Classroom Voting Questions: Statistics

The Sampling Distribution of a Sample Mean

1. Jamie randomly selects 25 houses that are for sale in the U.S. The shape of the distribution of their prices is probably

   (a) significantly right-skewed

   (b) approximately bell-shaped

   (c) significantly left-skewed

   Answer: (a). This is the first in a series of questions designed to solidify understanding of the sampling distribution of the sample mean and the Central Limit Theorem. This first question is review. House prices have a right-skewed distribution. A sample distribution approximates the distribution from which it is drawn. So we expect the sample of 25 prices to be right-skewed.

   by David A. Huckaby

   STT.05.02.010

   AS DH 3321 010 F16: **84**/13/3 time 1:00
   AS DH 1342 010 F17: **77**/13/9 time 1:00
   AS DH 1342 020 F18: **43**/57/0 time 1:30
   AS DH 1342 040 S19: **53**/47/0 time 1:40
   AS DH 1342 030 F19: **55**/45/0 time 2:00
   AS DH 1342 030 S20: **69**/31/0 time 2:20

2. Jamie randomly selects 100 houses that are for sale in the U.S. The shape of the distribution of their prices is probably

   (a) significantly right-skewed

   (b) approximately bell-shaped

   (c) significantly left-skewed

   Answer: (a). House prices have a right-skewed distribution. A sample distribution approximates the distribution from which it is drawn. So we expect the sample of 100 prices to be right-skewed. [Follow-up question: How is taking a sample of 100 prices different from taking a sample of 25 prices? Answer: The larger the sample, the better the sample distribution approximates the population distribution.]

   by David A. Huckaby

STT.05.02.020

AS DH 3321 010 F16: **90**/10/0 time 1:00
AS DH 1342 040 S19: **79**/21/0 time 2:30
AS DH 1342 030 F19: **28**/72/0 time 2:00
CC KC MA315 S20: **7**/93/0

3. Jamie asks 25 of her friends to each randomly select 4 houses that are for sale in the U.S. and average their 4 prices together. The distribution of these 25 averages is probably

   (a) significantly right-skewed

   (b) approximately bell-shaped

   (c) significantly left-skewed

   Answer: (a). House prices have a right-skewed distribution. According to the Central Limit Theorem, the sampling distribution of the sample mean will more closely approximate a normal distribution as the sample size increases. For very small sample sizes, the sampling distribution of the sample mean will look more like the population distribution than a normal distribution. A very rough rule of thumb: The sampling distribution is approximately normal for sample sizes greater than 30. Here the sample size is 4, so we expect the sampling distribution of the sample mean to be right-skewed, though not to the extent that the population distribution is. [Follow-up question: Is the distribution as right-skewed as the population distribution? Another follow-up question: Consider the case in which Jamie asks 25 of her friends to each randomly select 1 house that is for sale in the U.S. and, ahem, to average its price. What will the shape of the distribution be? Why?]

   by David A. Huckaby

   STT.05.02.030

   AS DH 3321 010 F16: **52**/45/3 time 2:00
   AS DH 1342 020 F18: **86**/7/7 time 1:50

4. Jamie asks 100 of her friends to each randomly select 4 houses that are for sale in the U.S. and average their 4 prices together. The distribution of these 100 averages is probably

   (a) significantly right-skewed

   (b) approximately bell-shaped

   (c) significantly left-skewed

Answer: (a). House prices have a right-skewed distribution. According to the Central Limit Theorem, the sampling distribution of the sample mean will more closely approximate a normal distribution as the sample size increases. For very small sample sizes, the sampling distribution of the sample mean will look more like the population distribution than a normal distribution. A very rough rule of thumb: The sampling distribution is approximately normal for sample sizes greater than 30. Here the sample size is 4, so we expect the sampling distribution of the sample mean to be right-skewed, though not to the extent that the population distribution is. [Follow-up question: How is this different from Jamie asking 25 of her friends to each randomly select 4 houses that are for sale in the U.S. and to average their 4 prices together? Answer: By taking a sample of 100 averages-of-four-prices, Jamie is attempting to approximate the distribution of all averages-of-four-prices, i.e., the population distribution of all averages-of-four-prices. By taking 100 averages rather than 25, the distribution she obtains will better approximate the distribution of all averages-of-four-prices. (Perhaps Jamie constructs histograms to visualize these distributions.) This principle was treated in the two questions prior to the previous one.]

by David A. Huckaby

STT.05.02.040

AS DH 3321 010 F16: **55**/32/13 time 2:00
AS DH 1342 010 F17: **22**/78/0 time 2:00
AS DH 1342 020 F18: **50**/50/0 time 2:40
AS DH 1342 040 S19: **7**/93/0 time 2:00
AS DH 1342 030 F19: **40**/60/0 time 2:40
AS DH 1342 030 S20: **38**/62/0 time 2:30

5. Jamie asks 25 of her friends to each randomly select 50 houses that are for sale in the U.S. and average their 50 prices together. The distribution of these 25 averages is probably

   (a) significantly right-skewed

   (b) approximately bell-shaped

   (c) significantly left-skewed

   Answer: (b). House prices have a right-skewed distribution. According to the Central Limit Theorem, the sampling distribution of the sample mean will more closely approximate a normal distribution as the sample size increases. For large sample sizes, the sampling distribution of the sample mean will approximate a normal curve. A very rough rule of thumb: The sampling distribution is approximately normal for sample sizes greater than 30. Here the sample size is 50, so we expect the sampling distribution of the sample mean to be approximately normal. [Follow-up question: But Jamie asks only 25 friends! Why does this fact not play a major role in the answer to the question?]

by David A. Huckaby

STT.05.02.050

AS DH 3321 010 F16: 13/**87**/0 time 1:00
AS DH 1342 010 F17: 13/**84**/3 time 1:30
AS DH 1342 020 F18: 3/**97**/0 time 1:40
AS DH 1342 040 S19: 27/**73**/0 time 1:30
AS DH 1342 030 F19: 0/**100**/0 time 1:20
AS DH 1342 030 S20: 0/**100**/0 time 1:10

6. Jamie asks 100 of her friends to each randomly select 50 houses that are for sale in the U.S. and average their 50 prices together. The distribution of these 100 averages is probably

   (a) significantly right-skewed

   (b) approximately bell-shaped

   (c) significantly left-skewed

   Answer: (b). House prices have a right-skewed distribution. According to the Central Limit Theorem, the sampling distribution of the sample mean will more closely approximate a normal distribution as the sample size increases. For large sample sizes, the sampling distribution of the sample mean will approximate a normal curve. A very rough rule of thumb: The sampling distribution is approximately normal for sample sizes greater than 30. Here the sample size is 50, so we expect the sampling distribution of the sample mean to be approximately normal. [Follow-up question: How is this different from Jamie asking 25 of her friends to each randomly select 50 houses that are for sale in the U.S. and to average their 50 prices together? Answer: A similar follow-up question was suggested two questions previous and four questions previous.]

   by David A. Huckaby

   STT.05.02.060

   AS DH 3321 010 F16: 28/**72**/0 time 2:10
   AS DH 1342 010 F17: 33/**44**/22 time 2:00

7. Why is there a μ in the symbol $\mu_{\bar{x}}$, which is used to denote the mean of the sampling distribution of the sample mean?

   (a) Strictly speaking, the correct symbol is $\bar{x}_{\bar{x}}$, but $\mu_{\bar{x}}$ is used for simplicity.

   (b) The μ refers to a parameter of the original population.

   (c) The distribution whose mean is being taken consists of all sample means.

Answer: (c). The μ confuses some students who think, "Aren't we dealing with samples?" This question can help clarify both what the sampling distribution of the sample mean is and what it means to take its mean.

by David A. Huckaby

STT.05.02.070

AS DH 3321 010 F16: 0/24/**76** time 2:20

8. A physical therapy class has 10 students. The lightest student weights 110 pounds, the heaviest student weighs 240 pounds, the median weight of the 10 students is 140 pounds, and the mean weight of the 10 students is 160 pounds. Every student in the class pairs up with another student. In each pair, the two students find the mean of their two weights and then enter the mean into a spreadsheet on a computer in the classroom. Then the students pair off with different partners and again find the mean weight of their pair and type the mean into the spreadsheet. The students keep doing this until each student has been partnered with every other student. What is the most precise thing that can be said about the mean of all the numbers that were typed into the spreadsheet?

   (a) It is between 110 pounds and 240 pounds.

   (b) It is between 110 pounds and 240 pounds, but is most likely between 140 pounds and 160 pounds.

   (c) It is 160 pounds.

   Answer: (c). This question drives home the fact that $\mu_{\bar{x}} = \mu$. [Follow-up question: How many numbers were typed into the spreadsheet? Another follow-up question: Is it true that $\mu_{\bar{x}} = \mu$ even if the distribution is skewed? Answer: Yes. In fact, note that the distribution in this problem appears to have some skew. (Why?)]

   by David A. Huckaby

   STT.05.02.080

   AS DH 3321 010 F16: 19/23/**58** time 2:30
   AS DH 1342 010 F17: 0/16/**84** time 3:30
   AS DH 1342 020 F18: 10/27/**63** time 3:20
   AS DH 1342 040 S19: 0/73/**27** time 4:00
   AS DH 1342 030 F19: 0/0/**100** time 3:50
   CC KC MA315 S20: 21/45/**34**
   AS DH 1342 030 S20: 8/31/**62** time 4:00

9. A physical therapy class has 10 students. The lightest student weights 110 pounds, the heaviest student weighs 240 pounds, the median weight of the 10 students is 140

pounds, and the mean weight of the 10 students is 160 pounds. Every student in the class pairs up with another student. In each pair, the two students find the mean of their two weights and then enter the mean into a spreadsheet on a computer in the classroom. Then the students pair off with different partners and again find the mean weight of their pair and type the mean into the spreadsheet. The students keep doing this until each student has been partnered with every other student. The standard deviation of all the numbers that were typed into the spreadsheet is calculated. Then the entire experiment is repeated, except this time the students get into groups of five, taking the mean of all five weights, and keep doing this until every possible group of five students has recorded its mean weight. How will the standard deviation of all of the 5-student means compare to the standard deviation of all the 2-student means?

(a) It will be smaller.

(b) It will be the same.

(c) It will be larger.

(d) We would need to know the weights of all 10 students to answer this.

Answer: (a). The formula is $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$. Intuitively, a larger sample is expected to provide a closer approximation to μ.

by David A. Huckaby

STT.05.02.090

AS DH 3321 010 F16: **88**/9/3/0 time 1:50
AS DH 1342 010 F17: **100**/0/0/0 time 3:00
AS DH 1342 020 F18: **97**/0/0/3 time 2:00
AS DH 1342 040 S19: **87**/0/13/0 time 3:20
AS DH 1342 030 F19: **100**/0/0/0 time 2:30
CC KC MA315 S20: **71**/6/13/10
AS DH 1342 030 S20: **92**/0/8/0 time 3:00

10. Your statistics professor says to you, "If you can guess a certain quantity within 7 points of its true value, I will give you some extra credit." Which quantity would you prefer to guess?

(a) the score on next week's exam of a randomly selected student

(b) the mean of the scores of all the students in the class on next week's exam

Answer: (b). [Follow-up question: Beyond intuition—which is great—can you provide a more rigorous justification for the answer to this question? Answer: The fact that the standard deviation of the sampling distribution of the sample mean decreases with increasing sample size. What are the two sample sizes involved?]

by David A. Huckaby

STT.05.02.100

AS DH 3321 010 F16: 20/**80** time 1:50
CC KC MA315 S20: 4/**96**

11. The finishing times in a certain race are normally distributed with a mean of 25 minutes and a standard deviation of 4 minutes. What percentage of the samples of 4 finishers have means less than 21 minutes?

    (a) Less than 1%

    (b) 2.5%

    (c) 16%

    (d) None of the above is even close.

    Answer: (b). The standard deviation of the sampling distribution of the sample mean for samples of size $n = 4$ is $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{4}{\sqrt{4}} = 2$. So 21 is two standard deviations below the mean. The empirical rule tells us that roughly 95% of the observations lie within two standard deviations of the mean. Hence, 2.5% of the observations lie more than two standard deviations below the mean. Answer (c) will be chosen by students who simply use $\sigma = 4$ as the pertinent standard deviation. [Follow-up questions: When we said that 21 is two standard deviations below the mean, shouldn't we be using the mean of $\bar{x}$ and not the mean of x? Answer: Remember that $\mu_{\bar{x}} = \mu$. Were we justified in using the empirical rule in this question? Answer: Yes. Since the population distribution x is normal, so is the sampling distribution of the sample mean $\bar{x}$. If it were not stated in the question that the finishing times were normally distributed, would we be justified in using the empirical rule? Answer: No. With such a small sample size ($n = 4$), the sampling distribution of the sample mean might be far from normal if the original population is far from normal.]

    by David A. Huckaby

    STT.05.02.110
    AS DH 3321 010 F16: 5/**45**/40/10 time 3:10
    CC KC MA315 S20: 0/**86**/10/3