

PSYCH 10 · Fall 2026 · Worksheet for Lecture 8: Modeling relationships · 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 · A mean for each group

In the Week 2 survey, 148 students reported their commute time. The mean of all 148 is 19 minutes. The 120 students who walked or biked have a mean of 13 minutes; the 28 who drove or took the shuttle have a mean of 43 minutes. (All three means are rounded to whole minutes.)

a. One student walked and took 22 minutes. The empty model predicts 19 minutes for every student. Compute this student's residual under the empty model, in minutes, with its sign.

a. *22 - 19 = +3 minutes. The residual is the actual time minus the model's prediction, so this student took 3 minutes longer than the empty model predicts.*

b. The group model predicts each student's time with their own group's mean. Compute the same student's residual under the group model, in minutes, with its sign.

b. *22 - 13 = +9 minutes, since the group model predicts 13 minutes for everyone who walked or biked.*

c. For this one student, which model has the smaller residual? Write *empty model* or *group model*.

c. *Empty model: its residual for this student is 3 minutes, against 9 for the group model. The model with less total error over all 148 students can still have the larger residual for one student.*

d. Part c compared the two models for one student. To compare them for all 148 students, you would compute one quantity for each model. Name that quantity.

d. *The sum of squared residuals (the SS from Lecture 6), computed over all 148 students under each model. The model with the smaller sum fits better.*

Part 2 · How much error did the groups remove?

Eight students from one discussion section reported their commute times. The four who walked: **8, 12, 12, and 16** minutes. The four who took the shuttle: **16, 20, 20, and 24** minutes.

a. Compute three means: the walkers' mean, the shuttle riders' mean, and the mean of all eight students. Write all three, in minutes. Then fill in the rest of the table on your sheet; the first row is done.

a. *Walkers 12, shuttle riders 20, and all eight 16.*

b. Compute SS Total: add up the eight squared deviations from the mean of all eight students (the table's fourth column).

b. *The deviations from 16 are -8, -4, -4, 0, 0, 4, 4, and 8; their squares are 64, 16, 16, 0, 0, 16, 16, and 64, so SS Total = 192.*

c. Compute SS Error: add up the eight squared deviations from each student's own group mean (the table's last column).

c. *Within each group the deviations are -4, 0, 0, and 4, with squares 16, 0, 0, and 16, so each group contributes 32 and SS Error = 64.*

d. Compute SS Model = SS Total - SS Error.

d. *SS Model = 192 - 64 = 128. As a check, every student's group mean lies 4 minutes from the mean of all eight, and $8 \times 4^2 = 128$.*

e. Compute the proportional reduction in error, PRE = SS Model / SS Total. Write it as a fraction and as a decimal.

e. *PRE = 128/192 = 2/3, about 0.67: how the student got to campus accounts for two thirds of the variation in these eight commute times.*

f. Suppose the four shuttle riders had taken 36, 40, 40, and 44 minutes instead (each time 20 minutes longer). For each of SS Total, SS Error, and SS Model, write *larger*, *smaller*, or *unchanged*.

f. *SS Total larger, SS Error unchanged, SS Model larger. SS Error stays at 64 because the spread within each group is the same. The group means become 12 and 40, each 14 minutes from the new overall mean of 26, so $SS\ Model = 8 \times 14^2 = 1,568$, $SS\ Total = 1,632$, and PRE rises to about 0.96.*

Part 3 • Reading the model's two numbers

In the NHANES survey, 7,182 adults had their height measured. A group model of height by sex, fitted in R, gives these two numbers (in inches, rounded):

(Intercept)	64
Gendermale	5

and these sums of squares: **SS Model 52,000 • SS Error 61,000 • SS Total 113,000.**

a. What is the mean height of the men, in inches?

a. *$64 + 5 = 69$ inches. The second number is the amount added to the women's mean to reach the men's mean, not the men's mean itself.*

b. Write the model's prediction for a man in the form $b_0 + b_1X_i$, with numbers in place of b_0 , b_1 , and X_i .

b. *$64 + 5 \times 1 = 69$ inches, with $X_i = 1$ for a man. For a woman, $X_i = 0$ and the prediction is $64 + 5 \times 0 = 64$.*

c. What proportion of the variation in height does sex account for? Write the answer two ways: as the fraction SS Model / SS Total with the two numbers filled in, and as a decimal.

c. *$PRE = 52,000 / 113,000$, about 0.46: sex accounts for about 46% of the variation in adult height.*

d. An AI assistant summarized this output in three sentences: "The intercept of 64 is the mean height of the women. The coefficient of 5 for Gendermale is the mean height of the men. $PRE = 0.46$ means that the model predicts 46% of adults' heights correctly." Two of the three sentences are not quite right. Say which two, and write a corrected version of each sentence.

d. *Sentences two and three. Corrected second sentence: the coefficient of 5 is the difference between the men's mean and the women's mean, so the men's mean height is $64 + 5 = 69$ inches. Corrected third sentence: $PRE = 0.46$ means that the group model removes 46% of the total variation in height (SS Total), not that it predicts 46% of adults correctly; the model predicts no one's height exactly. The first sentence is correct.*

e. Sex accounts for about 46% of the variation in height. Name one other variable that could account for some of the other 54%.

e. *Any variable that differs among adults of the same sex and goes with height is a reasonable answer: age (adults lose a little height after about 50), nutrition in childhood, the parents' heights, or country of origin. Sex itself is not an answer, since its share is the 46% already accounted for.*