PSY 201: Statistics in Psychology

Lecture 21
Estimation of population mean
How tall is the room?

Greg Francis

Purdue University

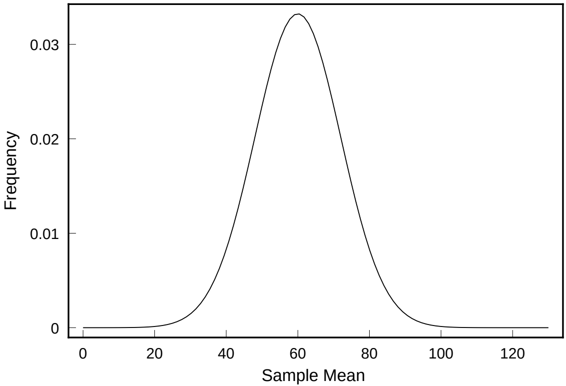
Fall 2023

LAST TIME

- we know how to check if a sample mean, $\overline{X}$, is statistically significantly different from a hypothesized population mean, μ.
- but sometimes we have no idea what μ is!
- we would like to be able to **estimate** μ using the sample data we have
- statistical estimation

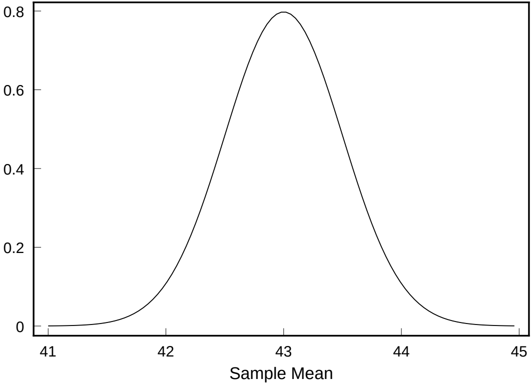
POINT ESTIMATION

- single value that represents the best estimate of a population value
- when we want to estimate μ, the best point estimate is the sample mean $\overline{X}$
- but the estimate depends on which sample we select!



INTERVAL ESTIMATION

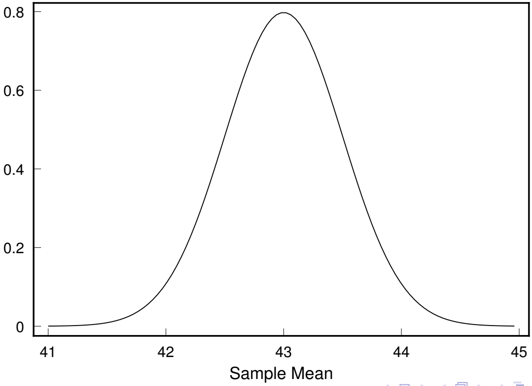
- we get a better idea of the value of μ by considering a **range** of values that are likely to contain μ
- we will show how to build up **confidence intervals** using the properties of the sampling distribution of the mean



σ KNOWN

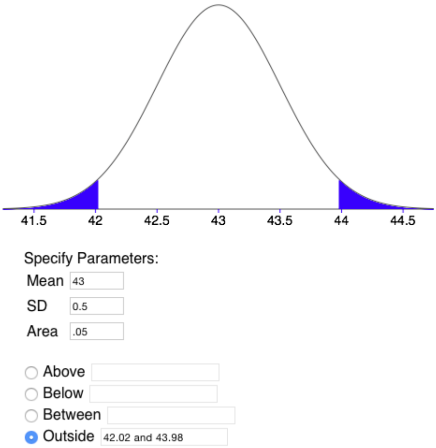
- to demonstrate our technique, suppose we have a population of scores with $\mu = 43$, $\sigma = 10$
- from the population we get the sampling distribution for samples of size $n = 400$ with

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = 0.50$$



INTERVAL ESTIMATION

- with the sampling distribution we can calculate (using the online Inverse Normal Distribution Calculator) that 95% of all sample means will lie between 42.02 and 43.98
- but since we do not really know the value of μ, we must **estimate** it



CONFIDENCE INTERVALS

- construct an interval around the observed statistic, $\overline{X}$

$$\text{CI} = \text{statistic} \pm (\text{critical value})\ (\text{standard error of the statistic})$$

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

- where
  - $\overline{X}$ is the sample mean
  - t_{cv} is the critical value using the appropriate t distribution for the desired level of confidence
  - $s_{\overline{X}}$ is the estimated standard error of the mean

$$s_{\overline{X}} = \frac{s}{\sqrt{n}}$$

LEVEL OF CONFIDENCE

- degree of confidence that computed interval contains μ
- usually complement of level of significance, α
- level of confidence is $(1 - \alpha)$
- calculating the critical value t_{cv} is the same!
- e.g., for $\alpha = 0.05$, $(1 - \alpha) = 0.95$, and

$$t_{cv} = 1.9659$$

- (using the Inverse t Distribution calculator with df=399)

CONFIDENCE INTERVAL

- suppose we calculate $\overline{X} = 44.6$
- the confidence interval is then

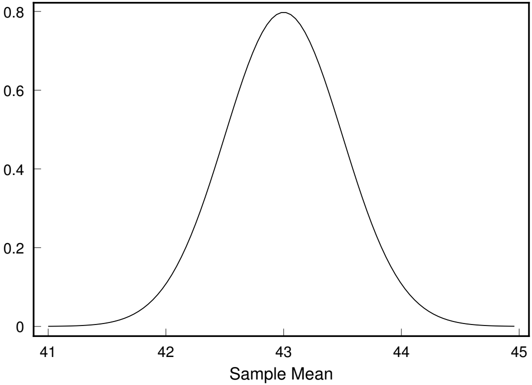
$$CI = \overline{X} \pm (t_{cv})(s_{\overline{X}})$$

$$CI_{95} = 44.6 \pm (1.9659)(0.50)$$

$$CI_{95} = (43.62, 45.58)$$

CONFIDENCE INTERVAL

- this means we are 95% confident that the interval (43.62, 45.58) contains the unknown value μ
  - Our procedure for producing the interval contains μ 95% of the time
- note: if $\mu = 43$ like was said originally, we are **wrong**!
  - CI does not contain μ (no way to avoid error completely)!



EXAMPLE

- Guess the height of this room in feet, and write down your guess on a piece of paper
- Now go around the room and get 10 guesses from other random people
- Then, tell me your guess
- Calculate the mean and standard deviation for your sample (use the on-line calculator for a one-sample t test)

$$\overline{X} = \frac{\sum X_i}{n}$$

$$s = \sqrt{\frac{\sum_i X_i^2 - [(\sum_i X_i)^2/n]}{n-1}}$$

- I'll calculate the *population* mean for the class
- each of you will calculate a confidence interval, for your sample, with $\alpha = 0.05$

CONFIDENCE INTERVAL

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

- Calculate standard error of the mean

$$s_{\overline{X}} = \frac{s}{\sqrt{n}} = \frac{s}{\sqrt{10}} =$$

- we have

$$\text{d.f.} = n - 1 = 10 - 1 = 9$$

- so from the Inverse t Distribution Calculator, we find that

$$t_{cv} = 2.262$$

CONFIDENCE INTERVALS

- thus

$$CI_{95} = \overline{X} \pm (t_{cv})(s_{\overline{X}})$$

$$CI_{95} = \overline{X} \pm (2.262)(s_{\overline{X}})$$

$$CI_{95} = (\quad , \quad)$$

WHAT DOES THIS MEAN?

- we conclude with 95% confidence that your interval contains μ
- this is a probabilistic statement about the **interval**
- μ is a **parameter**, a fixed number

$$\mu =$$

- different samples produce different confidence intervals, but 95% of the time the interval would contain μ
- check

CONFIDENCE

- we **never** say that a specific confidence interval contains μ with probability 0.95
- either the interval contains μ or it does not
- we **can** say that the procedure of producing CI's produce intervals that contain μ with probability 0.95
- we do talk about the **confidence** that an interval includes μ
- we would say that the confidence interval contains μ with confidence of 0.95
- the confidence is in the **procedure** of calculating CIs

CONCLUSIONS

- estimation
- confidence intervals
- t distribution
- interpretation

NEXT TIME

- more on estimation
- relationship between confidence intervals and hypothesis testing
- statistical precision

Less than 5% of published psychological research should be wrong (and why that probably isn't true).