Psych 10: Intro to Statistics

Lecture 15 · Confidence intervals · Mon Nov 16, 2026

Prof. Judith Fan · Fall 2026

Reminders & Announcements

- **The exam is Wednesday, November 18**, in this room at the usual time. It covers CourseKata Chapters 1 to 12. You will have the formula sheet, which is on the Resources page now; bring a pen or pencil.
- Today's worksheet is the last one before the exam. There is no worksheet on Wednesday.
- **Friday, 1:30 to 2:50 pm**: TA office hours in your section's tutorial room, for your project; attendance credit is automatic.
- Section this week: **Milestone 8**, the report and the poster, due Sunday, November 22.
- CourseKata **Chapter 13** is assigned after Thanksgiving.

Today

Lecture 15: Confidence intervals

1. The sampling distribution of an estimate, and the standard error
2. The confidence interval: what it is, and what it is not
3. Confidence intervals and model comparison

A poll two weeks before the election

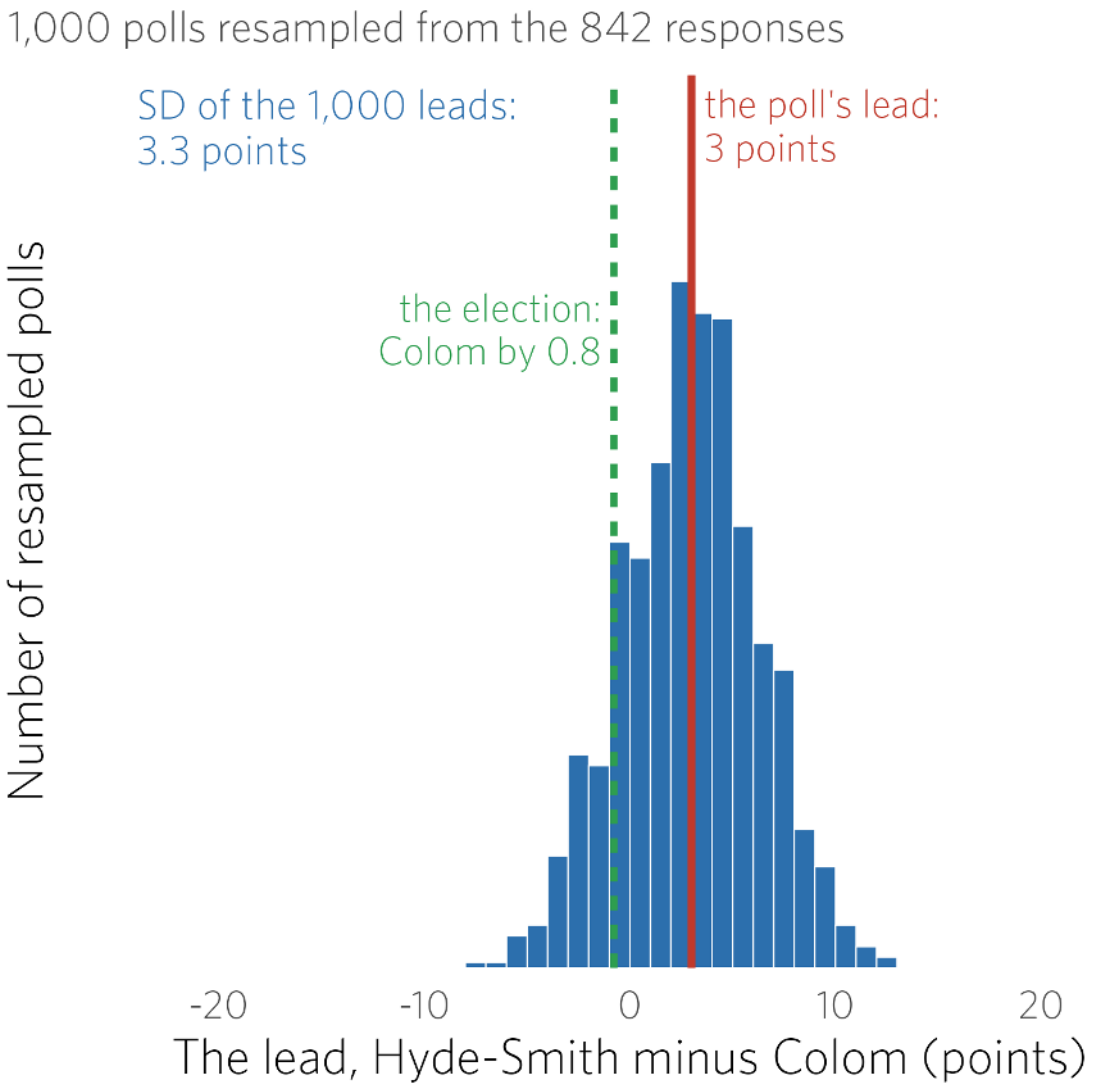
Mississippi's U.S. Senate race was decided on November 3, 2026, the closest Senate race of the year: Senator Cindy Hyde-Smith, the Republican incumbent since 2018, against Scott Colom, a Democrat and district attorney. Forecasters had rated the seat likely Republican. Colom won, and the race was called on November 11.

The last public poll before the vote, by Change Research (a Democratic-aligned firm), October 21 and 22, 2026, 842 likely voters:

Hyde-Smith	Colom	Another candidate	Undecided
43%	40%	6%	9%

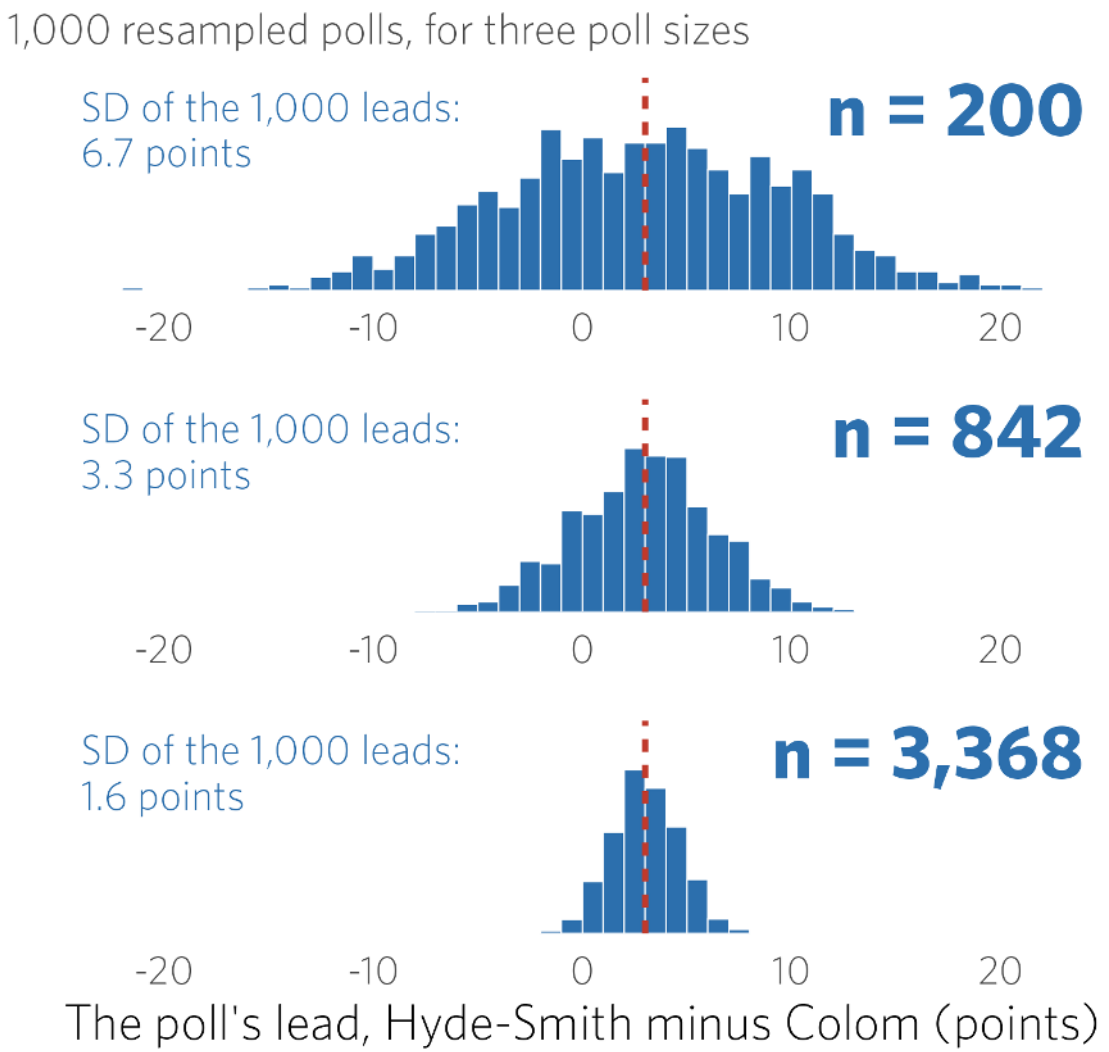
- The number to follow through the hour: the poll's **lead** for Hyde-Smith over Colom, 3 percentage points, as of October 22.
- About 970,000 Mississippians voted twelve days later. The poll reached 842 of those voters.

The same poll run a thousand times



- The pollster has one sample of 842 responses and no population. We treat those responses as a stand-in for the population of voters.
- **Bootstrap resampling:** draw 842 responses from the 842 at random, **with replacement**, so the new poll can contain one response several times and leave out other responses. Compute the lead. Repeat 1,000 times.
- The 1,000 leads approximate the **sampling distribution** of the lead. Its SD is the **standard error**: the typical distance between one poll's lead and the true lead.

Standard error against poll size



- A bigger poll gives a narrower sampling distribution: the standard error shrinks with the size of the sample.
- It shrinks with the square root of the sample size: four times the voters, half the standard error.
- Running more simulated polls adds bars to the same histogram; only the size of each poll changes its width.

A definition of the standard error

- The **standard error** of an estimate is the standard deviation of its sampling distribution: the typical distance between the estimate from one sample and the true value.
- For a sample mean: $SE = \frac{s}{\sqrt{n}}$, the SD of the data divided by the square root of the sample size.
- For a sample proportion p, such as the share of a poll's respondents for one candidate: $SE = \sqrt{\frac{p(1-p)}{n}}$, at most $0.5/\sqrt{n}$. Every form shrinks with the square root of n.

The standard error by resampling

- The same procedure gives the standard error of any estimate. For the species difference b_1 in the penguin dataset:

```
sdob1 <- do(1000) * b1(kg ~ species, data = resample(penguins))
sd(~ b1, data = sdob1)
```

- `resample(penguins)` draws 274 penguins from the 274 at random with replacement; `do(1000) *` fits the model to each resampled set and keeps its b_1; `sd()` takes the SD of the 1,000 values of b_1: the standard error.
- `resample()` keeps every case intact and mimics the process that produced the data; `shuffle()` deals the labels out at random and mimics a process with no effect, for the p-value.

Two spreads that are easy to confuse

	The SD of the data	**The standard error of an estimate**
What varies	one case's value, from case to case	the estimate, from sample to sample
The poll	one voter's answer, from voter to voter	the poll's lead, from poll to poll
The penguins	the masses, from penguin to penguin	the species difference b_1, from sample to sample
As the sample grows	stays about the same	shrinks, in proportion to 1 over the square root of n

- Both are standard deviations.
- The SD of the data measures how much the cases differ from one another.
- The standard error measures how much the estimate would change from one sample to the next, and so how uncertain the estimate is.

The standard error in R's output

- R computes the standard error from the formula and prints it beside each estimate, in the `Std. Error` column:

```
summary(species_model)$coef
                 Estimate  Std. Error
(Intercept)         3.701       0.039
speciesGentoo       1.375       0.058
```

- From a simulated sampling distribution, the formula sheet's `sd(~ b1, data = sdob1)` gives the same number: the SD of the simulated estimates.

Worksheet · Part 1

Two weeks before the election, Change Research polled 842 Mississippi voters about the U.S. Senate race between Senator Cindy Hyde-Smith (Republican) and Scott Colom (Democrat): Hyde-Smith 43%, Colom 40%. The estimate is Hyde-Smith's lead over Colom, 3 points, and its standard error is about 3 points.

a. The standard error is the standard deviation of a distribution. Which distribution: the individual responses in this poll, or the leads of many polls of this size? Write *responses* or *leads*, then one sentence saying what the standard error is a measure of.

b. Suppose the pollster had reached four times as many voters. Would the standard error be about the same, about half, about a quarter, or about double? Select one, then one sentence giving the reason.

c. A poll of the same size in California, a state with about ten times as many voters, would have a larger standard error. Is that right or wrong? Write *right* or *wrong*, then one sentence giving the reason.

Part 1a–c · Standard error and sample size

- **a.** *Responses or leads.* Leads: the standard error is the SD of the leads from many polls of 842, the typical distance between one poll's lead and the true lead.
- **b.** *Four times the voters.* About half. The standard error shrinks with the square root of n, and the square root of 4 is 2.
- **c.** *A state with ten times the voters.* Wrong. The standard error depends on the number of voters polled, not on the size of the state, as long as the poll is a small part of the state's voters.

Today

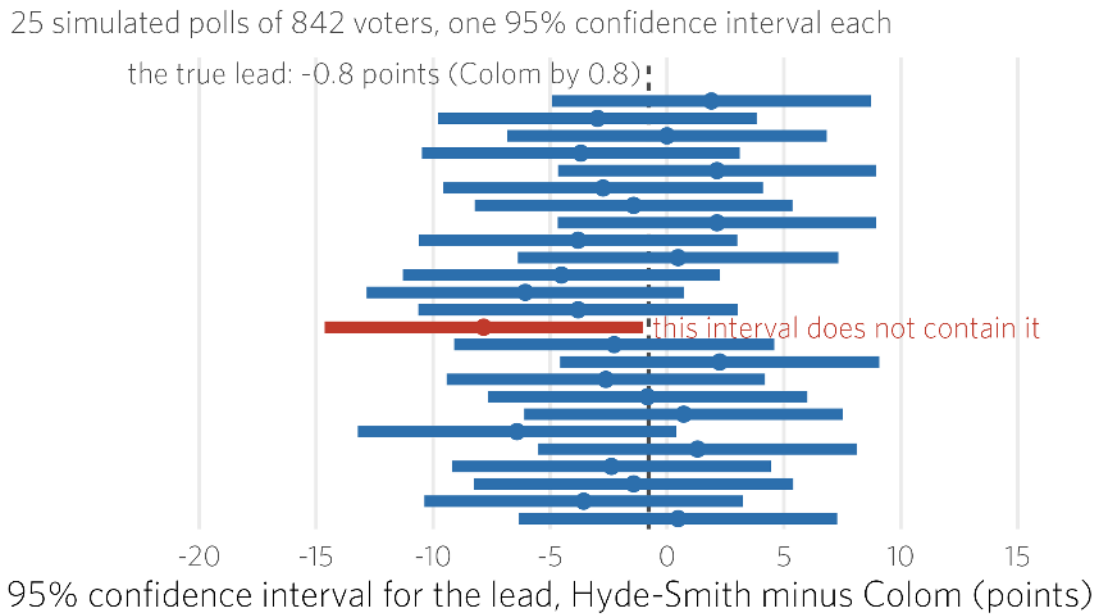
Lecture 15: Confidence intervals

1. The sampling distribution of an estimate, and the standard error
2. The confidence interval: what it is, and what it is not
3. Confidence intervals and model comparison

The best single guess and its likely error

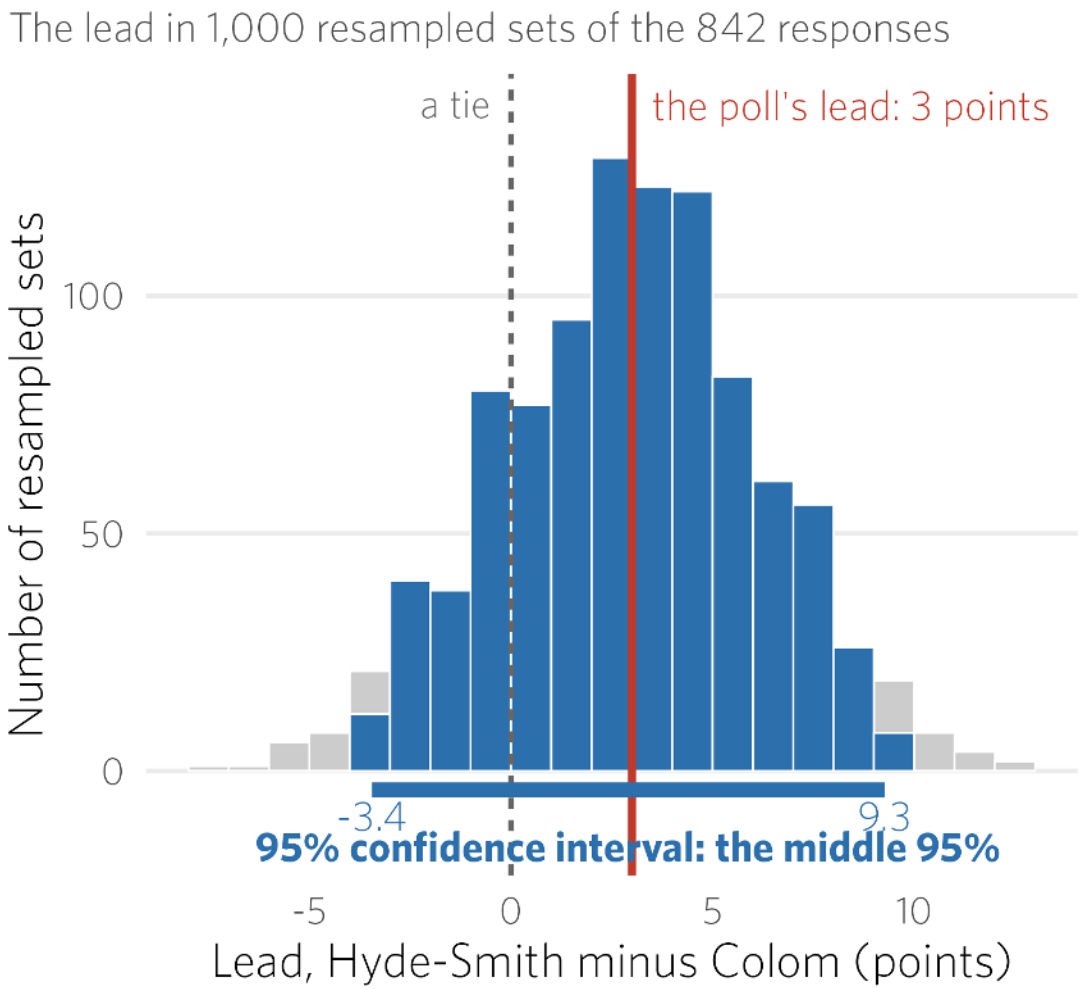
- The estimate is the best single guess at the true value: the truth is as likely to lie above the estimate as below the estimate.
- It is almost certainly not exactly right, and the standard error gives the typical size of that error.
- An estimate reported alone invites the reading that it is the truth. A statement of the estimate together with its uncertainty is called a **confidence interval**.

Confidence intervals



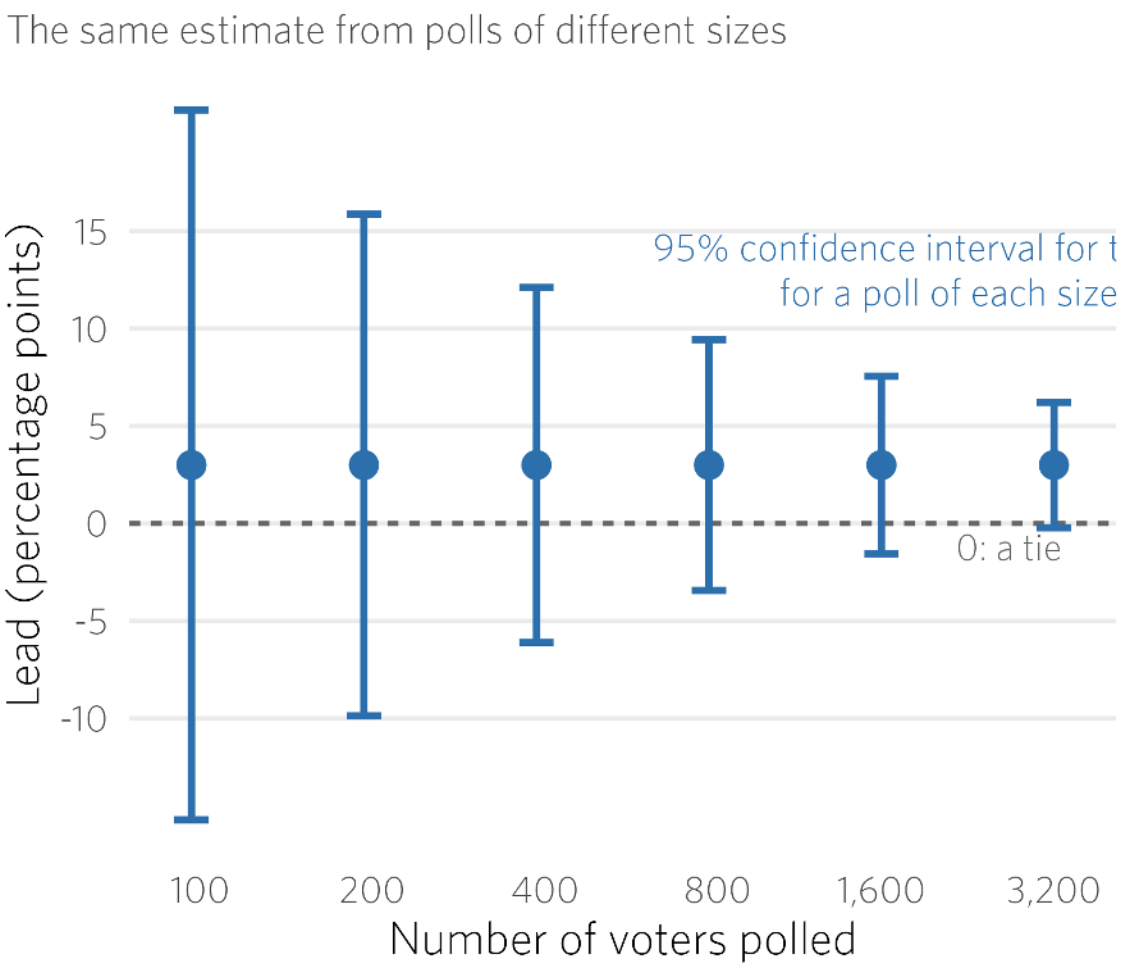
- A **confidence interval** is a range of values, with a lower bound and an upper bound, constructed to contain the true value of what was estimated a stated share of the time: for a **95% confidence interval**, 95% of the time.
- The 95% describes the method. In the figure, 25 polls were drawn from an electorate whose true lead is known, with a confidence interval constructed from each; in the long run, 95 in 100 such confidence intervals contain the true value.
- Any particular confidence interval either contains the true value or does not; with a real sample, which of the two is unknown.

Constructing a confidence interval by resampling



- Resample the data with replacement, many times, and compute the estimate each time: a sampling distribution of the estimate, built from the sample alone.
- The lower bound of the 95% confidence interval is the 2.5th percentile of that distribution, and the upper bound is the 97.5th percentile: the middle 95% of the estimates.
- The shortcut on the formula sheet, the estimate plus and minus two standard errors, gives nearly the same confidence interval when the sampling distribution is close to normal.

Confidence interval length depends on sample size



- Because the standard error decreases with sample size, the confidence interval narrows with the size of the sample.
- The narrowing slows: four times the sample gives half the width, the square root again. Beyond a few thousand cases, extra data make little difference to the width.
- For the poll, a few thousand voters would have moved 0 outside the confidence interval.

How a confidence interval is reported

The estimate first, then the confidence interval, with its level:

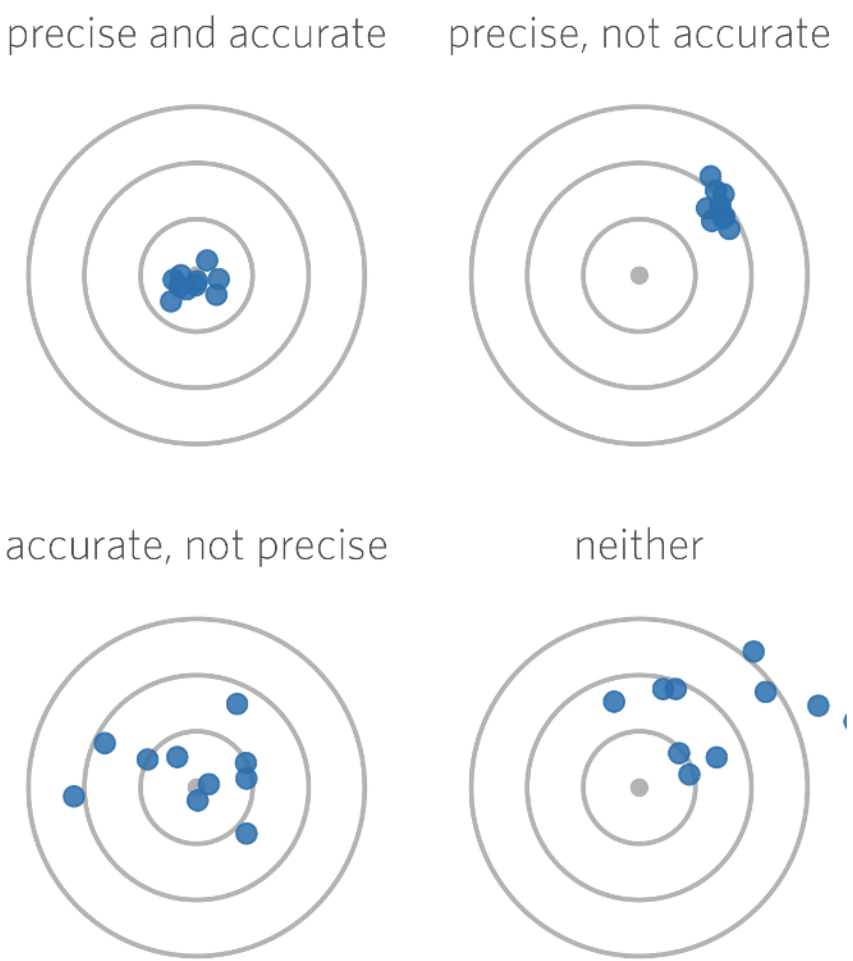
	Estimate	95% confidence interval
Mississippi poll, 842 voters	lead of 3 points	-3 to 9 points
Species, 274 penguins	b_1 = 1.38 kg	1.26 to 1.49 kg

- In a sentence: "Hyde-Smith led Colom 43% to 40% among 842 likely voters, a lead of 3 points, 95% CI [-3, 9]."
- In speech: "about 3 points, and the true lead is plausibly anywhere from about -3 to 9."

What a confidence interval is and is not

- It **is** a statement about one unknown number, the true value of what was estimated: the range of values for that number that remain plausible given the data.
- It is **not** a statement about individual cases. The confidence interval does not say that 95% of the cases fall inside it; the cases spread far more widely than the estimate.
- It is **not** a statement about future samples. The confidence interval does not predict the next sample's estimate.
- It is **not** a probability for this one confidence interval. The true value is fixed; the confidence interval either contains that value or misses it, and the 95% belongs to the method.

Accuracy and precision



- Each target shows ten estimates of the same quantity. The true value is at the center.
- **Precision** is the spread of the estimates around their own average. The standard error measures that spread; a larger sample gives a smaller spread.
- **Accuracy** is how close the center of the estimates lies to the true value. A systematic error, or bias, lowers accuracy, and a larger sample does not remove the bias.
- A confidence interval describes precision. It says nothing about accuracy: a narrow confidence interval can sit around the wrong value.

What the confidence interval leaves out

- A confidence interval is a statement about **precision**: how far the estimate might sit from the truth because of which cases landed in the sample.
- It says nothing about **accuracy**. A sample chosen from the wrong population, people who declined to answer, or a measure that captures the wrong quantity all shift the estimates away from the truth, and no confidence interval width reports that shift.
- A larger sample improves precision and narrows the confidence interval. It does not improve accuracy when the sample is biased.

Worksheet · Part 2

The poll's 95% confidence interval for Hyde-Smith's lead runs from about -3 to 9 points.

a. Each sentence below offers a reading of that confidence interval. Write *justified* or *not justified* beside each sentence's number.

1. The true lead is plausibly anywhere in this range, and values outside the range are implausible.
2. If the poll were repeated with new voters, the new lead would fall in this range 95% of the time.
3. There is a 95% probability that the true lead lies in this range.
4. Because the range includes 0, a tie remains plausible.

b. Suppose the poll had found a larger lead, with the same standard error. Would the confidence interval be wider, narrower, or the same width? Would it sit in the same place? One sentence.

c. A second pollster reports a confidence interval for the same race that is twice as wide. Give two different reasons that could explain the wider confidence interval.

Part 2a–c · Reading the confidence interval

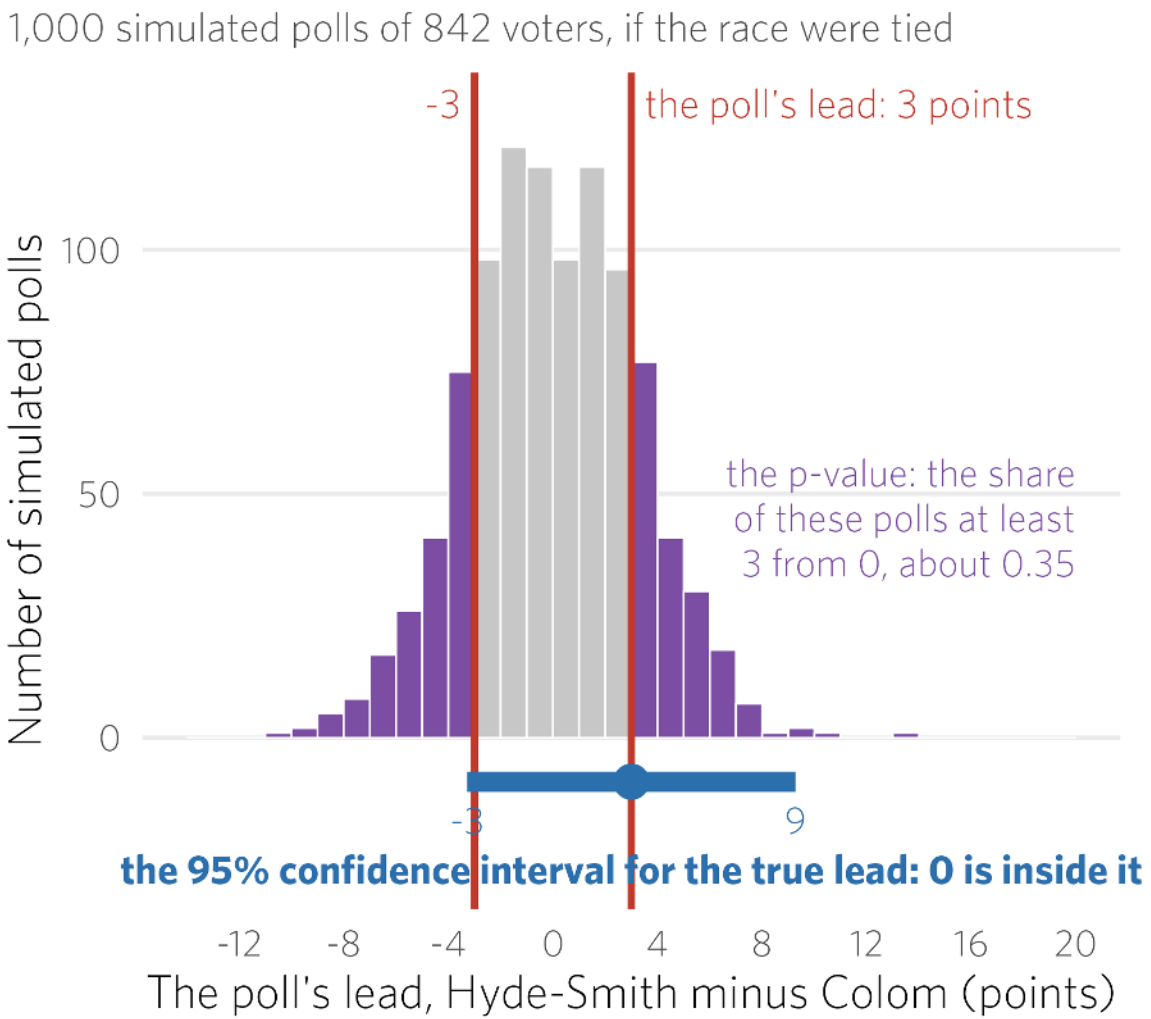
- **a.** *Four readings.* 1 justified: the definition. 2 not justified: the confidence interval is about the true lead, not about future polls. 3 not justified in the strict sense: the true lead is one fixed number, inside this confidence interval or not; the 95% describes the method. 4 justified: 0 is inside, so a tie remains plausible.
- **b.** *A larger lead, the same standard error.* The same width, because the width comes from the standard error; the confidence interval moves up with the estimate, and a large enough lead would leave 0 outside the confidence interval.
- **c.** *A confidence interval twice as wide.* A smaller poll, about a quarter the size, since the width shrinks with the square root of n; or a higher confidence level, such as 99%. Those are the two things that set a confidence interval's width.

Today

Lecture 15: Confidence intervals

1. The sampling distribution of an estimate, and the standard error
2. The confidence interval: what it is, and what it is not
3. Confidence intervals and model comparison

Confidence intervals and the empty model

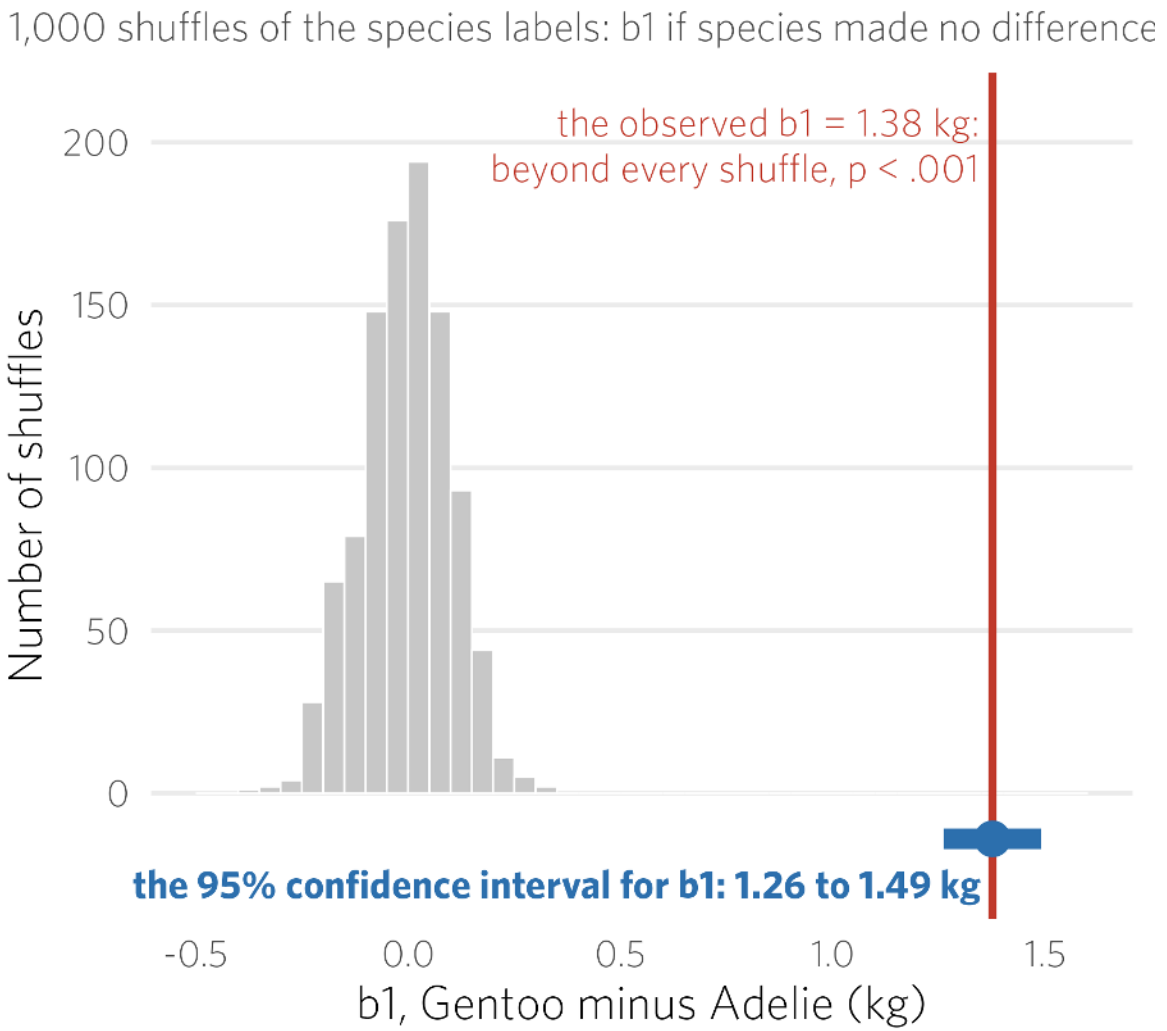


- For a poll, the empty model is a tie. We simulate polls of this size from a tied race; the **p-value** is the share of them at least as far from 0 as the poll's lead: about one in three, so the tie is not rejected.

- The **confidence interval** asks the same question from the other side: which true leads could have produced the poll's lead? The tie, 0, is among them.

- 0 inside the confidence interval and a p-value above .05 are the same result, read from the same sampling distribution.

The same logic in the ANOVA table



- The empty model says that $\beta_1 = 0$. The F test asks whether data like ours could have come from that model. The p-value is the answer.
- The confidence interval asks a broader question: which values of β_1 could have produced the observed b_1?
- Both answers come from the same sampling distribution, so they agree. When 0 lies outside the confidence interval, the p-value is below .05 and F is large. When 0 lies inside, the p-value is above .05.

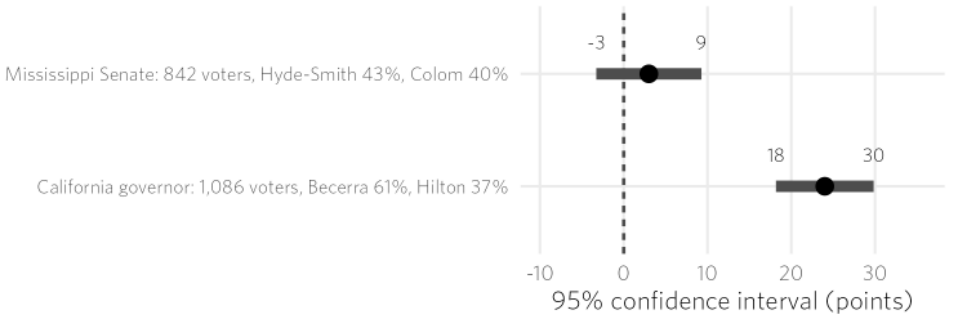
The effect size and our uncertainty about it

- The **effect size** is how large the effect is: the difference between two groups, or the slope of a line. The point estimate, b_1, is our single best estimate of the effect size.
- The confidence interval reports the effect size together with our uncertainty about that estimate: the point estimate at its center, and around it the range of effect sizes the data leave plausible.
- The hypothesis test answers a narrower question: could the effect be zero? A wide confidence interval that includes 0 leaves the effect size open, which is not the same as showing no effect. A narrow confidence interval far from 0 says that the effect is present and gives its approximate size.
- Report the confidence interval alongside the test statistic and the p-value.

What makes a confidence interval narrow

- A smaller standard error: more cases, since the standard error shrinks with the square root of n, or less spread in the outcome from case to case.
- A lower confidence level: a 90% confidence interval is narrower than a 95% confidence interval, at the price of missing the true value more often.
- Not more simulation runs. Repeating the simulation more times estimates the same sampling distribution more precisely; the spread of that distribution, and so the width of the confidence interval, depends on the sample size.

Worksheet · Part 3



The figure shows the 95% confidence interval for the leader's lead in two polls: the Mississippi Senate poll, and a California governor poll in which Xavier Becerra (Democrat) led Steve Hilton (Republican) by a wide margin.

a. For each poll, say in one sentence what its confidence interval tells you about the leader's true lead: how large it could plausibly be, and how small.

b. Suppose the Mississippi pollster had reached ten times as many voters and found the same lead. Write what you could then say about the true lead, hedged as carefully as the evidence warrants, and name one thing the bigger poll would not change.

c. A news report said that the Mississippi poll "showed Hyde-Smith ahead." Write a more carefully hedged sentence that says what the poll supports about who was ahead.

Part 3a–c · Careful conclusions from two polls

- **a.** *What each confidence interval says about the lead.* Mississippi: plausibly anywhere from a small Colom lead to a clear Hyde-Smith lead of about 9 points, so the size and even the direction are open. California: plausibly from about 18 to 30 points, large in any case.
- **b.** *Ten times the voters, the same lead.* The confidence interval would shrink to a few points around the same lead: Hyde-Smith most likely led, by somewhere between about 1 and 5 points. The estimate itself would not change, and nor would who answers polls or who votes.
- **c.** *"Showed Hyde-Smith ahead."* A hedged version: the poll's best estimate favored Hyde-Smith by a few points, but a small Colom lead remained plausible, so the poll could not tell who was ahead. The election went to Colom by less than a point.

Before Wednesday

- **The exam**: Wednesday, November 18, in this room. Chapters 1 to 12; the formula sheet is provided; bring a pen or pencil.
- The closest guide is the practice exam and its key on the Resources page, then the pre-practice questions, your worksheets, and the tutorials.
- Friday: TA office hours, 1:30 to 2:50 pm, for your project. **Milestone 8** is due Sunday, November 22.

Hand in your worksheet on the way out

On the last two lines, write down something new you learned today. Thanks!