Confidence Intervals

Grinnell College

November 10^{th}, 2025

Where We are at....

We are using normal distributions to help approximate the sampling distributions of common statistics.

- 1 mean (w/ known variance)
  - Hypothesis Test (Z-test)
  - Confidence Interval
- 1 mean (w/ unknown variance)
  - HT (t-test) and CI
- Difference for 2 means (HT and CI)
- Regression Coefficients (HT and CI)
- 1 and 2 proportions (HT and CI)

We then have other tests to discuss

- ANOVA (testing difference of more than 2 means)
- χ^2-test (testing for association between two categorical variables)

Background 1

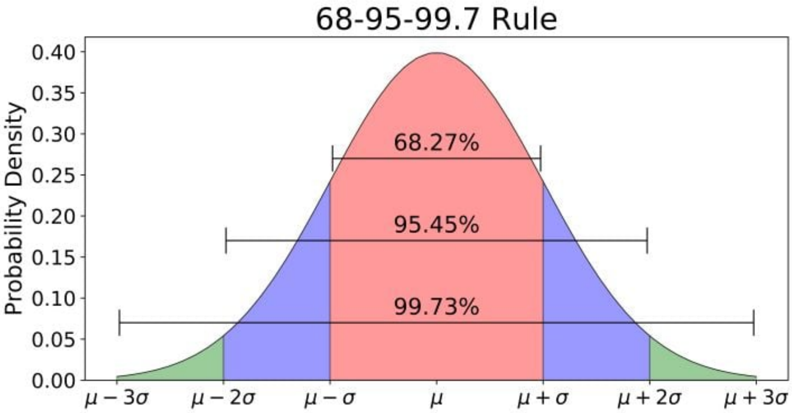
On the dice lab, I talked about rolling the dice twice being "safer" than doubling the value of the dice.

The idea was that most of our dice roll sums will be in the middle more often than the small (or large) values, in the long run

The "middle" was most probable

Background 2

Previously (and on your homework...) we discussed the 68-95-99.7% rule which works if we know the normal distribution parameters



Goals

We can talk about the same thing using our sampling distribution for our sample mean, given our assumptions are met (which are...?)

- N(μ, σ^2/n)
- Still assuming we know σ^2
- Instead of the sum of dice roll values we have the sample mean
- Can talk about where likely sample means are going to be
- Our population is all possible sample means
  - Similar to permutation test's all possible juries
  - Here we assume the number of possible combinations is infinite

Recall the Sampling Distribution

If our assumptions are met (which are....?) then we can talk about the long term behavior of the means from random samples.

- We have already seen calling BS via hypothesis tests

We will now talk about where most of the sample means will be

- Report things like where middle 95% of sample means will be over the long run
  - Similar to 68-95-99.7% rule
- Uses the normal probability distribution
- And we will formalize it with math and interpretations

Motivating Example

My niece's 1^{st} birthday is next month and I *might* get some Duplo legos.

I'm trying to get an idea of how much this is going to cost me, so what is the average cost of Duplo legos?

Motivating Example Solution 1

My niece's 1^{st} birthday is next month and I *might* get some Duplo legos.

I'm trying to get an idea of how much this is going to cost me, so what is the average cost of Duplo legos?

Naive Solution 1: One strategy is to just report the mean for the cost of the Duplo legos which is $34.26.....but I probably won't pay that exactly

Motivating Example Solution 2

My niece's 1^{st} birthday is next month and I *might* get some Duplo legos.

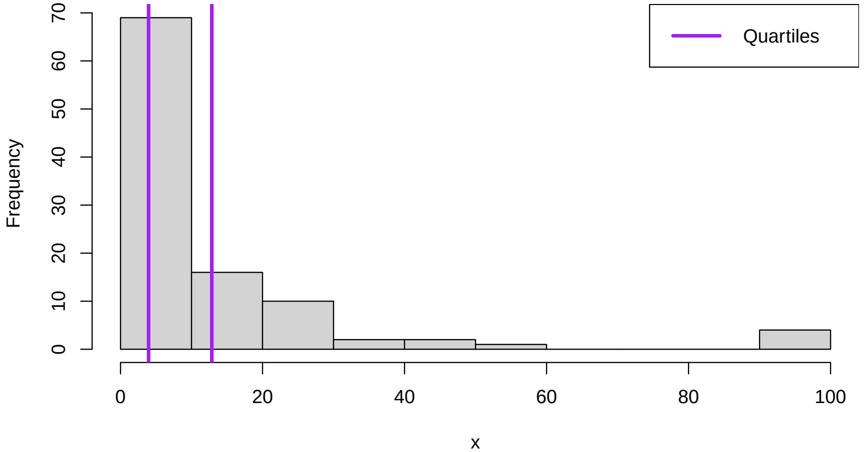
I'm trying to get an idea of how much this is going to cost me, so what is the average cost of Duplo legos?

A better strategy is to report a range where we think the mean might be

- Need an upper and lower value
- Naive Solution 2: report Q_1 and Q_3 or some other pair of percentiles near the "middle" from our sample

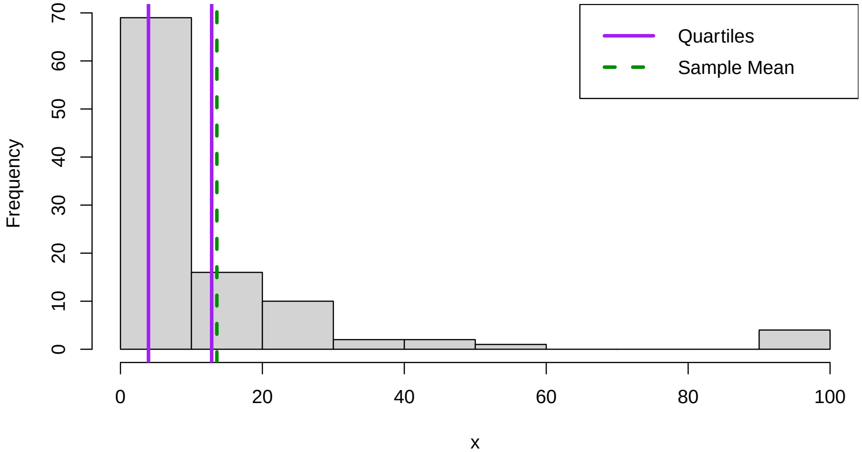
Motivating Example Solution 2 Problems

Our sample distribution has observations drawn from the population, these aren't averages we are looking at



Motivating Example Solution 2 Problems

Our sample distribution has observations drawn from the population, these aren't averages we are looking at



Motivating Example Solution 3

My niece's first birthday is next month (12/3) and I *might* get her some Duplo legos. I'm trying to get an idea of how much this is going to cost me, so what is the average cost of Duplo legos?

Let's combine the last two solutions.

The best(?) strategy is to report where we think random sample means are most likely going to be (in the long run)

- Need an upper and lower value
- from a distribution that describes means
- Ideas?

Motivating Example Solution 3

My niece's first birthday is next month (12/3) and I *might* get her some Duplo legos. How much is that going to cost me on average?

Let's combine the last two solutions.

A better strategy is to report where we think random sample means are most likely going to be (in the long run)

- Need an upper and lower value
- from a distribution that describes means
- Use the sampling distribution
  - and it's percentiles
  - Eg Q_1 and Q_3 are valid choices

Confidence Interval

This is called a confidence interval (often abbreviated CI). A **confidence interval** gives the location of the middle (BLANK) % of the sampling distribution.

- Suggests where we think a randomly drawn sample mean will be
- Also suggests where we think the population mean will be
- The sampling distribution is almost the same as for hypothesis tests
  - With the same assumptions

Sampling Distribution's location

Wait, wasn't there a whole thing about not knowing where the sampling distribution was actually located?

Sampling Distribution's location

Wait, wasn't there a whole thing about not knowing where the sampling distribution was *actually* centered?

Correct! Under hypothesis testing we assumed we knew the population's mean's true value (H_0)

Sampling Distribution's location

Wait, wasn't there a whole thing about not knowing where the sampling distribution was *actually* centered?

Correct! Under hypothesis testing we assumed we knew the population's mean's true value (H_0)

For confidence intervals we actually just plug in our sample mean

- Sample Dist $\approx$ Pop Dist (we hope)
- Sample mean $\approx$ Pop mean
    - $\bar{X} \approx \mu$

Example 1: Penguins!!

Body measurements were taken on 340 of penguins, randomly selected, on some islands off of Antarctica with a mean flipper length of 200mm. Find a 95% confidence interval for the mean flipper length

Example 1: Penguin Assumptions

Are the observations random?

- Yes, it says so

Are the observations independent and identically distributed?

- Maybe? There'll be genetic relatives in each colony
- And each colony could be different with different means

Is the population normally distributed or is the sample size large?

- The population is not normally distributed (because the sample isn't)
- The sample size is....large with over 300 observations

Despite possible concerns with independence let's move forward

Ex: Penguin Sampling Distribution

The resulting sampling distribution then is N(μ, σ^2/n) which is

$$\bar{X} \sim N(200, 180/340)$$

where

- μ is the mean of our population
  - Plug in our estimated sample mean (200)
  - We could hypothesize it's value (last slide deck)
- σ^2 is our (pop) variance
  - 180 for this example
  - Assumed known before we ever started talking about penguins
  - Very unrealistic short of an ornithologist savant
- n is our sample size, here 340

Alright, let's see the confidence interval formula!

CI Formula

A (1-α) **confidence interval** is calculated as....

$$estimate \pm (distributional\ value)(standard\ error)$$

$$sample\ mean \pm z_{1-\alpha/2}(\sqrt{\sigma^2/n})$$

$$200 \pm z_{1-\alpha/2}(\sqrt{180/340})$$

$$200 \pm z_{1-\alpha/2}(0.7276)$$

Okay but what is $z_{\alpha/2}$?

$z_{1-\alpha/2}$

This value comes from the standard normal distribution and represents the 1 - $\alpha/2$-percentile.

- $\alpha/2$ because we want α probability in both tails combined
- So we cut α in half for symmetry
- Can look it up with qnorm() in R
- Common choices given below
- No choice is more common than 95%

Confidence Level (= 1 - α)	80	90	95	99
Distributional Value (= $z_{1-\alpha/2}$)	1.28	1.645	1.96	2.58

Technically the $z_{\alpha/2}$ are actually negative but since we are doing $\pm$ z it washes out. Alternatively we could look up it's symmetric counterpart $z_{(1-\alpha/2)}$ and get the same result that way.

Example 1: Confidence Interval

So, we wanted a 95% confidence interval so we need to plug in our values...

$$
\begin{aligned}
200 &\pm z_{1-\alpha/2}(0.7276) \\
200 &\pm 1.96^{*}(0.7276) \\
200 &\pm 1.43
\end{aligned}
$$

$$
\begin{aligned}
\text{Lower Bound} &= 200 - 1.43 = 198.57 \\
\text{Upper Bound} &= 200 + 1.43 = 201.43
\end{aligned}
$$

CI Interpretation

" We are (BLANK) % confident the true (STATISTIC) is between (LOWER BOUND) and (UPPER BOUND)"

- BLANK = confidence level
  - Here 95%
- STATISTIC = statistic we are interested in
  - Here it's a mean
- Lower and Upper Bounds that were calculated last slide
  - Lower gets reported first!!
  - Safety check: should both be same distance from the mean

Example 1: CI Interpretation

" We are 95% confident the true mean flipper length is between 198.57 and 201.43 mm"

- BLANK = 95% = (1 - α)
- STATISTIC = mean flipper length
- Lower and Upper Bounds: 198.57 and 201.43 from our calculations

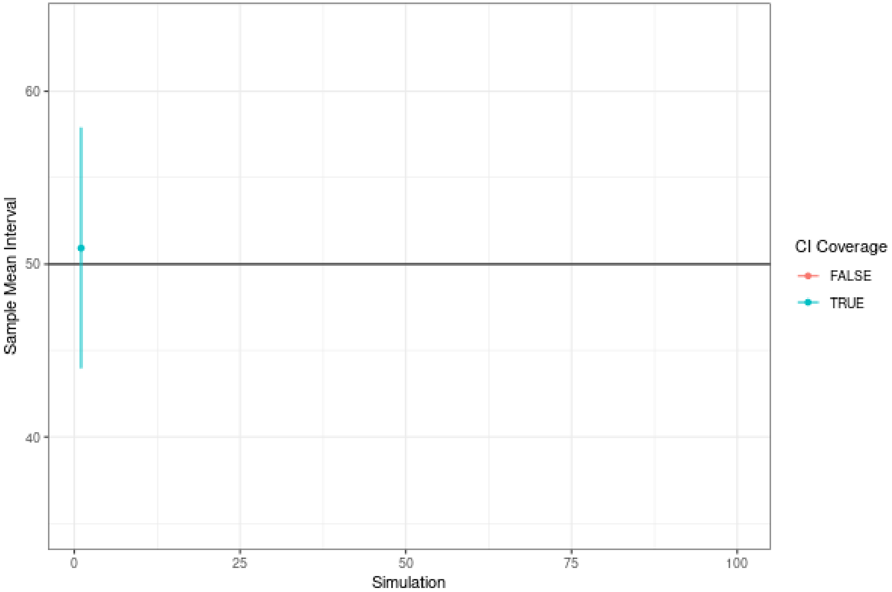
Okay but what is confidence? And what does it tell us?

Confidence does NOT mean probability. Instead, its the idea that if

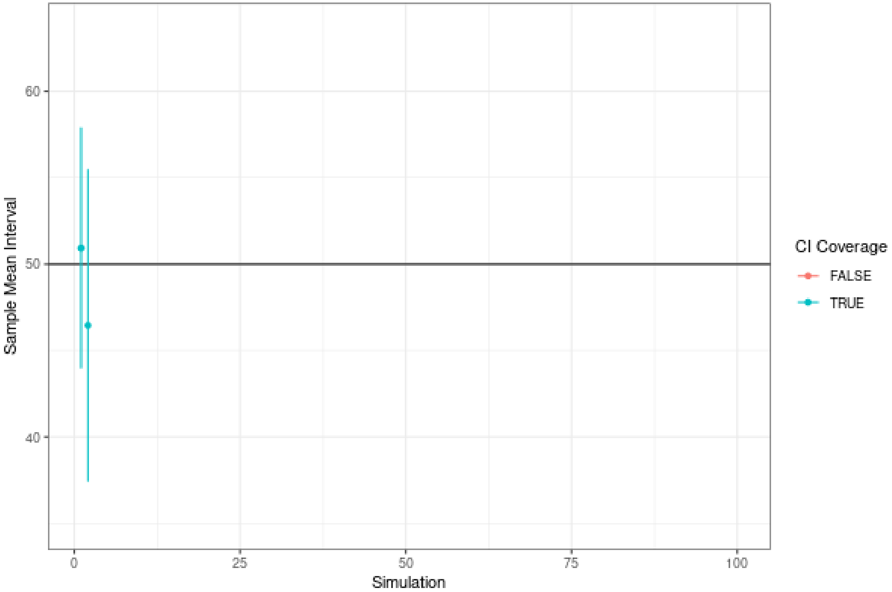
1. We take an infinite number of samples
2. And build a CI for each at 95% confidence (say)
3. About 95% of our intervals will include the "truth"

There is a famous picture for this.....

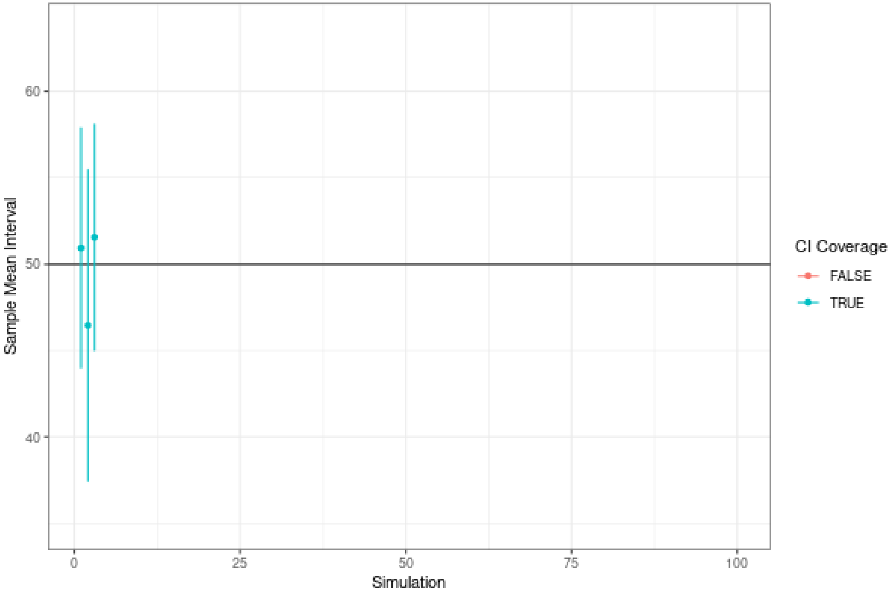
Confidence Visual: 1 CI



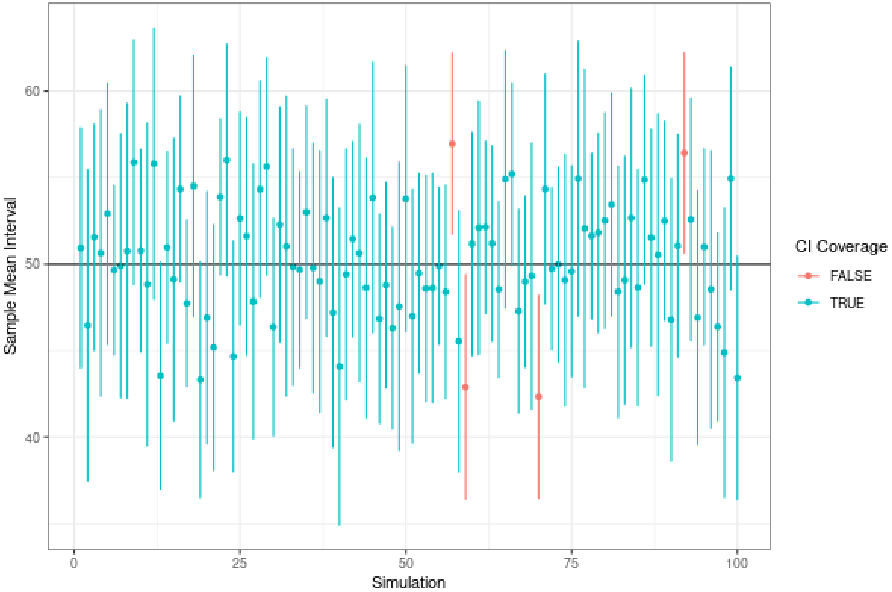
Confidence Visual: 2 CI



Confidence Visual: 3 CI



Confidence Visual: 100 CI



CI Interpretation: What it's not

A super common mistake is to think that the next observation is most likely going to be inside our interval. It's not.

- The CI are built using our sampling distribution for the sample means
- It doesn't talk about our sample directly per se
- If the distribution was created with means, then we talk about means
- CI are most useful for explanation/description of a statistic
- and almost useless for predictions

The interval that does this is called a *prediction interval*

Another Example

Lumber from cherry trees can be quite beautiful running from a pinkish brown to a reddish color. A common goal for woodworkers is to get long planks which is a function of the height of the tree. In a random sample of 31 ($= n$) trees from around a forest, the average height was 73 ($= \bar{x}$) feet with a known population variance of 45 ($= \sigma^2$)

Make a 90% confidence interval for the mean height of cherry trees

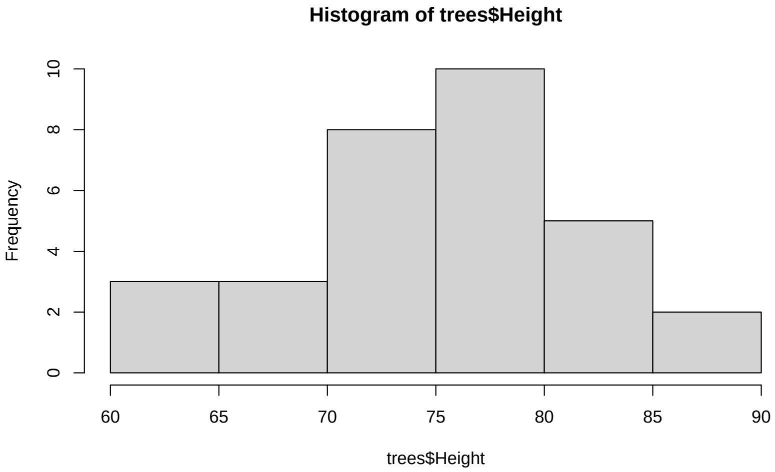
(HINT: You'll need to check assumptions to build a sampling distribution)

Another Example: Assumptions

From z-tests we know we are good but to reiterate...

- Random?
  - Yes, it says so
- Independent and Identically Distributed?
  - Yes although trees could compete maybe? Relatives?
  - I'm not too concerned
- Population distribution is normal or large n?
  - Sample distribution is normal(-ish) so we are good to go

Another Example: Visualize



Another Example: CI

$$\textit{estimate} \pm (\textit{distributional value})(\textit{standard error})$$

which leads to...

$$73 \pm (1.645)(\sqrt{45/31})$$
$$73 \pm 1.98$$

(71.02, 74.98)

Another Example: CI Interpretation

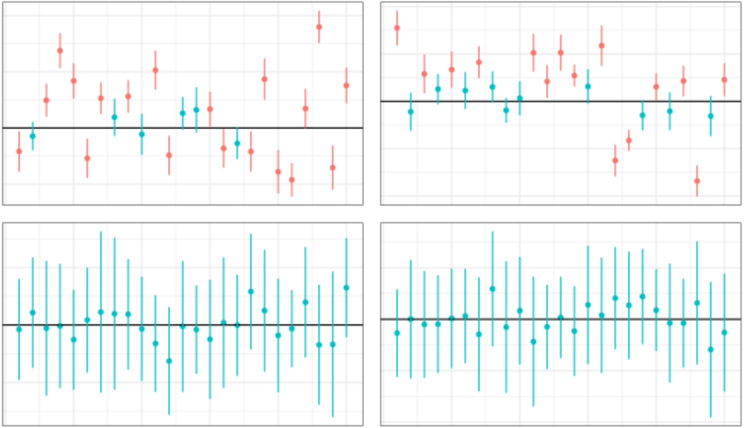
" We are (BLANK) % confident the true (STATISTIC) is between (LOWER BOUND) and (UPPER BOUND)"

- BLANK = 90%
- STATISTIC = mean cherry tree height
- Lower = 71.02
- Upper = 74.98

"We are 90% confident the true mean cherry tree height is between 71.02 and 74.98 inches"

Confidence Visual: Trade Offs

Smaller confidence leads to a tight interval with more “misses”. Larger confidence leads to wide intervals but few “misses”. It’s a trade off



Conclusion

Broadly, for any statistic there are two main types of inferences we can do

- Hypothesis test
  - We can see if a current working hypothesis seems reasonable
  - Check the probability we'd see a result as extreme as we saw IF THE NULL HYPOTHESIS IS TRUE
  - We assume the population mean in H_0
- Confidence Intervals
  - Gives the range that we have confidence in that the true parameter is located in
  - We assume the population mean *is* our sample mean for building the sampling distribution