Probability and Statistics Review plus Quiz 1

09 Sep 2026 | Chapters 3–4

ECO 351 – Introduction to Econometrics

Shi Qiu – Fall 2026

Quiz 1: begin here

Time

The quiz occupies the first 15-20 minutes of class. I will announce the stopping time.

1. Open Quiz 1 in Blackboard and answer all six questions.
2. The entire quiz is worth 4 points.
3. Follow the directions shown in Blackboard and submit before time is called.

Quiz 1 debrief: probability and moments

For one day, $P(X = 0) = .50$, $P(X = 1) = .30$, $P(X = 2) = .15$, and $P(X = 3) = .05$.

$$P(X \geq 2) = .15 + .05 = .20$$

$$E(X) = .75, \qquad E(X^2) = 1.35$$

$$\text{Var}(X) = 1.35 - .75^2 = .7875, \qquad SD(X) = .887$$

Interpret the mean

Across many comparable days, the distributor averages 0.75 late deliveries per day. A long-run average need not be a possible count on one day.

Quiz 1 debrief: the sample average has less variation

For 64 independent and identically distributed days, let $\bar{X}$ be the average number of late deliveries per day.

$$E(\bar{X}) = .75, \qquad SE(\bar{X}) = \frac{.887}{\sqrt{64}} = .111$$

$$\bar{X} \overset{approx}{\sim} N(.75, .111^2)$$

The distinction that matters

$SD(X) = .887$ describes variation across individual days. $SE(\bar{X}) = .111$ describes variation across repeated 64-day averages.

Quiz 1 debrief: turn the cutoff into a standard-normal value

$$z = \frac{1.00 - .75}{.111} = 2.25$$

$$2.25 > 1.645 \quad \Longrightarrow \quad P(\bar{X} > 1.00) < .05$$

Why this is an upper-tail comparison

The event asks whether the average exceeds 1.00. Standardizing preserves that direction, so compare with the supplied 5 percent upper-tail critical value.

A new question needs more than a correlation

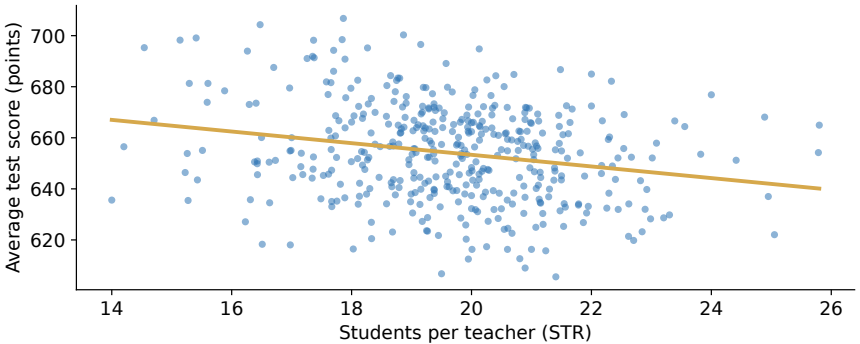
Return to the California school districts. Lecture 5 found a downward association between student-teacher ratio (STR) and average test score, with correlation about -0.226.

Question

How many test-score points does the fitted relationship associate with reducing STR from 20 to 18?

Correlation cannot answer because it has no units. We need a line whose slope is measured in score points per additional student per teacher.

A fitted line summarizes the district cloud



Each point is one district. The line summarizes the linear association but does not pass through every observation.

Write the population relationship as line plus remainder

$$Score_i = \beta_0 + \beta_1 STR_i + u_i$$

1. β_0: the line's intercept.
2. β_1: the line's change in score for a one-unit increase in STR.
3. u_i: determinants of district i's score not represented by the line.

Population quantities

β_0 and β_1 are unknown. The sample will provide estimates $\widehat{\beta}_0$ and $\widehat{\beta}_1$.

The fitted line creates a prediction and a residual

$$\widehat{Score}_i = \hat{\beta}_0 + \hat{\beta}_1 STR_i$$

$$\hat{u}_i = Score_i - \widehat{Score}_i$$

1. The fitted value is the point on the estimated line at the district's STR.
2. The residual is the observed score minus that fitted value: a vertical distance from the line.

Do not merge the notation

u_i is an unobserved population error; $\hat{u}_i$ is calculated after fitting the sample line.

OLS chooses the line with the smallest squared residuals

For any proposed intercept b_0 and slope b_1, add the squared vertical mistakes:

$$SSR(b_0, b_1) = \sum_{i=1}^{n} (Score_i - b_0 - b_1 STR_i)^2$$

$$(\widehat{\beta}_0, \widehat{\beta}_1) = \underset{b_0, b_1}{\arg\min}\ SSR(b_0, b_1)$$

Why squares?

Positive and negative residuals cannot cancel, and a large miss receives more weight than a small miss.

The OLS slope converts covariance into outcome units

$$\widehat{\beta}_1 = \frac{\sum_i (STR_i - \overline{STR})(Score_i - \overline{Score})}{\sum_i (STR_i - \overline{STR})^2} = \frac{s_{STR,Score}}{s^2_{STR}}$$

$$\widehat{\beta}_0 = \overline{Score} - \widehat{\beta}_1 \overline{STR}$$

Connection to Lecture 5

Covariance supplies the direction. Dividing by the variance of STR supplies the scale, so the slope has units of score points per unit of STR.

Work one line from three observations

For constructed observations $(X, Y) = (1, 2), (2, 4), (3, 5)$,

$$\bar{X} = 2, \qquad \bar{Y} = 11/3$$

$$\widehat{\beta}_1 = \frac{(-1)(-5/3) + 0(1/3) + (1)(4/3)}{(-1)^2 + 0^2 + 1^2} = 1.5$$

$$\widehat{\beta}_0 = 11/3 - 1.5(2) = 2/3$$

Fitted line

$\widehat{Y} = 2/3 + 1.5X$. A one-unit increase in X corresponds to 1.5 fitted Y units.

Stata: fit the school regression

Before running the regression, predict the sign of the STR coefficient from the scatterplot.

```
do "setup.do" schools
twoway (scatter score str) (lfit score str)
regress score str
```

Read the result

Stata estimates $\widehat{Score} = 698.93 - 2.280 STR$. The slope is in score points per one additional student per teacher; the sample R^2 is about 0.051.

Read the coefficient table from the research question

$$\widehat{Score} = 698.93 - 2.280 STR$$

1. The STR coefficient is negative: higher STR is associated with lower fitted scores.
2. A one-unit increase in STR corresponds to 2.280 fewer fitted score points.
3. The intercept is the fitted score at STR $= 0$, far outside the observed school setting.

What not to say

The slope is not a correlation and not a percentage. Its units come from score divided by students per teacher.

Answer the opening two-student comparison

$$\widehat{Score}(18) - \widehat{Score}(20) = \widehat{\beta}_1(18 - 20)$$

$$= (-2.280)(-2) = 4.56 \text{ score points}$$

Interpretation

The fitted score is 4.56 points higher at STR 18 than at STR 20. The intercept cancels because both predictions use the same fitted line.

Use the fitted line for one district

Using $\widehat{Score} = 698.93 - 2.280STR$, find the fitted score at STR $= 19$. If the district's actual score is 658, calculate and interpret its residual.

Solution: Use the fitted line for one district

$$\widehat{Score}(19) = 698.93 - 2.280(19) = 655.61$$

$$\widehat{u} = 658 - 655.61 = 2.39$$

The district scores 2.39 points above the fitted line, so its residual is positive.

A fitted association is not automatically a policy effect

Districts with different STR may also differ in family resources, language composition, teacher experience, and other score determinants.

What the regression establishes

In this sample, lower STR is associated with higher fitted scores along the OLS line.

What remains open

Would changing class size within a district change scores by the fitted amount? That requires assumptions about the omitted determinants in u_i.

From probability to the regression line

Conclusion

Quiz 1 joined a daily distribution to the sampling distribution of an average. We then used the same habit of defining a target and an estimator to begin regression: OLS fits a line, and its slope translates an STR change into score points.

Why can the fitted 4.56-point difference be numerically correct yet still fail to equal a class-size policy effect?