Classroom Voting Questions: Statistics

Inference for a Single Proportion

1. To estimate the proportion of students at a university who watch reality TV shows, a random sample of 50 students was selected and resulted in a sample proportion of .3. A 95% confidence interval for the proportion that watches reality TV would be ______ a 90% confidence interval.

   (a) narrower than

   (b) the same width as

   (c) wider than

   Answer: (c).

   by Roxy Peck for the textbooks: Roxy Peck and Jay Devore, Statistics: The Exploration and Analysis of Data, 6th Edition, Brooks/Cole Cengage Learning 2008 and Roxy Peck, Chris Olsen and Jay Devore, Introduction to Statistics and Data Analysis, 3rd Edition, Brooks/Cole Cengage Learning 2008.

   STT.08.01.010

   CC HZ MA207 F09: 31/0/**69**
   CC KC MA207 F09: 37/10/**53** time 2:30
   AS DH MA3321 Su12: 0/0/**100** time 1:20
   CC KC MA207 F15: 0/0/**100** time 1:00
   CC KC MA207 F18: 14/7/**79**

2. Suppose we wish to estimate the percentage of students who smoke marijuana at each of several liberal arts colleges. Two such colleges are StonyCreek (enrollment 5,000) and Whimsy (enrollment 13,000). The Dean of each college decides to take a random sample of 10% of the entire student population. The margin of error for a simple random sample of 10% of the population of students at each school will be

   (a) smaller for Whimsy than for StonyCreek.

   (b) smaller for StonyCreek than for Whimsy.

   (c) the same for each school.

   (d) insufficient information

Answer: (a). Note: Margin of error is another term for standard error.

(A)* correct The margin of error (calculated by $\sqrt{p * q/n}$) primarily depends on the sample size (n) because a large sample size gives more information, which leads to less uncertainty about the estimation (smaller variability).

(B) Students probably dont understand that the primary factor influencing the magnitude of the margin of error is the sample size.

(C) The sample sizes are different for each school so the standard errors are different.

(D) It is sufficient to know the sample sizes.

Two topics came up for discussion in the field-testing:

1. How extreme can p be (relative to the size of the sample) before it has an effect on the standard error?

2. How large does the population size have to be for it to be large enough for sampling theory to apply?

by Murphy, McKnight, Richman, and Terry

STT.08.01.020

CC HZ MA207 F09: **61**/31/8/0
CC KC MA207 F09: **48**/24/16/12 time 2:30
AS DH MA3321 Su12: **92**/8/0/0 time 1:50
CC KC MA207 F15: **40**/20/0/40 time 2:30
CC KC MA207 F18: **86**/7/7/0

3. Suppose we wish to estimate the percentage of people who speed while driving in a college town. We choose to sample the populations of Austin, TX (University of Texas) and Norman, OK (University of Oklahoma). We know that both cities have populations over 100,000 and that Austin is approximately 5 times bigger (in population) than Norman. We also expect the rates of speeding to be about the same in each city. Suppose we were to take a random sample of 1000 drivers from each city. The margin of error for a simple random sample of the population of drivers from each city will be

   (a) smaller for the Austin sample than the Norman sample.
   (b) smaller for the Norman sample than the Austin sample.
   (c) the same for both samples.
   (d) not possible to determine without more precise information about the population sizes.

Answer: (c). (A), (B) Margins of error are dependent upon samples and not populations.

(C)* correct Margin of error depends on sample size, which is the same for both cities.

(D) Information about population sizes is not relevant because margins of error are dependent on samples.

by Murphy, McKnight, Richman, and Terry

STT.08.01.030

CC HZ MA207 F09: 15/8/**77**/0
AS DH MA3321 Su12: 8/0/**83**/8 time 2:10
CC KC MA207 F15: 30/0/**70**/0 time 3:30

4. A parachute manufacturer is concerned that the failure rate of 0.1% advertised by his company may in fact be higher. What is the null hypothesis for the test he would run to address his worries.

   (a) $H_0 : \mu = 0.001$
   (b) $H_0 : p > 0.001$
   (c) $H_0 : \mu < 0.001$
   (d) $H_0 : p = 0.001$

   Answer: (d).

   by Jack Oberweiser

   STT.08.01.040

   CC HZ MA207 F09: 8/0/0/**92** time 1:20
   AS DH MA3321 Su12: 17/0/0/**83** time 1:10
   AS DH 1333 010 S13: 6/56/0/**38** time 1:20
   AS DH 1333 020 S14: 31/7/0/**62** time 1:30 ,
   AS DH 3321 010 S14: 50/0/7/**43** time 1:20 ,
   AS DH 1333 010 F14: 76/0/0/**24** time 2:10 ,
   AS DH 3321 010 F14: 0/0/0/**100** time 1:30 ,
   AS DH 1333 020 S15: 60/20/10/**10** time 1:30 ,
   AS DH 1333 020 F15: 95/0/5/**0** time 1:30 ,
   AS DH 3321 010 F15: 70/10/0/**20** time 1:40 ,
   CC KC MA207 F15: 11/0/0/**89** time 2:00
   AS DH 1342 010 F17: 0/4/0/**96** time 1:00
   AS DH 1342 020 F18: 0/3/0/**97** time 1:30
   AS DH 1342 040 S19: 0/0/0/**100** time 1:30
   AS DH 1342 030 F19: 0/0/0/**100** time 1:30

5. A parachute manufacturer is concerned that the failure rate of 0.1% advertised by his company may in fact be higher. A hypothesis test was run and the result was a P-value of 0.03333. The most likely conclusion the manufacturer might make is:

(a) My parachutes are safer than I claim.

(b) My parachutes are not as safe as I claim them to be.

(c) I can make no assumption of safety based on a statistical test.

(d) The probability of a parachute failure is 0.03333.

(e) Both (b) and (d) are true.

Answer: (b).

by Jack Oberweiser

STT.08.01.050

CC HZ MA207 F09: 8/**0**/15/0/77 time 1:20
AS DH MA3321 Su12: 9/**18**/27/9/36 time 3:30
AS DH 3321 010 F14: 4/**19**/0/12/65 time 3:10 ,
AS DH 3321 010 F15: 3/**7**/7/13/70 time 2:30 ,
CC KC MA207 F15: 0/**90**/10/0/0 time 2:00
CC KC MA207 F18: 20/**45**/15/0/20

6. To explain the meaning of a P-value of 0.033, you could say:

(a) There is approximately a 96.7% chance of obtaining my sample results.

(b) Assuming the null hypothesis is accurate, results like those found in my sample should occur only 3.3% of the time.

(c) We can't say anything for sure without knowing the sample results.

(d) There is approximately a 3.3% chance of obtaining my sample results.

Answer: (b).

by Jack Oberweiser

STT.08.01.060

CC HZ MA207 F09: 39/**39**/15/8 time 1:00
AS DH MA3321 Su12: 14/**57**/0/29 time 2:00

7. Suppose we have the results of a Gallup survey (simple random sampling) which asks participants for their opinions regarding their attitudes toward technology. Based on 1500 interviews, the Gallup report makes confidence statements about its conclusions. If 64% of those interviewed favored modern technology, we can be 95% confident that the percent of those who favored modern technology is

(a) 95% of 64%, or 60.8%

(b) 95% $\pm$ 3%

(c) 64%

(d) 64% ± 3%

Answer: (d). (A) Students are attending to the surface features of the problem, doing calculations with the numbers that are given.

(B) 95% is the confidence level not the point estimate for the population parameter.

(C) 64% is the point estimate, but this answer does not contain a margin of error (and thus does not give an interval estimate).

(D)* correct This answer has both the correct point estimate (64%) and the interval estimate (based on a margin of error). Note that 64% is the center of this interval, which is a feature that students should recognize about confidence intervals.

by Murphy, McKnight, Richman, and Terry

STT.08.01.070

AS DH MA3321 Su12: 8/0/8/**83** time 1:20

8. A confidence interval for a proportion is constructed using a sample proportion of 0.5. If the sample proportion was 0.9 instead of 0.5, what would happen to the width of the resulting confidence interval?

   (a) The new CI would be narrower.

   (b) The new CI would have the same width.

   (c) The new CI would be wider.

   Answer: (a). Verify this computationally, computing $1.96\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}}$ for the actual sample proportion, 0.5, and a hypothetical sample proportion of 0.9. Why? If the sample proportion is fairly high, then the population proportion is likely high, as well. That means that the binomial random variable X, the number of students in the sample that favor Obama, will have a relatively small standard deviation. Verify this using the population standard deviation, $\sqrt{np(1-p)}$. This means that the sample proportion will have less variability, which means the confidence interval based on this sample proportion will have more precision.

   by Derek Bruff

   STT.08.01.080

   AS DH MA3321 Su12: **75**/17/8 time 3:00

9. A sample needs to taken to answer the question Have you ever shoplifted? Assuming a random sample can be found, how many people would need to be polled to insure a margin of error of no more than 3% with 90% confidence?

(a) 1068

(b) 752

(c) 23

(d) None of the above

Answer: (b). Assuming the sample proportion is 0.5, the required sample size is $\frac{1}{4}(\frac{z^*}{m})^2 = \frac{1}{4}(\frac{1.645}{0.03})^2 = 751.7$. Rounding up gives us a required sample size of 752.

by Jack Oberweiser

STT.08.01.090

AS DH MA3321 Su12: 0/**92**/0/8 time 5:00
AS DH 3321 010 S14: 5/**75**/5/15 time 4:30 ,
AS DH 3321 010 F14: 5/**90**/0/5 time 3:40 ,

10. A sample needs to be taken to answer the question, "Have you jaywalked in the past month?" Assume that based on past studies we can be almost certain that the actual percentage of the population that jaywalks in any given month is between 70% and 90%. Assuming a random sample can be obtained, how many people would need to be polled to insure a margin of error of no more than 3% with 95% confidence?

   (a) 1068

   (b) 897

   (c) 385

   (d) None of the above

   Answer: (b). The formula is $n = p_g(1 - p_g)\left(\frac{z_\alpha}{E}\right)^2$, where p_g is chosen to maximize $p_g(1-p_g)$ on the interval of possible values of p. (The idea is that we want the smallest possible sample size that will still yield the desired margin of error and confidence level.) The graph of $p_g(1 - p_g)$ is a downward-opening parabola with vertex at (0.5, 0.25). If nothing is known in advance about p, then p_g is chosen to be 0.5, as this value yields the absolute maximum of $p_g(1-p_g)$, namely 0.25. In the present problem, the possible values of p lie on the interval $[0.7, 0.9]$. The number in this interval that maximizes $p_g(1 - p_g)$ is clearly 0.7, the number in the interval closest to 0.5. So we have $0.7(1-0.7)\left(\frac{1.96}{0.03}\right)^2 = 896.4$. Rounding up gives us the required sample size of 897.

   (a) This answer is obtained by assuming we have no advance knowledge of p: $0.5(1-0.5)\left(\frac{1.96}{0.03}\right)^2 = 0.25\left(\frac{1.96}{0.03}\right)^2 = 1067.1$. Rounding up gives 1068.

   (b) Correct. (See above.)

   (c) This answer is obtained by taking $p_g = 0.9$: $0.9(1-0.9)\left(\frac{1.96}{0.03}\right)^2 = 384.2$. Rounding up gives 385.

   (d) Incorrect. (b) is the correct answer. (See above.)

by David A. Huckaby (patterned on STT.08.01.090)

STT.08.01.095
AS DH 3321 010 F16: 13/**88**/0/0 time 3:40

11. A sample needs to be taken to answer the question, "Have you voted in a local election in the past 10 years?" Assume that based on past studies we can be almost certain that the actual population percentage is between 45% and 75%. Assuming a random sample can be obtained, how many people would need to be polled to insure a margin of error of no more than 3% with 99% confidence?

    (a) 1844
    (b) 1825
    (c) 1383
    (d) None of the above

    Answer: (a). The formula is $n = p_g(1-p_g)\left(\frac{z_\alpha}{E}\right)^2$, where p_g is chosen to maximize $p_g(1-p_g)$ on the interval of possible values of p. (The idea is that we want the smallest possible sample size that will still yield the desired margin of error and confidence level.) The graph of $p_g(1-p_g)$ is a downward-opening parabola with vertex at (0.5, 0.25). If nothing is known in advance about p, then p_g is chosen to be 0.5, as this value yields the absolute maximum of $p_g(1-p_g)$, namely 0.25. In the present problem, the possible values of p lie on the interval $[0.45, 0.75]$. The number in this interval that maximizes $p_g(1-p_g)$ is clearly 0.5, since it gives the absolute maximum. So we have $0.5(1-0.5)\left(\frac{2.576}{0.03}\right)^2 = 1843.3$. Rounding up gives us a required sample size of 1844.

    (a) Correct. (See above.)
    (b) This answer is obtained by taking $p_g = 0.45$: $0.45(1-0.45)\left(\frac{2.576}{0.03}\right)^2 = 1824.8$. Rounding up gives 1825.
    (c) This answer is obtained by taking $p_g = 0.75$: $0.75(1-0.75)\left(\frac{2.576}{0.03}\right)^2 = 1382.5$. Rounding up gives 1383.
    (d) Incorrect. (a) is the correct answer. (See above.)

    by David A. Huckaby

    STT.08.01.096 (patterned on STT.08.01.090)
    AS DH 3321 010 F16: **50**/44/6/0 time 3:10

12. The margin of error is computed for a poll with a sample size of 50. Approximately what sample size would you need if you wanted to cut the margin of error in half?

    (a) 25

(b) 100
(c) 200
(d) 400

Answer: (c). This can be seen from the formula for margin of error. Thanks to the square root in the formula, halving the margin of error requires quadrupling the sample size.

by Derek Bruff

STT.08.01.100

AS DH MA3321 Su12: 0/25/**75**/0 time 3:00

13. Which of the following does *not* result in a larger margin of error?

    (a) Increasing the confidence level
    (b) Decreasing the sample size
    (c) Having a larger population size

    Answer: (c). Choices (a) and (b) can be explained using the formula for the margin of error. While one might think that the larger the population, the greater our margin of error (and conversely, the smaller the population, the lesser our margin of error), margin of error doesn't depend on the population size. As long as the population size is sufficiently greater than the sample size (so that the binomial assumption we made way back when holds), it doesn't matter what the population size is. This means that we can take a random sample of 1200 people from the much, much larger population of all registered voters, and still have a margin of error of 2.83%, even though our sample size is much smaller than our population size.

    by Derek Bruff

    STT.08.01.110

    AS DH MA3321 Su12: 0/9/**92** time 1:00
    AS DH MA1333 010 F12: 8/15/**77** time 2:00
    AS DH MA1333 020 F12: 29/21/**50** time 2:50
    AS DH 1333 010 S13: 17/6/**78** time 3:50
    AS DH 1333 020 S14: 17/39/**44** time 3:00 ,
    AS DH 3321 010 S15: 4/16/**80** time 2:00 ,
    AS DH 1333 020 F15: 0/0/**100** time 2:30 ,
    AS DH 3321 010 F15: 3/3/**93** time 2:40 ,
    CC KC MA207 F15: 0/0/**100** time 1:00
    AS DH 1342 040 S19: 14/0/**86** time 2:30

14. Terrance hypothesizes that less than 20% of the population has a favorable view of the latest flavor-of-the-month boy band. He conducts a hypothesis test and gets $P = 0.0001$. Which of the following might be a reasonable summary of the results?

    (a) Accept the null hypothesis. The data provide sufficient evidence to conclude that less than 20% of the population has a favorable view of the band ($P = 0.0001$).

    (b) Accept the null hypothesis. The data do not provide sufficient evidence to conclude that less than 20% of the population has a favorable view of the band ($P = 0.0001$).

    (c) Reject the null hypothesis. The data provide sufficient evidence to conclude that less than 20% of the population has a favorable view of the band ($P = 0.0001$).

    (d) Reject the null hypothesis. The data do not provide sufficient evidence to conclude that less than 20% of the population has a favorable view of the band ($P = 0.0001$).

    Answer: (c).

    (a) One of the first steps of a hypothesis test is to assume that the null hypothesis is true. Therefore, "accepting" the null hypothesis cannot be a result of the test. (For more commentary, see, for example, questions STT.07.01.070 and STT.07.01.080 in this collection.)

    (b) See part (a).

    (c) Correct. The very small P-value indicates statistically significant evidence for Terrance's hypothesis, the alternative hypothesis ($p < 0.20$, or equivalently, $\pi < 0.20$). (The null hypothesis, $p = 0.20$, or equivalently, $\pi = 0.20$, is rejected.)

    (d) See part (c). The data *do* provide statistically significant evidence for the alternative hypothesis.

    by David A. Huckaby

    STT.08.01.120

    AS DH 3321 010 F16: 0/0/**97**/3 time 2:30