Introduction to the t-test

Topic 8 · Thursday, February 4 and Tuesday, February 9

EDUC 15: Statistics for Education Research
UCI School of Education

Where are we?

Description

Summarize data

Visualize data

Prediction

We are here



Explanation

Peter the Anteater at a UC Irvine basketball game.

The four steps stay. Step 3 changes

Step 1: State a hypothesis about the population: H_0

Step 2: Set the criteria: choose α and find the critical region

Step 3: Observe a sample and compute a test statistic

Step 4: Reject or maintain H_0

Plan for today

One tail or two?

When we don't know σ

The t-test

On Tuesday: how big is the effect, and how sure are we?

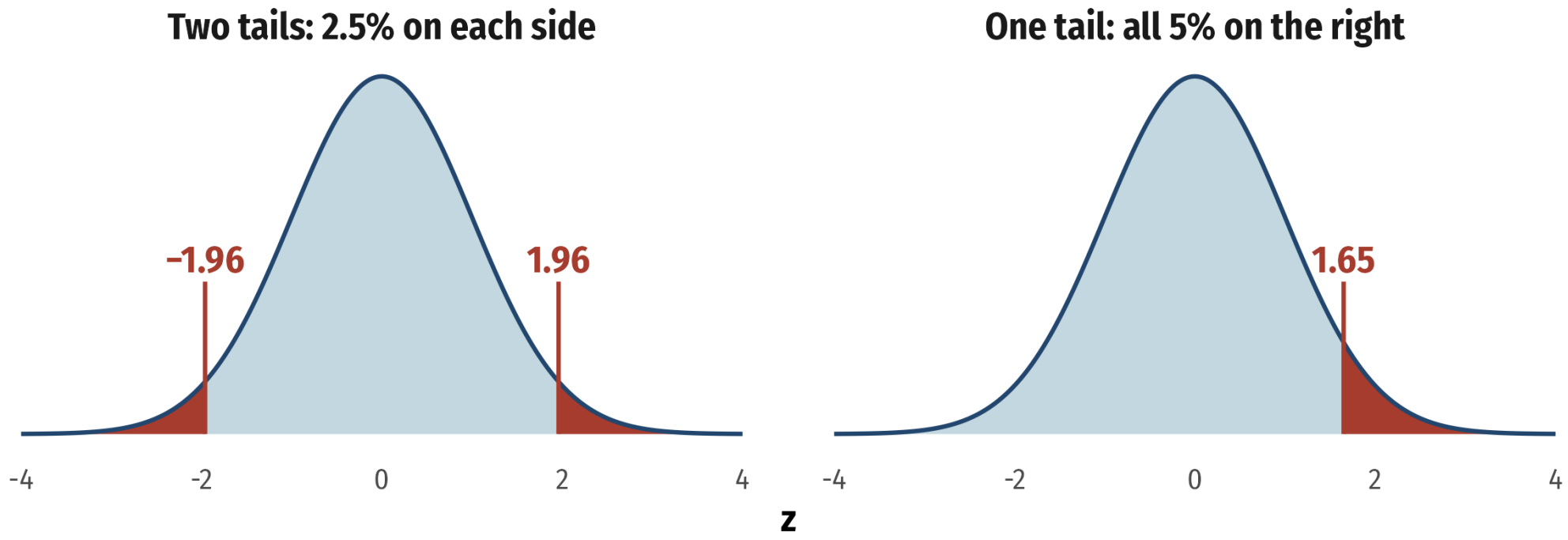
One tail or two?

A directional hypothesis picks a side

	Non-directional H_0	**Directional H_0**
In symbols	$\mu = 100$	$\mu \leq 100$
Laptops in class	Laptops don't affect learning	Laptops don't improve learning
Cat videos	Cat videos don't change your happiness	Cat videos don't make you happier

A directional test says which way the mean should move, not only that it moves

Same α, one tail: the cutoff moves in



With α = 0.05 in one tail, the critical z is 1.65, not 1.96. Type I errors stay at 5%

Predict it: One tail, and a real effect

H_0 says $\mu = 100$, with $\sigma = 15$ and $n = 10$. In truth, $\mu = 113.4$, so $d = 0.89$. A two-tailed test has 80% power

We switch to a one-tailed test on the right, and keep $\alpha = 0.05$. What happens to power?

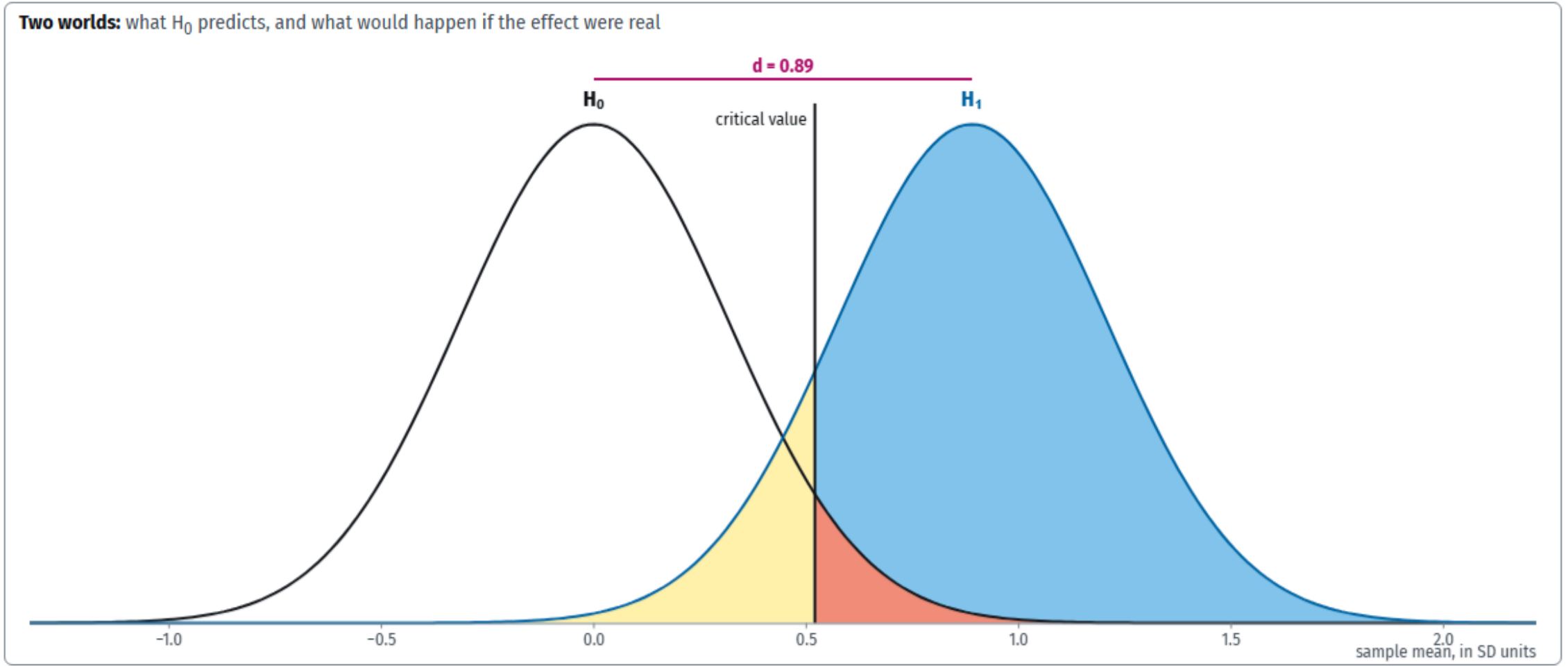
A It goes up

B It goes down

C It stays the same

D We can't say without the data

One tail adds power on that side



Simulator after Kristoffer Magnusson, rpsychologist.com.

So why not always use one tail?

A one-tailed test is blind on the other side. If the program hurts students, it can't say so

You must pick the side before you see the data. Picking it afterward is really a test at $\alpha = 0.10$

Two tails is our default from now on, unless a problem says otherwise

When we don't know σ

So far, z has needed σ

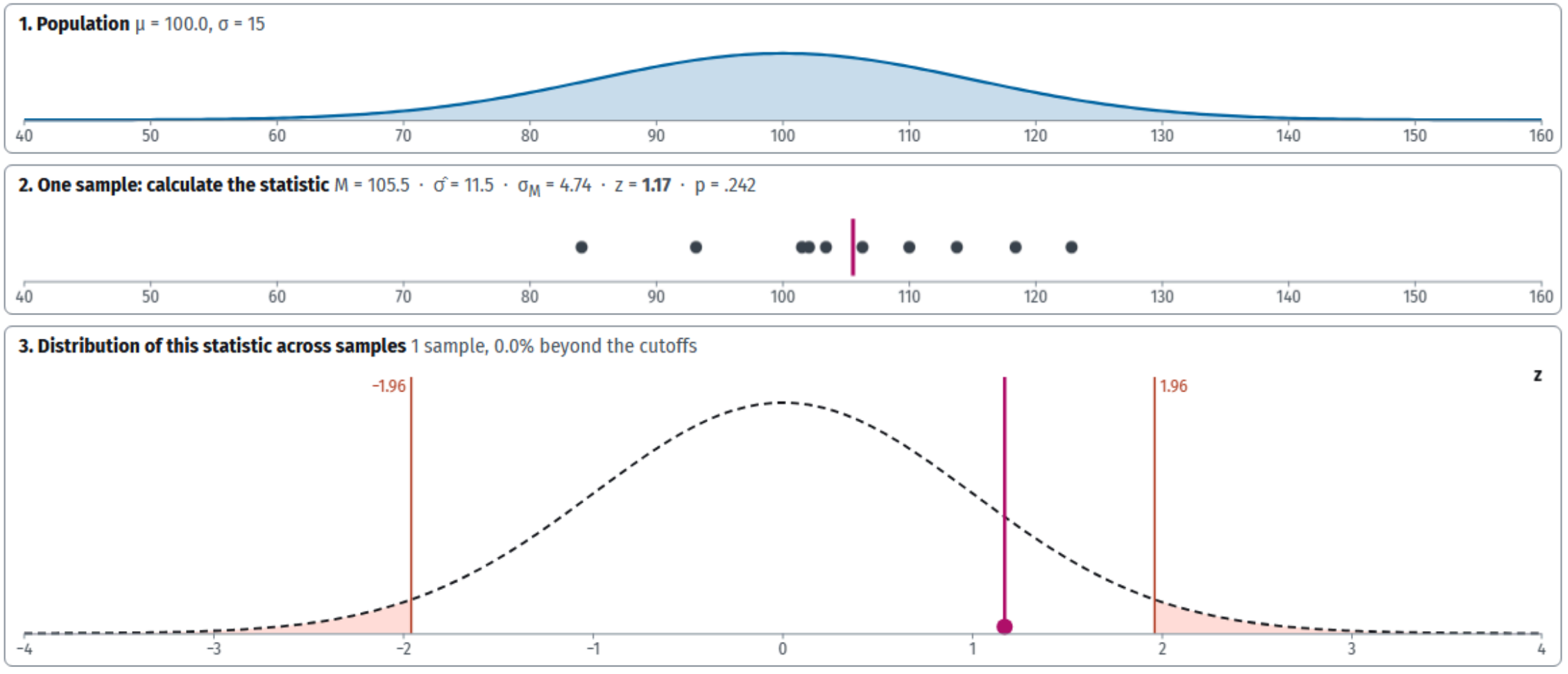
$$z = \frac{M - \mu}{\sigma_M} \qquad \text{where} \qquad \sigma_M = \frac{\sigma}{\sqrt{n}}$$

Top: how far M is from the mean under H_0

Bottom: how far a mean typically falls

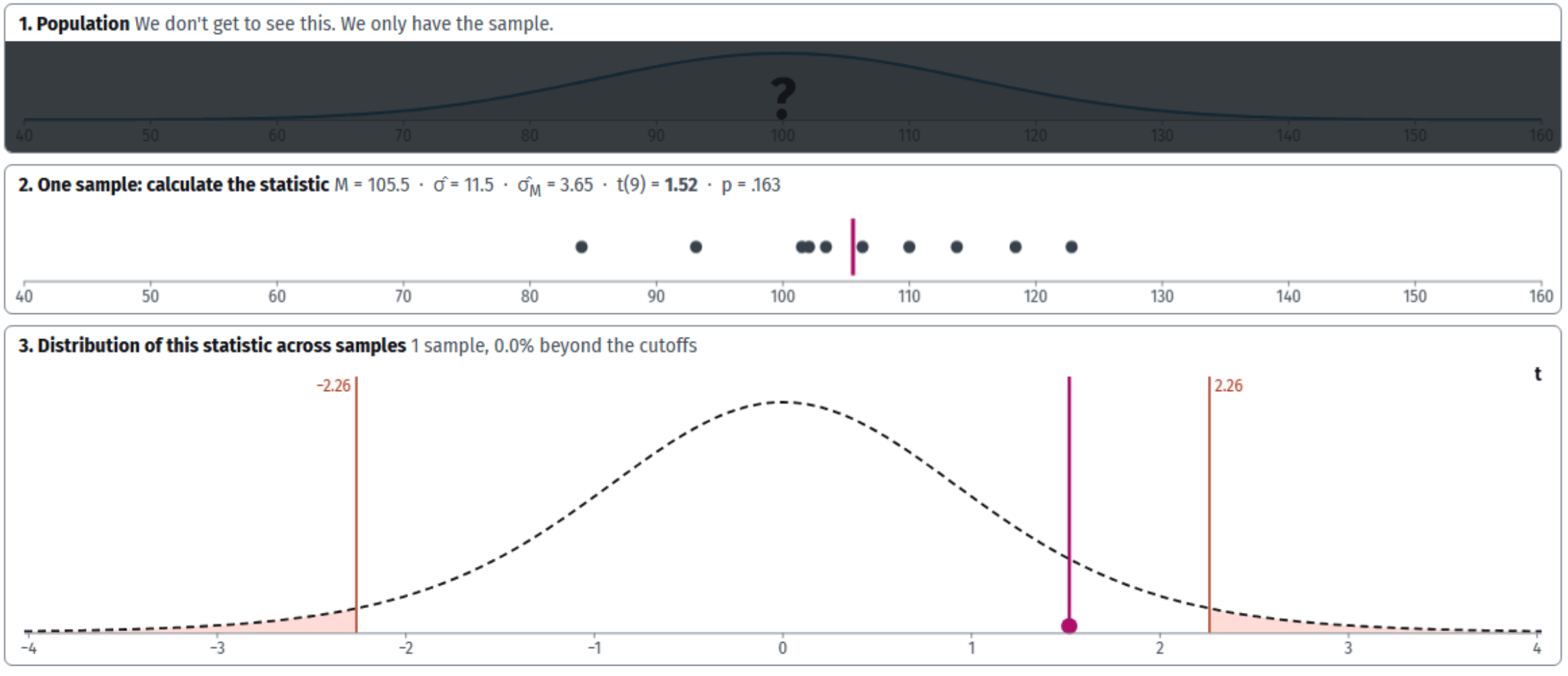
All quarter, someone has told us σ = 15. Who told us?

Here's one sample of 10, with σ known



Simulator after Kristoffer Magnusson, rpsychologist.com.

In real research, panel 1 goes dark



Simulator after Kristoffer Magnusson, rpsychologist.com.

So $\hat{\sigma}$ takes the place of σ

$$z = \frac{M - \mu}{\sigma_M} \qquad \text{becomes} \qquad t = \frac{M - \mu}{\hat{\sigma}_M}$$

Only the bottom changes: $\hat{\sigma}_M = \hat{\sigma} / \sqrt{n}$. The hat says it's an estimate

The new name warns us: this score behaves a little differently

In the book: $\hat{\sigma}$ is written s, and $\hat{\sigma}_M$ is written s_M

Dividing by n – 1 fixes the variance

Samples of 5 from our 619 students	Average across samples
The sample's own variance, SS / n	12.13
Our estimate, $\hat{\sigma}^2$ = SS / (n – 1)	15.17
The truth, σ^2	15.22

SS / n tends to underestimate σ^2. Dividing by n – 1 gives an unbiased estimate of the variance

Then $\hat{\sigma}$ is its square root. From here on, n – 1 is the degrees of freedom: df

Data: Muralidharan, Singh, and Ganimian (2019). 100,000 samples.

Predict it: Does swapping in $\hat{\sigma}$ matter?

Respond at **PollEv.com/adb**

Treat our 619 students as a population. Draw samples of 5, and use $\hat{\sigma}$ in place of σ. With the true σ, 5% of samples land beyond ± 1.96. With $\hat{\sigma}$?

A Much less than 5%

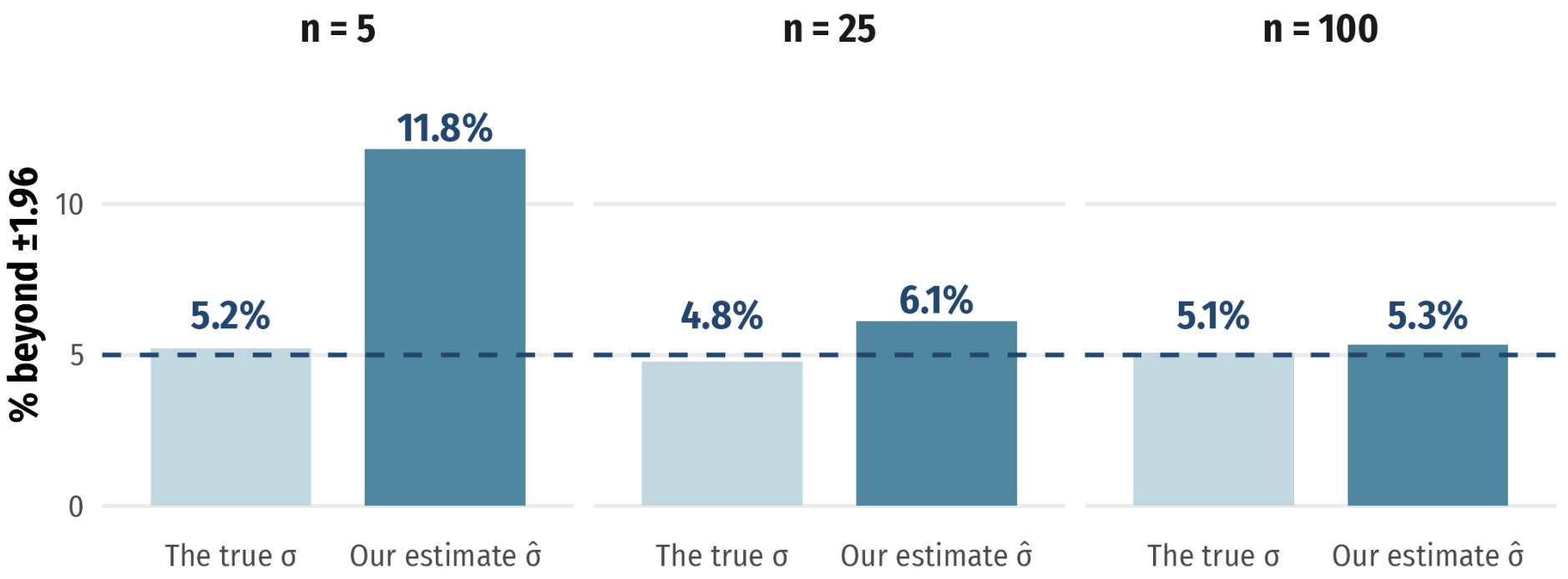
B About 5%, the same

C More than 5%

D It depends on μ

Data: Muralidharan, Singh, and Ganimian (2019).

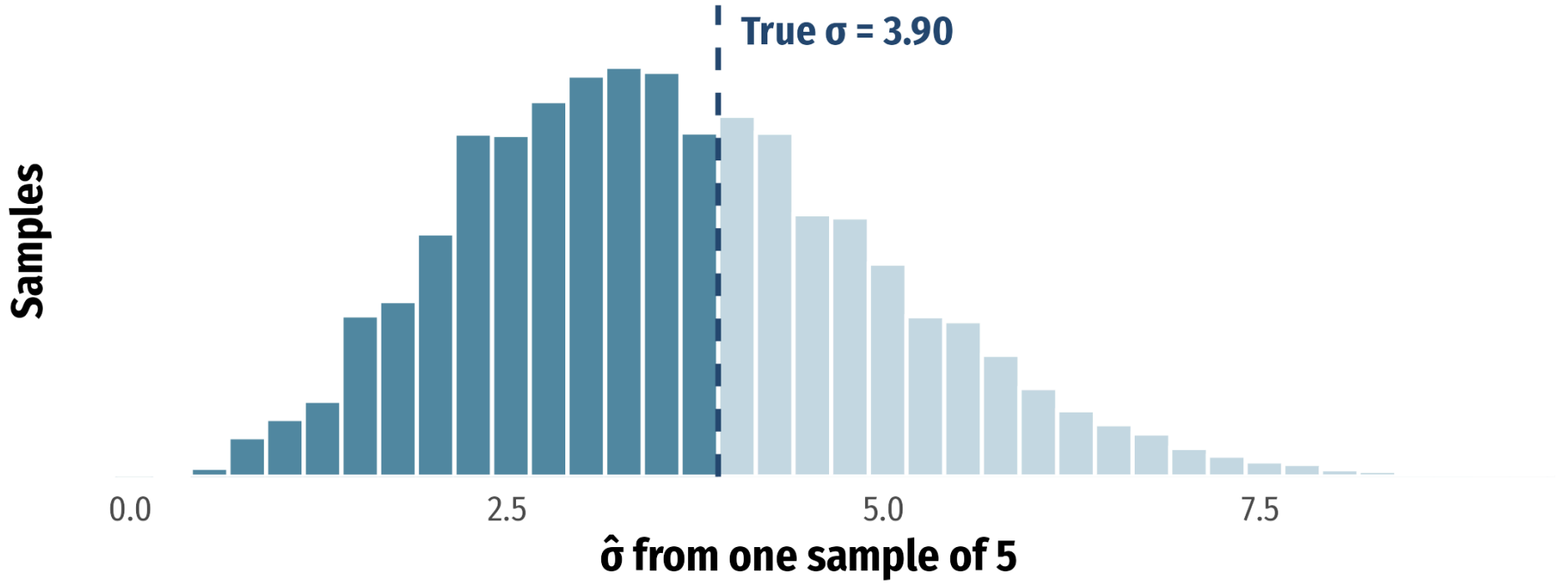
Small samples bring too many surprises



As n grows, $\hat{\sigma}$ gets closer to σ, and the problem fades

Data: Muralidharan, Singh, and Ganimian (2019). 100,000 samples of each size.

Why? $\hat{\sigma}$ bounces, and often lands low

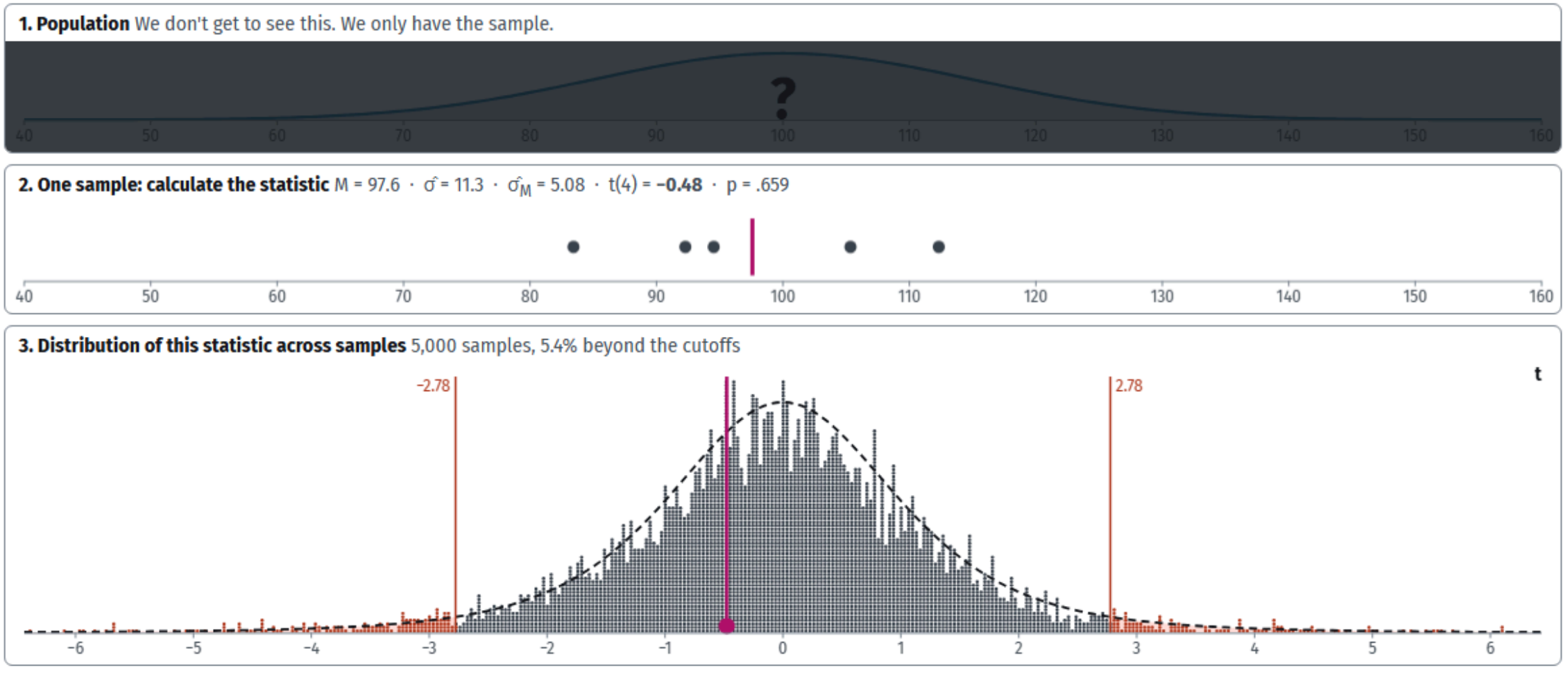


60% of samples give a $\hat{\sigma}$ below σ. Then the bottom shrinks and the score grows

Data: Muralidharan, Singh, and Ganimian (2019).

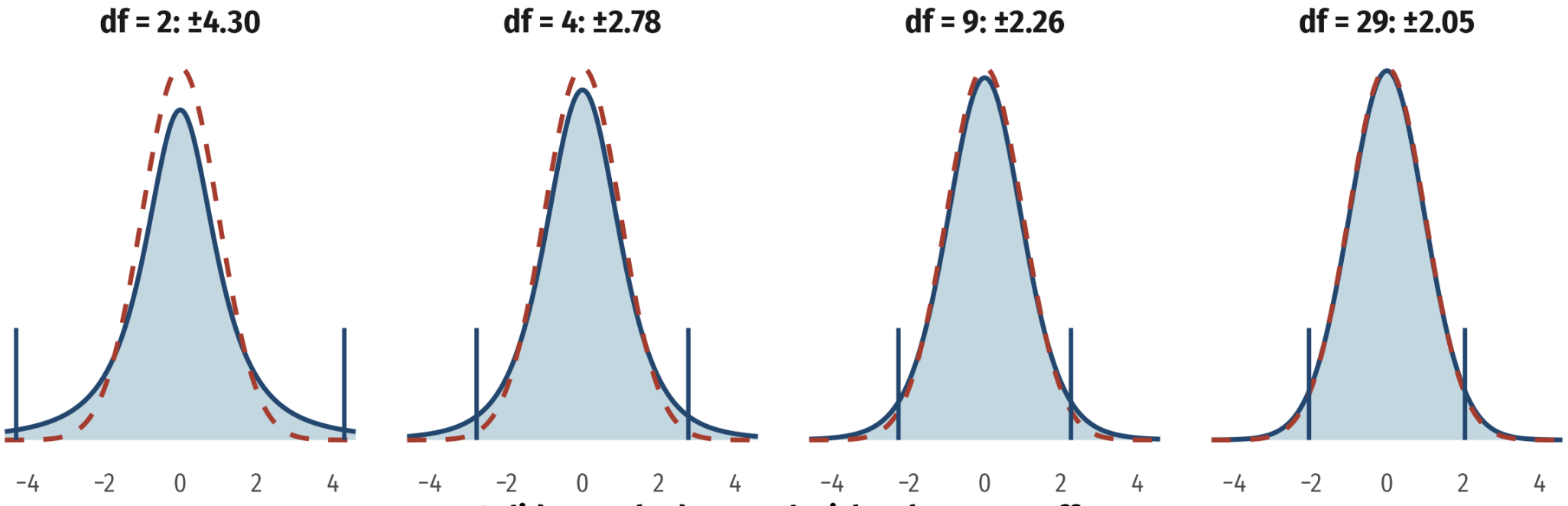
The t-test

With n = 5, the pile has fatter tails



Simulator after Kristoffer Magnusson, rpsychologist.com.

Fewer degrees of freedom, fatter tails



Solid: t. Dashed: normal. Ticks: the 5% cutoffs

A family of curves, one for each df. Near df = 30, it's hard to tell from z

Each df has its own cutoff

df	Two tails, α = 0.05	One tail, α = 0.05	Two tails, α = 0.01
4	2.776	2.132	4.604
9	2.262	1.833	3.250
14	2.145	1.761	2.977
24	**2.064**	**1.711**	**2.797**
60	2.000	1.671	2.660
120	1.980	1.658	2.617
Very large (z)	1.960	1.645	2.576

In practice, people use t at any n. With large samples it barely differs from z

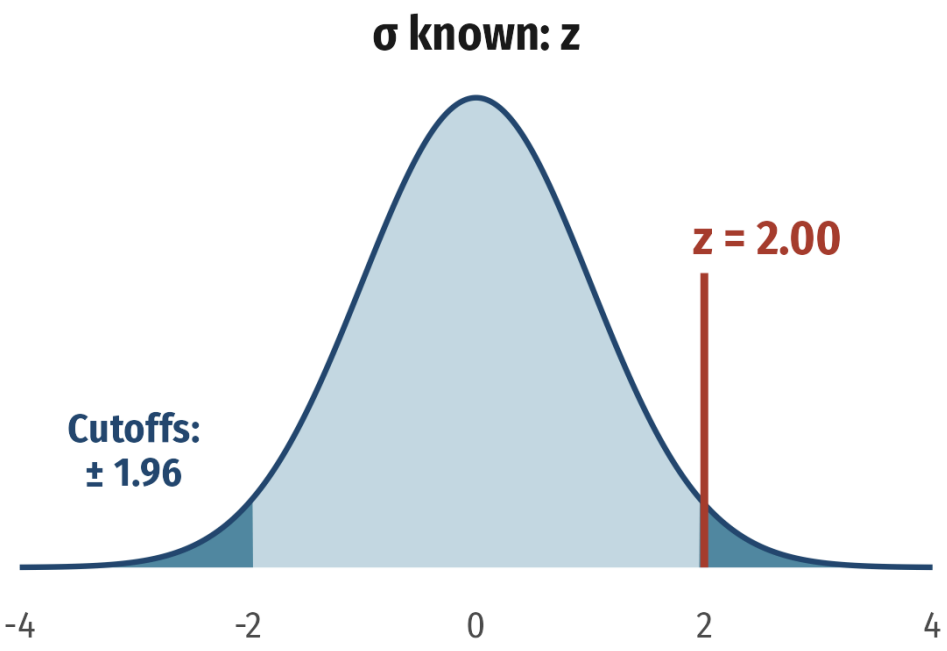
The full t table is in the book's appendix, and comes with the exams

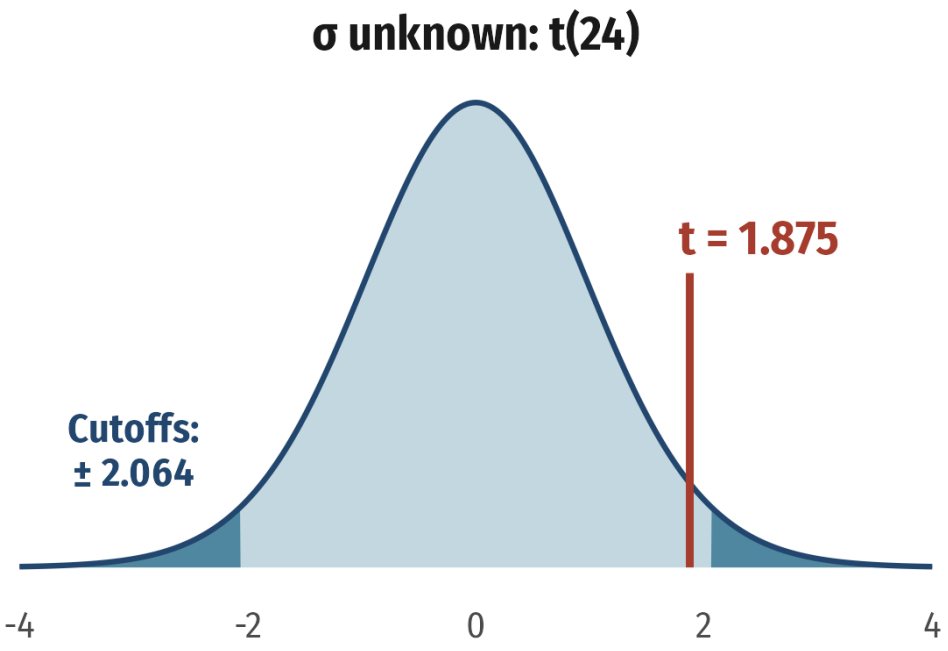
One example, with and without σ

n = 25, M = 106, SS = 6,144. H_0: μ = 100. α = 0.05, two tails

Step	If we know σ = 15	If we don't
1. State H_0	μ = 100	μ = 100
2. Set the criteria	z beyond ±1.96	t(24) beyond ±2.064
3. Find the spread	σ_M = 15 / √25 = 3.00	$\hat{\sigma}$ = √(6,144 / 24) = 16 $\hat{\sigma}_M$ = 16 / √25 = 3.20
3. Compute the score	z = 6 / 3 = 2.00	t = 6 / 3.2 = 1.875
4. Decide	Reject H_0: 2.00 > 1.96	Maintain H_0: 1.875 < 2.064

Same sample, different decisions





The score fell: $\hat{\sigma} = 16$ is above 15

The cutoff rose: 2.064 is above 1.96

We report: $t(24) = 1.88$, $p = 0.07$. We maintain H_0

Try it: A real group of students

24 Mindspark students take 2 to 3 hours a week of extra math lessons, called tuition in India. At baseline: M = 12.21, SS = 282.0

Test H_0: $\mu = 11.12$, the mean of all 619 students. Use $\alpha = 0.05$ and two tails. The cutoff is $t(23) = \pm 2.069$

Find $\hat{\sigma}$, $\hat{\sigma}_M$ and t, then decide. Work alone, compare with a neighbor, then we vote

Data: Muralidharan, Singh, and Ganimian (2019).

Try it: Answer

$\hat{\sigma} = \sqrt{(282.0 \;/\; 23)} = \mathbf{3.50}$

$\hat{\sigma}_M = 3.50 \;/\; \sqrt{24} = \mathbf{0.715}$

$t = (12.21 - 11.12) \;/\; 0.715 = \mathbf{1.52}$

1.52 is inside ±2.069: **maintain H_0**

We report: t(23) = 1.52, p = 0.14. These 24 students aren't surprisingly far from 11.12

Data: Muralidharan, Singh, and Ganimian (2019).

Before Tuesday

Friday, February 5: Lab 5, power and the one-sample t-test. Bring a calculator and a laptop

Sunday, February 7 at 6 PM: Mastery and Problem Set for Chapter 8

Read section 9-3 for Tuesday

Session 2: Tuesday, February 9

Thursday in one line: when we estimate σ, z becomes t, and cutoffs depend on df

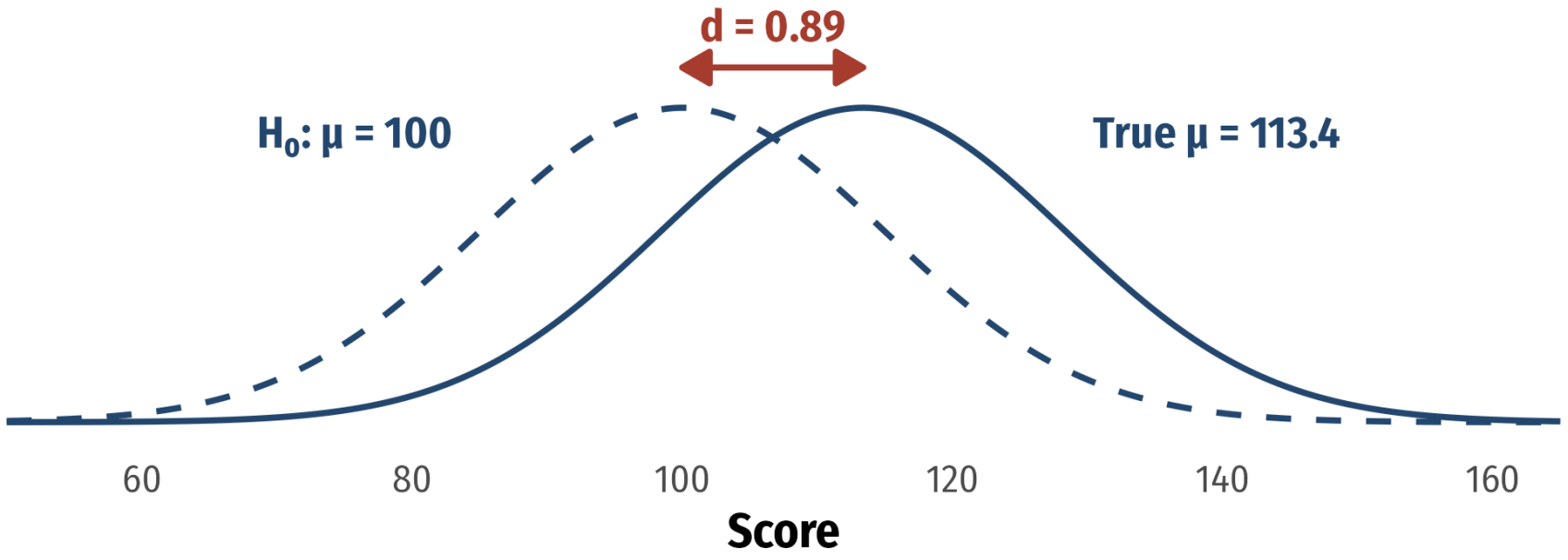
How big is the effect?

Confidence intervals

Intervals and tests

How big is the effect?

d says how far apart, in SDs



Cohen's d = (true μ − H₀ μ) / σ = (113.4 − 100) / 15 = 0.89. It's a parameter: it doesn't depend on n

Hedges' g is the estimate of d

$$d = \frac{\mu_{\text{true}} - \mu}{\sigma} \quad \text{is estimated by} \quad g = \frac{M - \mu}{\hat{\sigma}}$$

In our example: g = (106 − 100) / 16 = 0.375, so the mean sits about four tenths of an SD above H_0

Cohen's rough guide: 0.2 is small, 0.5 is medium, and 0.8 is large

In the book: g is called the estimated Cohen's d. The book also covers r^2 here: skip it for now, we come back to it

Predict it: Four times the people

Respond at **PollEv.com/adb**

Our example has n = 25, M = 106 and $\hat{\sigma}$ = 16, so t = 1.875 and g = 0.375. Now keep M and $\hat{\sigma}$, and make n = 100. What happens?

A t and g both go up

B t goes up, and g stays the same

C g goes up, and t stays the same

D Neither changes

Similar effect, different decision

Group	n	M	g	t	p	Decision
2 to 3 hours of tuition	24	12.21	0.31	1.52	0.142	Maintain H_0
More than 4 hours	81	12.46	0.35	3.11	0.003	Reject H_0

Both groups sit about a third of an SD above 11.12. Only the larger group is significant

Significant means "hard to explain by sampling error". It doesn't mean big

These families chose tuition, so neither test shows that tuition works

Data: Muralidharan, Singh, and Ganimian (2019). H_0: $\mu = 11.12$, $\alpha = 0.05$, two tails.

Confidence intervals

You already know the margin of error

A poll: 52% say yes, ± 3 points

Election polls report one, and charts draw it as a band around the line

It's about two standard errors: how far off this estimate might be

That's the standard error's first job from Topic 6. Now, for a mean

An interval: estimate ± margin of error

$$M \pm z\,\sigma_M \qquad \text{or} \qquad M \pm t\,\hat{\sigma}_M$$

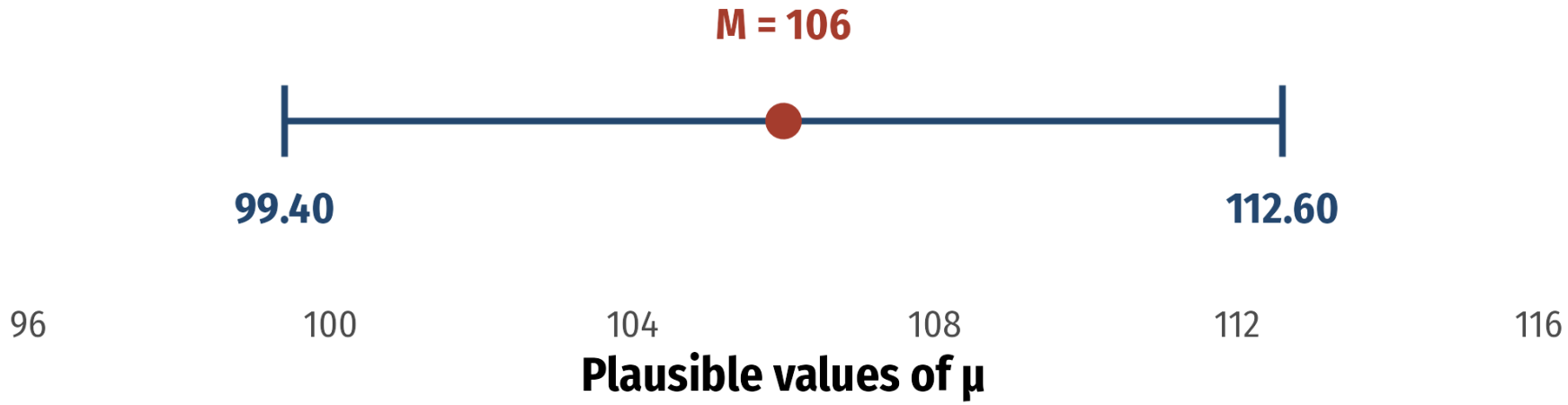
σ known: z = 1.96 for a 95% interval

σ unknown: t from the table, n − 1 df

Use the two-tailed 5% cutoff. Margin of error = cutoff × standard error

In the book: the second one is $M \pm t\ s_M$

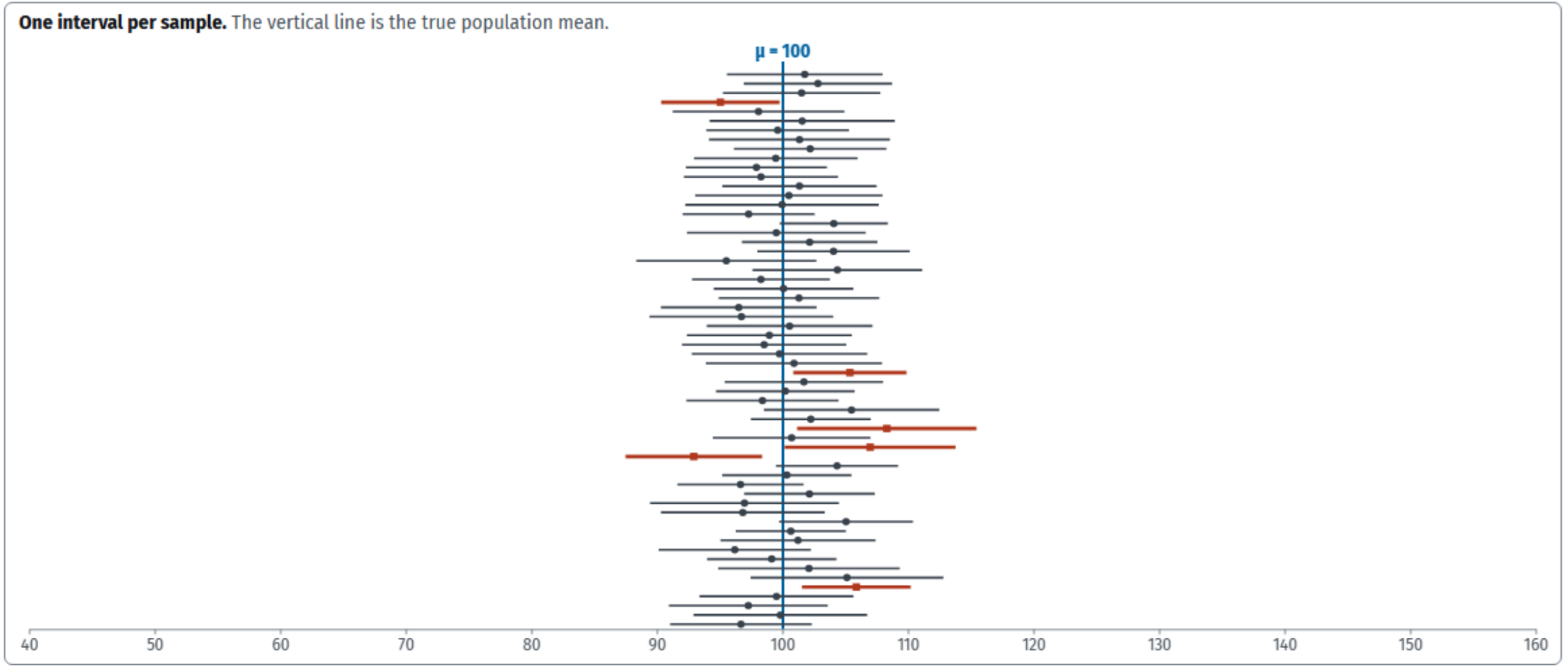
Our example: from 99.40 to 112.60



$$106 \pm 2.064 \times 3.2 = 106 \pm 6.60$$

We're 95% confident that μ is between 99.40 and 112.60

About 95% of the intervals catch μ



Simulator after Kristoffer Magnusson, rpsychologist.com.

What the 95% describes

The method: across many samples, about 95% of the intervals contain μ

μ stays put. It's the interval that moves from sample to sample

Our one interval either holds μ or it doesn't. We can't tell which

On real scores: 94.9% of 5,000 intervals from samples of 25 contain our 619 students' μ = 11.12

Data: Muralidharan, Singh, and Ganimian (2019).

Predict it: Ask for more confidence

We have a 95% interval from 99.40 to 112.60. We want 99% instead, with the same sample

What happens to the interval?

A It gets narrower

B It gets wider

C It stays the same

D It shifts to the right

Predict it: More confidence costs width

Confidence	n	Cutoff t	Margin of error	Interval
90%	25	1.711	5.47	[100.53, 111.47]
95%	25	2.064	6.60	[99.40, 112.60]
99%	25	2.797	8.95	[97.05, 114.95]
95%	100	1.984	3.17	[102.83, 109.17]

The answer is B. To be more sure of catching μ, cast a wider net

More people buy the width back: the last row is 95% with n = 100, same M and $\hat{\sigma}$

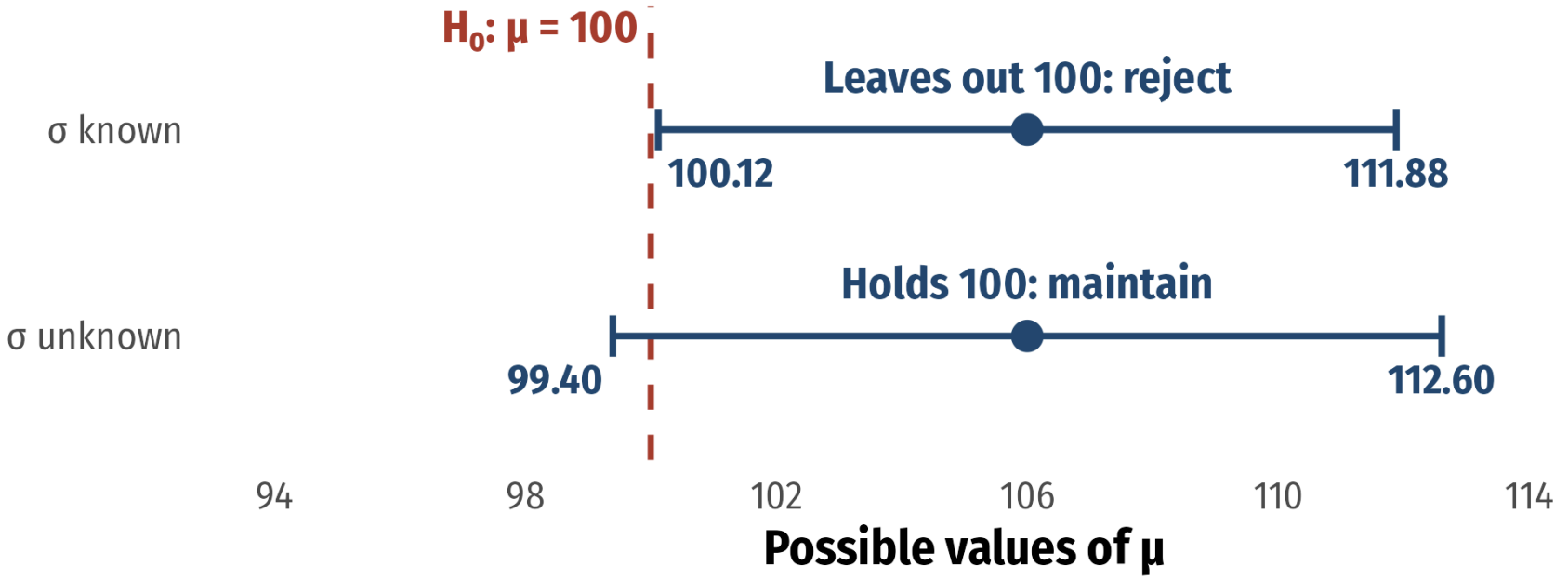
Intervals and tests

An interval can run the test for you

Step	If we know σ = 15	If we don't
1. State H_0	μ = 100	μ = 100
2. Set the criteria	95% interval, z = 1.96	95% interval, t(24) = 2.064
3. Build the interval	106 ± 1.96 × 3.00 [100.12, 111.88]	106 ± 2.064 × 3.20 [99.40, 112.60]
4. Decide	Reject H_0: 100 isn't inside	Maintain H_0: 100 is inside

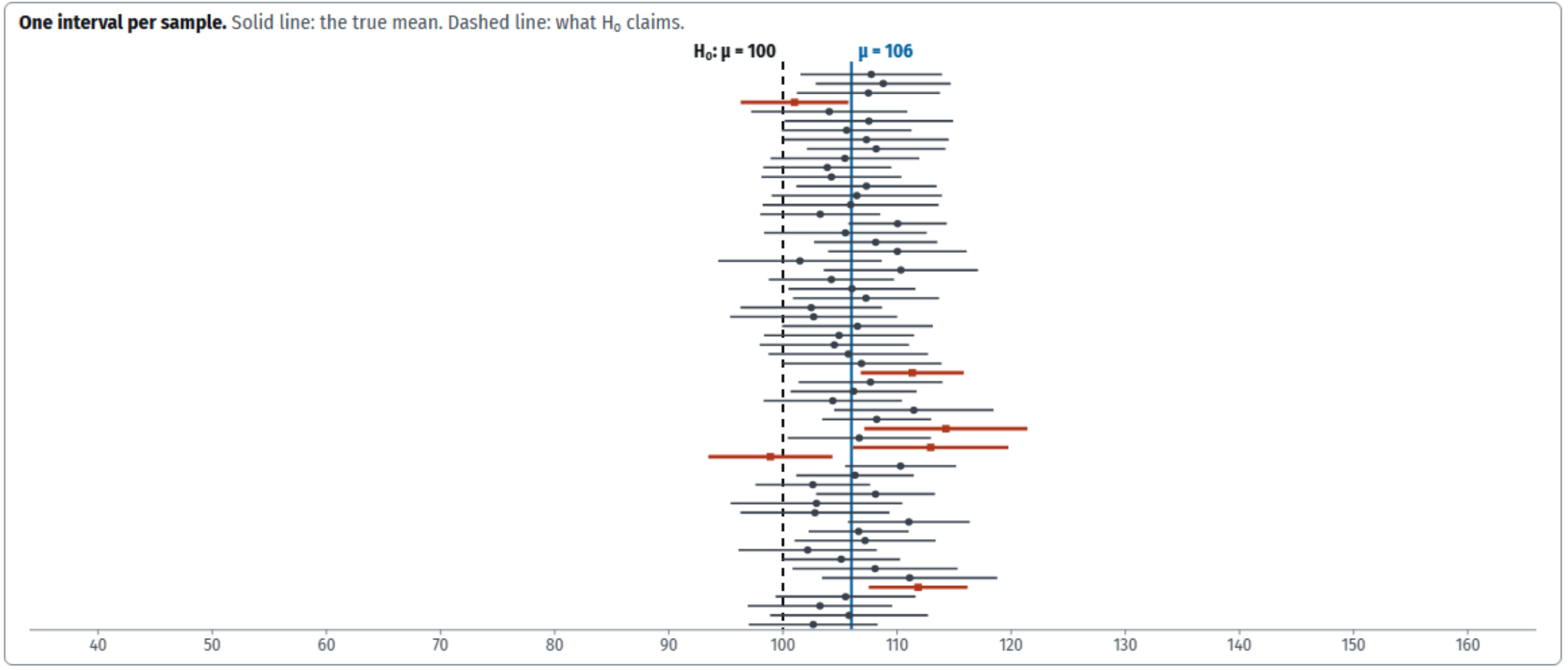
A 95% interval that leaves out the H_0 mean means rejecting H_0 at α = 0.05, two tails

Only one interval leaves out 100



The t interval is wider, for two reasons: a bigger $\hat{\sigma}$ and a bigger cutoff

With a real effect, intervals miss H_0



Simulator after Kristoffer Magnusson, rpsychologist.com.

Try it: The 24 students again

$n = 24$, $M = 12.21$, $\hat{\sigma} = 3.50$, $\hat{\sigma}_M = 0.715$, H_0: $\mu = 11.12$, cutoff $t(23) = 2.069$

Find g, then a 95% confidence interval

Does the interval hold 11.12? Does that match Thursday's decision? Work alone, compare with a neighbor, then we vote

Data: Muralidharan, Singh, and Ganimian (2019).

Try it: Answer

$g = (12.21 - 11.12) / 3.50 = \mathbf{0.31}$

$12.21 \pm 2.069 \times 0.715 = \mathbf{[10.73, 13.69]}$

The interval holds 11.12, so we maintain H_0, the same decision as Thursday

The interval also says what the test didn't: μ could be a bit below 11.12, or as much as 2.6 points above it

Data: Muralidharan, Singh, and Ganimian (2019).

One sample, three answers

The t-test

Is this mean surprising under H_0?

$t(23) = 1.52$

Effect size

How big is the gap, in SDs?

$g = 0.31$

The interval

Which values of μ fit the data?

$[10.73, 13.69]$

Report all three. Each one says something the others don't

Data: Muralidharan, Singh, and Ganimian (2019).

The blueprint so far

Blueprint step	Statistic	Bottom of the fraction	In the book
1. One score	z	σ	σ
2. One mean, σ known	z	σ_M	σ_M
3a. One mean, σ unknown	t	$\hat{\sigma}_M$	s_M
4. Two means: next	t	Topic 9	

Same three panels: population, one sample, the statistic across samples

What you can do now

- Choose one tail or two, and say what each costs
- Say why $\hat{\sigma}$ gives us t, not z, and why df matters
- Run a one-sample t-test by hand, from n, M and SS
- Compute Hedges' g, and tell significance from size
- Build a confidence interval, and use it to test H_0

The midterm: what to expect

Tuesday, February 16, in class, 80 minutes

It covers Chapters 1 to 9, which are Topics 1 to 8

It's free response: show your work and explain your thinking

Open book: bring any print materials, and your own notes

We attach tables and a formula sheet, in both sets of symbols

A calculator is the only device you can use

Next steps

Thursday, February 11: Midterm review

Practice problems in pairs. Bring a calculator and your questions

Friday, February 12: Lab 6

Quiz, as midterm review. Bring a calculator

Sunday, February 14 at 6 PM: Mastery and Problem Set for Chapter 9. The midterm is Tuesday, February 16