DALITE Q4 - Normal Curve Calculations and Confidence Intervals. Solutions.

EPIB607 - Inferential Statistics[a]

[a]Fall 2018, McGill University

This version was compiled on September 27, 2018

This DALITE quiz will cover the normal curve calculations and confidence intervals.

Normal calculations | Confidence intervals | Central Limit Theorem (CLT)

1. Normal Calculations

Cholesterol levels among fourteen-year-old boys are roughly Normal with mean 170 and standard deviation 30 milligrams per deciliter (mg/dl). In a SRS of 4 fourteen-year-old boys, the probability that the average cholesterol level is 200 mg/dl or more is close to (simply provide the corresponding `R` code used to answer this question in your rationale)

a. **0.023 (Correct)**
b. 0.159
c. 0.977

1.1. Correct rationales.

- `stats::pnorm(q=200, mean=170, sd=(30/sqrt(4)), lower.tail = FALSE)`

1.2. Incorrect rationales.

- `stats::pnorm(q= 200, mean=170, sd= 30)`
- `stats::pnorm(q=200, mean = 170, sd = 30, lower.tail = FALSE)`
- `1-mosaic::xpnorm(q=200, mean = 170, sd=30)`

2. Normal calculations 2

Suppose that the distribution of heights of all male students on your campus is Normal with mean 70 inches and standard deviation 2.8 inches. How large a simple random sample (SRS) do you need to reduce the standard deviation of the mean to 0.5?

a. 31.36
b. **32 (Correct)**
c. 6
d. 12

2.1. Correct rationales.

- If we want to reduce the standard deviation of the sample mean to 0.5 inches, then we must choose n to satisfy $2.8/\sqrt{n} = 0.5$. Solving for n gives $n = (2.8/0.5)^2 = 31.36$. So we need 32 people (cant have 31.36 people).
- Standard deviation of the mean is equal to standard deviation of the population over the square root of n. $0.5 = 2.8/\sqrt{n}$. Solve for n. $n = 31.36$ Round up to nearest whole person to get standard deviation of the mean equal to 0.5.
- We cannot take $n = 31$ because then our standard deviation of the mean would not be at least 0.5

2.2. Incorrect rationales.

- $2.8/\sqrt{n} = 0.5 \rightarrow n = (2.8/0.5)^2 \rightarrow n = 31.36$

3. Confidence Interval 1

A study reports the mean change in HDL of adults eating raw garlic six days a week for six months. The margin of error for a 95% confidence interval is given as plus or minus 6 milligrams per deciliter of blood (mg/dl). This means that

a. we can be certain that the study result is within 6 mg/dl of the truth about the population
b. we could be certain that the study result is within 6 mg/dl of the truth about the population if the conditions for inferences were satisfied
c. **the study used a method that gives a result within 6 mg/dl of the truth about the population in 95% of all samples (Correct)**

3.1. Correct rationales.

- Confidence interval comments about the method of obtaining the result, not the result itself. A 95% confidence interval implies that we can be certain that if this method was repeated, 95% of the time it will be within 6 mg/dl of the population mean.
- You cannot determine whether the true mean was within or outside the CI. We do not know the true mean. The confidence interval's purpose is to provide a certain level of confidence on the method.
- The confidence interval of 95% means that the test will give accurate results (within 6 milligrams per decilitre of blood) 95% of the time
- The confidence level states the probability that the method will give a correct result. Since the confidence level is 95%, we can only be certain that 95% of the time the method will correctly capture the true mean.

3.2. Incorrect rationales.

- If conditions were satisfied(all statistics such as mean, sd, and n were calculated), the result interprets that 95% of the data will fall within the CI range.
- A 95% confidence interval indicates that 95% of observations will fall within the given margin
- As per the definition of confidence interval
- 1.96 x sigma/square root of n
- 95% of the population distribution is contained in the confidence interval

4. Confidence Intervals 2

A laboratory scale is known to have a standard deviation of $\sigma = 0.001$ gram in repeated weighings. Scale readings in repeated weighings are Normally distributed, with mean equal to the true weight of the specimen. Three weighings of a specimen on this scale give 3.412, 3.416 and 3.414 grams. Answer both questions below:

i) A 95 % confidence interval for the true weight is
ii) The margin of error for a 99% confidence interval would be

a. i) 3.414 +/- 0.00113
   ii) smaller

b. i) 3.414 +/- 0.00113
   ii) about the same

c. **i) 3.414 +/- 0.00113 ii) larger (Correct)**

d. i) 3.414 +/- 0.00065
   ii) larger

e. i) 3.414 +/- 0.00196
   ii) larger

4.1. Correct rationales.

- i) $3.414 \pm 1.96 \times 0.001/\sqrt{3} = 3.414 \pm 0.00113$ ii) larger because 99% corresponds to a z value of 2.58
- The answer is C because the true weight for a 95% confidence interval type is 3.414 ± 0.00113, the margin of error for a 99% confidence interval would be larger because for me to be right 99% of the time the interval will have to be a lot larger than if i were to be right 95% of the time

4.2. Incorrect rationales.

- ii) About the same since we are still only using n=3
- ii) Would be larger, as including more people

5. Confidence Intervals 3

You calculate a 95% confidence interval of 27 ± 2 centimeters (cm) for the mean needle length of Torrey pine trees. You ask a friend to explain this result. He believes it means that "95% of all Torrey pine needles have lengths between 25 and 29 cm." Is he right? or wrong? Explain your answer in the rationale.

a. He is right
b. **He is wrong (Correct)**

5.1. Correct rationales.

- We can't know that. the mean is either 27 ± 2cm or it's not. What the 95% confidence interval means is that 95% of the time, 27 ± 2cm will contain the true value of the mean needle length
- The 95% confidence interval suggests that we are 95% certain that the true mean of the population is between those two numbers - not that 95% of the individual values will fall between these numbers.
- It means that this confidence interval has a 95% chance to capture the population mean needle length of Torrey pine trees. It cannot indicate the population distribution.

5.2. Incorrect rationales.

- The parameter in this case is the mean needle length of all Torrey pine trees. By the definition of a CI, this has a probability of 0.95 of being within the interval. So he is right.
- Because if they repeatedly took samples we would see that 95% of them would contain the population mean of 27.
- 95% of samples trees will have lengths between 25 and 29 cm
- He is wrong because a number that lies within one standard deviation of the mean which would be 25 and 29 is a 68% confidence interval while numbers lying 2 standard deviations from the mean is a 95% confidence interval (meaning a range of 23 to 31)
- 25cm-29cm would be the 68% confidence interval
- Assuming the inferences are met (normally distributed, etc.), we have said with a 95% confidence interval (i.e. using a method that is right 95% of the time) that 95% of all lengths of Torrey pine needles are between 25 and 29cm in length.

6. Confidence Intervals 4

A New York Times poll on women's issues – which interviewed 1025 women randomly selected from the United States excluding Alaska and Hawaii– in which 47% of the women said they do not get enough time for themselves; the poll reported a margin of error of ± 3 percentage points for 95% confidence in the conclusions.

Which of the following statements best explains what "95% confidence" means.

a. This poll is accurate 19 times out of 20. (NO. This poll is either accurate or its not)
b. 95% chance that the info is correct for between 44 and 50% of women. (NO. 95% confidence in the procedure that produced the interval 44-50)
c. In 95 of 100 comparable polls, expect 44 - 50% of women will give the same answer. (NO. Same answer? as what?)
d. If this same poll were repeated many times, then 95 of every 100 such polls would give a range that included 47%. (NO. Estimate will be between $\mu -$ margin and $\mu +$ margin in 95% of applications.)
e. It means that 47% give or take 3% is an accurate estimate of the population mean 19 times out of 20 such samplings. (NO. 95% of applications of CI give correct answer. How can the same interval 47% ± 3 be accurate in 19 but not in the other 1?)

6.1. Correct rationales.

6.2. Incorrect rationales.

- 95% confidence translates to saying this poll is accurate 95% of the time
- There is 95% probability that the calculated confidence interval will include the mean (47%).
- 95 out of 100 times, 47% will be in this range–this is what the confidence interval states.
- the answer is D because "95% confidence" means that you are 95% confident that the value you have will be right
- 95% of the samples will contain the true parameter value (which in this case is 47%).
- Definition of confidence interval is that with 95% confidence, our true mean lies within the margin of error, and 95% of the time, our sample mean will also lie within that margin of error.
- the answer is D because "95% confidence" means that you are 95% confident that the value you have will be right
- 95% confidence indicates confidence level- reflects that method will give these same results 95% of the time
- when you do a poll of this size 100 times, 95% will contain the value 47% +/- 3% points
- While we can not be sure whether the poll is accurate at reflecting the true mean, we can be 95% sure that our methodology if we repeat the poll many times, 95 of 100 polls would include a range that include 47%.
- If we were to repeat the same sample many times, we are using the same methodology, 95 of every 100 polls would give a mean of 47% +/- 3% which includes 47 by default. Given the margin of error of +/- 3%, any number found between the minimum 44% and maximum 50% would also include 47% in their range by default.
- E explains the best because 95% of the times the mean would be between 47 ± 3