LAST TIME

construct an interval around an observed statistic, $\bar{X}$

$CI = \bar{X} \pm (t_{cv})(s_{\bar{X}})$

CONFIDENCE

we never say that a specific 95% confidence interval contains $\mu$ with probability 0.95

either the interval contains $\mu$ or it does not

we can say that the procedure of producing CIs produces intervals that contain $\mu$ with probability 0.95

we do talk about the confidence that an interval includes $\mu$

we would say that the confidence interval contains $\mu$ with confidence of 0.95

the confidence is in the procedure of calculating CIs

CONFIDENCE INTERVAL

given our data, we could also compute confidence intervals around $\bar{X} = 535$

$t_{cv} = \pm 1.96$

$CI = \bar{X} \pm (t_{cv})(s_{\bar{X}})$

$CI_{95} = 535 \pm (1.96)(8.33)$

$CI_{95} = (518.67, 551.33)$

COMPARISON

note: the rejected $H_0: \mu = 455$ is consistent with the CI

455 is not in the 95% confidence interval (518.67, 551.33)

CI contains only tenable values of $\mu$, given the sampled data
CI AND HYPOTHESIS TESTS

CIs ask: which values of \( \mu \) would it be reasonable for me to get the value of \( X \) that I found?

Hypothesis tests ask: is the value of \( X \) I found consistent with a hypothesized value of \( \mu \)?

“reasonable” and “consistent” are defined relative to Type I error (\( \alpha \)), and confidence (1-\( \alpha \))

HYPOTHESIS TESTING

constructing a CI is like testing a large number of non-directional hypotheses simultaneously:

\[
H_0 : \mu = 435 \\
H_0 : \mu = 22 \\
H_0 : \mu = 522 \\
H_0 : \mu = 549 \\
H_0 : \mu = 563
\]

anything in the CI (518.67, 551.33) would be not be rejected, anything not in the CI would be rejected

EXAMPLE

On the papers going around the room, write down the number of math-based courses you have taken at college (include physics, engineering, and computer science, if it was largely math-based)

Now go around the room and sample this information from 6 other people

Calculate the mean and standard deviation for your sample

\[
\bar{x} = \frac{\sum X_i}{n} \\
S = \sqrt{\frac{\sum X_i^2 - (\sum X_i)^2/n}{n-1}}
\]

I’ll calculate the population mean for the class

CONFIDENCE INTERVAL

\[
CI = \bar{X} \pm (t_{cv})(s_X) \\
\]

Calculate standard error of the mean

\[
s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{8}{\sqrt{6}} =
\]

\[
t = \frac{\bar{x} - 3}{s_{\bar{x}}}
\]

use the \( t \) Distribution calculator with \( df = n - 1 = 5 \) to compute

\[
p =
\]

(4) Make your decision:

COMPARISON

Who rejected \( H_0 \)?

Who have the value 3 outside their CI?

Should be similar!

Now repeat everything for \( H_0 : \mu = 4 \)

notice what is required in the new calculations!
STATISTICAL PRECISION

consider the equation for confidence intervals
\[ CI = \bar{X} \pm (t_{cv})(s_{\bar{X}}) \]
where
• \( \bar{X} \) is the sample mean
• \( t_{cv} \) is the critical value using the appropriate \( t \) distribution for the desired level of confidence
• \( s_{\bar{X}} \) is the estimated standard error of the mean
\[ s_{\bar{X}} = \frac{s}{\sqrt{n}} \]
smaller \( t_{cv} \) or \( s_{\bar{X}} \) produce narrower widths

PUBLISHED DATA

most researchers in the behavioral sciences use \( \alpha = 0.05 \)
this means that they make a Type I error only 5% of the time (or less)
no way to completely avoid making mistakes
this makes it quite likely that some of the data in published journals is wrong
it is important in science to double (and triple) check everything
if a bit of data is tremendously important, better replicate the experimental finding

PROBLEMS

Researchers often use statistical significance as a way of identifying what findings should be published
If only findings with \( p < .05 \) are published, then journals can be filled with findings where \( H_0 \) is actually true
even if \( H_0 \) is true, around 5% of samples will produce a significant \( p \) value
If non-significant findings are not published, then it becomes hard to interpret the findings that actually are published (publication bias)

STATISTICAL PRECISION

since
\[ s_{\bar{X}} = \frac{s}{\sqrt{n}} \]
increasing the sample size \( n \) produces narrower widths of CI
narrower widths imply greater precision about where \( \mu \) is located
increasing \( n \) also modifies \( t_{cv} \) by changing degrees of freedom
\[ df = n - 1 \]
larger \( df \) leads to smaller \( t_{cv} \)
(see the Inverse \( t \) Calculator)

STATISTICAL PRECISION

we can also change \( t_{cv} \) by changing the level of confidence
larger level of confidence, implies smaller \( \alpha \), which implies larger \( t_{cv} \), which implies larger width of CI
makes sense, we become more confident the interval includes \( \mu \) by broadening the interval
of course, then we are less sure about the value \( \mu \)

PROBLEMS

Suppose you run a study with \( n = 20 \) subjects and get \( p = .07 \). This does not meet the \( \alpha = 0.05 \) criterion.
It is tempting to add an additional 10 subjects (for a total of \( n = 30 \)) and do the analysis again
This is a problem because you have given yourself an extra chance to get a significant outcome. Your Type I error is bigger than the \( \alpha = 0.05 \) that you intended.
Cannot add subjects to an experiment and re-analyze.
Nor can you stop data collection when you get a significant result (data peeking, optional stopping).
The sampling distribution is only valid for a fixed sample size. In the above cases, the sample size is not fixed.
To avoid these problems, you have to plan your experiment carefully in advance.
CONCLUSIONS

- estimation
- confidence intervals
- relationship with hypothesis testing
- statistical precision
- potential problems

NEXT TIME

- more hypothesis testing
- tests for correlation

Is there a correlation between homework and exam grades?