Psych 10: Intro to Statistics

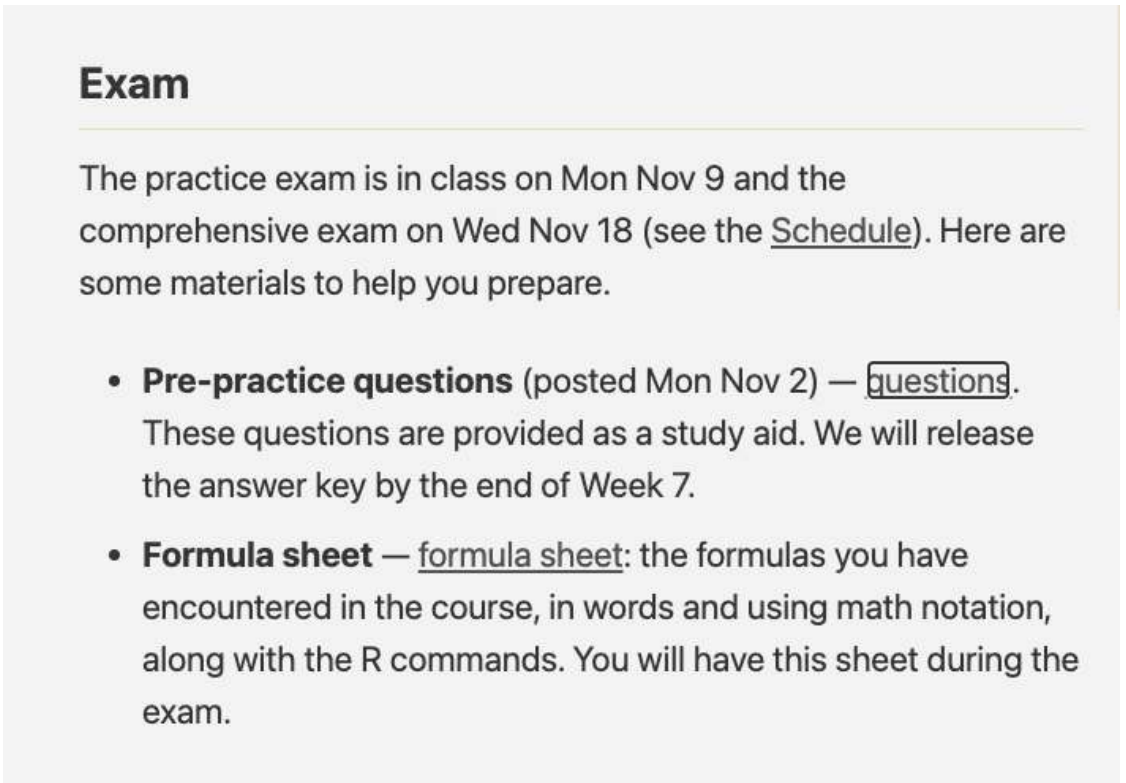
Lecture 12 · Every model so far, and the logic of inference · Mon Nov 2, 2026

Prof. Judith Fan · Fall 2026

Reminders & Announcements

- **Mid-quarter survey**: please complete it by end of day this Wednesday, November 4, for the 1% extra credit. The link is on the course website's syllabus page.
- **Review session** next Wednesday, November 11: please suggest topics with the form in your section's Slack channel, also by this Wednesday.
- CourseKata **Chapters 10 and 11** are due Friday.
- Section this week: **Milestone 6**, due Sunday, November 8.
- Friday: **Case #6**, in your tutorial session. Check your section's Slack channel for your time slot. Bring a pen.
- This Wednesday: **Ji Son, author of CourseKata,** will be visiting. Bring your laptops for a live coding activity. It will be awesome!
- The **practice exam** is Monday, November 9, in class. The exam is Wednesday, November 18.

Study materials for the exam are on the course website



Exam

The practice exam is in class on Mon Nov 9 and the comprehensive exam on Wed Nov 18 (see the Schedule). Here are some materials to help you prepare.

- **Pre-practice questions** (posted Mon Nov 2) — questions. These questions are provided as a study aid. We will release the answer key by the end of Week 7.
- **Formula sheet** — formula sheet: the formulas you have encountered in the course, in words and using math notation, along with the R commands. You will have this sheet during the exam.

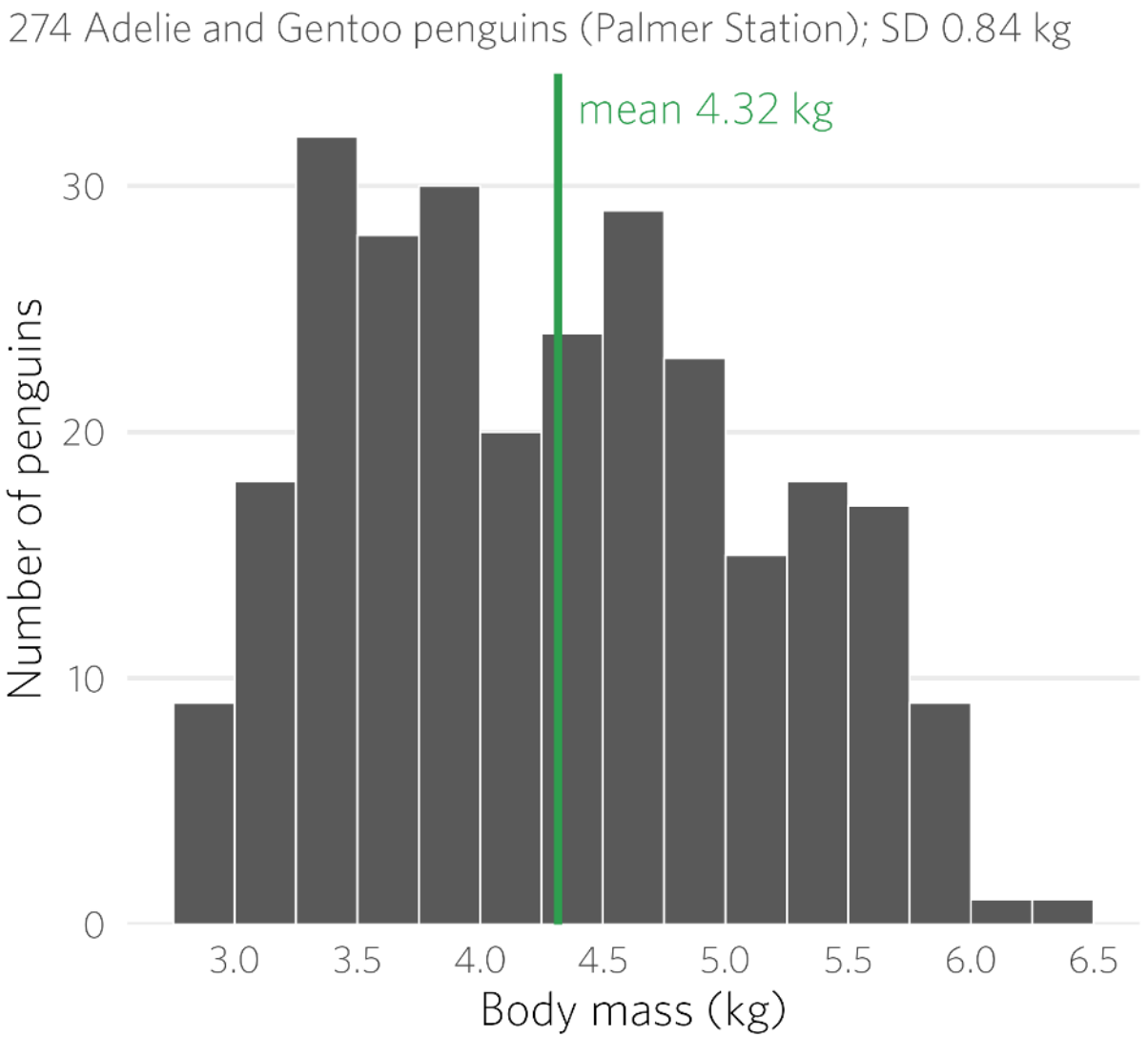
- I've posted **pre-practice exam study questions** on the Resources page of the course website, with the answer key to follow by the end of this week.
- Alongside them is a **preview of the formula sheet**: the formulas you have encountered in the course, in words and using math notation, with the R commands. You will have this sheet during the exam.

Today

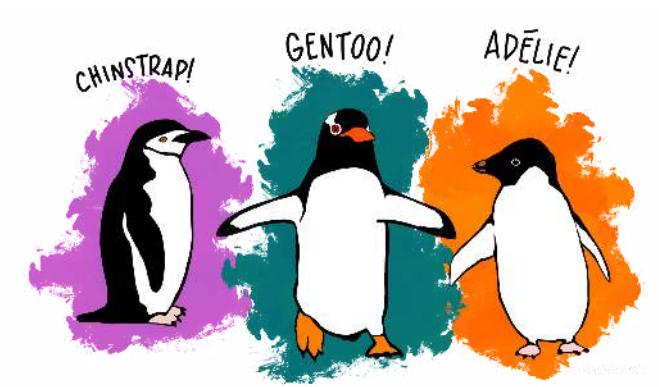
Every model so far, and the logic of inference

1. Every model so far on one dataset
2. From the sample to the process behind it
3. Could the empty model have produced these data?

274 penguins from Palmer Station



- Researchers at Palmer Station, Antarctica, weighed 274 Adelie and Gentoo penguins.
- One dataset for every model of the last four weeks.
- The empty model predicts the mean; a residual is mass minus the mean.



Gorman, Williams and Fraser (2014). Artwork: Allison Horst (CC0).

SS Error, the variance, and the SD

$$SS_{\text{Error}} = \sum (\text{mass} - 4.32)^2$$

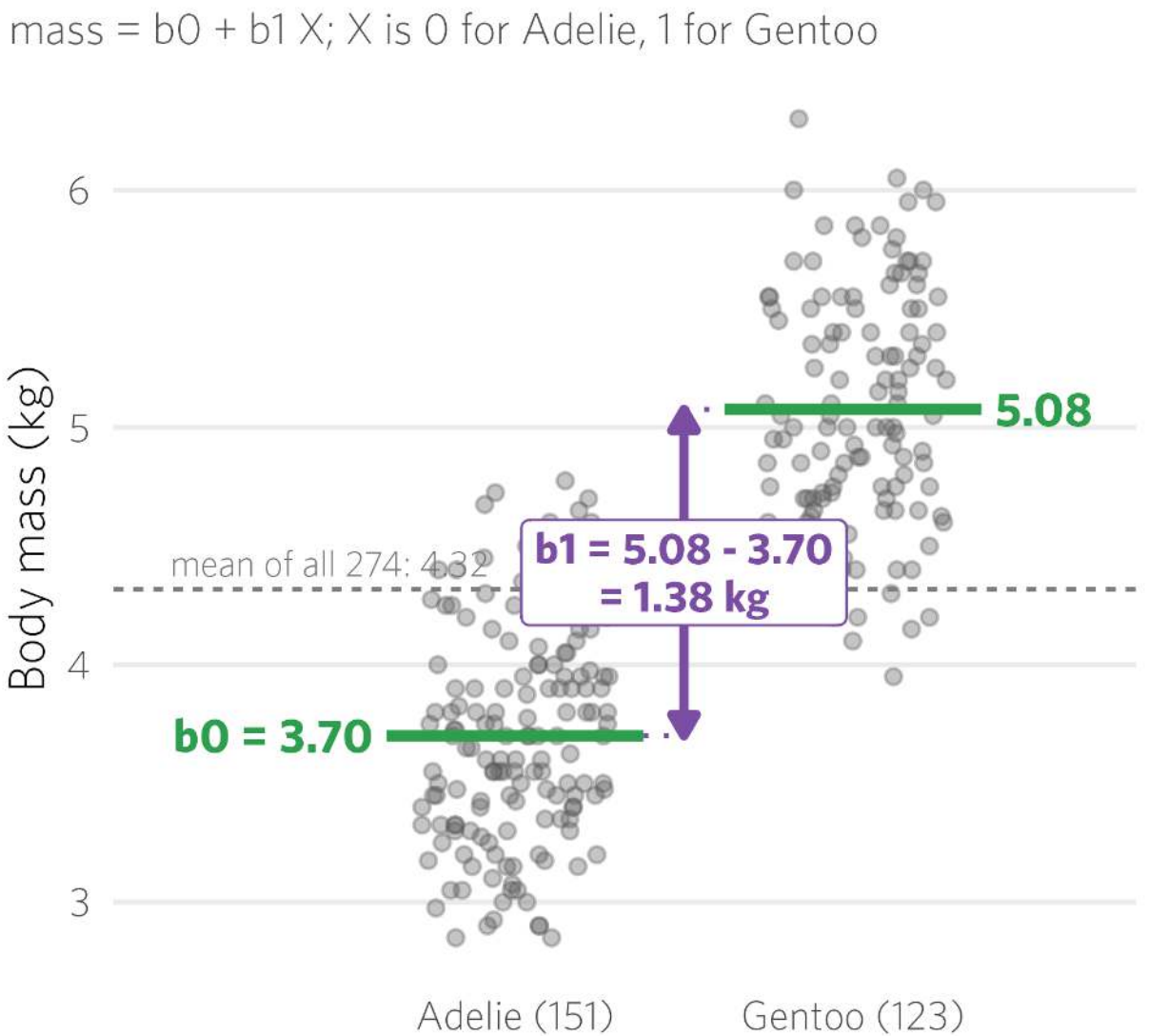
$$= 190.8 \text{ kg}^2$$

$$\text{variance} = \frac{SS_{\text{Error}}}{\text{df}} = \frac{190.8}{273} = 0.70$$

$$SD = \sqrt{0.70} = 0.84 \text{ kg}$$

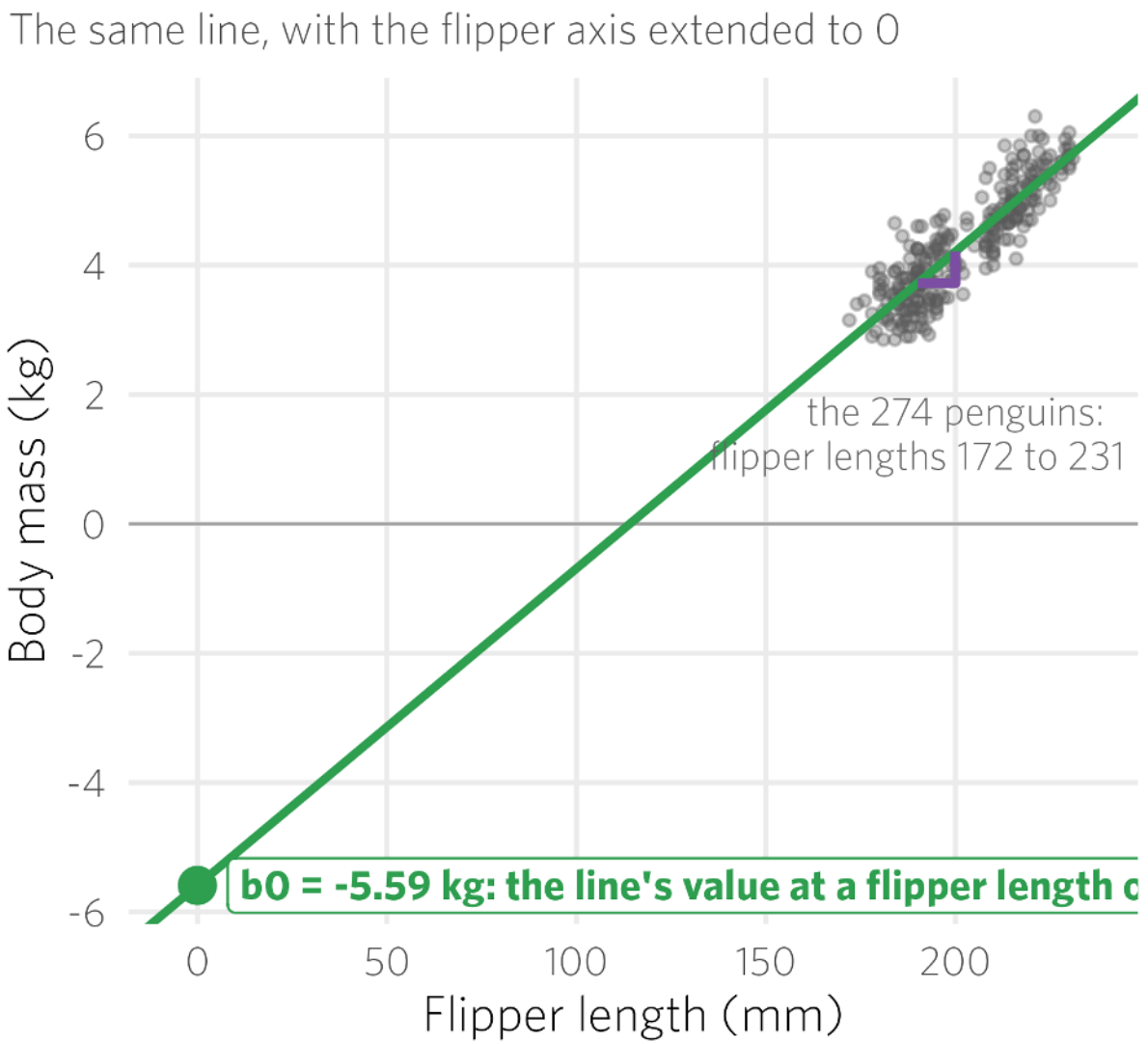
- Square each **residual**, mass minus the **mean**, and add the 274 squares: the empty model's total error, **SS Error**. For the empty model it is also SS Total.
- Divide by the **degrees of freedom**, 274 numbers minus the one mean they are measured from: the error per degree of freedom, the **variance**, in kg^2.
- Its square root is back in kilograms: the **SD**, the typical residual. A typical penguin is about 0.84 kg from the mean.

The group model: one mean per species



- A model with one mean per species has less error than the model with one mean for all.
- b_0 is the Adelie mean; b_1 is the Gentoo mean minus the Adelie mean.
- Residuals from the species means are smaller; PRE is about two thirds.

Predicting body mass from flipper length

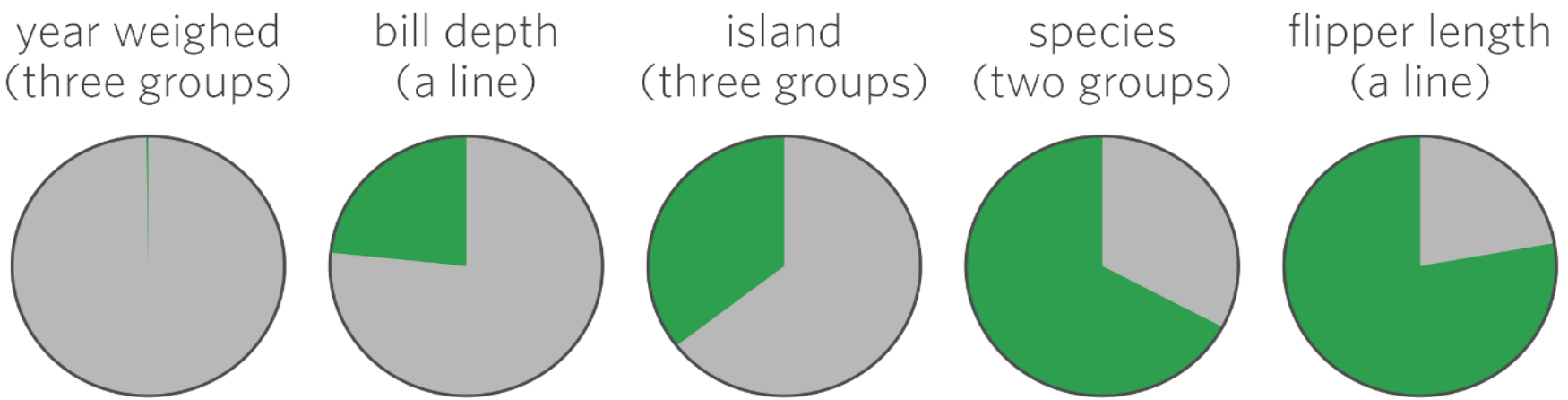


- Flipper length is a quantity, so the model is a line.
- b_1 is the slope: about 50 grams more mass per millimeter.
- Residuals are the vertical gaps to the line; PRE = r^2, about three quarters.

- b_0 is the line's value at a flipper length of 0, far outside the data.

A model's share of SS Total: PRE

Five models of body mass on the same 274 penguins. Green: SS Model. Grey: SS Error.



The ANOVA table for each model

The species model

	SS	df	MS	F	PRE
Model	128.2	1	128.2	558	0.67
Error	62.5	272	0.230		
Total	190.8	273			

The flipper-length model

	SS	df	MS	F	PRE
Model	148.5	1	148.5	955	0.78
Error	42.3	272	0.155		
Total	190.8	273			

- **df**: df Total = 274 - 1 = 273 (the masses minus the one mean); df Model = 1 (one per parameter added to the empty model, here b_1); df Error = 273 - 1 = 272.
- **MS** = **SS** / **df**: error per degree of freedom. The square root of MS Error is the typical residual, 0.48 kg under the species model.
- **F** = **MS Model** / **MS Error**: error removed per parameter, over error left per degree of freedom; near 1 for a parameter no better than chance.
- **PRE** = **SS Model** / **SS Total**: the share of the error removed.

Mean, residual, SS Error, variance, and SD

Quantity	What it is	Computed as	In R
the mean	the empty model's prediction for everyone	$\bar{Y} = \frac{\sum Y_i}{n}$	`mean(penguins$mass)` `empty_model <- lm(mass ~ NULL, data = penguins)`
a residual	actual minus predicted	$e_i = Y_i - \hat{Y_i}$	`resid(empty_model)`
SS Error	a model's total error	$SS_{\text{Error}} = \sum e_i^{\,2}$	`sum(resid(empty_model)^2)`
the variance	error per degree of freedom	$s^2 = \frac{SS_{\text{Error}}}{n-1}$	`var(penguins$mass)`
the SD	the typical residual	$s = \sqrt{s^2}$	`sd(penguins$mass)`

Two-parameter models and PRE

Quantity	What it is	Computed as
the group model	one mean per group	$Y_i = b_0 + b_1 X_i + e_i$, with X_i = 0 or 1 b_0 is the mean of group 0; b_1 is the difference between the two group means
the regression model	a line	$Y_i = b_0 + b_1 X_i + e_i$, with X_i a quantity b_0 is the line's value at $X = 0$; b_1 is its slope
PRE	the share of the error removed	$PRE = \frac{SS_{\text{Model}}}{SS_{\text{Total}}}$, where $SS_{\text{Model}} = SS_{\text{Total}} - SS_{\text{Error}}$

Degrees of freedom, MS, and F

Quantity	What it is	Computed as
degrees of freedom	three numbers, not one: the values free to vary in the total, in the model, and in the error	$df_{\text{Total}} = n - 1$ $df_{\text{Model}} = \text{parameters} - 1$ $df_{\text{Error}} = n - \text{parameters}$
MS	error per degree of freedom	$MS_{\text{Model}} = \frac{SS_{\text{Model}}}{df_{\text{Model}}}$ $MS_{\text{Error}} = \frac{SS_{\text{Error}}}{df_{\text{Error}}}$
F	error removed per parameter, over error left per degree of freedom	$F = \frac{MS_{\text{Model}}}{MS_{\text{Error}}}$

Worksheet · Part 1

The two ANOVA tables below are for two models of the same outcome, the body masses of the 274 penguins: one with species as the explanatory variable, the other with flipper length. The questions ask for comparisons between the tables, not for arithmetic.

Model: species

	SS	df	MS	F	PRE
Model	128.2	1	128.2	558	0.67
Error	62.5	272	0.230		
Total	190.8	273			

Model: flipper length

	SS	df	MS	F	PRE
Model	148.5	1	148.5	955	0.78
Error	42.3	272	0.155		
Total	190.8	273			

a. SS Total is 190.8 in both tables. In one sentence, say why the two models have the same SS Total.

b. Which model is the better fit to the data? Write its name, then two quantities in the tables that show the better fit, with their values for both models.

c. Species is categorical and flipper length is quantitative, yet df Model is 1 in both tables. In one sentence, say what the 1 counts in each table.

Worksheet · Part 1, continued

d. For each statement, write *true* or *false*, then a few words giving your reason.

1. The flipper-length model removes more of SS Total than the species model.
2. The two models have the same degrees of freedom, so the one with the smaller SS Error has the larger F.
3. Fit to 548 penguins instead of 274, with the same PRE, the species model would have about the same F.
4. Together the two models account for 0.67 + 0.78 = 1.45 of SS Total.

Formulas for Part 1

$$\text{mean} = \frac{\sum \text{values}}{n}$$

$$SS = \sum (\text{deviation})^2$$

$$\text{variance} = \frac{SS_{\text{Total}}}{n - 1}$$

$$SD = \sqrt{\text{variance}}$$

$$b_1 = \text{mean}_{\text{Gentoo}} - \text{mean}_{\text{Adelie}}$$

$$SS_{\text{Model}} = SS_{\text{Total}} - SS_{\text{Error}}$$

$$PRE = \frac{SS_{\text{Model}}}{SS_{\text{Total}}}$$

$$\text{df}_{\text{Model}} = \text{parameters} - 1$$

$$\text{df}_{\text{Error}} = n - \text{parameters}$$

$$MS = \frac{SS}{\text{df}}$$

$$F = \frac{MS_{\text{Model}}}{MS_{\text{Error}}}$$

n is the number of penguins, 274 here; a residual is actual minus predicted.

Part 1a–d · Two models, one SS Total

- **a.** *Why SS Total is the same in both tables.* **SS Total** is the empty model's error, the 274 masses' squared deviations from their mean. It depends on the outcome alone, so every model of these masses starts from 190.8.
- **b.** *Which model fits better.* The flipper-length model: **SS Error** 42.3 against 62.5, and **PRE** 0.78 against 0.67. **F**, 955 against 558, shows it too, because the two models have the same degrees of freedom.
- **c.** *What df Model = 1 counts.* The one parameter added to the empty model, b_1: a slope for a quantity, or a difference between two group means. **df Model** counts parameters, not the kind of predictor.
- **d.** *True or false.* 1 true and 2 true. 3 false: with 548 penguins and the same PRE, df Error doubles and F about doubles, to about 1,100. 4 false: the PREs are shares of one SS Total and overlap, since species and flipper length carry much of the same information; shares of a whole cannot sum past 1.

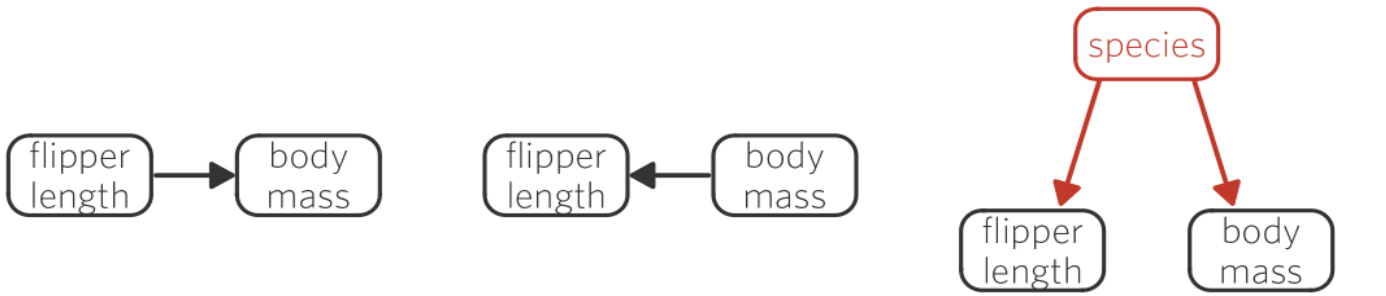
What a parameter is

- A **parameter** is a number in a model that is estimated from the data.

Model	Its parameters	Number	df Model
the empty model	b_0, the mean of all 274 masses	1	0
the species model	b_0, the Adelie mean; b_1, the Gentoo mean minus the Adelie mean	2	1
the flipper-length model	b_0, the line's value at a flipper length of 0; b_1, its slope	2	1

- Each parameter estimated from the data uses one degree of freedom. df Model counts the parameters beyond the empty model's one.
- b_0 and b_1 are the estimates from this sample. The corresponding values in the process behind the data are written β_0 and β_1.

Three hypotheses about what causes what



hypothesis 1: flipper length affects body mass

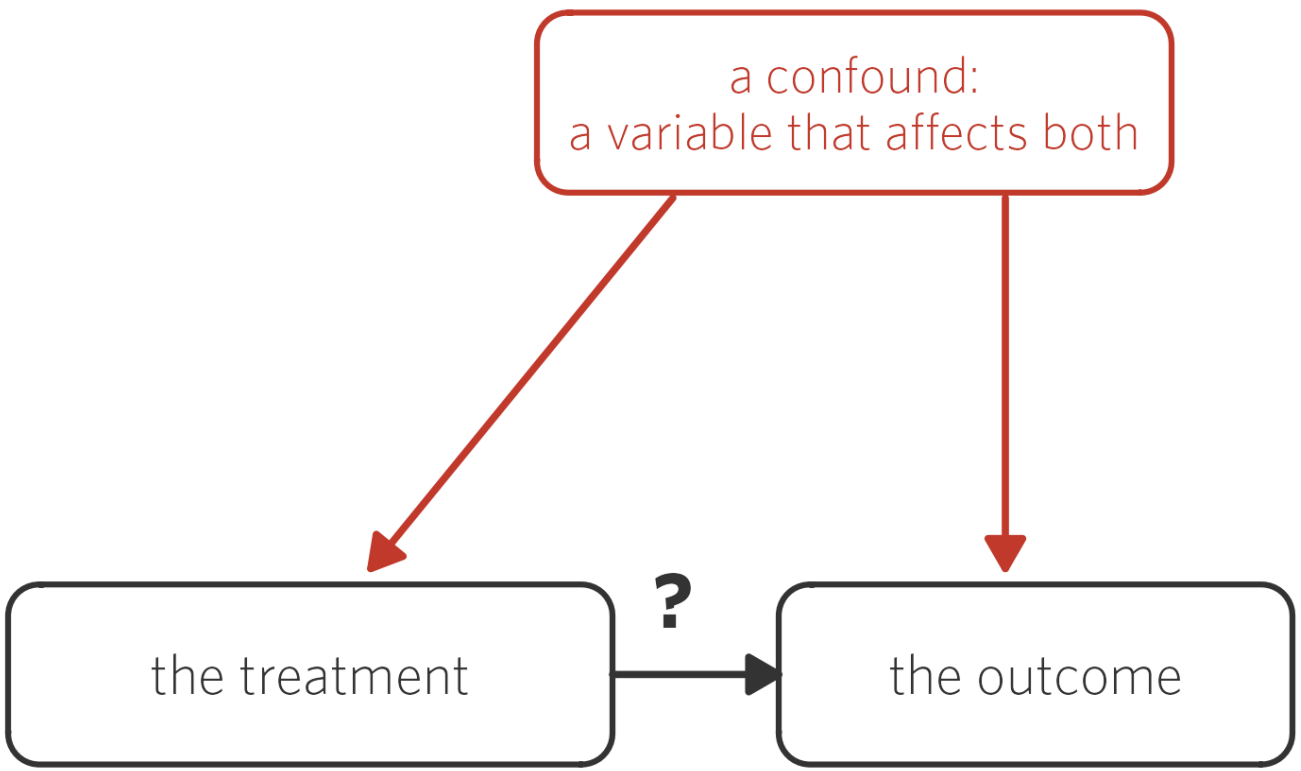
hypothesis 2: body mass affects flipper length

hypothesis 3: species affects both, a confound

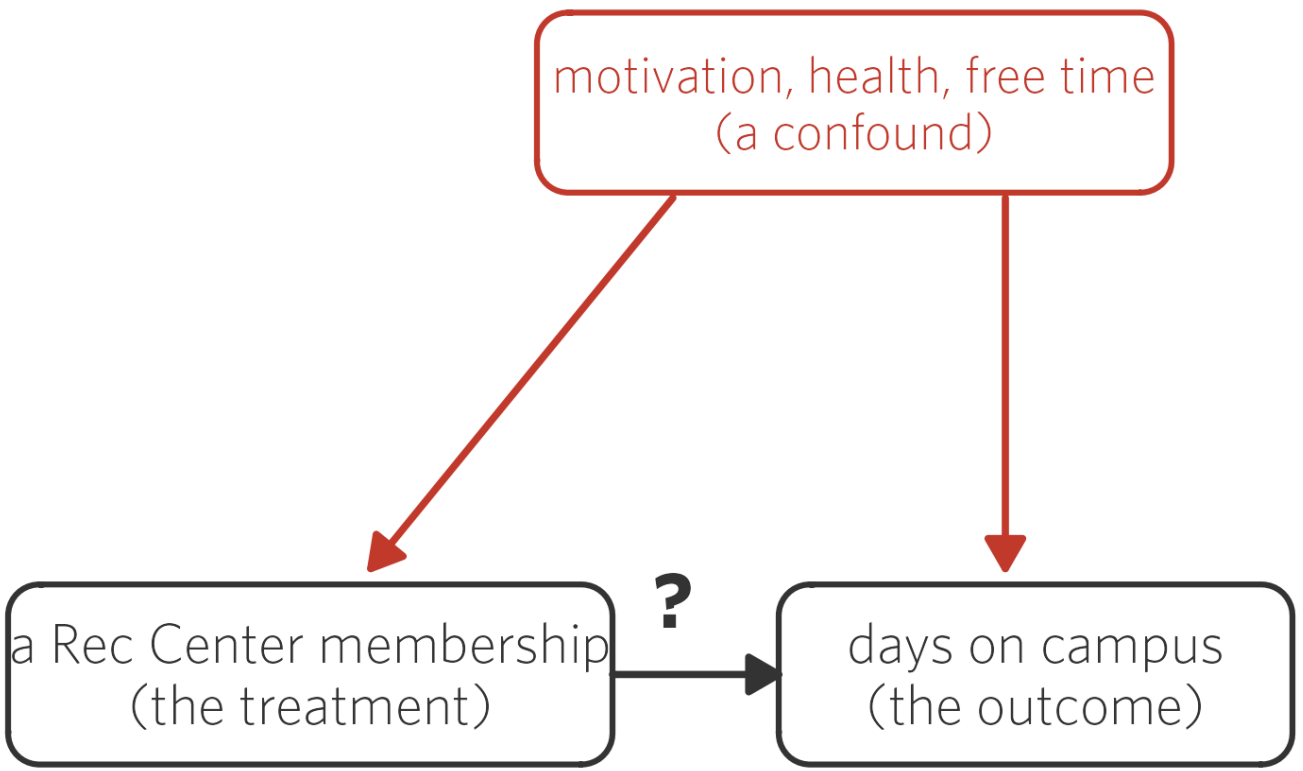
- The flipper-length model predicts body mass well, but a good prediction does not identify the cause. Each diagram is a hypothesis, not a finding: an arrow means that one variable affects the other, and all three are consistent with the same scatterplot.

- Under the third, species affects both variables and neither affects the other, so the two are correlated without either affecting the other. A variable that affects both predictor and outcome is a **confound**.

A confound affects both the treatment and the outcome

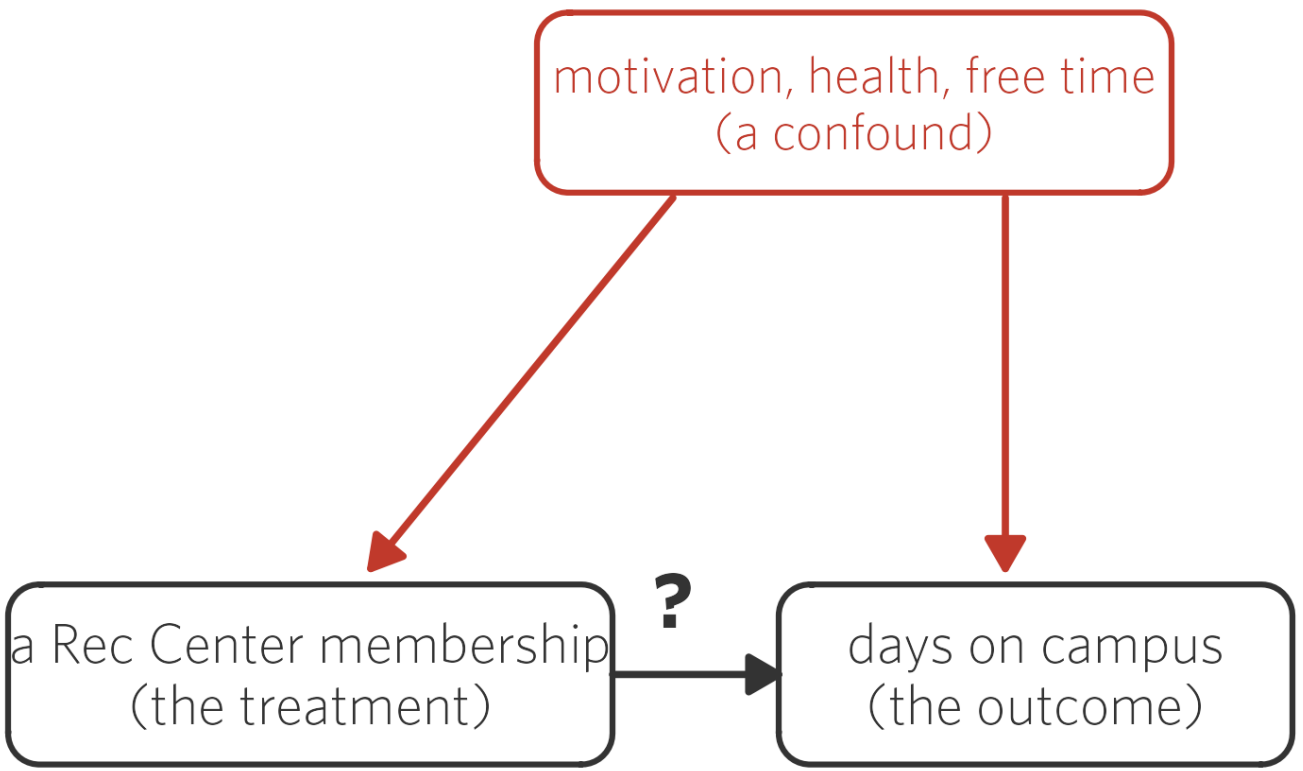


An example: does a Rec Center membership increase days on campus?



Friday's case: the Rec Center had 150 memberships for the 400 staff who signed up; the outcome is each person's days on campus per week

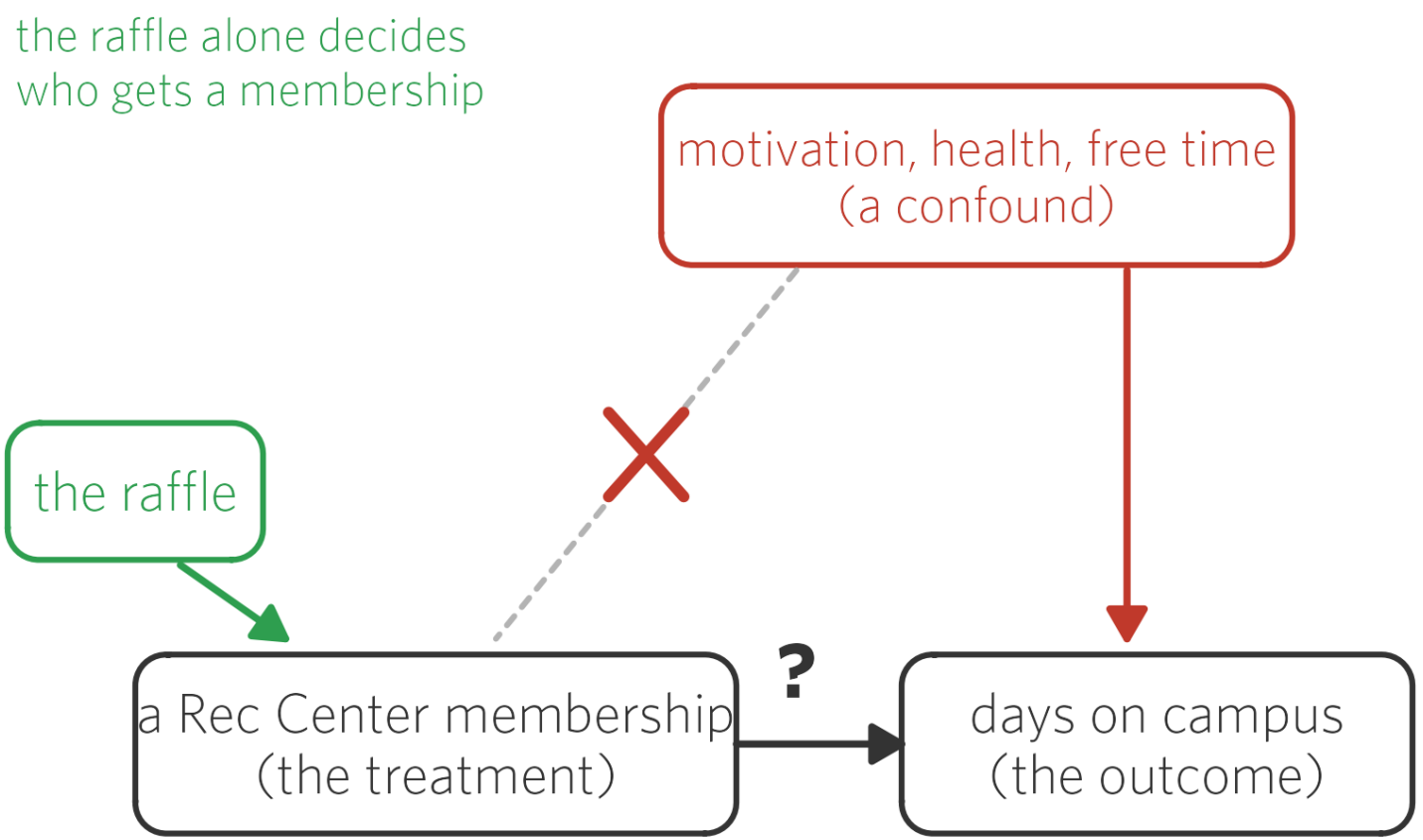
If each person chose whether to join, the confound would affect both



Friday's case: the Rec Center had 150 memberships for the 400 staff who signed up; the outcome is each person's days on campus per week

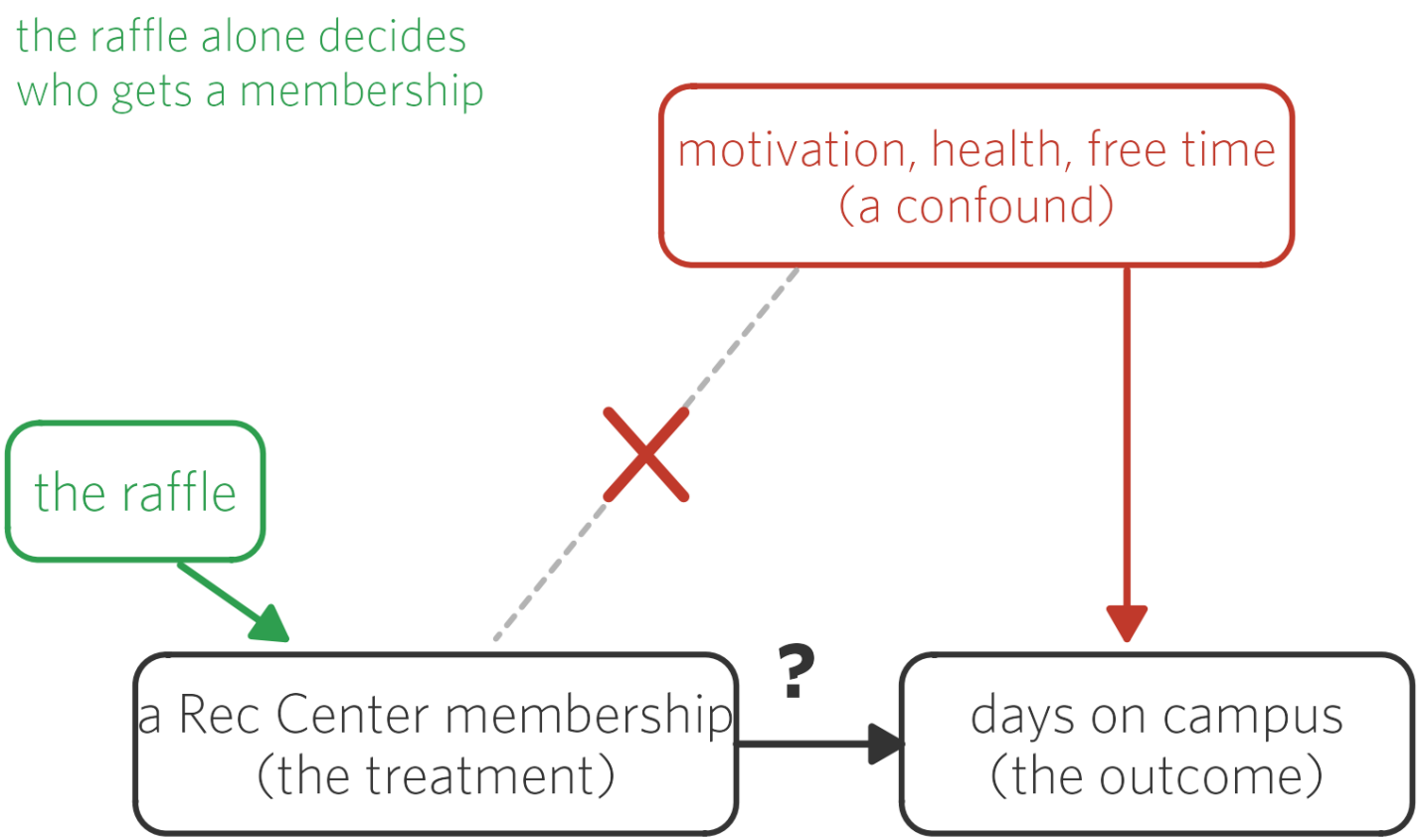
The raffle alone decided who got a membership

So the confound no longer affects who gets the treatment



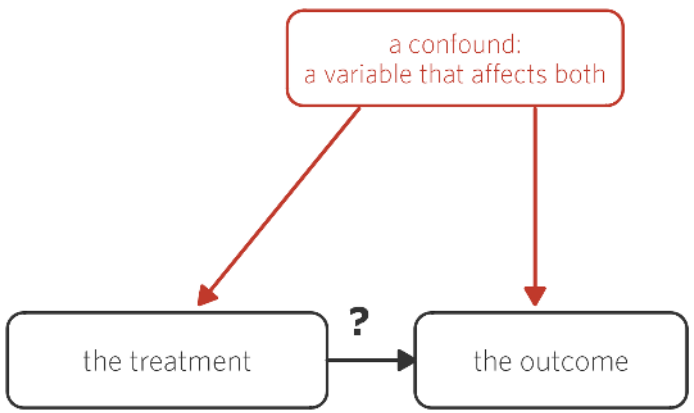
Friday's case: the Rec Center had 150 memberships for the 400 staff who signed up; the outcome is each person's days on campus per week

Two explanations remain for a difference: the membership, or chance



a difference in days on campus between the two groups
has two possible sources: the membership, or chance in the raffle

When random assignment would not be ethical



- Some assignments **would harm people**: to smoke, to lose sleep, to go without a treatment that works. No knowledge gained would justify that harm.
- Research with people follows the **Belmont Report's** three principles: **respect for persons** (informed consent), **beneficence** (benefits outweigh risks), and **justice** (burdens and benefits fairly shared). An **institutional review board** applies them before any data are collected.
- Where assignment would be unethical, researchers study **people who chose for themselves**. The confound remains, and the study shows **an association only**.
- Where the treatment is a benefit in short supply, **a lottery among those who asked for it** is both fair and a randomized experiment.

Today

Every model so far, and the logic of inference

1. Every model so far on one dataset
2. From the sample to the process behind it
3. Could the empty model have produced these data?

The sample's mean and the population's mean

Every number so far describes the 274 penguins that were weighed. The researchers' question is about Adelie and Gentoo penguins in general.

The sample: the 274 penguins weighed at Palmer Station. Its mean is a number within reach:

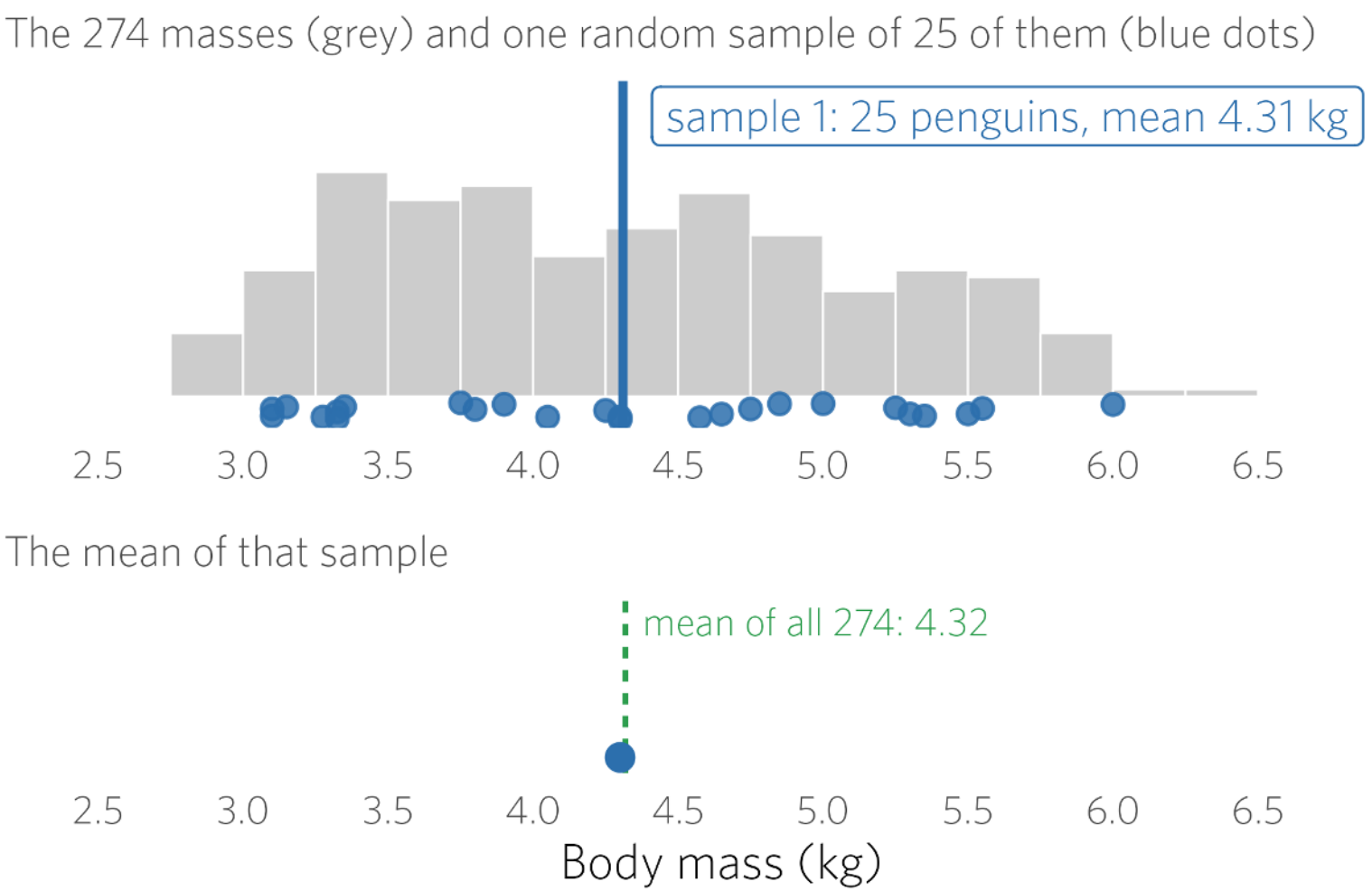
$$\bar{X} = \frac{\sum x_i}{n} = 4.32 \text{ kg}$$

The population: every Adelie and Gentoo penguin. Its mean is the number we want, beyond reach:

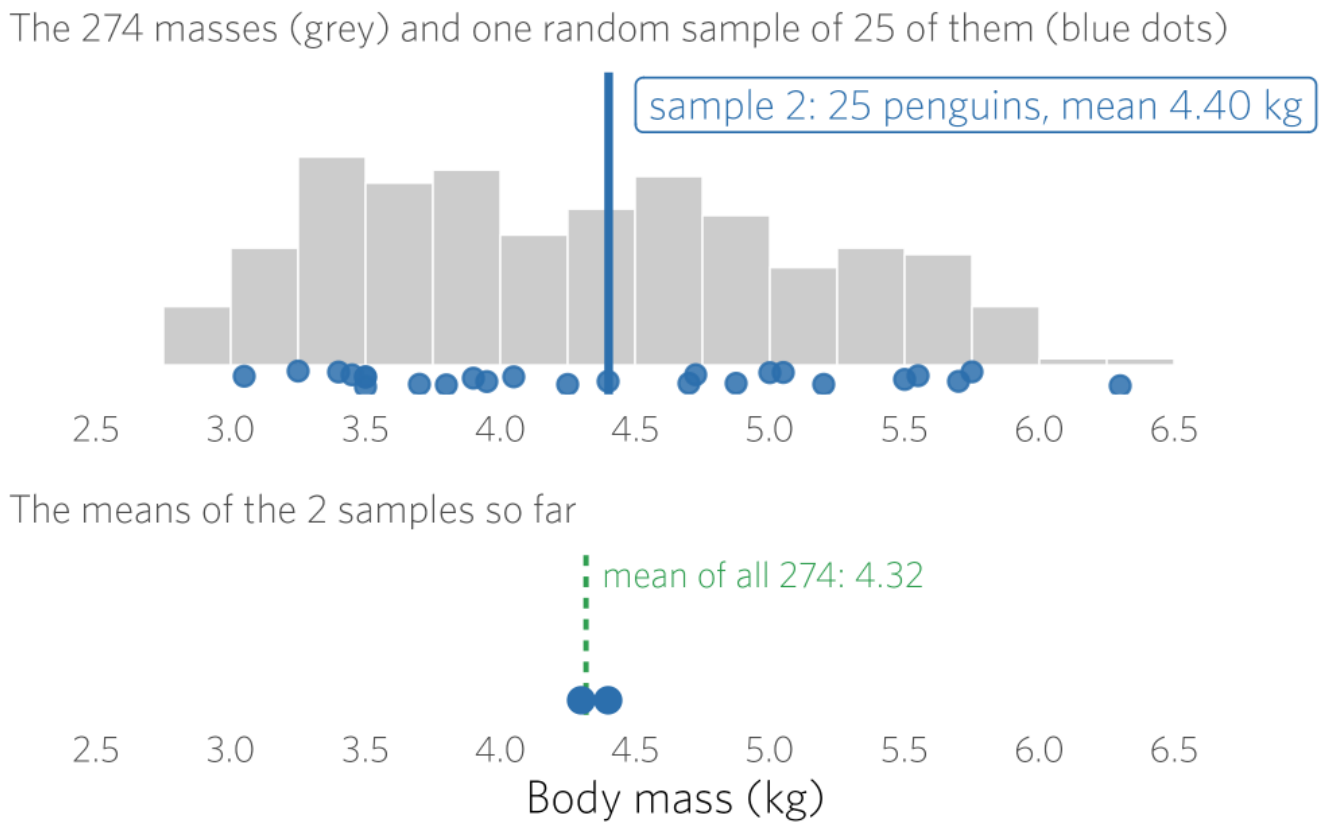
$$\mu = \frac{\sum x_i}{N}$$

- Using the sample's mean to learn about the population's mean is **statistical inference**. CourseKata calls the population the data generating process. The same reasoning carries b_1 to β_1.

One random sample of 25 penguins has one mean

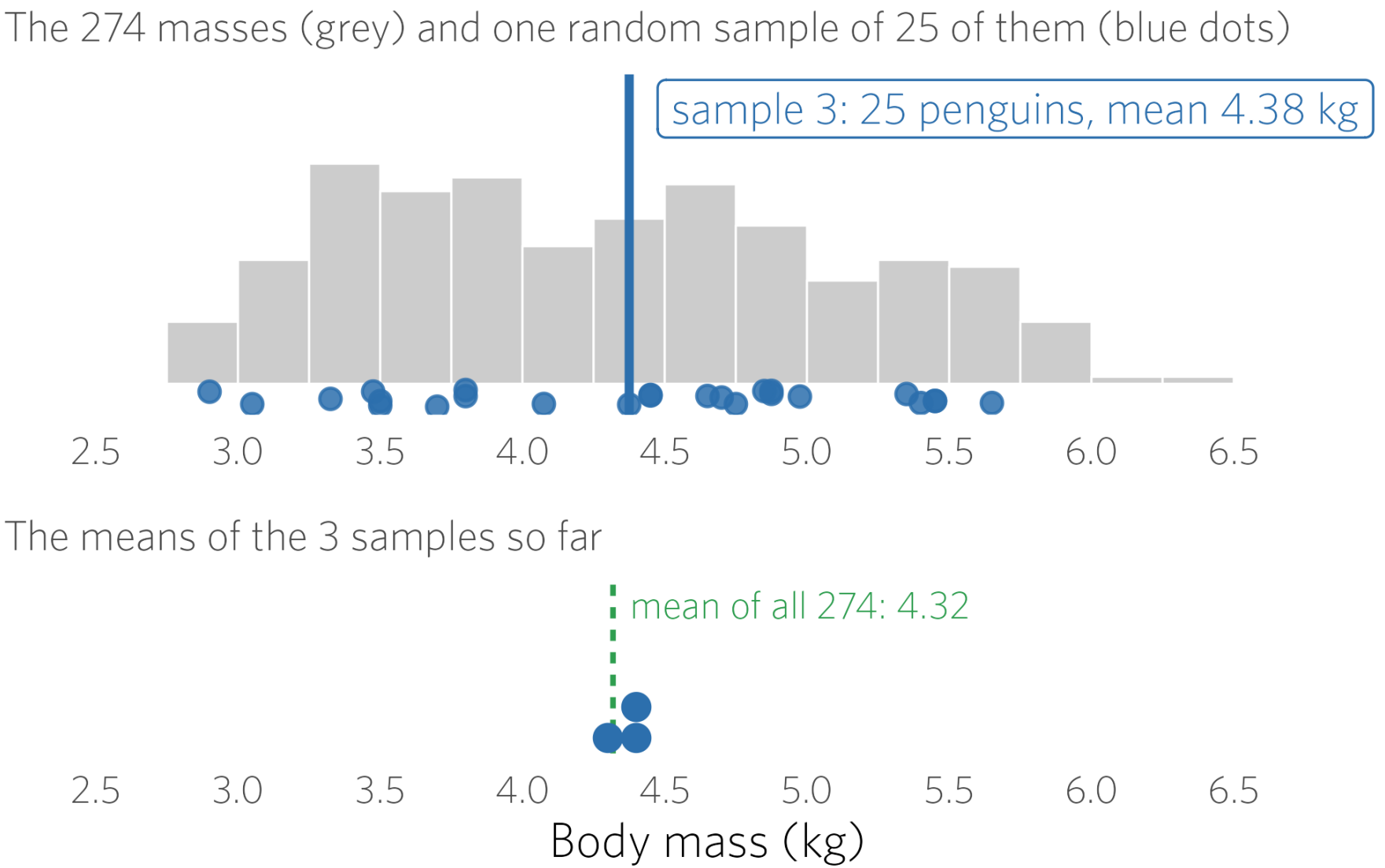


A second sample of 25 has a different mean

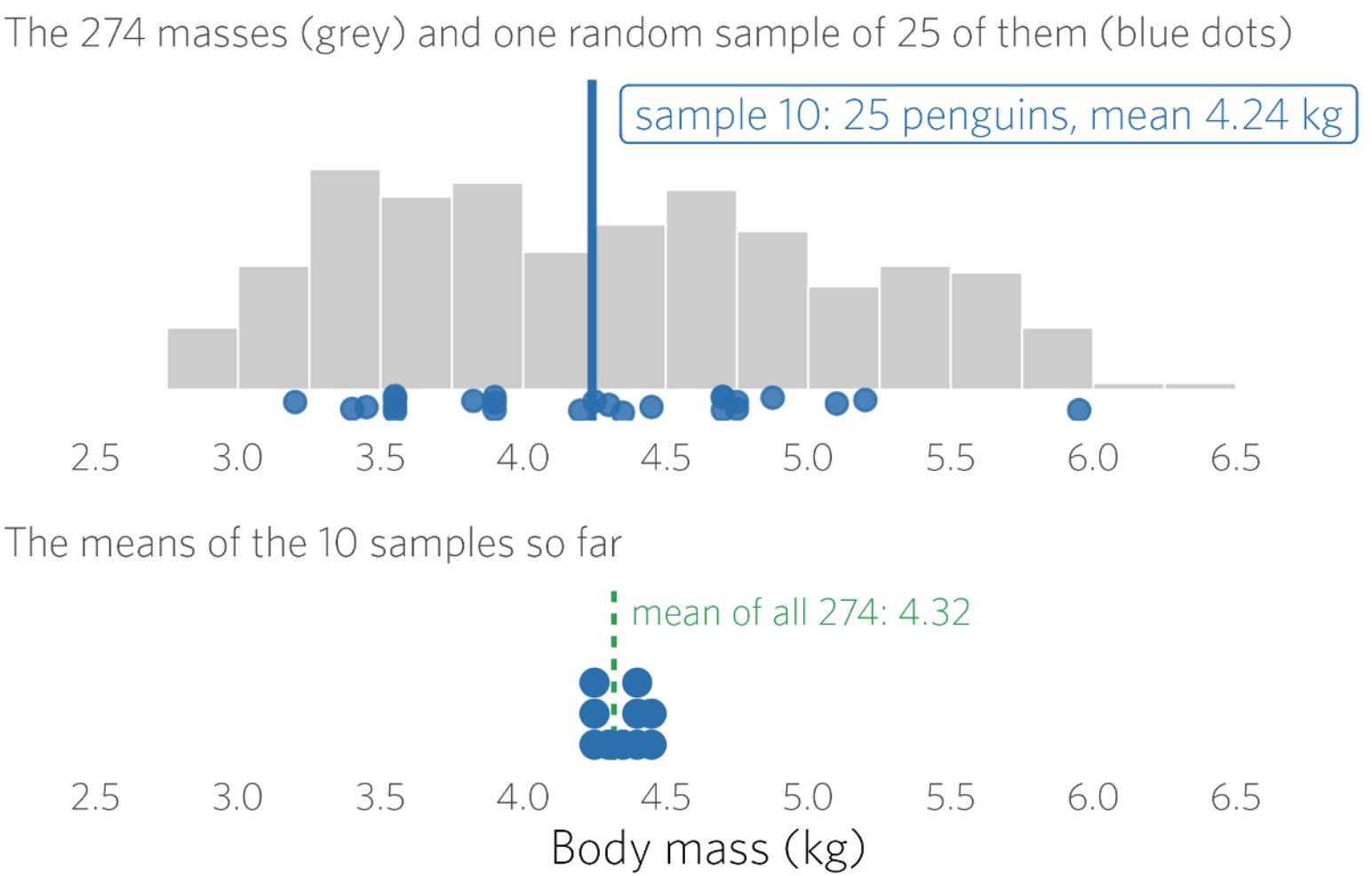


Sampling variation: a second random sample draws different penguins, so its 25 masses and their mean will almost certainly differ from the first sample's. Getting exactly the same mean again is possible, but very unlikely.

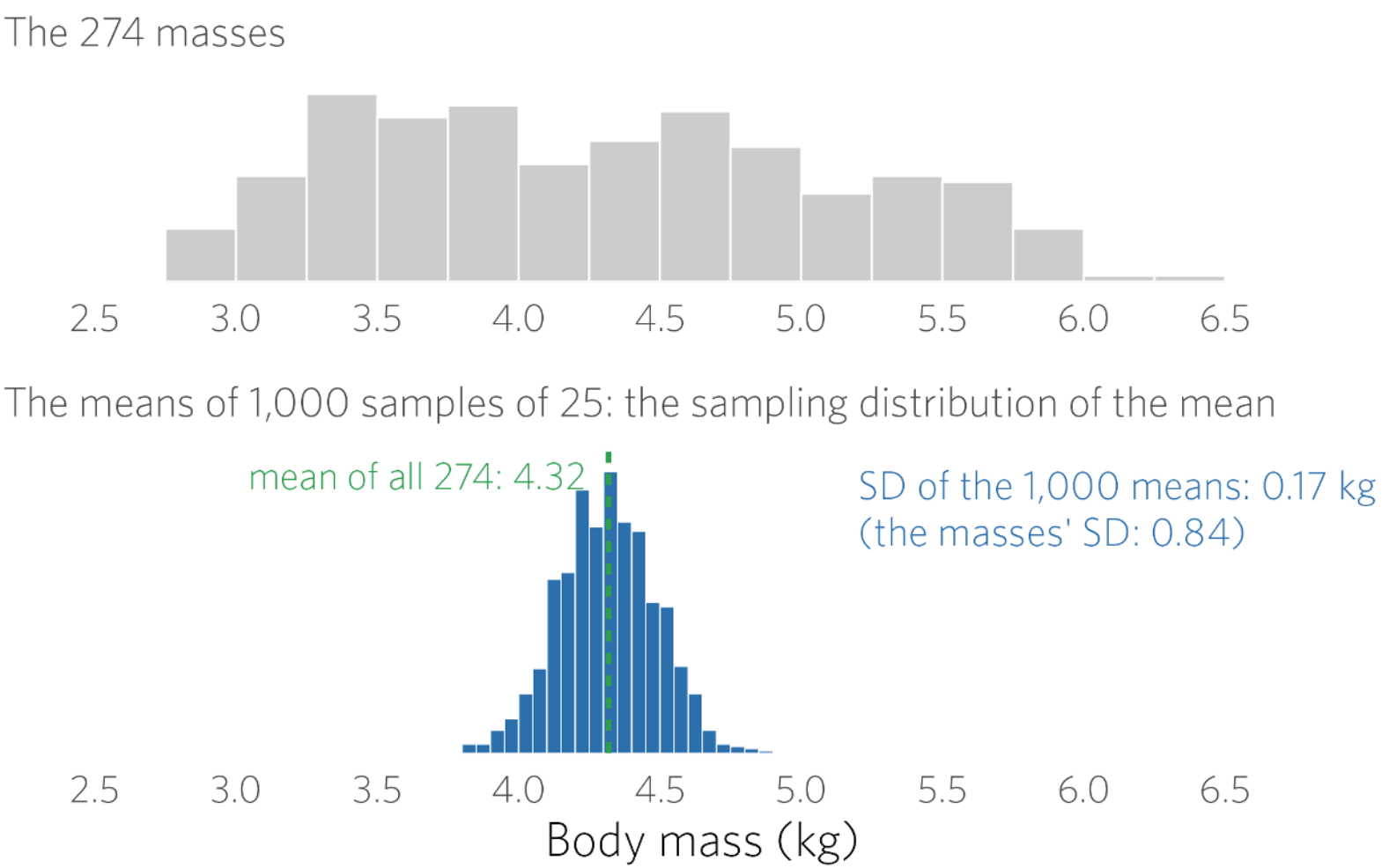
Each new sample gives a new mean



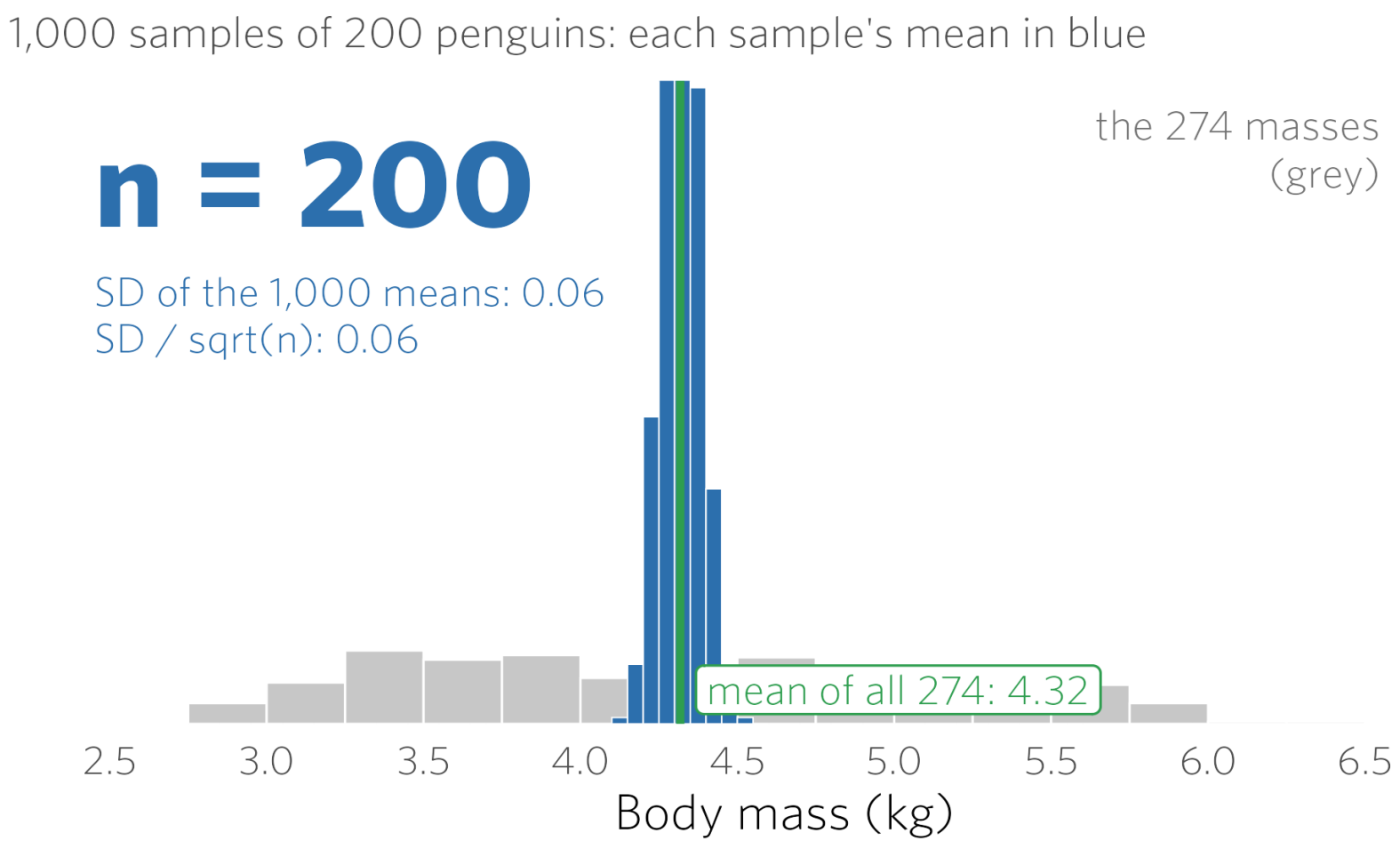
After ten samples, the means form a distribution



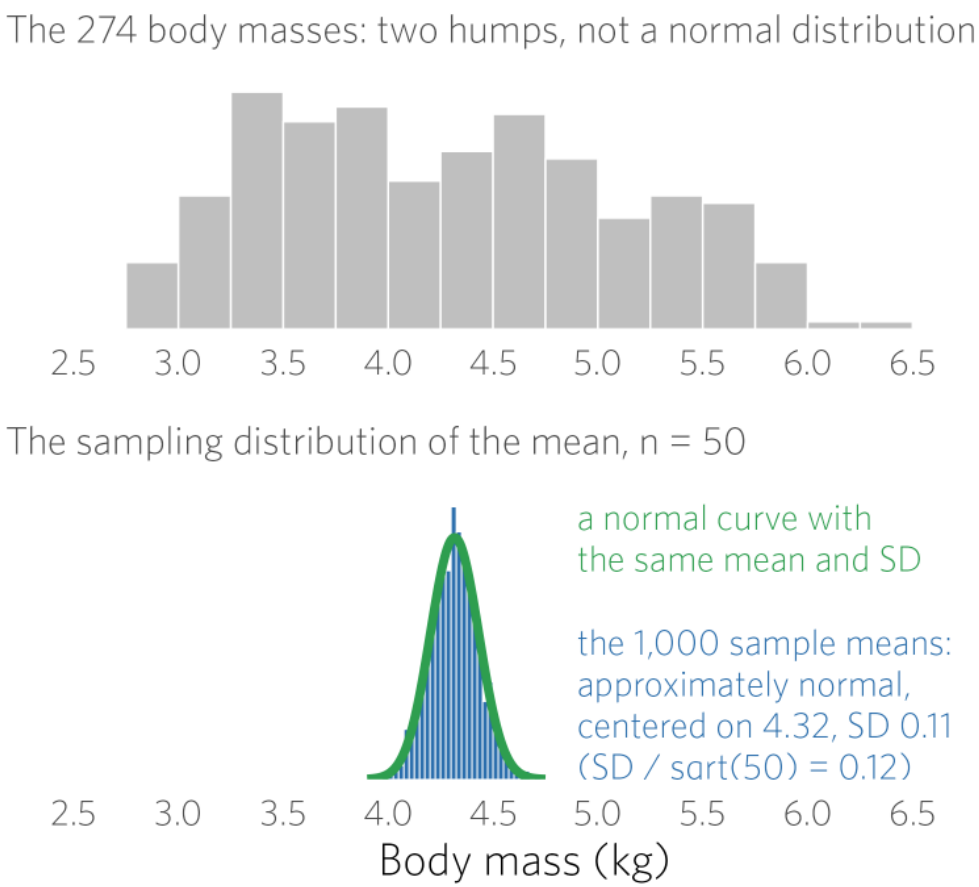
A thousand samples: the sampling distribution of the mean



The sampling distribution narrows as the sample grows

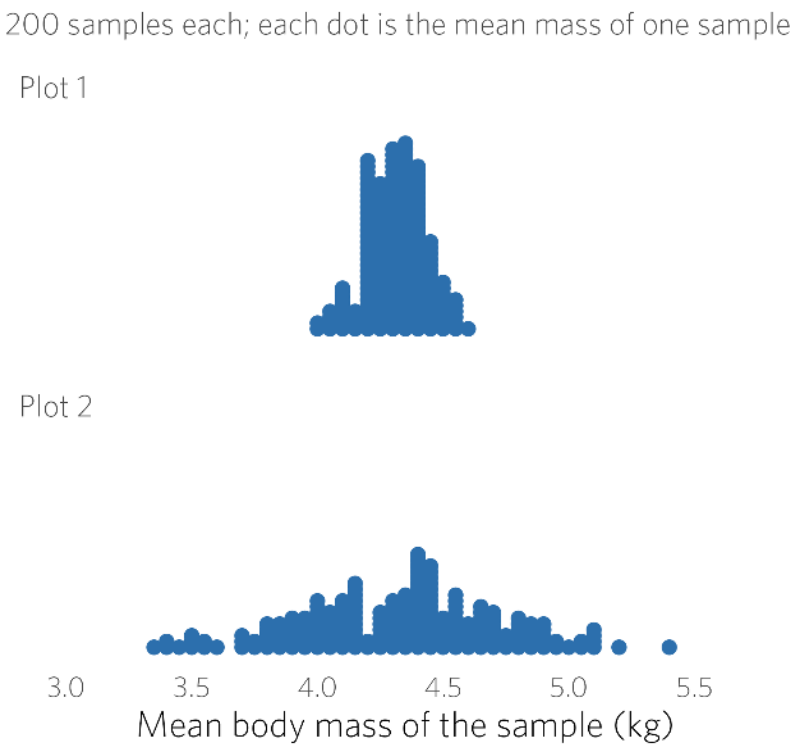


The sampling distribution of the mean approaches normality, whatever the shape of the data distribution



- The 274 masses have two humps, one per species. The data are not normally distributed.
- The means of 1,000 random samples of 50 are approximately normal all the same, centered on the population mean with SD equal to the masses' SD divided by the square root of n. This is the **central limit theorem**.
- It holds when the observations are drawn at random and independently from the same population, and the sample is large enough. For data as far from normal as these, 50 is enough.

Worksheet · Part 2

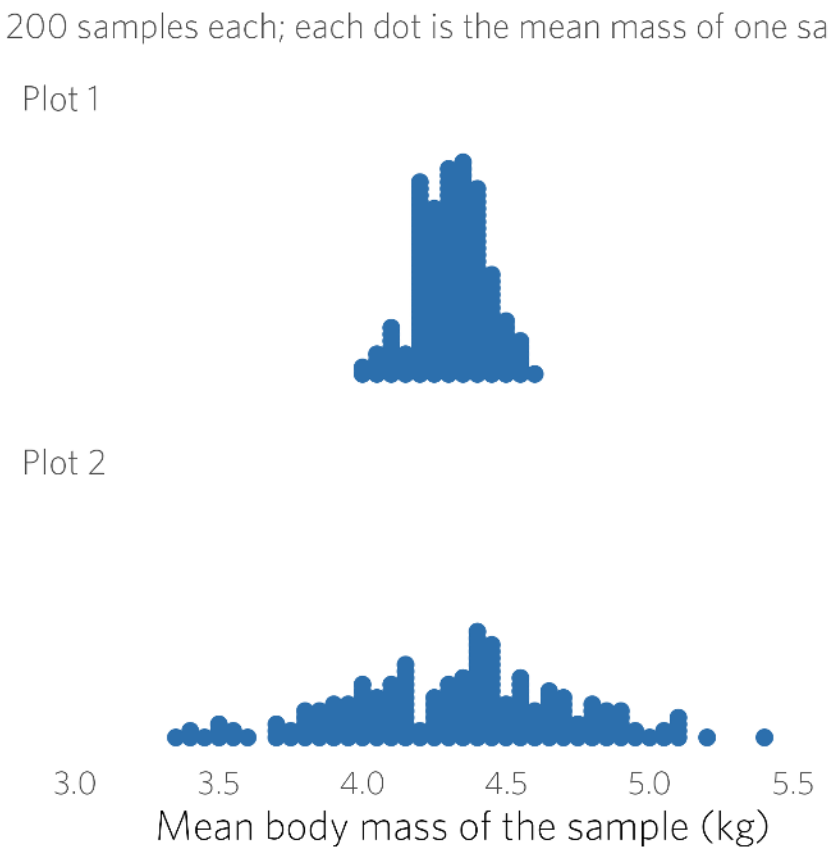


Treating the 274 penguins as the population, a computer drew 200 random samples of 5 penguins and 200 random samples of 50 penguins, and computed the mean mass of each sample. Each dot is the mean mass of one sample; one plot shows the samples of 5, the other the samples of 50.

a. Which plot shows the samples of 50 penguins? Select one: *Plot 1* or *Plot 2*. Then write one sentence giving the reason.

b. A researcher weighs 500 penguins, all from one Adelie colony. Is the mean of that sample a better estimate of the mean mass of all Adelie and Gentoo penguins than the mean of a random sample of 50 penguins drawn from both species? Write *yes* or *no*, then one sentence giving the reason.

Part 2a–b · Which plot, and one colony



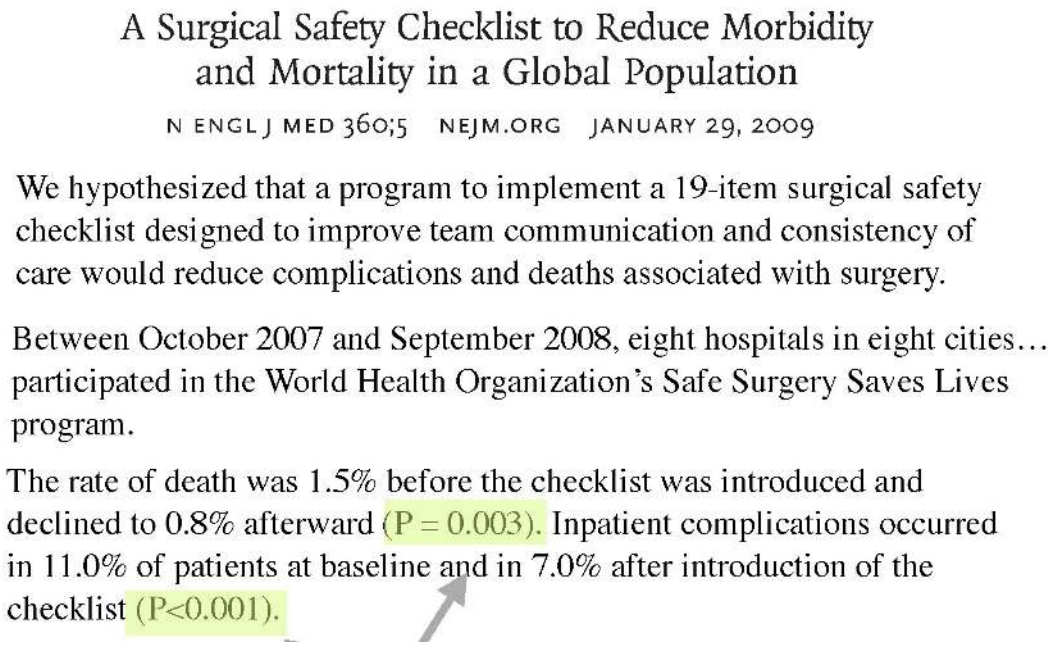
- **a.** *Which plot shows the samples of 50?* Plot 1. In a sample of 50, heavy and light penguins largely cancel out, so the sample means stay within a narrow band around the population mean. In a sample of 5 they often do not, so those means spread widely. Both plots are centered on the same value; they differ in spread.
- *A thousand samples of 5 instead of 200?* The spread would be unchanged. Drawing more samples adds more dots to the plot; only the number of penguins in each sample determines the spread of the means.
- **b.** *500 penguins from one colony.* No. Every sample from one colony has a mean near that colony's mean, so the estimate misses the population mean by about the same amount in every sample. A larger sample does not correct a sample drawn from the wrong penguins.

Today

Every model so far, and the logic of inference

1. Every model so far on one dataset
2. From the sample to the process behind it
3. Could the empty model have produced these data?

Two sentences from published studies



A Surgical Safety Checklist to Reduce Morbidity and Mortality in a Global Population

N ENGL J MED 360;5 NEJM.ORG JANUARY 29, 2009

We hypothesized that a program to implement a 19-item surgical safety checklist designed to improve team communication and consistency of care would reduce complications and deaths associated with surgery.

Between October 2007 and September 2008, eight hospitals in eight cities… participated in the World Health Organization's Safe Surgery Saves Lives program.

The rate of death was 1.5% before the checklist was introduced and declined to 0.8% afterward (P = 0.003). Inpatient complications occurred in 11.0% of patients at baseline and in 7.0% after introduction of the checklist (P<0.001).

Haynes et al. (2009), *New England Journal of Medicine* 360: 491–499.

"We are unable to reject the null hypotheses that BWCs have no effect on police use of force, citizen complaints, policing activity, or judicial outcomes."

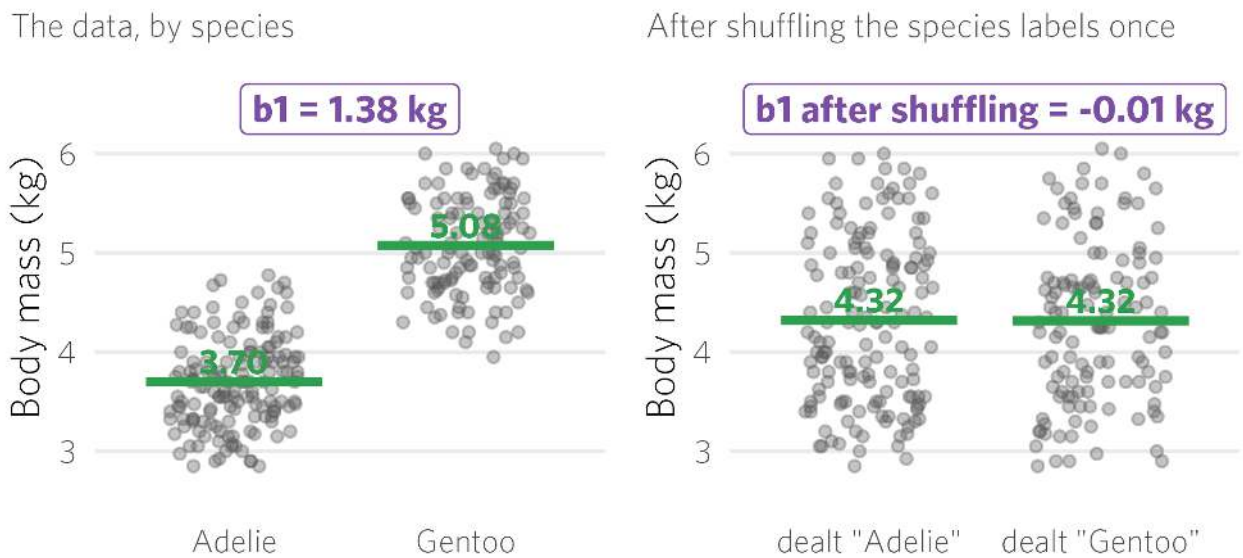
Yokum, Ravishankar and Coppock (2017), *Evaluating the Effects of Police Body-Worn Cameras: A Randomized Controlled Trial*, working paper; 2,224 officers randomly assigned.

What is P = 0.003? What is a null hypothesis, and what does it mean to be unable to reject one? Both sentences report the result of the same procedure, a hypothesis test.

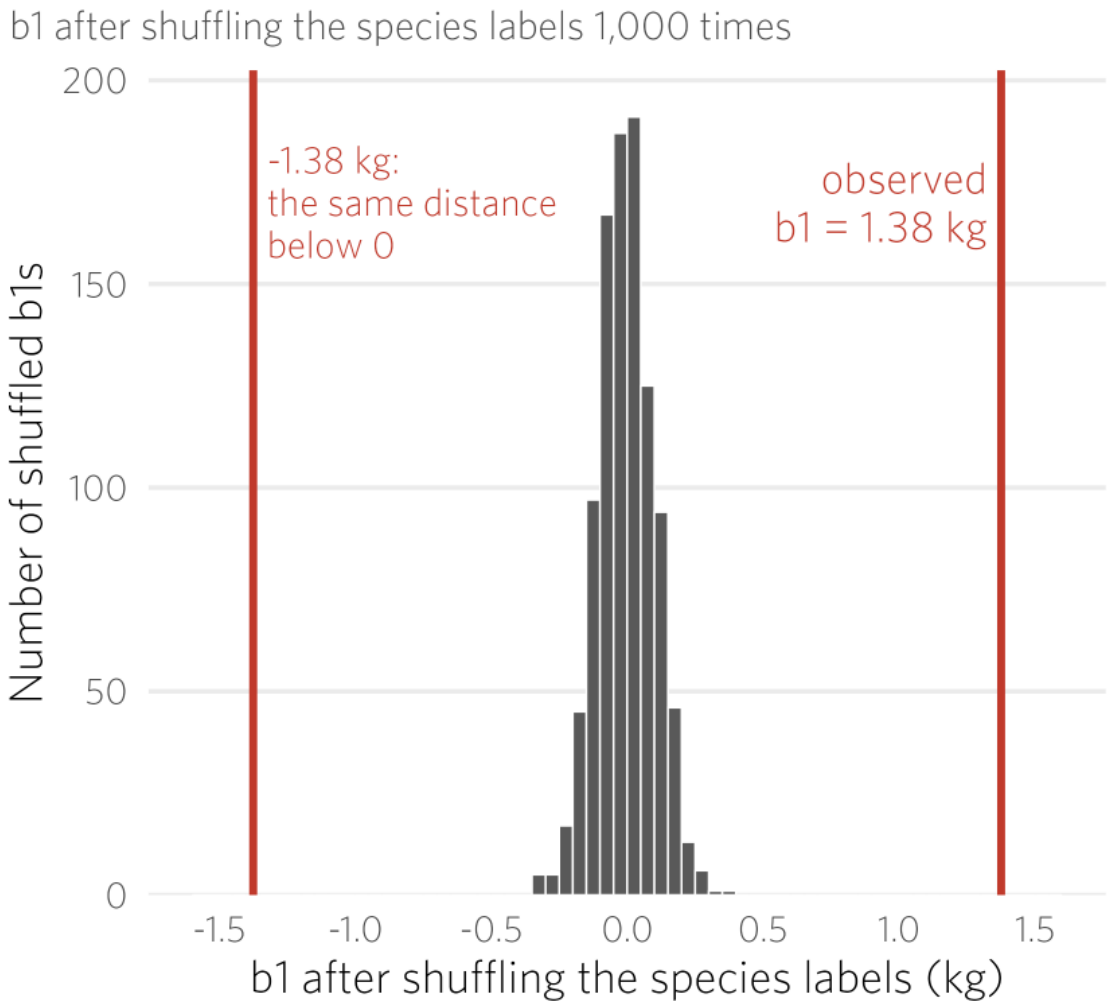
How a hypothesis test works

- **The hypothesis.** Gentoo penguins are heavier than Adelie penguins: species affects body mass.
- **State the opposite claim.** The test is run on the **null hypothesis**: species makes no difference to mass. In this course's terms, the null hypothesis is the empty model, $\beta_1 = 0$.
- **Collect the data.** The 274 penguins give $b_1 = 1.38$ kg, the sample's estimate of the process's β_1.
- **Ask how likely data like these would be if the null hypothesis were true.** Unlikely: reject the empty model. Not unlikely: keep it. The question is about the data given the model, not about the model given the data.

What shuffling does to the data

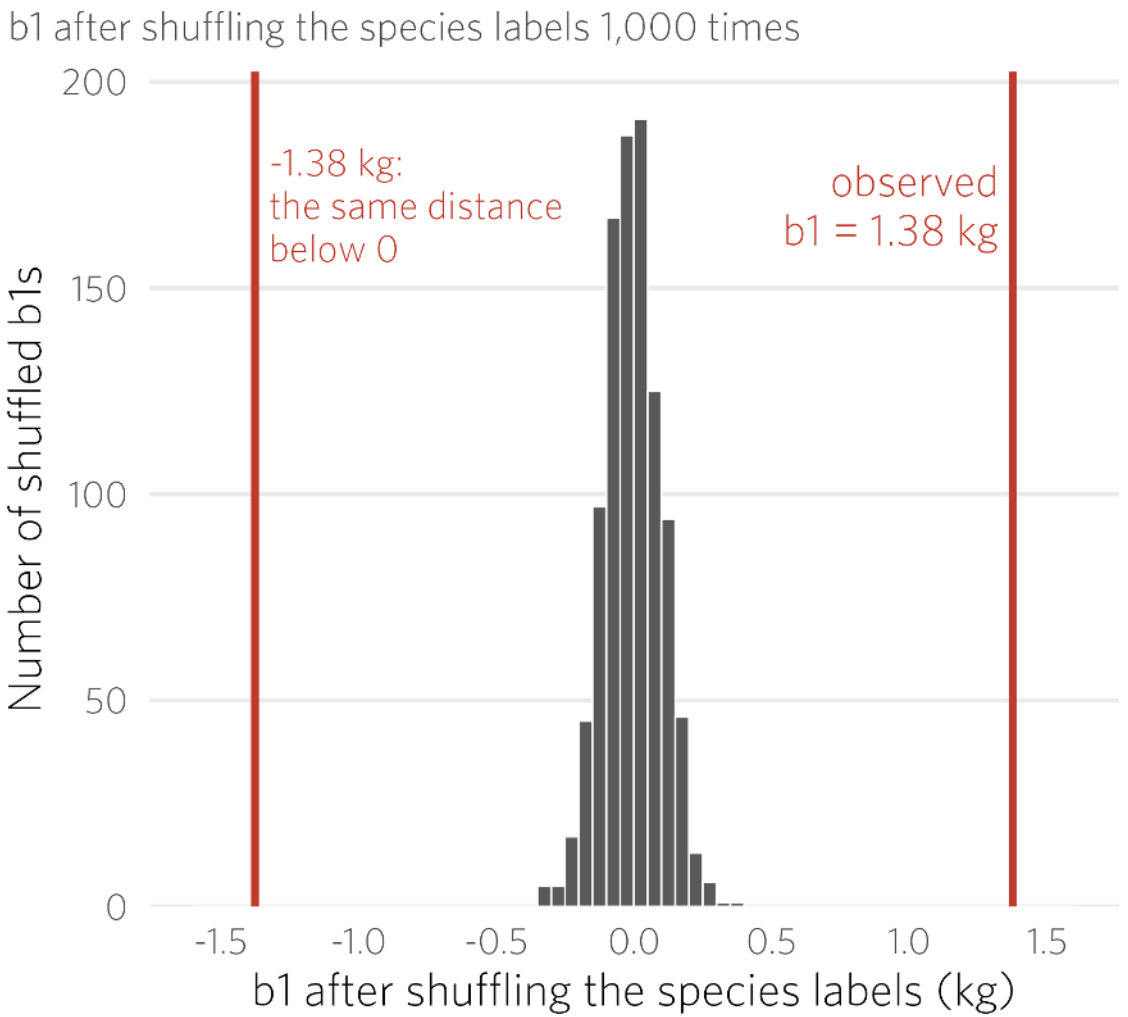


Shuffling a thousand times: b_1 when species makes no difference



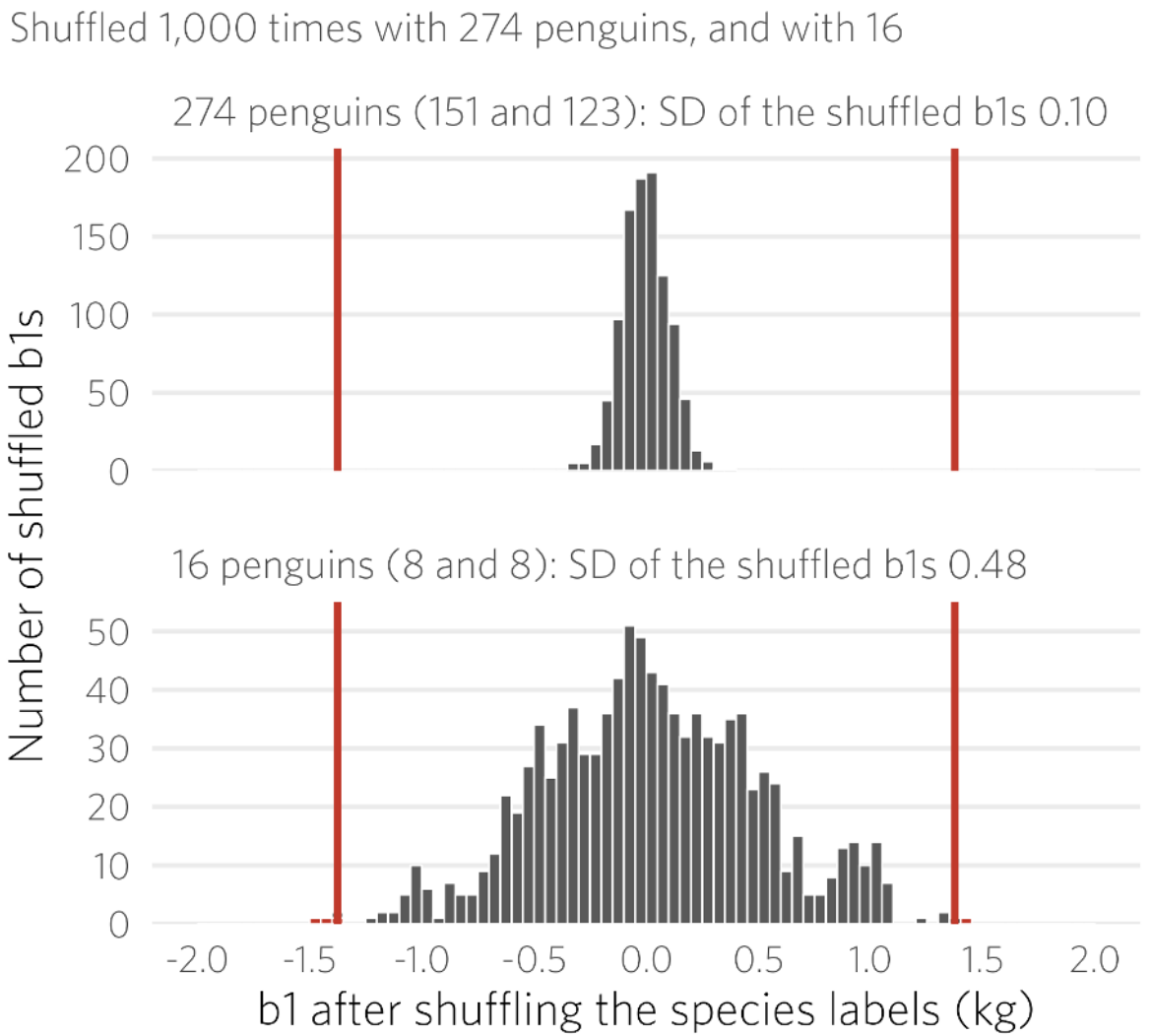
- Each shuffled dataset gives one b_1, from species labels that carry no information about mass. The first ten are the dots.
- Shuffling a thousand times gives the distribution of b_1 under the empty model. It is centered on 0 and shaped like the normal distribution.
- The red lines mark the observed 1.38 and its mirror image, -1.38. No shuffled b_1 reached either line.

The p-value



- The p-value: if the empty model were true, the probability of a b_1 at least as far from 0 as the observed one, in either direction.
- Here none of the 1,000 shuffled b_1s came that far from 0, so p is below 1 in 1,000, written $p < 0.001$.
- A small p-value means that data like these would be rare if the empty model were true: a statement about the data, not the probability that the empty model is true.
- The checklist study's P = 0.003 is the same quantity, for its before-and-after difference in deaths.

Sixteen penguins instead of 274



- Had the researchers weighed 16 penguins instead of 274, each shuffled mean would come from eight masses, and the shuffled b_1s would spread to almost five times the width.
- A few of these 1,000 shuffled b_1s reach 1.38, where none did with 274 penguins: the same b_1 has a larger p-value in a smaller sample.
- The spread of the shuffled b_1s shrinks as the sample grows, in proportion to one over the square root of n, in the same way as the sample means.

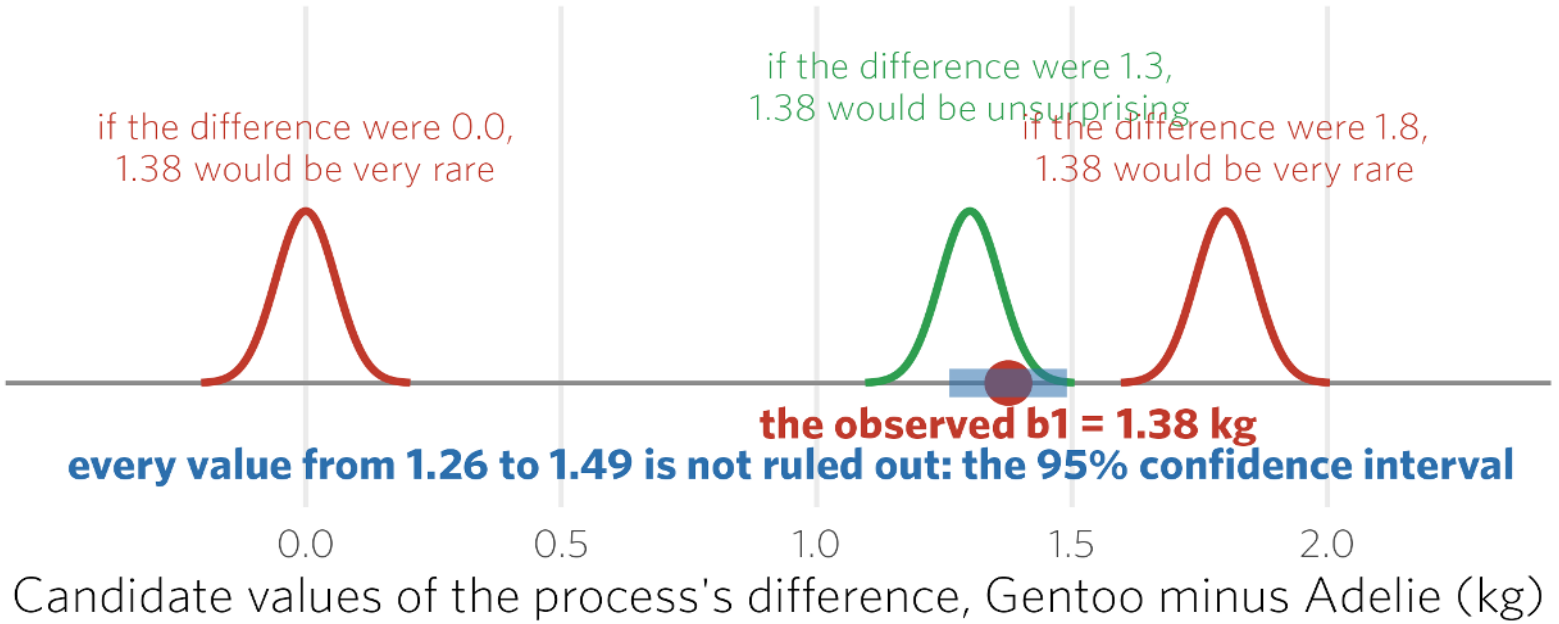
How small must a p-value be? The 5% convention

- The p-value here is below 0.001. How small must a p-value be before the empty model is rejected?
- By a long-standing convention, an outcome that chance produces in fewer than 5 of 100 tries counts as unusual: the line is drawn at $\alpha = 0.05$.
- The line dates from the 1920s and was proposed as a convenient cutoff, not a fixed rule. Some fields now ask for 0.005 before a result counts as a discovery.

The decision, and what it does not say

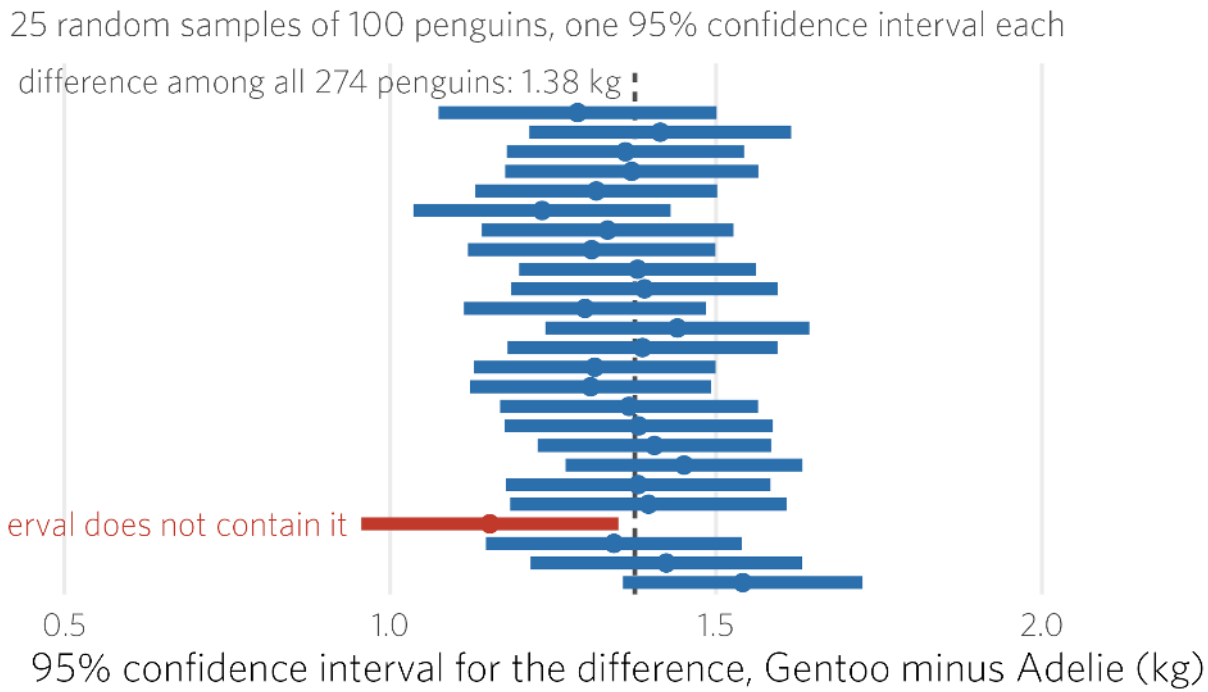
- p below 0.05: the data count as rare under the empty model, so we **reject** the empty model.
- p at or above 0.05: we **keep** the empty model, because chance remains a plausible account of the data.
- Rejecting the empty model says only that the data would be rare if species made no difference. It does not say why the species differ, that β_1 equals 1.38, or how large the difference is.

The confidence interval



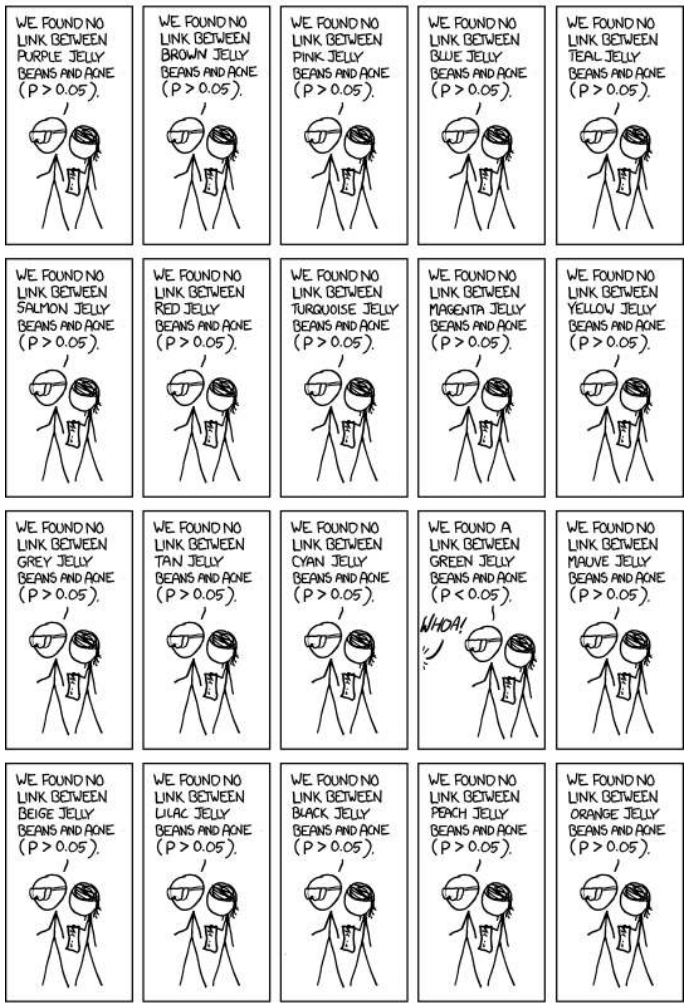
- **Confidence interval**: the range of values of the process's difference under which the observed b_1 would not be rare. Here, 1.26 to 1.49 kg.
- The p-value tests one value of that difference, zero. The confidence interval tests every value and keeps the ones the data do not rule out.

What a confidence interval does not say



- It does not say that 95% of penguins differ by 1.26 to 1.49 kg, nor that the process's difference has a 95% chance of lying in this one interval.
- Of intervals built this way from repeated samples, about 95% contain the process's difference. This one either does or does not.

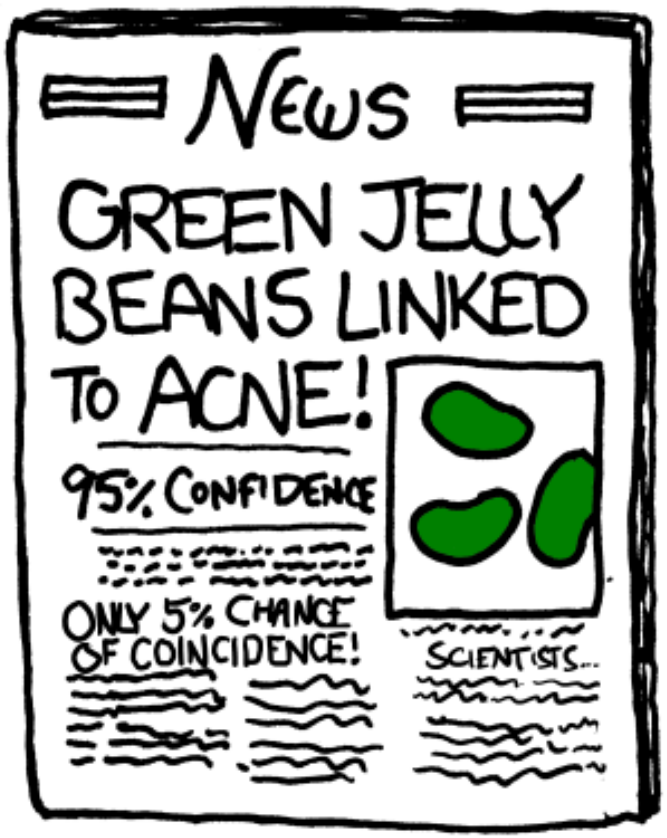
A 5% false-positive rate, and many tests



- $\alpha = 0.05$ is a tolerated false-positive rate: when the empty model is true, 1 test in 20 rejects it anyway, by chance.
- Run 20 tests of effects that do not exist, and one false positive is expected; the chance of at least one is 64%.
- Searching many tests, subgroups, or analyses for one with $p < 0.05$ therefore produces a "finding" whether or not any effect exists. In the comic, twenty colors of jelly bean are tested against acne, and one test gives $p < 0.05$: the green one.

xkcd 882, "Significant", by Randall Munroe (CC BY-NC 2.5).

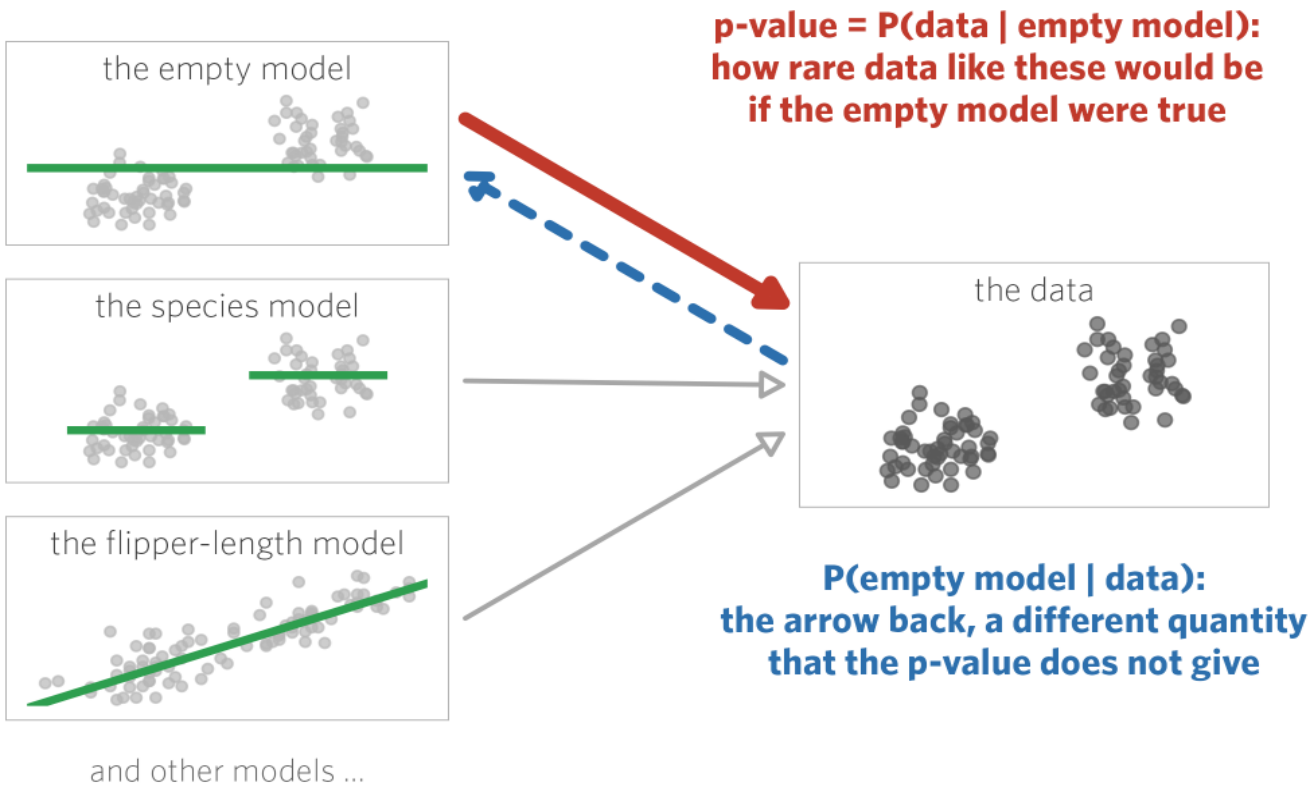
A result selected from many shrinks on retest



xkcd 882, "Significant", by Randall Munroe (CC BY-NC 2.5).

- The reported result is the one that crossed 0.05; the nineteen that did not go unmentioned.
- A p-value answers one question asked before the data were seen. Every extra test, subgroup, or way of analyzing the data is another draw from the chance distribution, and "95% confidence" is true of one test chosen in advance, not of the best of twenty.
- Tested again on new people, the selected result usually shrinks, because chance contributed to its selection. A replication is the proper check of a result selected as the best of many.

The p-value measures one arrow



The p-value is $P(\text{data} \mid \text{empty model})$: the one arrow in red.

$P(\text{empty model} \mid \text{data})$, the arrow back, is a different quantity; a p-value does not provide that probability.

Hypothesis testing is model comparison

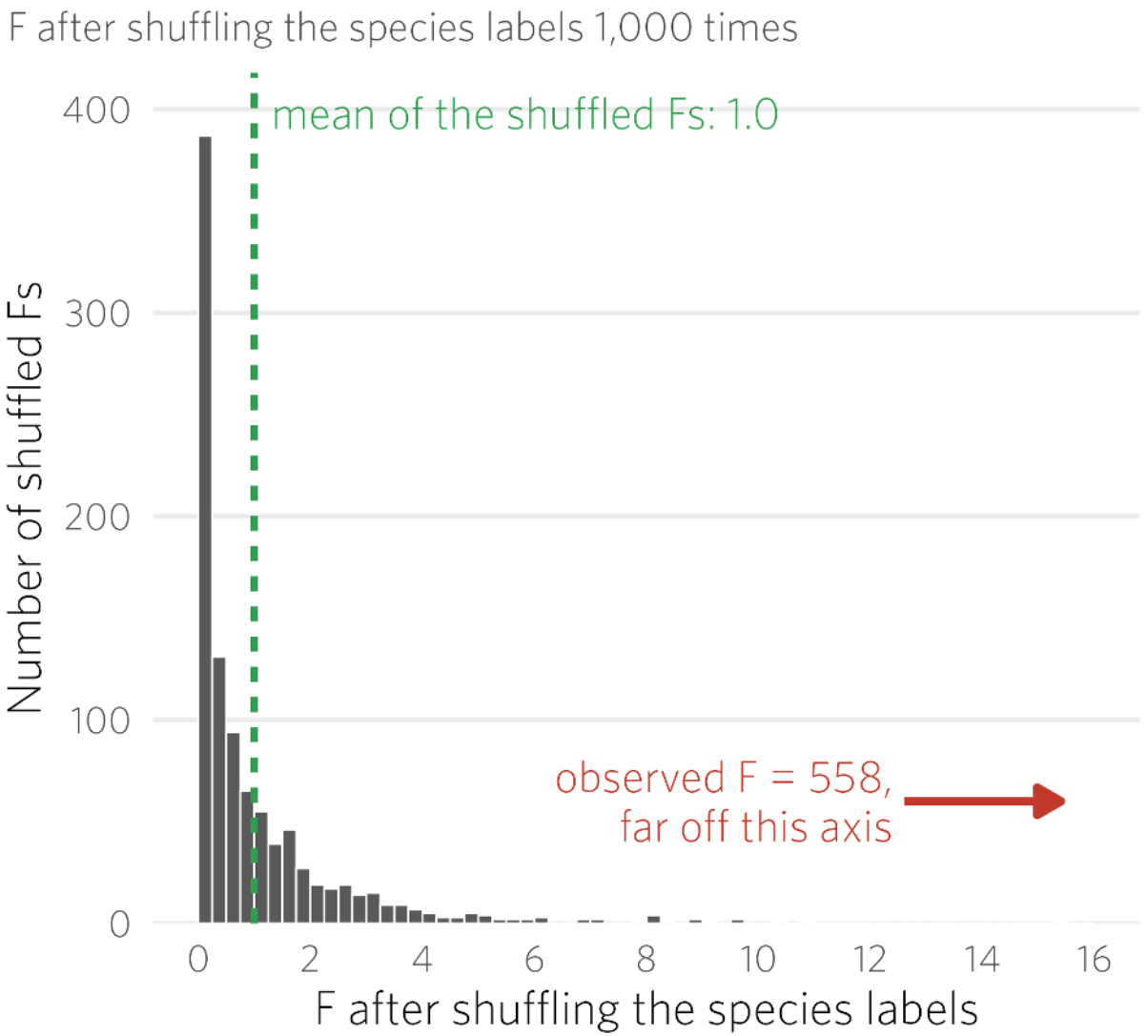
$$F = \frac{SS_{\text{Model}} \,/\, \text{df}_{\text{Model}}}{SS_{\text{Error}} \,/\, \text{df}_{\text{Error}}} = 558 \text{ for the species model}$$

Error removed per parameter added, over error left per degree of freedom.

- The same question, could the empty model have produced these data, asked as a comparison of two models: does the species model's added parameter remove more error than a label dealt from a hat?
- Would F be near 0, near 1, or far above 1 when the species model removes a great deal of error per parameter?

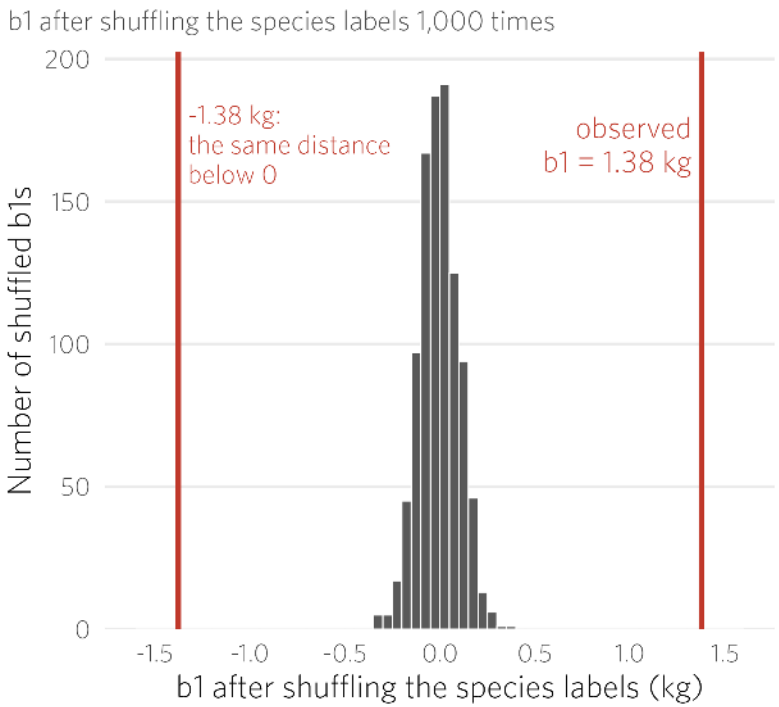
- Far above 1. Here F = 558: the species model removes 558 times as much error per parameter as remains per degree of freedom. Labels dealt from a hat would remove about as much error per parameter as remains per degree of freedom, so F would be near 1.

Shuffle the labels and recompute F



- Every shuffled fit gives an F as well as a b_1: 1,000 Fs from a world in which species makes no difference. They cluster near 1, with one tail to the right, since F is never below 0.
- The observed F lies beyond every one of them, far off this axis: the same p-value, below 1 in 1,000.
- The same comparison applies to any pair of models, the empty model against a line or against several groups; Chapter 11 calls it the F-test.

Worksheet · Part 3



The species model of the 274 penguins' body masses gave b_1 = 1.38 kg, the Gentoo mean minus the Adelie mean. A computer then shuffled the species labels across the 274 masses 1,000 times, computing b_1 each time; the histogram shows those 1,000 shuffled b_1s.

a. In one sentence, say what process produced one shuffled b_1.

b. About how many of the 1,000 shuffled b_1s were at least as far from 0 as 1.38, in either direction? Write a whole number.

c. By the 5% convention, is the empty model rejected or kept? Write *reject* or *keep*, then a one-sentence reason that uses your answer to b.

Worksheet · Part 3, continued

d. A study compares two groups, finds a difference between them, and reports p = 0.01. Each of the four sentences below offers an interpretation of that p-value. Write *justified* or *not justified* beside each sentence's number.

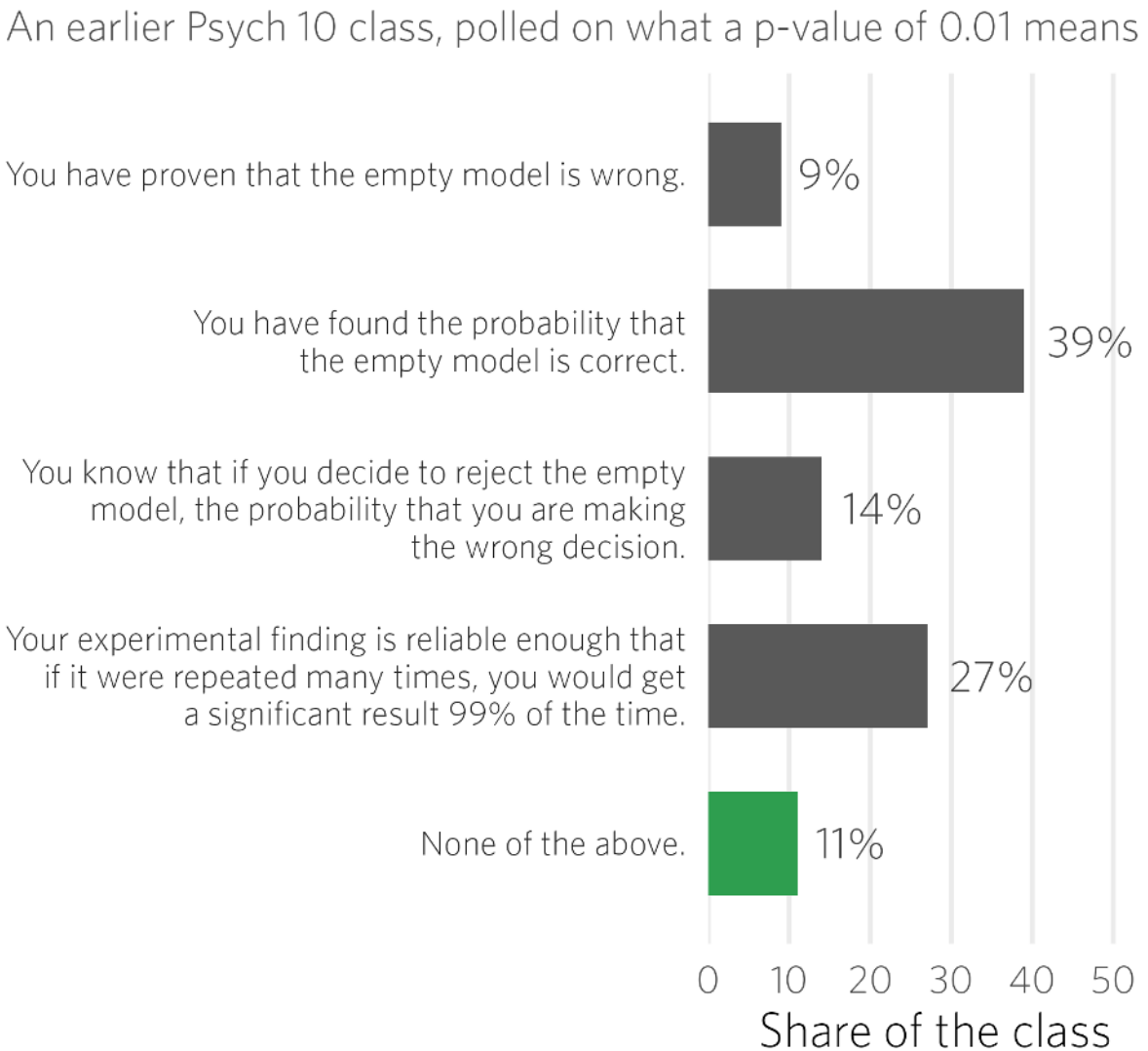
1. The study has proven that the empty model is wrong.
2. The probability that the empty model is correct is 1%.
3. If the study were repeated many times, 99% of the repeats would find a difference.
4. If the empty model were true, a difference at least this large would arise in 1% of samples.

e. A randomized study assigned 2,224 police officers to wear a body camera or not, then compared the two groups' use of force. The authors wrote: "We are unable to reject the null hypothesis that body-worn cameras have no effect on police use of force." (The null hypothesis is their name for the empty model.) Does this result show that the cameras have no effect? Write *yes* or *no*, then one sentence giving the reason.

Part 3a–c · Rejecting the empty model

- **a.** *What one shuffled b_1 represents.* One b_1 from a world in which species makes no difference: the labels were dealt at random, so the difference that came out is chance alone.
- **b.** *How many of the 1,000 reached 1.38.* None. Every shuffled b_1 stayed within about 0.4 kg of 0.
- **c.** *Reject or keep?* Reject: p is below 1 in 1,000, far below the 5% line. The data would be rare under a process with no difference.

Part 3d–e · Three misreadings of a p-value



- **d.** *Which sentences are justified?*
  Only the fourth sentence, the definition. The other three were choices on this poll, and three quarters of an earlier class chose one of those three.
- A p-value is the probability of the data given the empty model: not proof, not the probability of the empty model, not the share of repeats with the same result, not the size of the effect.
- **e.** *Unable to reject: no effect?*
  No. The data are not rare under the empty model, which a small real effect produces too; the study cannot tell no effect from a small effect.

Before Friday

- Finish CourseKata **Chapters 10 and 11**.
- Wednesday: **Ji Son**, author of CourseKata, in this room.
- Friday: **Case #6**, in your tutorial session. Check your Slack channel for your time slot.
- Milestone 6 is due Sunday.

Hand in your worksheet on the way out

On the last line, write down one thing you learned today.

Thanks!