**Date** 11-02-2026

# PSYCH 10 · Fall 2026 · Worksheet for Lecture 12: Every model so far, and the logic of inference · Answer key

*The answers are in blue. Where a question has more than one defensible answer, the key says so and describes what a good answer contains.*

## Part 1 · Two models of body mass, one ANOVA table each

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.

**a.** *SS Total is the empty model's error: the sum of the 274 masses' squared deviations from their mean. It depends on the outcome alone and not on the explanatory variable, so every model of these 274 masses starts from the same 190.8.*

**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.

**b.** *The flipper-length model. Its SS Error is smaller, 42.3 against 62.5, and its PRE is larger, 0.78 against 0.67. F, 955 against 558, shows the same thing, because the two models have the same degrees of freedom. A weaker answer names F alone.*

**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.

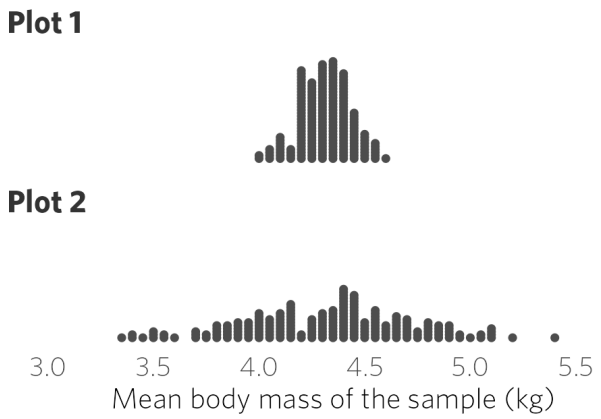
**c.** *The one parameter each model adds to the empty model, $b_1$: a slope for a quantity, or a difference between two group means. df Model counts parameters, not the kind of explanatory variable; a model with three species would spend 2.*

**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.

*d. 1 true: SS Model 148.5 against 128.2, PRE 0.78 against 0.67. 2 true: F = (SS Model / 1) / (SS Error / 272) for both, so with equal degrees of freedom the smaller SS Error gives the larger F. 3 false: with 548 penguins df Error is about 546, and F ≈ (0.67 / 1) / (0.33 / 546) ≈ 1,100, about double; the same PRE is stronger evidence in a larger sample. 4 false: the two PREs are shares of one SS Total and they overlap, because species and flipper length carry much of the same information (Gentoo penguins have the longer flippers); shares of a whole cannot sum past 1.*

**Part 2 · From the sample to the population**




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.

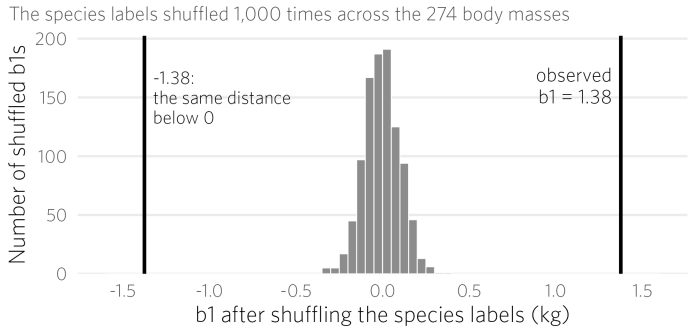
**a.** Which plot shows the samples of 50 penguins? Select one: *Plot 1* or *Plot 2*. Then write one sentence giving the reason.

**a.** *Plot 1. Its means bunch tightly together: in a sample of 50, heavy and light penguins offset each other far more completely than in a sample of 5, so the means of samples of 50 land within a narrow band. Both plots center in about the same place; the difference is the spread. A weaker answer picks Plot 2 "because it has more variety": the variety is among penguins, and more of it in a sample is exactly what makes that sample's mean more stable.*

**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.

**b.** *No. Every sample from one colony lands near that colony's mean, about 3.7 kg, so the estimate misses the mean of all Adelie and Gentoo penguins, 4.32 kg, by about the same amount every time, whatever the size of the sample. A large sample makes an estimate more precise; only a random draw from the whole population makes it representative. A weaker answer says "yes, 500 is a large sample" or "no, 500 is too few."*

**Part 3 · Could the empty model have produced these data?**




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 $b_1$s.

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

**a.** *One $b_1$ from a world in which species makes no difference to body mass: the 274 species labels were dealt out at random, so any difference between the two means is chance alone.*

**b.** About how many of the 1,000 shuffled $b_1$s were at least as far from 0 as 1.38, in either direction? Write a whole number.

**b.** *0. None of the shuffled $b_1$s reached 1.38 in either direction; they run from about -0.33 to 0.39.*

**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.

**c.** *Reject. None of the 1,000 shuffled $b_1$s was as far from 0 as 1.38, so the p-value is below 1 in 1,000, far below the 5% line; the empty*

*model is a poor account of these data. (Rejecting it does not say what causes the difference between the species, or that the difference in the process is exactly 1.38.)*

**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.

*d. Only sentence 4 is justified; it is the definition of the p-value. Sentence 1 is not: a p-value gives evidence against the empty model, never proof. Sentence 2 is not: p is the probability of the data given the empty model, not the probability of the empty model given the data. Sentence 3 is not: a p-value says nothing about how often a repeat of the study would come out the same way.*

**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.

**e.** *No. Keeping the empty model means the data were not rare enough under it to reject it, and a small real effect produces data like that too, so the study cannot tell "no effect" from "an effect too small to show up in these data." A strong answer says that the result leaves open both no effect and a small effect; a weaker answer restates "no effect," or argues from the sample size alone (2,224 officers is not a small sample).*