Like this course? Become an expert by joining the Methods and Statistics in Social Sciences Specialization (/specializations/social-science).

Upgrade

(https://accounts.coursera.org/i/zendesk/courserahelp?return_to=https://learner.coursera.help/hc)

Course Forums

All Course Discussions (/learn/inferential-statistics/discussions?sort=lastActivityAtDesc&page=1)

General Discussion (/learn/inferential-statistics/forum/LzonxtzlEeWd6hLHI8FYwQ/discussions?sort=lastActivityAtDesc&page=1)

Meet and Greet (/learn/inferential-statistics/forum/LzgFGtzlEeW81Q4r-1eYEQ/discussions?sort=lastActivityAtDesc&page=1)

Module Forums

Before we get started... (/learn/inferential-statistics/module/vL4vb/discussions?sort=lastActivityAtDesc&page=1)

Comparing two groups (/learn/inferential-statistics/module/7YoGo/discussions?sort=lastActivityAtDesc&page=1)

Categorical association (/learn/inferential-statistics/module/wnZZK/discussions?sort=lastActivityAtDesc&page=1)

Simple regression (/learn/inferential-statistics/module/pT4zG/discussions?sort=lastActivityAtDesc&page=1)

Multiple regression (/learn/inferential-statistics/module/uRWxz/discussions?sort=lastActivityAtDesc&page=1)

Analysis of variance (/learn/inferential-statistics/module/rLqez/discussions?sort=lastActivityAtDesc&page=1)

Non-parametric tests (/learn/inferential-statistics/module/sHH98/discussions?sort=lastActivityAtDesc&page=1)

Exam time! (/learn/inferential-statistics/module/jyCOE/discussions?sort=lastActivityAtDesc&page=1)

All Course Discussions

(/learn/inferential-statistics/discussions?sort=lastActivityAtDesc&page=1&q=)

DW (/learn/inferential-statistics/profiles/def8b145896c8d3cbbe538e80e9b18ef)

Quiz 3, question 14 mistake?

David Whiting Multiple regression (/learn/inferential-statistics/module/uRWxz/discussions) · 2 days ago (/learn/inferential-statistics/discussions/rtJyteu6EeWfwAohgaM63Q)

The wikipedia section on interpretation of a confidence interval seems to contradict the answer of question 14, quiz 3. In particular the section under misunderstandings on wikipedia. I don't want to be more specific as then I will be pretty much stating what the current answer is.

0 Upvote · Follow 3 · Reply to David Whiting

Earliest (/learn/inferential-statistics/discussions/rtJyteu6EeWfwAohgaM63Q?sort=createdAtAsc&page=1)

Top (/learn/inferential-statistics/discussions/rtJyteu6EeWfwAohgaM63Q?sort=upvotesDesc&page=1)

Most Recent (/learn/inferential-statistics/discussions/rtJyteu6EeWfwAohgaM63Q?sort=createdAtDesc&page=1)

ZB (/learn/inferential-statistics/profiles/3cdb3ac214a020ebfbe64f1e50b03eac)

Zara Bamdad Teaching Staff · a day ago (/learn/inferential-statistics/discussions/rtJyteu6EeWfwAohgaM63Q/replies/efS_0OwrEeWcTQpIg9bO1w)

Hi David,

Thanks for posting! I had a look at the wikipedia page, but I can't quite see what's wrong with the 'correct' answer on the quiz. Can you be slightly more specific with which of the misunderstandings reflect the 'correct' answer?

0 Upvote · Reply

DW (/learn/inferential-statistics/profiles/def8b145896c8d3cbbe538e80e9b18ef)

David Whiting · a day ago (/learn/inferential-statistics/discussions/rtJyteu6EeWfwAohgaM63Q/replies/WNfxF-xvEeWkUAr_DMJ5Lw)

Wikipedia say "A particular confidence interval of 95% calculated from an experiment does not mean that there is a 95% probability of a sample mean from a repeat of the experiment falling within this interval.".

I was am making the assumption that the confidence interval in question 14 was calculated based on a sample. Therefore you can't guarantee it will contain 95% of repeated sample means. Maybe I am misunderstanding something or maybe that assumption was wrong?

1 Upvote · Hide 7 Replies

DS (/learn/inferential-statistics/profiles/1f7f043688de1cdf42d40ee185779133)

david scott · a day ago (/learn/inferential-statistics/discussions/rtJyteu6EeWfwAohgaM63Q/replies/WNfxF-xvEeWkUAr_DMJ5Lw/comments/9jc7X-x9EeWfwAohgaM63Q)

I thought the language was a little loose on that question also. As I recall, the answer that auto-grades as correct is a far better interpretation than any of the alternatives though.

0 Upvote

david scott · a day ago (/learn/inferential-statistics/discussions/rtJyteu6EeWfwAohgaM63Q/replies/WNfxF-

DS

xvEeWkUAr_DMJ5Lw/comments/NqTJfOyGEeWB0QpuSDkq-Q)

(/learn/inferential-statistics/profiles/1f7f043688de1cdf42d40ee185779133)

I had time to write a quick routine, and David Whiting is more right about the impact of the language discrepancy than I hypothesized.

The code sets a seed, takes 50 draws from N(0,1), calculates the sample mean and 95% CI, and repeats 20,000 times.

For this seed, 4.9% of the 95% CI fail to contain the population mean. But, on average, 16.2% of the sample means fail to fall in any given 95% CI.

See below, and set a different seed and/or improve the code as needed:

```
set.seed(0317162028)
nRuns <- 20000
nPer <- 50
sampMeans <- rep(0,nRuns)
sampCI <- matrix(data=0,nrow=nRuns,ncol=2)
critT <- abs(qt(0.025, df=(nPer-1), lower.tail=TRUE))

for (intCtr in 1:nRuns) {
    myRands <- rnorm(nPer)
    sampMeans[intCtr] <- mean(myRands)
    mySE <- sd(myRands)/sqrt(nPer)
    sampCI[intCtr,] <- mean(myRands) + c(-1,1) * critT * mySE
}

## Percentage of CI that miss the true population mean of 0
sum(sampCI[,1]>0 | sampCI[,2]<0) / nRuns ## 0.0491

## Cumulative CI misses for each sampMean looking at each CI
nMiss <- 0
for (intCtr in 1:nRuns) {
    myVal <- sampMeans[intCtr]
    nMiss <- nMiss + sum(sampCI[,1]>myVal | sampCI[,2]<myVal)
}

## Percentage of misses for sampMean vs. CI (nRuns*nRuns checks)
nMiss / nRuns^2 ## 0.1617
```

▲ 0 Upvote

DW

David Whiting · 21 hours ago (/learn/inferential-statistics/discussions/rtJyteu6EeWfwAohgaM63Q/replies/WNfxF-xvEeWkUAr_DMJ5Lw/comments/Flxxk-yUEeWqSQ7kx0rpLQ) · Edited

(/learn/inferential-statistics/profiles/def8b145896c8d3cbbe538e80e9b18ef)

I thought answer "a" makes more sense but it depends how picky you are on the language. As I understand it, technically its not right to say that there is a 95% probability the calculated CI contains the population mean as the CI either does or it doesn't. If it does there is a 100% probability it contains it, if it doesn't there is a 0% chance it contains it. Personally I think this is a bit pedantic as you can say, out of all possible calculated CI, there is a 95% probability that the calculated CI contains the population mean. Therefore, you can be 95% "sure" (but technically not 95% probable) that the CI contains the population mean. Thats as I understand it anyway!

Your script is interesting. I find it difficult to have an intuitive feel for the how repeated sample means fall into confidence interval. I would be interested if their is any math the outlines the answer to the problem i.e. predicts the 16.17%

▲ 1 Upvote

DW

David Whiting · 21 hours ago (/learn/inferential-statistics/discussions/rtJyteu6EeWfwAohgaM63Q/replies/WNfxF-xvEeWkUAr_DMJ5Lw/comments/pdL44uyUEeWnOBISxifLdw)

(/learn/inferential-statistics/profiles/def8b145896c8d3cbbe538e80e9b18ef)

Also, David Scott, how do the percentages change as you increase nPer? Maybe it gets closer to 5%.

▲ 0 Upvote

DS

david scott · 20 hours ago (/learn/inferential-statistics/discussions/rtJyteu6EeWfwAohgaM63Q/replies/WNfxF-xvEeWkUAr_DMJ5Lw/comments/UMisp-ycEeWPKQpBKpg03Q)

(/learn/inferential-statistics/profiles/1f7f043688de1cdf42d40ee185779133)

I think this is a lucky random seed that fits well with theory. If you use mySE * sqrt(2) to calculate each CI (two draws from the same population) and re-run the code, then 95% of sample means fall in the average CI.

Attempt at theory (hand-waving to assume known population sigma=1):

- Draw 1 - mean m1, SE=1/sqrt(N), df=N-1
- Draw 2 - mean m2, SE=1/ sqrt(N), df=N-1
- Difference in means m1 - m2 with SE=sqrt(2)/sqrt(N), df = 2*N - 2
- CI if using N=50 on Draw 1 is m1 +/- 2.01 * 1/sqrt(50) = m1 +/- 0.284
- Actual SE for m1-m2 on N=50 is sqrt(2)/sqrt(50) = 0.20, df=98
- We can only miss by 0.284 / 0.20 = 1.42 SE and still have m1-m2 contain 0
- 2 * pt(1.42, df=98, lower.tail=FALSE) = 0.159

So, 15.9% of the time, the CI for m1-m2 (using the one-sample CI approach) will not contain zero, I believe the same as m2 will not fall in the m1 CI.

▲ 0 Upvote

DS

david scott · 20 hours ago (/learn/inferential-statistics/discussions/rtJyteu6EeWfwAohgaM63Q/replies/WNfxF-xvEeWkUAr_DMJ5Lw/comments/hI9OH-yeEeWlJhJpbKSYeQ)

(/learn/inferential-statistics/profiles/1f7f043688de1cdf42d40ee185779133)

Per your question on changing nPer:

- Original Run - seed 0317162028, nPer 50

```
## Using mySE <- sd(myRands)/sqrt(nPer)
sum(sampCI[,1]>0 | sampCI[,2]<0) / nRuns  ## 0.0491
nMiss / nRuns^2 ## 0.1617

## Using mySE <- sd(myRands)/sqrt(nPer) * sqrt(2)
sum(sampCI[,1]>0 | sampCI[,2]<0) / nRuns  ## 0.0061
nMiss / nRuns^2 ## 0.0497
```

- Additional Run - seed 0318160006, nPer 500

```
## Using mySE <- sd(myRands)/sqrt(nPer)
sum(sampCI[,1]>0 | sampCI[,2]<0) / nRuns  ## 0.0510
nMiss / nRuns^2 ## 0.1674

## Using mySE <- sd(myRands)/sqrt(nPer) * sqrt(2)
sum(sampCI[,1]>0 | sampCI[,2]<0) / nRuns  ## 0.0062
nMiss / nRuns^2 ## 0.0512
```

- Small-N Run at/below CLT threshold - seed 0318160010, nPer 10

```
## Using mySE <- sd(myRands)/sqrt(nPer)
sum(sampCI[,1]>0 | sampCI[,2]<0) / nRuns  ## 0.0546
nMiss / nRuns^2 ## 0.1482

## Using mySE <- sd(myRands)/sqrt(nPer) * sqrt(2)
sum(sampCI[,1]>0 | sampCI[,2]<0) / nRuns  ## 0.0113
nMiss / nRuns^2 ## 0.0524
```

▲ 0 Upvote

DW

(/learn/inferential-statistics/profiles/def8b145896c8d3cbbe538e80e9b18ef)

David Whiting · 19 hours ago (/learn/inferential-statistics/discussions/rtJyteu6EeWfwAohgaM63Q/replies/WNfxF-xvEeWkUAr_DMJ5Lw/comments/Rgmm6uyjEeWqSQ7kx0rpLQ)

Kudos to you David Scott. Took me a while to figure out your explanation but makes sense. You can do above using summation of variances which shows the 15.9% is a constant.

SE of sample mean = sd/sqrt(N)

sd of m1 - m2 = sqrt(sd^2 + sd^2) = sqrt(2)*sd

SE of m1 - m2 = sqrt(2)*sd/sqrt(N)

Taking the division of the SE gives you the sqrt(2) or 1.42 SE as you stated.

▲ 0 Upvote

SD

(/learn/inferential-statistics/profiles/f34069ce8df6de7dbebfedfb7e760d9f)

Reply

Reply

DS

(/learn/inferential-statistics/profiles/1f7f043688de1cdf42d40ee185779133)

david scott · 20 hours ago (/learn/inferential-statistics/discussions/rtJyteu6EeWfwAohgaM63Q/replies/j7wlkuyhEeWWORKcDoY1lQ)

David Whiting, would you mind re-posting the portion of your comment on 95% CI "probable" vs. "sure" (answer "a")? I probably buried the point between a large chunk of code and output. I think it is an interesting topic in its own right.

I agree with your perspective, but it seems every stats major fights to the death that 95% CI absolutely does not mean a 95% chance the population mean is in the CI. I assumed it had to be a wrong answer as such. But, I think I want to change my mind on which answer is "least wrong" if the auto-grader answer allows for ~16% misses on an intended 95% CI.

▲ 0 Upvote · Hide 1 Reply

DW

(/learn/inferential-statistics/profiles/def8b145896c8d3cbbe538e80e9b18ef)

David Whiting · 19 hours ago (/learn/inferential-statistics/discussions/rtJyteu6EeWfwAohgaM63Q/replies/j7wlkuyhEeWWORKcDoY1lQ/comments/oaDb6-yeEeWeyQ6xu8k3Vw) · Edited

"I thought answer "a" makes more sense but it depends how picky you are on the language. As I understand it, technically its not right to say that there is a 95% probability the calculated CI contains the population mean as the CI either does or it doesn't. If it does there is a 100% probability it contains it, if it doesn't there is a 0% chance it contains it. Personally I think this is a bit pedantic as you can say, out of all possible calculated CI, there is a 95% probability that the calculated CI contains the population mean. Therefore, you can be 95% "sure" (but technically not 95% probable) that the CI contains the population mean. Thats as I understand it anyway!"

An example. Lets say the population mean is 0. I take a sample and get a CI of 0.2 to 1.2. Unfortunately I was unlucky and the CI doesn't contain 0. Therefore there is 0% probability that this CI contains the population mean. It doesn't make sense to say the CI 0.2 to 1.2 has a 95% probability of containing the population mean. Its all right to say we are 95% confident that the CI 0.2 to 1.2 contains the population mean but its a no no to say 95% probable. I think if anything this makes it confusing and the reality is 95% of the time the calculated CI will contain the population mean and therefore I would consider that

95% of the time the calculated CI will contain the population mean and therefore I would consider that the equivalent of a CI having 95% probability of population mean. Statistical semantics if you ask me!

▲ 0 Upvote

SD (/learn/inferential-statistics/profiles/f34069ce8df6de7dbebfedfb7e760d9f)

Reply

Reply

ZB (/learn/inferential-statistics/profiles/3cdb3ac214a020ebfbe64f1e50b03eac)

Zara Bamdad Teaching Staff · 12 hours ago (/learn/inferential-statistics/discussions/rtJyteu6EeWfwAohgaM63Q/replies/7k9Fw-zgEeWnOBISxifLdw)

It seems like you were both pretty thorough with this! Where do we stand on the issue now?

Regarding the 'loose language' and 'probable' vs. 'sure' discussion - the word "likely" is added to an answer option because, as you have discussed here, it's risky to speak in terms of certainties.

▲ 0 Upvote · Reply

DW (/learn/inferential-statistics/profiles/def8e145896c8dbebe538eb0e9b18afd)

David Whiting · 3 hours ago (/learn/inferential-statistics/discussions/rtJyteu6EeWfwAohgaM63Q/replies/GTkLeu0wEeWCWg4sBTB7QQ)

The grader says b is the answer but it should be a.

▲ 0 Upvote · Reply

1

SD (/learn/inferential-statistics/profiles/f34069ce8df6de7dbebfedfb7e760d9f)

Reply

Reply