ compare with the shape of the sampling distribution of \bar{p} with $n = 3$ (displayed in Figure 18-1)?
- (i) How does the shape of the sampling distribution of \bar{p} with $n = 4$ compare with the shape of the sampling distribution of \bar{p} with $n = 10$ (displayed in Figure 18-2)?

Since we have demonstrated that a sample proportion \bar{p} is really a sample mean \bar{x} calculated from zeros (0s) and ones (1s), we should not be surprised that properties of the sampling distribution of \bar{x} also apply to the sampling distribution of \bar{p} . A parent population consisting only of zeroes (0s) and ones (1s) cannot possibly have a normal distribution; however, The Central Limit Theorem tells us that the sampling distribution of \bar{p} can be treated as a normal distribution with a sufficiently large sample size n .

In order to apply The Central Limit Theorem, we need the mean μ and standard deviation σ from the parent population. We have already seen that the mean for the parent population is $\mu = \lambda$, where λ is the population proportion of ones. It is possible to prove that the standard deviation for the parent population consisting only of zeroes (0s) and ones (1s) is $\sigma = \sqrt{\lambda(1-\lambda)}$, and we shall accept this fact without proof. This fact implies that the standard deviation of the sampling distribution of \bar{p} must be $\sigma_{\bar{p}} = \sigma / \sqrt{n} = \sqrt{\lambda(1-\lambda)} / \sqrt{n} = \sqrt{\lambda(1-\lambda)/n}$. We may now state a version of The Central Limit Theorem for sample proportions.

The Central Limit Theorem for sample proportions. When selecting a simple random sample from a parent population which can be treated as an infinite population consisting only of 1s (ones) and 0s (zeros), where λ is the proportion of 1s (ones), then the following are true:

- (1) The mean of the sampling distribution of \bar{p} is $\mu_{\bar{p}} = \lambda$.
- (2) The standard deviation of the sampling distribution of \bar{p} is $\sigma_{\bar{p}} = \sqrt{\frac{\lambda(1-\lambda)}{n}}$.
- (3) If $n\lambda \geq 5$ and $n(1-\lambda) \geq 5$, then the sampling distribution of \bar{p} with samples of size n can be accurately approximated by a normal distribution.

The statement of the Central Limit Theorem for sample proportions gives us a rule for deciding whether the sample size n is large enough so that the sampling distribution of \bar{p} can be treated as a normal distribution; however, for the general version of the Central Limit Theorem concerning the sampling distribution of \bar{x} , there is no universal rule for deciding whether the sample size is sufficiently large. Notice that in order to use the rule stated in the Central Limit Theorem for sample proportions to decide whether n is sufficiently large, we must know the value of λ ; in situations where we have no knowledge about the value of λ , we can simply use the "worst case scenario" which is to assume $\lambda = 1/2 = 0.5$. All the previous comments concerning the distinction between infinite and finite populations apply to the sampling distribution of \bar{p} the same way they apply to the sampling distribution of \bar{x} .

Looking at Figures 18-1 and 18-2, we see how the sampling distribution of the proportion of red sides observed in rolls of the cube looks more like a normal distribution as the sample size n (number of rolls) increases. However, we find that with $\lambda = 2/3 = 0.667$ and n as small as 10, it is not true that $n(1 - \lambda) \geq 5$. This suggests that we should not use a normal distribution to approximate either of the sampling distributions displayed in Figures 18-1 and 18-2. It is easy to check that in order for us to have both $n\lambda \geq 5$ and $n(1 - \lambda) \geq 5$, we must have n at least equal to 15.

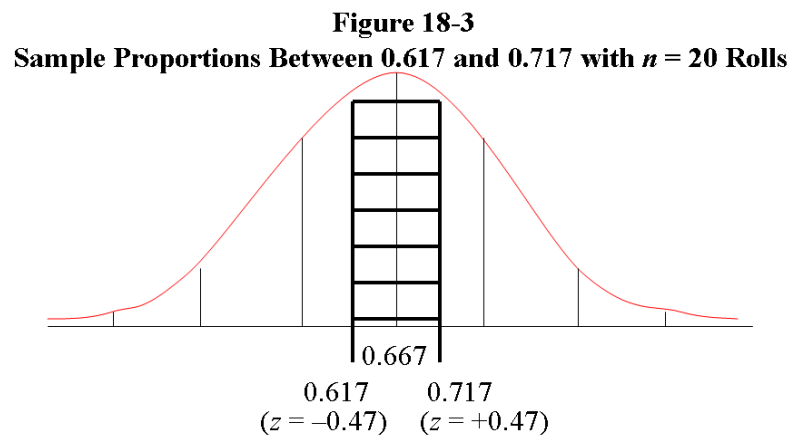
Let us now pose the following question: What is the probability that the value of \bar{p} obtained from $n = 20$ rolls of the cube will be between 0.617 and 0.717? In other words, if we did not know that $\lambda = 2/3 = 0.667$, we would be asking for the probability that the value of \bar{p} , our estimate of λ , obtained from $n = 20$ rolls of the cube is not more than a distance of 0.05 below or above 0.667. We shall obtain this probability by using the fact that the sampling distribution can be treated as a normal distribution with mean $\mu_{\bar{p}} = \lambda = 2/3 = 0.667$ and standard deviation $\sigma_{\bar{p}} = \sqrt{(2/3)(1 - 2/3)/20} = 0.1054$. To find this probability, we first find the z -score of 0.617 to be

$$\frac{0.617 - 0.667}{0.1054} = -0.47,$$

and find the z -score of 0.717 to be

$$\frac{0.717 - 0.667}{0.1054} = +0.47.$$

The probability that \bar{p} from a simple random sample of $n = 20$ rolls will be between 0.617 and 0.717 is the proportion of shaded area in Figure 18-3. The unshaded area below 0.617 in Figure 18-3

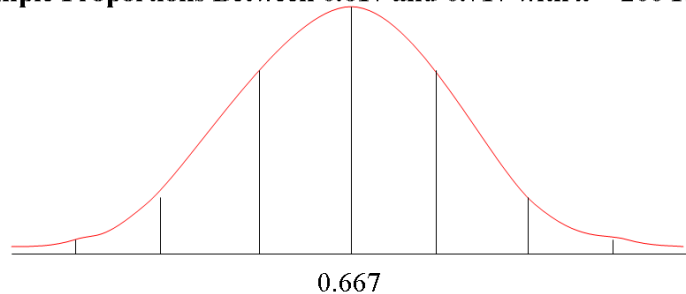


is a mirror image of the shaded area in the figure at the top of Table A.2; also, the unshaded area above 0.717 in Figure 18-3 corresponds exactly to the shaded area in the figure at the top of Table A.2. Since a normal distribution is symmetric, we can find the total unshaded area by doubling the entry of Table A.2 in the row labeled 0.4 and the column labeled 0.07; we then obtain the desired area by subtracting this unshaded area from 1. The probability that \bar{p} from a simple random sample of $n = 20$ rolls will be between 0.617 and 0.717 is $1 - (0.3192 + 0.3192) = 0.3616$ (or 36.16%).

As another illustration, we shall have you find the probability that the sample proportion \bar{p} from a simple random sample of $n = 200$ rolls will be between 0.617 and 0.717. Find the z -score for each of 0.617 and 0.717, and label the values 0.617 and 0.717 on the horizontal axis in Figure 18-4. Then, shade the desired area under the normal curve, and use Table A.2 to obtain the desired probability. (You should find that the z -scores are -1.50 and $+1.50$, and that the probability that \bar{p} from a simple random sample of $n = 200$ rolls will be between 0.617 and 0.717 is 0.8664 (or 86.64%).)

It should be clear that the probability that \bar{p} will be between 0.617 and 0.717 increases as the sample size increases. One could ask how large the simple random sample size must be in order that the probability of obtaining \bar{p} between 0.617 and 0.717 is very large, say around 95%. The answer to this particular question is that the simple random sample size must be at least $n = 342$ rolls. In general, to find the simple random sample size needed to insure that the probability of obtaining a sample

Figure 18-4
Sample Proportions Between 0.617 and 0.717 with $n = 200$ Rolls



proportion no more than d away from λ is at least 95%, one can solve the equation $d = 1.96\sqrt{\lambda(1-\lambda)/n}$ for n where $\lambda(1-\lambda)$ is set equal to a value depending on our knowledge about λ (i.e., $\lambda = 1/2$ implies we have no knowledge about λ). The value 1.96 is in the equation because the z -scores between which lies 95% of the area under a standard normal density curve are -1.96 and $+1.96$.

Self-Test Problem 18-2. In a certain large city, 30% of all the passengers who ride the city bus are senior citizens.

- If \bar{p} is the proportion of senior citizens in a simple random sample of n passengers, is it reasonable to treat the sampling distribution of \bar{p} with $n \geq 100$ as a normal distribution? Why or why not?
- What is the smallest sample size n with which we could be reasonably certain that the sampling distribution of \bar{p} is approximately a normal distribution?
- Draw a sketch illustrating the probability that the proportion of senior citizens in a simple random sample of 200 passengers is within 0.04 of the population proportion of senior citizens, and find this probability.
- Draw a sketch illustrating the probability that the proportion of senior citizens in a simple random sample of 500 passengers is within 0.04 of the population proportion of senior citizens, and find this probability.
- Draw a sketch illustrating the probability that the proportion of senior citizens in a simple random sample of 500 passengers is more than 0.02 away from (below or above) the population proportion of senior citizens, and find this probability.
- Draw a sketch illustrating the probability that the proportion of senior citizens in a simple random sample of 400 passengers is greater than 0.28, and find this probability.
- Draw a sketch illustrating the probability that the proportion of passengers who are not senior citizens in a simple random sample of 400 passengers is less than 0.72, and find this probability.
- Is it very unlikely to select a simple random sample of 200 passengers with a sample proportion of senior citizens more than 0.03 away from (below or above) the population proportion of senior citizens? Why or why not?
- Is it very unlikely to select a simple random sample of 2000 passengers with a sample proportion of senior citizens more than 0.03 away from (below or above) the population proportion of senior citizens? Why or why not?

Self-Test Problem 18-3. Fifteen hundred voters in a state are selected and polled about how they intend to vote in the upcoming election for governor, in order to make a conclusion about whether or not the proportion of voters intending to vote for candidate Jones gives her a majority of all voters.

- Identify the sample, the population of interest, the parameter of interest, and the statistic used to estimate the parameter.
- Suppose the sample of voters is selected by polling voters at various shopping malls until 1500 are obtained. Is it reasonable to treat this as a simple random sample? Why or why not?

Answers to Self-Test Problems

- 18-1** (a) The entries in the second column are correct. (b) The respective entries in the third column are 0.75, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.25, 0.25, 0.25, 0.25, 0. (c) 1296 (d) The histogram will have a bar or line with a height of 1.32% at 0, a bar or line with a height of 9.88% at 0.25, a bar or line with a height of 29.63% at 0.5, a bar or line with a height of 39.51% at 0.75, and a bar or line with a height of 19.75% at 1. (e) $\mu_{\bar{p}} = \lambda = 2/3 = 0.667$ (f) The sampling distribution of \bar{p} with $n = 4$ shows less variation than the sampling distribution of \bar{p} with $n = 3$. (g) The sampling distribution of \bar{p} with $n = 4$ shows more variation than the sampling distribution of \bar{p} with $n = 10$. (h) The sampling distribution of \bar{p} with $n = 4$ is more symmetric and bell-shaped than sampling distribution of \bar{p} with $n = 3$. (i) The sampling distribution of \bar{p} with $n = 4$ is less symmetric and bell-shaped than sampling distribution of \bar{p} with $n = 10$.
- 18-2** (a) We may treat the sampling distribution of \bar{p} with $n \geq 100$ as a normal distribution, because $(100)(0.3) = 30$ and $(100)(0.7) = 70$ are both greater than 5. (b) Since $n(0.3)$ and $n(0.7)$ are both greater than or equal to 5 for $n = 17$ but not for $n = 16$, then $n = 17$ is the smallest sample size n with which we could be reasonably certain that the sampling distribution of \bar{p} approximately a normal distribution. (c) 0.7814 or 78.14% (d) 0.9488 or 94.88% (e) 0.6730 or 67.30% (f) 0.8078 or 80.78% (g) 0.8078 or 80.78% (h) Since 35.24% of all simple random samples of 200 passengers have a sample proportion more than 0.03 away from the population proportion, we cannot consider this very unlikely. (i) Since only 0.34% of all simple random samples of 2000 passengers have a sample proportion more than 0.03 away from the population proportion, we should consider this very unlikely.
- 18-3** (a) The sample is the 1500 voters selected and polled, the population of interest is all voters in the state, the parameter of interest is the proportion of all voters in the state intending to vote for candidate Jones, and the statistic used to estimate the parameter is the proportion of the 1500 voters sampled intending to vote for candidate Jones. (b) If one assumes that all voters are equally likely to shop at the malls, then one might reasonably consider the voters selected and polled to be a simple random sample; however, it is probably not true that all voters are equally likely to shop at the malls.

Summary

The *sampling distribution of \bar{p}* based on simple random sampling is a frequency distribution displaying the possible values of \bar{p} together with the corresponding relative frequency of times each value occurs with random sampling. Since a sample proportion \bar{p} is really a sample mean \bar{x} calculated from zeros (0s) and ones (1s), the properties of the sampling distribution of \bar{x} also apply to the sampling distribution of \bar{p} . The Central Limit Theorem for sample proportions says that when selecting a simple random sample from a parent population which can be treated as an infinite population consisting only of 1s (ones) and 0s (zeros), where λ is the proportion of 1s (ones), then the following are true:

- (1) The mean of the sampling distribution of \bar{p} is $\mu_{\bar{p}} = \lambda$.
- (2) The standard deviation of the sampling distribution of \bar{p} is $\sigma_{\bar{p}} = \sqrt{\frac{\lambda(1-\lambda)}{n}}$.
- (3) If $n\lambda \geq 5$ and $n(1-\lambda) \geq 5$, then the sampling distribution of \bar{p} with samples of size n can be accurately approximated by a normal distribution.

In order to use the rule for deciding whether the sample size is large enough so that the sampling distribution of \bar{p} can be treated as a normal distribution, we must know the value of λ . In situations where we have no knowledge about the value of λ , we can simply use the "worst case scenario" which is to assume $\lambda = 1/2 = 0.5$. All the previous comments concerning the distinction between infinite and finite populations apply to the sampling distribution of \bar{p} the same way they apply to the sampling distribution of \bar{x} .