Date 10-26-2026

PSYCH 10 · Fall 2026 · Worksheet for Lecture 10: Linear regression · 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 · The cost of a parameter

Eight students, four walkers and four shuttle riders, one mean per group: **SS Model = 72, SS Error = 36, SS Total = 108.** 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.

b. *MS Model = 72 / 1 = 72; MS Error = 36 / 6 = 6.*

c. Compute F = MS Model / MS Error and write it in the Model row.

c. *F = 72 / 6 = 12: the one parameter removed about twelve times as much error as remains per degree of freedom.*

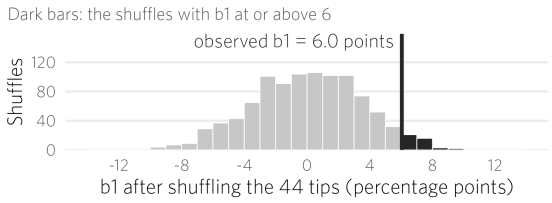
d. Now suppose each of the eight students is in a group of their own, so the model has eight means. Write that model's SS Error, PRE, and df Error, one number each.

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.*

Part 2 · Could the difference be chance?



In 1996 a server drew a smiley face on the checks of 22 tables chosen at random out of 44, and left the other 22 plain. Smiley-face tables tipped 33% of the bill on average and plain-check tables 27%, so b_1 = 6 percentage points. The histogram shows b_1 for 1,000 shuffles of the 44 tips across the two groups.

a. Add the heights of the bars to the right of the line: in about how many of the 1,000 shuffles was b_1 at least 6? Write a whole number.

a. *About 40 (the bars to the right of the line). The exact count in this run is 42 of 1,000.*

b. If the smiley face had no effect, in about how many of every 100 shuffles would a difference of 6 points or more arise?

b. *About 4 of every 100 shuffles, or 1 in 25.*

c. If the smiley face had no effect, would a difference of 6 points or more be common or rare? Write *common* or *rare*, and one sentence giving your reason, using your answer to b.

c. *Rare. A difference of 6 points or more arises in about 4 of every 100 shuffles with no smiley face involved; by the 5% convention from Lecture 7 that is unusual enough to deserve a closer look, although chance is not ruled out. (A good answer can say common or rare as long as it uses the 4 in 100.)*

Part 3 · The slope and the intercept



The figure shows the heights of 1,487 children aged 2 to 12 from the NHANES health survey, with the regression line of height on age. Rounded to simple numbers, the line is **height = 32 + 2.5 x age**, with height in inches and age in years.

a. Compute the line's predicted height for a 4-year-old and for an 8-year-old, in inches.

a. *32 + 2.5 × 4 = 42 inches for a 4-year-old and 32 + 2.5 × 8 = 52 inches for an 8-year-old. (The survey's own mean heights at those ages are 42.1 and 51.9 inches.)*

b. An 8-year-old in the survey is 49 inches tall. Compute this child's residual, in inches, with its sign.

b. *49 - 52 = -3 inches. The residual is the actual height minus the height the line predicts, so a child below the line has a negative residual.*

c. In one sentence: is this child taller or shorter than the line predicts, and by how many inches?

c. *Shorter, by 3 inches: the residual is negative, so this child is 3 inches below the line's prediction for an 8-year-old.*

d. The oldest children in these data are 12. Compute the line's predicted height for a 24-year-old, in inches.

d. *32 + 2.5 × 24 = 92 inches, which is about 7 ft 8 in. (The unrounded line gives 92.7.)*

e. Is that prediction trustworthy? Write *yes* or *no*, and one sentence giving your reason.

e. *No. The line was fitted to children aged 2 to 12, so a prediction at 24 is twelve years past the oldest child in the data, with no observations at that age. Growth also slows after about age 14 and stops in the late teens, which is why the survey's 13-to-17-year-olds fall farther below the line each year. Either reason is a full answer; the point is that the line's arithmetic works at any age while its evidence stops at 12.*