Chapter 7: Sampling Distributions

1. A news website claims that 30% of all Major League Baseball players use performance-enhancing drugs (“PEDs”) Indignant at this claim, league officials conducts a survey in which 200 randomly selected baseball players are tested. Of the tested players, 52, or 26%, test positive. Which of the following statements about this situation is *true*?

\*A. The number 26% is a statistic.

AR. Correct. The number 26% refers to the percentage of sampled players who used PEDs, and therefore it is a statistic.

B. The number 30% is a statistic.

BR. Incorrect. The number 30% is a parameter, since it is a characteristic of the entire population of interest, that is, the population of all Major League Baseball players.

C. The number 26% is a parameter.

CR. Incorrect. The number 26% is a statistic, since it is measured from the *sample* of 200 players and would very probably be different if the sample were 200 different players.

2. Suppose you take a sample of 50 students from your school and measure their height. Which one of the following is a random variable?

A. The true mean height of all students from your school.

AR. Incorrect. This is a population parameter: a fixed characteristic of the population that does not vary.

B. The mean of the sampling distribution of mean heights for samples of size 50.

BR. Incorrect. The sample mean is a unbiased estimator of the population mean, so the “mean of the sample means” is exactly equal to the population mean, and does not vary.

\*C. The mean of the sample data.

CR. Correct. This value will vary, because the particular students in the sample vary. Thus it is a random variable.

3. Suppose you take a sample of 50 students from your school and find the mean height of the students in your sample. Which one of the following does the sampling distribution of the mean describe?

A. The distribution of individual heights in your sample.

AR. Incorrect. The sampling distribution describes the values we can expect to get for the *mean* of samples of this size. In other words, it’s the probability distribution for the sample mean—a random variable whose value depends on the particular sample that is selected.

\*B. The distribution of means of all possible samples of size 50 that could be selected from the students in your school.

BR. Correct.

C. The distribution of means of all samples of size 50 that have actually been selected?

CR. Incorrect. In practice we can only take one sample. The sampling distribution describes the values we can expect to get for the mean of samples of this size. In other words, it’s the probability distribution for the sample mean—a random variable whose value depends on the particular sample that is selected.

4. In a simple random sample of 1000 Americans, it was found that 61% were satisfied with the service provided by the dealer from which they bought their car. In a simple random sample of 1000 Canadians, 58% said that they were satisfied with the service provided by their car dealer. Which of the following statements concerning the sampling variability of these statistics is *true*?

\*A. The sampling variability is about the same in both cases.

AR. Correct. As long as the population is much larger than the sample (say, at least 10 times as large), the spread of the sampling distribution for a sample of fixed size *n* will be approximately the same for any population size. Here, *n* = 1000, and the populations of car owners of both the United States and Canada are at least 10 times this large. Therefore, the sampling variability associated with “the proportion satisfied out of 1000” is about the same in both cases.

B. The sampling variability is much smaller for the statistic based on the sample of 1000 Canadians because the population of Canada is smaller than that of the United States and, therefore the sample is a larger proportion of the population.

BR. Incorrect. As long as the population is much larger than the sample (say, at least 10 times as large), the spread of the sampling distribution for a sample of fixed size *n* will be approximately the same for any population size. In this case, the population sizes are not relevant, since both populations are at least 10 times as large as the sample size *n* = 1000.

C. The sampling variability is much larger for the statistic based on the sample of 1000 Canadians, because Canada has a lower population *density* than the United States and having subjects living farther apart always increases sampling variability.

CR. Incorrect. The additional information supplied in this answer is irrelevant to the question of interest. The physical distance between subjects or units need not, in general, have any effect on sampling variability.

5. In a statistics class of 250 students, each student is instructed to toss a coin 20 times and record the value of , the sample proportion of heads. The instructor then makes a histogram of the 250 values of obtained. In a second statistics class of 200 students, each student is told to toss a coin 40 times and record the value of , the sample proportion of heads. The instructor then makes a histogram of the 200 values of  obtained. Which of the following statements regarding the two histograms of -values is *true*?

A. The first class’s histogram is more biased because it is derived from a smaller number of tosses per student.

AR. Incorrect. The sample proportion  is an unbiased estimator of the parameter *p*. This property holds regardless of the number of trials (tosses) used to get each estimate of .

\*B. The first class’s histogram has greater spread (variability) because it is derived from a smaller number of tosses per student.

BR. Correct. Larger samples (based on more trials or tosses) yield sampling distributions with smaller spreads. If the coin had been tossed 100 times by each student, the variability would have been even smaller.

C. The first class’s histogram has less spread (variability) because it is derived from a larger number of students.

CR. Incorrect. The number of students (200 or 250) is the size of each “simulation.” It doesn’t affect the variability of the estimate. Using more students would give a more accurate approximation of the sampling distribution. However, the variability of the estimator depends only on the number of trials (tosses) used to get each estimate, not on how large the “simulation” is.

6. As part of a promotion for a new type of cracker, free samples are offered to shoppers in a local supermarket. The probability that a shopper will buy a package of crackers after tasting the free sample is 0.2. Different shoppers can be regarded as independent trials. Let  be the sample proportion of the next 100 shoppers that buy a package of crackers after tasting a free sample. Which of the following *best* describes the sampling distribution of the statistic ?

A. It is approximately Normal with 

AR. Incorrect. The distribution of is approximately Normal with mean 0.2, but the standard deviation is not 0.0016. The variance of  is = 0.0016. You have neglected to take the square root.

\*B. It is approximately Normal with 

BR. Correct. The distribution of  is approximately Normal with mean p = 0.2 and standard deviation .

C. but the shape is non-Normal.

CR. Incorrect. Since *n* = 100 and *p* = 0.2 satisfy the conditions *np* ≥ 10 and *n*(1 – *p*) ≥ 10, we can use the Normal distribution to approximate the sampling distribution of .

7. A 2011 poll by the Pew Research Center found that 66% of young women ages 18 to 34 rated a high-paying career high on their list of life priorities. Let’s assume for this problem that 66% of the female students in your high school consider a high-paying career as a priority. You plan to take a sample of 75 female students. If you calculate the variance of the sampling distribution of proportions for a sample of 75 using , what assumption are you making about this situation?

\*A. You are assuming that there are at least 750 female students at your high school.

AR. Correct. You must satisfy the 10% condition for calculating the standard deviation of the sampling distribution using this formula.

B. You are assuming the population is Normally distributed.

BR. Incorrect. This makes no sense in this situation. The variable “rates a high-paying career as a priority is categorical!

C. You are assuming the central limit theorem applies.

BR. Incorrect. The central limit theorem refers to the shape of the sampling distribution of *means.* This situation involves the sampling distribution of proportions.

8. As part of a promotion for a new type of cracker, free samples are offered to shoppers in a local supermarket. The probability that a shopper will buy a package of crackers after tasting the free sample is 0.2. Different shoppers can be regarded as independent trials. Let  be the sample proportion of the next 100 shoppers that buy a package of crackers after tasting a free sample. The probability that fewer than 30% of these individuals buy a package of crackers after tasting a sample is closest to which of the following?

A. 0.3.

AR. Incorrect. You have mistaken the value 30% for the desired probability, P( ≤ 0.3). First decide what you know about the distribution of .

\*B. 0.9938.

BR. Correct. The statistic  has an approximately Normal distribution with  and . So 

C. 0.0062.

CR. Incorrect. You have found .

9. Suppose you take a sample of size 16 from a large Normally distributed population with mean μ = 64 and standard deviation σ = 10. What are the mean and standard deviation of the sampling distribution of ?

A. 

AR. Incorrect. You have the correct mean, but check the formula for the standard deviation of the sampling distribution of means.

B. 

BR. Incorrect. You have the correct mean, but the standard deviation of the sampling distributed is smaller than that of the population.

\*C. 

CR. Correct. Since the sample mean is an unbiased estimator of the population mean, . For independent samples (when the sample is less than 10% of the population), 

10. The American Community Survey estimates that the mean household income in the United States in 2011 was $69,800, with a standard deviation of $17,000. The distribution is strongly skewed to the right. Suppose we take a simple random sample of 36 households. Which of the following accurately describes the shape of the sampling distribution of the mean household income?

A. strongly skewed right—about as much as the population.

AR. Incorrect. For sufficiently large samples from a skewed population, the sampling distribution is not skewed. Review the central limit theorem!

B. skewed right, but not as strongly as the population.

BR. Incorrect. For sufficiently large samples from a skewed population, the sampling distribution is not skewed. Review the central limit theorem!

\*C. Approximately Normal

CR. Correct. Because the sample is relative large (*n* > 30), the sampling distribution of means is approximately Normal by the central limit theorem.

11. The American Community Survey estimates that the mean household income in the United States in 2011 was $69,800, with a standard deviation of $17,000. The distribution is strongly skewed to the right. Suppose we take a simple random sample of 36 households. Which of the following is closest to the probability that the sample mean is less than $65,000?

\*A. 0.0455

AR. Correct. The sampling distribution is approximately Normal with  and . So .

B. 0.3897

BR. Incorrect. It appears you used the standard deviation of the population in your calculation, not the standard deviation of the sampling distribution.

C. 0.9545

CR. Incorrect. This is the probability that the sample mean is *greater* than $65,000.

12. The American Community Survey estimates that the mean household income in the United States in 2011 was $69,800, with a standard deviation of $17,000. The distribution is strongly skewed to the right. Suppose we take a simple random sample of 36 households. Which of the following is closest to the 80th percentile for the sample mean?

A. $72,067

AR. Incorrect. You are using 0.80—the 80th percentile—as the standard score. Find the correct standard score for the 80th percentile using the Standard Normal probabilities table.

\*B. $72,180

BR. Correct. From the Standard Normal table, the 80th percentile has a standard score of 0.84. So the 80th percentile is 0.84 standard deviations above the mean. That is, .

C. $84,080

CR. Incorrect. It appears you used the standard deviation of the population in your calculation, not the standard deviation of the sampling distribution.