

PSYCH 10 · Fall 2026 · Worksheet for Lecture 9: Comparing group models · 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 · Reading the model's two numbers (Monday's Part 3, which the class did not reach)

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

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.*

Part 2 · A model with three means

A group model of program earnings by major was fitted to 1,247 programs from the Times' college guide: English, mechanical engineering, and computer science. R reports (in dollars, rounded to the nearest thousand):

(Intercept)	49,000
majorMechanical Engineering	43,000
majorComputer Science	57,000

a. What is the mean earnings of the computer science programs, in dollars?

a. *49,000 + 57,000 = 106,000 dollars. The computer science coefficient is added to the intercept, which is the English mean.*

- b. A reader computes the computer science mean as $49,000 + 43,000 + 57,000 = 149,000$. One of the three numbers belongs to a different major and should not be in the sum. Write that number and the name of its major.
- b. *The 43,000, which belongs to mechanical engineering. Each coefficient is added to the intercept, never to the other coefficient: mechanical engineering $49,000 + 43,000 = 92,000$, computer science $49,000 + 57,000 = 106,000$.*
- c. R used English as the reference group. If computer science were the reference group instead, what three numbers would the output show? Write all three, with their signs.
- c. *(Intercept) 106,000; majorEnglish -57,000; majorMechanical Engineering -14,000. The three predicted means are unchanged, since $106,000 - 57,000 = 49,000$ and $106,000 - 14,000 = 92,000$; only the way the output states them changes.*

Part 3 · The cost of a parameter

Eight students from the Lecture 8 worksheet, four walkers and four shuttle riders, one mean per group: **SS Model = 128, SS Error = 64, SS Total = 192**. Fill in the table on your sheet as you go.

- a. Fill in the df column: df Model, df Error, and df Total. (The empty model spends one of the eight degrees of freedom; the two-group model spends one more.)
- a. *df Model = 1 (two group means, one beyond the empty model's), df Error = $8 - 2 = 6$, df Total = $8 - 1 = 7$.*
- b. Fill in the MS column: MS Model and MS Error, each SS divided by its df. Round to one decimal.
- b. *MS Model = $128 / 1 = 128$; MS Error = $64 / 6 \approx 10.7$.*
- c. Compute $F = \text{MS Model} / \text{MS Error}$ and write it in the Model row, rounded to a whole number.
- c. *$F = 128 / 10.7 \approx 12$: the one parameter removed about twelve times as much error as remains per degree of freedom.*
- d. Now suppose each of the eight students is in a group of their own, so the model has eight means. Write that model's SS Error, PRE, and df Error, one number each.
- d. *SS Error = 0 (every prediction equals the student's own time), PRE = 1, df Error = $8 - 8 = 0$.*
- e. A ninth student from the same section is about to report a commute time. Which model gives a usable prediction for that student: the two-group model or the eight-group model? Write your choice, and one sentence giving your reason.
- e. *The two-group model. The eight-group model's eight means are the eight observed times, and nothing in the model says which one a ninth student gets; with df Error = 0 there is no estimate of the error around a prediction, and F cannot be computed (its denominator is $0 / 0$). A perfect fit to these eight students is not a model of the ninth.*