HE2003 Econometrics I

Tutorial 2 (Week 3) – Problems & Solutions

Academic Year 2026/2027, Semester 1

Quantitative Research Society @NTU

October 12, 2026

Overview

- **Lecture coverage:** Lecture 2 — *Linear Regression with One Regressor 1*: the simple linear regression model $Y_i = \beta_0 + \beta_1 X_i + u_i$; the OLS minimisation problem; the first-order-condition (FOC) derivation of the closed-form OLS estimators; the summation identities that convert the "$1/n$" form of $\hat{\beta}_1$ into the deviation-from-mean form; and the OLS predicted values and residuals.
- **Questions:** 5 questions from Tutorial Question Set No. 2.
- **Notation:** as in lectures — true unknown parameters β_0, β_1; generic candidate values b_0, b_1 (the arguments of the minimisation); OLS estimators $\hat{\beta}_0, \hat{\beta}_1$ (random variables, functions of the sample); true error u_i; OLS residual $\hat{u}_i$ (never e_i); ESS explained, SSR residual, TSS = ESS + SSR.
- **Question themes:** Q1 the least-squares objective; Q2 parameters vs candidate values vs estimators; Q3 the closed-form OLS estimators; Q4 two deviation-sum identities; Q5 predictions, predicted changes and the sample mean from a fitted line.

Useful rules (summation notation, restated verbatim from the question sheet). Let $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ be two sequences of numbers and let a be a constant. Then

$$\sum_{i=1}^{n} a x_i = a \sum_{i=1}^{n} x_i, \tag{R1}$$

$$\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i, \tag{R2}$$

$$\sum_{i=1}^{n} a = n \cdot a. \tag{R3}$$

These three rules are all that is needed to expand every square and cross-term appearing below.

For academic reference only. This document is provided for educational purposes and does not constitute an official question paper, answer key, or any formally authorised academic material. Its content is not guaranteed to be complete or accurate.

Question 1 (The idea of OLS estimation and the minimisation problem)

Problem

Explain the idea of OLS estimation and the related minimisation problem. *(Tutorial Question Set No. 2, Q1.)*

Solution

Method 1: Least-squares objective built from prediction errors

- The population model is $Y_i = \beta_0 + \beta_1 X_i + u_i$. Here β_0 (in units of Y) and β_1 (units of Y per unit of X) are fixed but unknown population parameters, and the error u_i collects every other factor that moves Y_i. Only the sample (X_i, Y_i), $i = 1, \ldots, n$, is observed, so the line must be chosen from the sample alone.
- Any candidate line $b_0 + b_1 X$ makes the prediction error $Y_i - (b_0 + b_1 X_i)$ at observation i. A positive error means the point lies above the candidate line; a negative error means it lies below.
- The plain sum of errors does not measure fit, because positive and negative errors cancel: every line through $(\bar{X}, \bar{Y})$ has errors summing to zero, since $\sum_{i=1}^{n}\big(Y_i - b_0 - b_1 X_i\big) = n\big(\bar{Y} - b_0 - b_1 \bar{X}\big) = 0$ for such a line. Squaring makes every miss count positively and penalises a large miss more than proportionally (one miss of 4 points costs 16 squared points, two misses of 2 points cost only 8).
- OLS therefore chooses the pair (b_0, b_1) that makes the sum of squared prediction errors as small as possible. The minimisers are the OLS estimators $(\hat{\beta}_0, \hat{\beta}_1)$, and the fitted line $\hat{\beta}_0 + \hat{\beta}_1 X$ is the sample estimate of the population line $\beta_0 + \beta_1 X$.

Write the objective function observation by observation:

$$\begin{aligned} S(b_0, b_1) &= \sum_{i=1}^{n} \big[Y_i - (b_0 + b_1 X_i)\big]^2 \\ &= \big[Y_1 - (b_0 + b_1 X_1)\big]^2 + \big[Y_2 - (b_0 + b_1 X_2)\big]^2 + \cdots + \big[Y_n - (b_0 + b_1 X_n)\big]^2. \end{aligned}$$

Expanding one squared term,

$$\big[Y_i - b_0 - b_1 X_i\big]^2 = Y_i^2 + b_0^2 + b_1^2 X_i^2 - 2b_0 Y_i - 2b_1 X_i Y_i + 2b_0 b_1 X_i.$$

Summing over i with (R1)–(R3),

$$\begin{aligned} S(b_0, b_1) = \sum_{i=1}^{n} Y_i^2 + n b_0^2 + b_1^2 \sum_{i=1}^{n} X_i^2 \\ - 2b_0 \sum_{i=1}^{n} Y_i - 2b_1 \sum_{i=1}^{n} X_i Y_i + 2b_0 b_1 \sum_{i=1}^{n} X_i. \end{aligned}$$

Thus S is a quadratic function of the two unknowns (b_0, b_1) whose coefficients are sample sums, and the minimisation problem is to find the lowest point of this bowl (Question 3 solves it).

For a hypothetical numerical illustration take three classrooms with class sizes $X = (20, 22, 24)$ students and average scores $Y = (651, 645, 627)$ points, so $\bar{X} = 22$ and $\bar{Y} = (651 + 645 + 627)/3 = 641$.
The horizontal line $b_0 = 641$, $b_1 = 0$ passes through $(\bar{X}, \bar{Y})$. Its errors are

$$651 - 641 = 10, \qquad 645 - 641 = 4, \qquad 627 - 641 = -14,$$

which sum to $10 + 4 - 14 = 0$, yet

$$S(641, 0) = 10^2 + 4^2 + (-14)^2 = 100 + 16 + 196 = 312.$$

The candidate line $b_0 = 760$, $b_1 = -5.5$ predicts

$$760 - 5.5 \times 20 = 650, \qquad 760 - 5.5 \times 22 = 639, \qquad 760 - 5.5 \times 24 = 628,$$

so its errors are $1,\ 6,\ -1$ and

$$S(760, -5.5) = 1^2 + 6^2 + (-1)^2 = 1 + 36 + 1 = 38.$$

The OLS line (formula of Question 3) uses $\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) = (-2)(10) + (0)(4) + (2)(-14) = -48$ and $\sum_{i=1}^{n}(X_i - \bar{X})^2 = 4 + 0 + 4 = 8$, so

$$\hat{\beta}_1 = \frac{-48}{8} = -6, \qquad \hat{\beta}_0 = 641 - (-6)(22) = 641 + 132 = 773.$$

It predicts $773 - 120 = 653$, $773 - 132 = 641$, $773 - 144 = 629$, so its errors are $-2,\ 4,\ -2$ and

$$S(773, -6) = (-2)^2 + 4^2 + (-2)^2 = 4 + 16 + 4 = 24.$$

Hence $24 < 38 < 312$ squared points: the OLS line has the smallest sum of squared errors, while the horizontal line has a zero error sum and the largest S. Therefore

$$\boxed{(\hat{\beta}_0, \hat{\beta}_1) = \underset{b_0, b_1}{\arg\min} \sum_{i=1}^{n} \big[Y_i - (b_0 + b_1 X_i)\big]^2.}$$

Method 2: Sample analogue of the population least-squares problem

(uses Least Squares Assumption 1 from Lecture 3 / beyond the Lecture 2 scope)

For any candidate (b_0, b_1) define the population mean squared prediction error

$$M(b_0, b_1) = E\big[(Y - b_0 - b_1 X)^2\big].$$

Substitute $Y = \beta_0 + \beta_1 X + u$:

$$Y - b_0 - b_1 X = u + g(X), \qquad g(X) = (\beta_0 - b_0) + (\beta_1 - b_1)X.$$

Square and take expectations:

$$M(b_0, b_1) = E(u^2) + 2E\big[u\, g(X)\big] + E\big[g(X)^2\big].$$

The middle term splits by linearity of expectation:

$$E\big[u\, g(X)\big] = (\beta_0 - b_0)E(u) + (\beta_1 - b_1)E(Xu).$$

Least Squares Assumption 1, $E(u \mid X) = 0$, and the law of iterated expectations give

$$E(u) = E\big[E(u \mid X)\big] = 0, \qquad E(Xu) = E\big[X\, E(u \mid X)\big] = 0.$$

Hence the middle term is zero and

$$M(b_0, b_1) = E(u^2) + E\big[g(X)^2\big] \;\geq\; E(u^2) = M(\beta_0, \beta_1),$$

because $g(X) \equiv 0$ when $(b_0, b_1) = (\beta_0, \beta_1)$. So the population parameters solve the population least-squares problem $\min_{b_0, b_1} M(b_0, b_1)$. This problem cannot be solved in practice, because M depends on the unknown joint distribution of (X, Y).

The sample analogue replaces the population mean $E[\,\cdot\,]$ by the sample average $\frac{1}{n}\sum_{i=1}^{n}(\,\cdot\,)$:

$$M_n(b_0, b_1) = \frac{1}{n}\sum_{i=1}^{n}\big[Y_i - (b_0 + b_1 X_i)\big]^2 = \frac{1}{n}\, S(b_0, b_1).$$

Since $\frac{1}{n} > 0$, for any two candidates

$$M_n(b_0, b_1) \leq M_n(b_0', b_1') \iff S(b_0, b_1) \leq S(b_0', b_1'),$$

so M_n and S have the same minimiser. In the three-classroom data of Method 1, $M_n(773, -6) = 24/3 = 8$ squared points, the smallest value among the three lines tried there. Therefore the OLS estimators are the sample counterpart of the population parameters:

$$\boxed{(\hat{\beta}_0, \hat{\beta}_1) = \arg\min_{b_0, b_1} \sum_{i=1}^{n}\big[Y_i - (b_0 + b_1 X_i)\big]^2.}$$

Method 3: Vector geometry (closest point of a plane)

This method should be used only if the course allows this interpretation.

Stack the data into column vectors of length n:

$$\mathbf{Y} = (Y_1, \ldots, Y_n)', \qquad \mathbf{1} = (1, \ldots, 1)', \qquad \mathbf{X} = (X_1, \ldots, X_n)'.$$

The candidate fitted vector $b_0\mathbf{1} + b_1\mathbf{X}$ has i-th entry $b_0 + b_1X_i$, so the error vector $\mathbf{Y} - b_0\mathbf{1} - b_1\mathbf{X}$ has i-th entry $Y_i - (b_0 + b_1X_i)$. Its squared Euclidean length is

$$\big\|\mathbf{Y} - b_0\mathbf{1} - b_1\mathbf{X}\big\|^2 = \sum_{i=1}^{n}\big[Y_i - (b_0 + b_1X_i)\big]^2 = S(b_0, b_1).$$

Hence $S(b_0, b_1)$ is the squared distance from the data vector $\mathbf{Y}$ to the point $b_0\mathbf{1} + b_1\mathbf{X}$ of the plane $P = \{b_0\mathbf{1} + b_1\mathbf{X} : \ b_0, b_1 \in \mathbb{R}\}$, and minimising S means finding the point of P closest to $\mathbf{Y}$. (P is a plane, not a line, when the X_i are not all equal, so that $\mathbf{X}$ is not a multiple of $\mathbf{1}$.)

Let $\hat{Y} = \hat{\beta}_0\mathbf{1} + \hat{\beta}_1\mathbf{X}$ be a point of P whose residual vector $\hat{u} = \mathbf{Y} - \hat{Y}$ is orthogonal to both $\mathbf{1}$ and $\mathbf{X}$:

$$\mathbf{1}'\hat{u} = \sum_{i=1}^{n}\hat{u}_i = 0, \qquad \mathbf{X}'\hat{u} = \sum_{i=1}^{n}X_i\hat{u}_i = 0.$$

Such a point exists: these two conditions are two linear equations in $(\hat{\beta}_0, \hat{\beta}_1)$, the normal equations of Question 3, which have a solution when the X_i are not all equal. For any other point $\mathbf{p} = b_0\mathbf{1} + b_1\mathbf{X}$ of P, the vector $\hat{Y} - \mathbf{p}$ lies in P, so $\hat{u}'(\hat{Y} - \mathbf{p}) = (\hat{\beta}_0 - b_0)\mathbf{1}'\hat{u} + (\hat{\beta}_1 - b_1)\mathbf{X}'\hat{u} = 0$. Therefore

$$\|\mathbf{Y} - \mathbf{p}\|^2 = \|\hat{u} + (\hat{Y} - \mathbf{p})\|^2 = \|\hat{u}\|^2 + 2\hat{u}'(\hat{Y} - \mathbf{p}) + \|\hat{Y} - \mathbf{p}\|^2 = \|\hat{u}\|^2 + \|\hat{Y} - \mathbf{p}\|^2 \ \geq \ \|\hat{u}\|^2,$$

which is Pythagoras' theorem: the orthogonal projection of $\mathbf{Y}$ onto P is the closest point. In the three-classroom data of Method 1 the OLS residuals are $\hat{u} = (-2, 4, -2)'$, and

$$\mathbf{1}'\hat{u} = -2 + 4 - 2 = 0, \qquad \mathbf{X}'\hat{u} = 20(-2) + 22(4) + 24(-2) = -40 + 88 - 48 = 0.$$

Hence the coefficients of the projection are the least-squares minimisers:

$$\boxed{(\hat{\beta}_0, \hat{\beta}_1) = \arg\min_{b_0, b_1} \ \sum_{i=1}^{n}\big[Y_i - (b_0 + b_1X_i)\big]^2.}$$

Question 2 (Notation: (β_0, β_1) vs. (b_0, b_1) vs. $(\hat{\beta}_0, \hat{\beta}_1)$)

Problem

Explain the difference of the notations (β_0, β_1), (b_0, b_1), and $(\hat{\beta}_0, \hat{\beta}_1)$ used in class. *(Tutorial Question Set No. 2, Q2.)*

Solution

Method 1: Role of each symbol in the model and in the minimisation problem

- (β_0, β_1) are the intercept and slope of the population regression line $\beta_0 + \beta_1 X$ in $Y_i = \beta_0 + \beta_1 X_i + u_i$. They are fixed, non-random constants that describe the whole population, and they are unknown: no finite sample reveals them exactly. In a test-score regression on class size, β_1 is measured in points per student.
- (b_0, b_1) are generic candidate values: the dummy arguments of the objective function $S(b_0, b_1) = \sum_{i=1}^{n} [Y_i - (b_0 + b_1 X_i)]^2$. They may be set to any pair of real numbers, they are not attached to any estimator, and they exist only so that the minimisation problem can be written down and solved.
- $(\hat{\beta}_0, \hat{\beta}_1)$ are the OLS estimators: the particular candidate values that minimise S. Because they are formulas in the sample $(X_i, Y_i)_{i=1}^n$, they are random variables (sample statistics); the numbers they take on one realised sample are the estimates of (β_0, β_1).
- The population pair (β_0, β_1) goes with the unobservable error $u_i = Y_i - \beta_0 - \beta_1 X_i$; the sample pair $(\hat{\beta}_0, \hat{\beta}_1)$ goes with the observable residual $\hat{u}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$.

For a numerical illustration use the three hypothetical classrooms $X = (20, 22, 24)$ students, $Y = (651, 645, 627)$ points of Question 1. The candidate $(b_0, b_1) = (700, -3)$ predicts

$$700 - 3 \times 20 = 640, \qquad 700 - 3 \times 22 = 634, \qquad 700 - 3 \times 24 = 628,$$

so its errors are $11,\ 11,\ -1$ and

$$S(700, -3) = 11^2 + 11^2 + (-1)^2 = 121 + 121 + 1 = 243.$$

The minimiser on this sample is $(\hat{\beta}_0, \hat{\beta}_1) = (773, -6)$, with

$$S(773, -6) = (-2)^2 + 4^2 + (-2)^2 = 24.$$

So $(700, -3)$ is one value of the candidate pair (b_0, b_1); $(773, -6)$ is the estimate, the value of the estimator $(\hat{\beta}_0, \hat{\beta}_1)$ on this sample; and (β_0, β_1) is not computed from the data at all.

(β_0, β_1): true unknown parameters (fixed); (b_0, b_1): candidate arguments of S; $(\hat{\beta}_0, \hat{\beta}_1)$: minimisers of S, random variables (estimates on one sample).

Method 2: Simulated data-generating process with two samples

Fix the population parameters at hypothetical known values, so that all three objects can be seen side by side:

$$\beta_0 = 760, \qquad \beta_1 = -5.5 \text{ points per student}, \qquad X = (20, 22, 24).$$

The population line gives

$$760 - 5.5 \times 20 = 650, \qquad 760 - 5.5 \times 22 = 639, \qquad 760 - 5.5 \times 24 = 628.$$

Draw two samples that differ only in their errors:

$$\begin{aligned} &\text{Sample A: } u = (1, 6, -1) && \implies Y = (651, 645, 627), \\ &\text{Sample B: } u = (4, -2, -2) && \implies Y = (654, 637, 626). \end{aligned}$$

In both samples $\bar{X} = 22$ and the deviations are $X_i - \bar{X} = (-2, 0, 2)$, so $\sum_{i=1}^{n}(X_i - \bar{X})^2 = 4 + 0 + 4 = 8$.
Sample A: $\bar{Y} = (651 + 645 + 627)/3 = 641$ and $Y_i - \bar{Y} = (10, 4, -14)$. Then

$$\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) = (-2)(10) + (0)(4) + (2)(-14) = -48,$$

$$\hat{\beta}_1 = \frac{-48}{8} = -6, \qquad \hat{\beta}_0 = 641 - (-6)(22) = 773.$$

Sample B: $\bar{Y} = (654 + 637 + 626)/3 = 639$ and $Y_i - \bar{Y} = (15, -2, -13)$. Then

$$\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) = (-2)(15) + (0)(-2) + (2)(-13) = -56,$$

$$\hat{\beta}_1 = \frac{-56}{8} = -7, \qquad \hat{\beta}_0 = 639 - (-7)(22) = 793.$$

Now read off the three objects.

- $(\beta_0, \beta_1) = (760, -5.5)$ is the same in both samples: it is a fixed population constant.
- $(\hat{\beta}_0, \hat{\beta}_1)$ moves from $(773, -6)$ in sample A to $(793, -7)$ in sample B: it is a function of the sample, hence a random variable.
- (b_0, b_1) is any point at which S can be evaluated. Evaluating each sample's S at the candidate $(b_0, b_1) = (760, -5.5)$ gives the squared errors u_i^2:

  $$S_A(760, -5.5) = 1^2 + 6^2 + (-1)^2 = 38,$$
  $$S_B(760, -5.5) = 4^2 + (-2)^2 + (-2)^2 = 24.$$

  At the estimates, sample A has fitted values $653, 641, 629$ and residuals $(-2, 4, -2)$; sample B has fitted values $793 - 7 \times 20 = 653$, $793 - 7 \times 22 = 639$, $793 - 7 \times 24 = 625$ and residuals $(654 - 653,\ 637 - 639,\ 626 - 625) = (1, -2, 1)$. Hence

  $$S_A(773, -6) = (-2)^2 + 4^2 + (-2)^2 = 24 < 38,$$
  $$S_B(793, -7) = 1^2 + (-2)^2 + 1^2 = 6 < 24.$$

Thus the true parameters, used as a candidate, do not minimise the sample objective; the minimiser is $(\hat{\beta}_0, \hat{\beta}_1)$, which differs from (β_0, β_1) because the realised errors do not satisfy $\sum_{i=1}^n u_i = 0$ and $\sum_{i=1}^n (X_i - \bar{X})u_i = 0$ (sample A: $\sum_{i=1}^n u_i = 6$ and $\sum_{i=1}^n (X_i - \bar{X})u_i = -2 + 0 - 2 = -4$).

(β_0, β_1): true unknown parameters (fixed);
(b_0, b_1): candidate arguments of S;
$(\hat{\beta}_0, \hat{\beta}_1)$: minimisers of S, random variables (estimates on one sample).

Unofficial - QRS@NTU - qrs.ntu.org

Question 3 (Deriving the OLS estimators from the FOCs)

Problem

Derive the formula of the OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ following the steps in the lecture note. *(Tutorial Question Set No. 2, Q3.)*

Solution

Method 1: First-order conditions and normal equations

Throughout, assume the X_i are not all equal, so that $\sum_{i=1}^{n}(X_i - \bar{X})^2 > 0$. Let

$$S(b_0, b_1) = \sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i)^2.$$

Step 1: first-order conditions. By the chain rule, each term differentiates as

$$\frac{\partial}{\partial b_0}(Y_i - b_0 - b_1 X_i)^2 = 2(Y_i - b_0 - b_1 X_i)(-1),$$
$$\frac{\partial}{\partial b_1}(Y_i - b_0 - b_1 X_i)^2 = 2(Y_i - b_0 - b_1 X_i)(-X_i).$$

Summing over i,

$$\frac{\partial S}{\partial b_0} = -2\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i),$$
$$\frac{\partial S}{\partial b_1} = -2\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i)X_i.$$

At the minimiser $(\hat{\beta}_0, \hat{\beta}_1)$ both derivatives are zero. Dividing each by -2,

$$\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0, \qquad \text{(FOC } b_0\text{)}$$
$$\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)X_i = 0. \qquad \text{(FOC } b_1\text{)}$$

Step 2: normal equations. Expand (FOC b_0) with (R2), then (R3) and (R1):

$$\sum_{i=1}^{n} Y_i - \sum_{i=1}^{n}\hat{\beta}_0 - \sum_{i=1}^{n}\hat{\beta}_1 X_i = \sum_{i=1}^{n} Y_i - n\hat{\beta}_0 - \hat{\beta}_1\sum_{i=1}^{n} X_i = 0.$$

Divide by n and use $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$, $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$:

$$\bar{Y} - \hat{\beta}_0 - \hat{\beta}_1\bar{X} = 0. \qquad (4)$$

Expand (FOC b_1) the same way:

$$\sum_{i=1}^{n} X_i Y_i - \hat{\beta}_0 \sum_{i=1}^{n} X_i - \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 = 0.$$

Divide by n:

$$\frac{1}{n}\sum_{i=1}^{n} X_i Y_i - \hat{\beta}_0 \bar{X} - \hat{\beta}_1 \frac{1}{n}\sum_{i=1}^{n} X_i^2 = 0. \tag{5}$$

Step 3: solve (4) for the intercept. Adding $\hat{\beta}_0$ to both sides of (4),

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}.$$

Step 4: substitute into (5) and isolate the slope. Replace $\hat{\beta}_0$ in (5) by $\bar{Y} - \hat{\beta}_1 \bar{X}$:

$$\frac{1}{n}\sum_{i=1}^{n} X_i Y_i - \big(\bar{Y} - \hat{\beta}_1 \bar{X}\big)\bar{X} - \hat{\beta}_1 \frac{1}{n}\sum_{i=1}^{n} X_i^2 = 0.$$

Expand the bracket, $(\bar{Y} - \hat{\beta}_1 \bar{X})\bar{X} = \bar{X}\bar{Y} - \hat{\beta}_1 \bar{X}^2$:

$$\frac{1}{n}\sum_{i=1}^{n} X_i Y_i - \bar{X}\bar{Y} + \hat{\beta}_1 \bar{X}^2 - \hat{\beta}_1 \frac{1}{n}\sum_{i=1}^{n} X_i^2 = 0.$$

Move the two $\hat{\beta}_1$ terms to the right-hand side:

$$\frac{1}{n}\sum_{i=1}^{n} X_i Y_i - \bar{X}\bar{Y} = \hat{\beta}_1 \Big(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2\Big).$$

By Question 4(a), $\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 > 0$, so we may divide by it:

$$\hat{\beta}_1 = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i Y_i - \bar{X}\bar{Y}}{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2}.$$

Step 5: deviation form. Multiply numerator and denominator by n, then use Question 4(a) and 4(b):

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

Multiplying numerator and denominator by the same constant $\frac{1}{n-1}$ leaves the ratio unchanged:

$$\hat{\beta}_1 = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{s_{XY}}{s_X^2},$$

with $s_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$ the sample covariance and $s_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ the sample variance of X.

Step 6: second-order condition. Differentiate the first derivatives once more:

$$\frac{\partial^2 S}{\partial b_0^2} = \frac{\partial}{\partial b_0}\Big[-2\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i)\Big] = -2\sum_{i=1}^{n}(-1) = 2n,$$
$$\frac{\partial^2 S}{\partial b_0\,\partial b_1} = \frac{\partial}{\partial b_1}\Big[-2\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i)\Big] = -2\sum_{i=1}^{n}(-X_i) = 2\sum_{i=1}^{n} X_i,$$
$$\frac{\partial^2 S}{\partial b_1^2} = \frac{\partial}{\partial b_1}\Big[-2\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i)X_i\Big] = -2\sum_{i=1}^{n}(-X_i)X_i = 2\sum_{i=1}^{n} X_i^2.$$

The Hessian is therefore the same at every (b_0, b_1):

$$H = 2\begin{pmatrix} n & \sum_{i=1}^{n} X_i \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 \end{pmatrix}.$$

Its leading principal minor is $2n > 0$. Its determinant, using $\sum_{i=1}^{n} X_i = n\bar{X}$ and Question 4(a), is

$$\begin{aligned}\det H &= (2n)\Big(2\sum_{i=1}^{n} X_i^2\Big) - \Big(2\sum_{i=1}^{n} X_i\Big)^2 = 4\Big[n\sum_{i=1}^{n} X_i^2 - \Big(\sum_{i=1}^{n} X_i\Big)^2\Big] \\ &= 4\Big[n\sum_{i=1}^{n} X_i^2 - n^2\bar{X}^2\Big] = 4n\Big[\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\Big] = 4n\sum_{i=1}^{n}(X_i - \bar{X})^2 > 0.\end{aligned}$$

Hence H is positive definite everywhere, S is strictly convex, and the stationary point is the unique global minimum. Therefore

$$\boxed{\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{s_{XY}}{s_X^2}, \qquad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}.}$$

Unofficial solutions, not endorsed by the course instructor

Method 2: Centred regressor

Assume the X_i are not all equal. For any candidate line, add and subtract $b_1\bar{X}$:

$$b_0 + b_1 X_i = (b_0 + b_1\bar{X}) + b_1(X_i - \bar{X}) = a + b_1(X_i - \bar{X}), \qquad a = b_0 + b_1\bar{X}.$$

Hence $S(b_0, b_1) = T(a, b_1)$ with

$$T(a, b_1) = \sum_{i=1}^{n} \big[Y_i - a - b_1(X_i - \bar{X})\big]^2.$$

The map $(b_0, b_1) \mapsto (a, b_1)$ is one-to-one with inverse $b_0 = a - b_1\bar{X}$, so minimising T over (a, b_1) and translating back gives the minimiser of S. First record the deviation-sum identity:

$$\sum_{i=1}^{n}(X_i - \bar{X}) = \sum_{i=1}^{n} X_i - n\bar{X} = n\bar{X} - n\bar{X} = 0.$$

FOC for a.

$$\frac{\partial T}{\partial a} = -2\sum_{i=1}^{n}\big[Y_i - a - b_1(X_i - \bar{X})\big].$$

Setting it to zero at $(\hat{a}, \hat{b}_1)$ and dividing by -2:

$$\sum_{i=1}^{n} Y_i - n\hat{a} - \hat{b}_1\sum_{i=1}^{n}(X_i - \bar{X}) = 0.$$

The last sum is zero, so

$$n\bar{Y} - n\hat{a} = 0 \implies \hat{a} = \bar{Y}.$$

FOC for b_1.

$$\frac{\partial T}{\partial b_1} = -2\sum_{i=1}^{n}\big[Y_i - a - b_1(X_i - \bar{X})\big](X_i - \bar{X}).$$

Setting it to zero at $(\hat{a}, \hat{b}_1) = (\bar{Y}, \hat{b}_1)$ and dividing by -2:

$$\sum_{i=1}^{n}\big[(Y_i - \bar{Y}) - \hat{b}_1(X_i - \bar{X})\big](X_i - \bar{X}) = 0.$$

Split the sum with (R1), (R2):

$$\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) - \hat{b}_1\sum_{i=1}^{n}(X_i - \bar{X})^2 = 0 \implies \hat{b}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

Second-order condition.

$$\frac{\partial^2 T}{\partial a^2} = 2n, \qquad \frac{\partial^2 T}{\partial a\,\partial b_1} = 2\sum_{i=1}^{n}(X_i - \bar{X}) = 0, \qquad \frac{\partial^2 T}{\partial b_1^2} = 2\sum_{i=1}^{n}(X_i - \bar{X})^2 > 0.$$

The Hessian $\begin{pmatrix} 2n & 0 \\ 0 & 2\sum_{i=1}^{n}(X_i - \bar{X})^2 \end{pmatrix}$ is diagonal with positive entries, hence positive definite, so $(\hat{a}, \hat{b}_1)$ is the unique global minimiser of T.

Translate back.

$$\hat{\beta}_1 = \hat{b}_1, \qquad \hat{\beta}_0 = \hat{a} - \hat{b}_1\bar{X} = \bar{Y} - \hat{\beta}_1\bar{X}.$$

Dividing numerator and denominator of $\hat{\beta}_1$ by $n-1$ gives the ratio s_{XY}/s_X^2, where

$$s_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}), \qquad s_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Therefore

$$\boxed{\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{s_{XY}}{s_X^2}, \qquad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}.}$$

Unofficial_QRS@NTU_github.io

Method 3: Completing the square (no derivatives)

Assume the X_i are not all equal, and write the deviation sums

$$D_{XX} = \sum_{i=1}^{n}(X_i - \bar{X})^2 > 0, \qquad D_{XY} = \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}), \qquad D_{YY} = \sum_{i=1}^{n}(Y_i - \bar{Y})^2.$$

Step 1: minimise over b_0 for fixed b_1. Fix b_1 and let $c_i = Y_i - b_1 X_i$, with mean $\bar{c} = \bar{Y} - b_1\bar{X}$. Then $S(b_0, b_1) = \sum_{i=1}^{n}(c_i - b_0)^2$. Add and subtract $\bar{c}$:

$$(c_i - b_0)^2 = \big[(c_i - \bar{c}) + (\bar{c} - b_0)\big]^2 = (c_i - \bar{c})^2 + 2(c_i - \bar{c})(\bar{c} - b_0) + (\bar{c} - b_0)^2.$$

Sum over i with (R1)–(R3):

$$S(b_0, b_1) = \sum_{i=1}^{n}(c_i - \bar{c})^2 + 2(\bar{c} - b_0)\sum_{i=1}^{n}(c_i - \bar{c}) + n(\bar{c} - b_0)^2.$$

Since $\sum_{i=1}^{n}(c_i - \bar{c}) = n\bar{c} - n\bar{c} = 0$,

$$S(b_0, b_1) = \sum_{i=1}^{n}(c_i - \bar{c})^2 + n(\bar{c} - b_0)^2 \;\geq\; \sum_{i=1}^{n}(c_i - \bar{c})^2,$$

with equality if and only if $b_0 = \bar{c} = \bar{Y} - b_1\bar{X}$.

Step 2: minimise the remaining function of b_1. Since $c_i - \bar{c} = (Y_i - \bar{Y}) - b_1(X_i - \bar{X})$, the value left after Step 1 is

$$P(b_1) = \sum_{i=1}^{n}\big[(Y_i - \bar{Y}) - b_1(X_i - \bar{X})\big]^2.$$

Expand the square and sum:

$$P(b_1) = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 - 2b_1\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) + b_1^2\sum_{i=1}^{n}(X_i - \bar{X})^2 = D_{YY} - 2b_1 D_{XY} + b_1^2 D_{XX}.$$

Complete the square in b_1:

$$\begin{aligned} P(b_1) &= D_{XX}\Big[b_1^2 - 2\frac{D_{XY}}{D_{XX}}b_1\Big] + D_{YY} \\ &= D_{XX}\Big[\Big(b_1 - \frac{D_{XY}}{D_{XX}}\Big)^2 - \frac{D_{XY}^2}{D_{XX}^2}\Big] + D_{YY} \\ &= D_{XX}\Big(b_1 - \frac{D_{XY}}{D_{XX}}\Big)^2 + D_{YY} - \frac{D_{XY}^2}{D_{XX}}. \end{aligned}$$

Because $D_{XX} > 0$, the first term is non-negative and is zero only at $b_1 = D_{XY}/D_{XX}$.

Step 3: combine. For every (b_0, b_1),

$$S(b_0, b_1) \;\geq\; P(b_1) \;\geq\; D_{YY} - \frac{D_{XY}^2}{D_{XX}},$$

with both equalities if and only if $b_1 = D_{XY}/D_{XX}$ and $b_0 = \bar{Y} - b_1\bar{X}$. So the global minimiser is unique and equals

$$\hat{\beta}_1 = \frac{D_{XY}}{D_{XX}} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}, \qquad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}.$$

Dividing D_{XY} and D_{XX} by $n-1$ gives $s_{XY} = \frac{1}{n-1}D_{XY}$ and $s_X^2 = \frac{1}{n-1}D_{XX}$, so

$$\boxed{\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{s_{XY}}{s_X^2}, \qquad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}.}$$

Unofficial QRS@NTU - qrs.ntu.org

Method 4: Orthogonality of residuals solved by Cramer's rule

This method should be used only if the course allows this interpretation.

Assume the X_i are not all equal. Least Squares Assumption 1, $E(u_i \mid X_i) = 0$, implies the population moment conditions $E(u_i) = 0$ and $E(X_i u_i) = 0$. Their sample analogues require the residuals $\hat{u}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$ to satisfy

$$\frac{1}{n}\sum_{i=1}^{n} \hat{u}_i = 0, \qquad \frac{1}{n}\sum_{i=1}^{n} X_i \hat{u}_i = 0.$$

Step 1: the linear system. Expand each condition with (R1)–(R3) and multiply by n:

$$\begin{aligned} n\,\hat{\beta}_0 + \Big(\sum_{i=1}^{n} X_i\Big)\hat{\beta}_1 &= \sum_{i=1}^{n} Y_i, \\ \Big(\sum_{i=1}^{n} X_i\Big)\hat{\beta}_0 + \Big(\sum_{i=1}^{n} X_i^2\Big)\hat{\beta}_1 &= \sum_{i=1}^{n} X_i Y_i. \end{aligned}$$

The determinant of the coefficient matrix, using $\sum_{i=1}^{n} X_i = n\bar{X}$ and Question 4(a), is

$$\Delta = n\sum_{i=1}^{n} X_i^2 - \Big(\sum_{i=1}^{n} X_i\Big)^2 = n\Big(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\Big) = n\sum_{i=1}^{n}(X_i - \bar{X})^2 > 0.$$

Step 2: Cramer's rule for the slope.

$$\hat{\beta}_1 = \frac{n\sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{\Delta} = \frac{n\sum_{i=1}^{n} X_i Y_i - (n\bar{X})(n\bar{Y})}{n\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

By Question 4(b) the numerator is $\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$, so

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

Step 3: Cramer's rule for the intercept.

$$\hat{\beta}_0 = \frac{\sum_{i=1}^{n} X_i^2 \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} X_i Y_i}{\Delta} = \frac{n\bar{Y}\sum_{i=1}^{n} X_i^2 - n\bar{X}\sum_{i=1}^{n} X_i Y_i}{\Delta}.$$

To compare with $\bar{Y} - \hat{\beta}_1 \bar{X}$, put that expression over the same denominator:

$$\begin{aligned} \bar{Y} - \hat{\beta}_1 \bar{X} &= \frac{\bar{Y}\,\Delta - \bar{X}\big(n\sum_{i=1}^{n} X_i Y_i - n^2 \bar{X}\bar{Y}\big)}{\Delta} \\ &= \frac{n\bar{Y}\sum_{i=1}^{n} X_i^2 - n^2\bar{X}^2\bar{Y} - n\bar{X}\sum_{i=1}^{n} X_i Y_i + n^2\bar{X}^2\bar{Y}}{\Delta} \\ &= \frac{n\bar{Y}\sum_{i=1}^{n} X_i^2 - n\bar{X}\sum_{i=1}^{n} X_i Y_i}{\Delta} = \hat{\beta}_0. \end{aligned}$$

Step 4: the solution minimises S (Pythagoras). Take any (b_0, b_1) and let $d_0 = \hat{\beta}_0 - b_0$, $d_1 = \hat{\beta}_1 - b_1$. Then

$$Y_i - b_0 - b_1 X_i = \hat{u}_i + d_0 + d_1 X_i.$$

Square and sum:

$$S(b_0, b_1) = \sum_{i=1}^{n} \hat{u}_i^2 + 2d_0 \sum_{i=1}^{n} \hat{u}_i + 2d_1 \sum_{i=1}^{n} X_i \hat{u}_i + \sum_{i=1}^{n} (d_0 + d_1 X_i)^2.$$

The two middle sums are zero by the orthogonality conditions, so

$$S(b_0, b_1) = \mathrm{SSR} + \sum_{i=1}^{n} (d_0 + d_1 X_i)^2 \;\geq\; \mathrm{SSR}, \qquad \mathrm{SSR} = \sum_{i=1}^{n} \hat{u}_i^2.$$

Equality needs $d_0 + d_1 X_i = 0$ for every i. If $d_1 \neq 0$ this forces $X_i = -d_0/d_1$ for every i, contradicting that the X_i are not all equal; so $d_1 = 0$, and then $d_0 = 0$. Hence $(\hat{\beta}_0, \hat{\beta}_1)$ is the unique global minimiser of S. With $s_{XY} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$ and $s_X^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$,

$$\boxed{\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2} = \frac{s_{XY}}{s_X^2}, \qquad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}.}$$

Unofficial QRS@NTU Tutorial

Question 4 (Two summation identities)

Problem

Show the following holds:

$$\frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}=\frac{\frac{1}{n}\sum_{i=1}^{n}X_iY_i-\bar{X}\bar{Y}}{\frac{1}{n}\sum_{i=1}^{n}X_i^2-\bar{X}^2}, \qquad \text{where } \bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i,\ \bar{Y}=\frac{1}{n}\sum_{i=1}^{n}Y_i.$$

(Hint: show $\sum_{i=1}^{n}(X_i-\bar{X})^2=\sum_{i=1}^{n}X_i^2-n\bar{X}^2$ *and* $\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})=\sum_{i=1}^{n}X_iY_i-n\bar{X}\bar{Y}$*.) (Tutorial Question Set No. 2, Q4.)*

Solution

Method 1: Direct expansion

Throughout, the definitions of the sample means give, after multiplying by n,

$$\sum_{i=1}^{n}X_i=n\bar{X}, \qquad \sum_{i=1}^{n}Y_i=n\bar{Y}.$$

The ratio requires the X_i not to be all equal, so that its denominator $\sum_{i=1}^{n}(X_i-\bar{X})^2$ is positive.

Part (a). Expand one squared deviation:

$$(X_i-\bar{X})^2=X_i^2-2\bar{X}X_i+\bar{X}^2.$$

Sum over i; split the sum with (R2), pull out the constant $-2\bar{X}$ with (R1), and sum the constant $\bar{X}^2$ with (R3):

$$\begin{aligned}\sum_{i=1}^{n}(X_i-\bar{X})^2&=\sum_{i=1}^{n}X_i^2+\sum_{i=1}^{n}(-2\bar{X}X_i)+\sum_{i=1}^{n}\bar{X}^2\\&=\sum_{i=1}^{n}X_i^2-2\bar{X}\sum_{i=1}^{n}X_i+n\bar{X}^2\\&=\sum_{i=1}^{n}X_i^2-2\bar{X}(n\bar{X})+n\bar{X}^2\\&=\sum_{i=1}^{n}X_i^2-2n\bar{X}^2+n\bar{X}^2\\&=\boxed{\sum_{i=1}^{n}X_i^2-n\bar{X}^2.}\end{aligned}$$

Part (b). Expand one cross-product:

$$(X_i-\bar{X})(Y_i-\bar{Y})=X_iY_i-\bar{Y}X_i-\bar{X}Y_i+\bar{X}\bar{Y}.$$

Sum over i with (R2), then (R1) for the constants $\bar{Y}$ and $\bar{X}$, and (R3) for the constant $\bar{X}\bar{Y}$:

$$\begin{aligned}\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y}) &= \sum_{i=1}^{n}X_iY_i-\bar{Y}\sum_{i=1}^{n}X_i-\bar{X}\sum_{i=1}^{n}Y_i+n\bar{X}\bar{Y}\\ &= \sum_{i=1}^{n}X_iY_i-\bar{Y}(n\bar{X})-\bar{X}(n\bar{Y})+n\bar{X}\bar{Y}\\ &= \sum_{i=1}^{n}X_iY_i-n\bar{X}\bar{Y}-n\bar{X}\bar{Y}+n\bar{X}\bar{Y}\\ &= \boxed{\sum_{i=1}^{n}X_iY_i-n\bar{X}\bar{Y}.}\end{aligned}$$

Ratio. Substitute (b) into the numerator and (a) into the denominator, then factor n out of each:

$$\begin{aligned}\frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sum_{i=1}^{n}(X_i-\bar{X})^2} &= \frac{\sum_{i=1}^{n}X_iY_i-n\bar{X}\bar{Y}}{\sum_{i=1}^{n}X_i^2-n\bar{X}^2}\\ &= \frac{n\Big(\frac{1}{n}\sum_{i=1}^{n}X_iY_i-\bar{X}\bar{Y}\Big)}{n\Big(\frac{1}{n}\sum_{i=1}^{n}X_i^2-\bar{X}^2\Big)}.\end{aligned}$$

Cancelling the common factor n,

$$\boxed{\frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}=\frac{\frac{1}{n}\sum_{i=1}^{n}X_iY_i-\bar{X}\bar{Y}}{\frac{1}{n}\sum_{i=1}^{n}X_i^2-\bar{X}^2}.}$$

Method 2: Using $\sum_{i=1}^{n}(X_i - \bar{X}) = 0$

First show that deviations from the mean sum to zero. By (R2), (R3) and $\sum_{i=1}^{n} X_i = n\bar{X}$,

$$\sum_{i=1}^{n}(X_i - \bar{X}) = \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \bar{X} = n\bar{X} - n\bar{X} = 0.$$

As in Method 1, the ratio requires the X_i not to be all equal.

Part (b). Split the second factor: $(X_i - \bar{X})(Y_i - \bar{Y}) = (X_i - \bar{X})Y_i - (X_i - \bar{X})\bar{Y}$. Sum with (R2) and pull out $\bar{Y}$ with (R1):

$$\begin{aligned}
\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) &= \sum_{i=1}^{n}(X_i - \bar{X})Y_i - \bar{Y}\sum_{i=1}^{n}(X_i - \bar{X}) \\
&= \sum_{i=1}^{n}(X_i - \bar{X})Y_i - \bar{Y} \cdot 0 \\
&= \sum_{i=1}^{n} X_i Y_i - \bar{X}\sum_{i=1}^{n} Y_i \\
&= \boxed{\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}.}
\end{aligned}$$

Part (a). The same steps with X_i in place of Y_i (so $\bar{X}$ in place of $\bar{Y}$):

$$\begin{aligned}
\sum_{i=1}^{n}(X_i - \bar{X})^2 &= \sum_{i=1}^{n}(X_i - \bar{X})X_i - \bar{X}\sum_{i=1}^{n}(X_i - \bar{X}) \\
&= \sum_{i=1}^{n}(X_i - \bar{X})X_i - \bar{X} \cdot 0 \\
&= \sum_{i=1}^{n} X_i^2 - \bar{X}\sum_{i=1}^{n} X_i \\
&= \boxed{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2.}
\end{aligned}$$

Ratio. Divide both results by n:

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i - \bar{X}\bar{Y}, \qquad \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2.$$

A ratio is unchanged when numerator and denominator are both divided by n, so

$$\frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

Substituting the two divided results,

$$\boxed{\frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i Y_i - \bar{X}\bar{Y}}{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2}.}$$

Method 3: Add-and-subtract decomposition with polarisation

As before, $\sum_{i=1}^n X_i = n\bar{X}$, $\sum_{i=1}^n Y_i = n\bar{Y}$, deviations sum to zero ($\sum_{i=1}^n (X_i - \bar{X}) = n\bar{X} - n\bar{X} = 0$), and the ratio requires the X_i not to be all equal.

Part (a). Write each observation as its deviation plus the mean, $X_i = (X_i - \bar{X}) + \bar{X}$, and square:

$$X_i^2 = (X_i - \bar{X})^2 + 2\bar{X}(X_i - \bar{X}) + \bar{X}^2.$$

Sum over i with (R1)–(R3):

$$\sum_{i=1}^n X_i^2 = \sum_{i=1}^n (X_i - \bar{X})^2 + 2\bar{X}\sum_{i=1}^n (X_i - \bar{X}) + n\bar{X}^2 = \sum_{i=1}^n (X_i - \bar{X})^2 + 0 + n\bar{X}^2.$$

Subtracting $n\bar{X}^2$ from both sides,

$$\boxed{\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2.}$$

This holds for any sequence of numbers, in particular for Y_i and for $Z_i = X_i + Y_i$.

Part (b) by polarisation. Let $Z_i = X_i + Y_i$, so $\bar{Z} = \bar{X} + \bar{Y}$ and $Z_i - \bar{Z} = (X_i - \bar{X}) + (Y_i - \bar{Y})$. Part (a) for Z reads

$$\sum_{i=1}^n (Z_i - \bar{Z})^2 = \sum_{i=1}^n Z_i^2 - n\bar{Z}^2.$$

Expand the left-hand side:

$$\sum_{i=1}^n (Z_i - \bar{Z})^2 = \sum_{i=1}^n (X_i - \bar{X})^2 + 2\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) + \sum_{i=1}^n (Y_i - \bar{Y})^2.$$

Expand the right-hand side, using $Z_i^2 = X_i^2 + 2X_iY_i + Y_i^2$ and $\bar{Z}^2 = \bar{X}^2 + 2\bar{X}\bar{Y} + \bar{Y}^2$:

$$\sum_{i=1}^n Z_i^2 - n\bar{Z}^2 = \Big[\sum_{i=1}^n X_i^2 - n\bar{X}^2\Big] + 2\Big[\sum_{i=1}^n X_iY_i - n\bar{X}\bar{Y}\Big] + \Big[\sum_{i=1}^n Y_i^2 - n\bar{Y}^2\Big].$$

By part (a) for X and for Y, the first bracket equals $\sum_{i=1}^n (X_i - \bar{X})^2$ and the third equals $\sum_{i=1}^n (Y_i - \bar{Y})^2$. Equating the two expansions and cancelling these equal terms from both sides leaves

$$2\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = 2\Big[\sum_{i=1}^n X_iY_i - n\bar{X}\bar{Y}\Big].$$

Dividing by 2,

$$\boxed{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^n X_iY_i - n\bar{X}\bar{Y}.}$$

Ratio. Substitute (b) and (a), and divide numerator and denominator by n:

$$\frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{\sum_{i=1}^n X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^n X_i^2 - n\bar{X}^2} = \frac{\frac{1}{n}\sum_{i=1}^n X_i Y_i - \bar{X}\bar{Y}}{\frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}^2}.$$

Therefore

$$\boxed{\frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{\frac{1}{n}\sum_{i=1}^n X_i Y_i - \bar{X}\bar{Y}}{\frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}^2}.}$$

Unofficial_QRS@NTU - qrs.ntu.org

Question 5 (Class size and test scores: numerical predictions)

Problem

Suppose that a researcher, using data on class size (CS) and average test scores from 50 third-grade classes, estimates the OLS regression

$$\widehat{\text{TestScore}} = 640.3 - 4.93 \times \text{CS}.$$

(a) A classroom has 25 students. What is the OLS regression's prediction for that classroom's average test score?

(b) Last year a classroom had 21 students, and this year it has 24 students. What is the OLS regression's prediction for the change in the classroom average test score?

(c) The sample average class size across the 50 classrooms is 22.8. What is the sample average of the test scores across the 50 classrooms? *(Hint: review the formulas of the OLS estimators.)*

(Tutorial Question Set No. 2, Q5.)

Solution

Throughout, the estimated intercept and slope are $\hat{\beta}_0 = 640.3$ points and $\hat{\beta}_1 = -4.93$ points per student, and $n = 50$ classrooms.

Part (a): prediction at $\text{CS} = 25$

Method 1: Direct substitution into the fitted line

The fitted value at $\text{CS} = 25$ is $\hat{\beta}_0 + \hat{\beta}_1 \times 25$. The slope term is

$$4.93 \times 25 = \frac{4.93 \times 100}{4} = \frac{493}{4} = 123.25.$$

The prediction is a point on the sample regression line $\hat{\beta}_0 + \hat{\beta}_1\text{CS}$, not on the unknown population line $\beta_0 + \beta_1\text{CS}$, and the classroom's actual average differs from it by a prediction error (its residual $\hat{u}_i$ if the classroom is one of the 50 in the sample). Hence the estimated line predicts, for a classroom of 25 students, an average score of

$$\widehat{\text{TestScore}} = 640.3 - 4.93 \times 25 \implies \boxed{\widehat{\text{TestScore}} = 640.3 - 123.25 = 517.05.}$$

Method 2: Mean-point (centred) form of the fitted line

Using $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}$ (Question 3), the fitted line can be rewritten around the sample means:

$$\widehat{\text{TestScore}} = \hat{\beta}_0 + \hat{\beta}_1\text{CS} = \big(\overline{\text{TestScore}} - \hat{\beta}_1\overline{\text{CS}}\big) + \hat{\beta}_1\text{CS} = \overline{\text{TestScore}} + \hat{\beta}_1\big(\text{CS} - \overline{\text{CS}}\big).$$

With $\overline{\text{CS}} = 22.8$ (given in part (c)), the sample mean score is

$$\overline{\text{TestScore}} = \hat{\beta}_0 + \hat{\beta}_1 \overline{\text{CS}} = 640.3 - 4.93 \times 22.8 = 640.3 - 112.404 = 527.896.$$

The deviation of the classroom from the mean class size is $25 - 22.8 = 2.2$ students, and

$$\hat{\beta}_1 \times 2.2 = -4.93 \times 2.2 = -10.846.$$

So a classroom 2.2 students larger than average is predicted to score 10.846 points below the average of 527.896 points:

$$\widehat{\text{TestScore}} = 527.896 - 10.846 \implies \boxed{\widehat{\text{TestScore}} = 517.05.}$$

Unofficial - QRS@NTU - qrs.ntu.org

Part (b): predicted change from CS = 21 to CS = 24

Method 1: predict both years, then subtract

The two slope terms are

$$4.93 \times 21 = 103.53, \qquad 4.93 \times 24 = 118.32.$$

So the two fitted values are

$$\widehat{\text{TestScore}}_{\text{last year}} = 640.3 - 4.93 \times 21 = 640.3 - 103.53 = 536.77,$$
$$\widehat{\text{TestScore}}_{\text{this year}} = 640.3 - 4.93 \times 24 = 640.3 - 118.32 = 521.98.$$

Subtracting last year from this year,

$$\Delta\widehat{\text{TestScore}} = 521.98 - 536.77 \implies \boxed{\Delta\widehat{\text{TestScore}} = -14.79 \text{ points.}}$$

Method 2: slope as marginal effect

Write $\text{CS}_1 = 21$ and $\text{CS}_2 = 24$. Subtract the two fitted values; the intercept cancels:

$$\begin{aligned}\Delta\widehat{\text{TestScore}} &= \left[\hat{\beta}_0 + \hat{\beta}_1 \text{CS}_2\right] - \left[\hat{\beta}_0 + \hat{\beta}_1 \text{CS}_1\right] \\ &= \hat{\beta}_0 - \hat{\beta}_0 + \hat{\beta}_1 \text{CS}_2 - \hat{\beta}_1 \text{CS}_1 \\ &= \hat{\beta}_1 \big(\text{CS}_2 - \text{CS}_1\big) = \hat{\beta}_1 \, \Delta\text{CS}.\end{aligned}$$

Here $\Delta\text{CS} = 24 - 21 = 3$ students, and each extra student moves the prediction by $\hat{\beta}_1 = -4.93$ points. So three more students lower the predicted average score along the estimated line by

$$\boxed{\Delta\widehat{\text{TestScore}} = \hat{\beta}_1 \, \Delta\text{CS} = -4.93 \times 3 = -14.79 \text{ points.}}$$

Part (c): sample average test score, given mean CS = 22.8

Method 1: Rearranging the intercept formula

From Question 3 the OLS intercept is $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$, here $\hat{\beta}_0 = \overline{\text{TestScore}} - \hat{\beta}_1 \overline{\text{CS}}$. Adding $\hat{\beta}_1 \overline{\text{CS}}$ to both sides,

$$\overline{\text{TestScore}} = \hat{\beta}_0 + \hat{\beta}_1 \, \overline{\text{CS}}.$$

The slope term is

$$4.93 \times 22.8 = 4.93 \times 22 + 4.93 \times 0.8 = 108.46 + 3.944 = 112.404.$$

Geometrically, the OLS line passes through the point of sample means $(\overline{\text{CS}}, \overline{\text{TestScore}})$. Hence

$$\overline{\text{TestScore}} = 640.3 - 4.93 \times 22.8 = 640.3 - 112.404$$

$$\boxed{\overline{\text{TestScore}} = 527.896 \text{ points.}}$$

Method 2: Averaging fitted values with zero-sum residuals

Each observed score splits into fitted value plus residual, $Y_i = \hat{Y}_i + \hat{u}_i$ with $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 \text{CS}_i$. Average over the $n = 50$ classrooms:

$$\overline{\text{TestScore}} = \frac{1}{50} \sum_{i=1}^{50} \hat{Y}_i + \frac{1}{50} \sum_{i=1}^{50} \hat{u}_i.$$

The first-order condition for b_0 in Question 3 states $\sum_{i=1}^{50} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 \text{CS}_i) = 0$, that is

$$\sum_{i=1}^{50} \hat{u}_i = 0.$$

The average fitted value follows from (R1)–(R3):

$$\frac{1}{50} \sum_{i=1}^{50} \left(\hat{\beta}_0 + \hat{\beta}_1 \text{CS}_i \right) = \frac{1}{50} \left(50 \hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{50} \text{CS}_i \right) = \hat{\beta}_0 + \hat{\beta}_1 \overline{\text{CS}}.$$

Hence the sample mean score of the 50 classrooms (a sample object, which estimates but need not equal the population mean score) is

$$\overline{\text{TestScore}} = \hat{\beta}_0 + \hat{\beta}_1 \overline{\text{CS}} + 0 = 640.3 + (-4.93)(22.8) = 640.3 - 112.404,$$

$$\boxed{\overline{\text{TestScore}} = 527.896 \text{ points.}}$$

Final Answers

1. $(\hat{\beta}_0, \hat{\beta}_1) = \arg\min_{b_0,b_1} \sum_{i=1}^{n} [Y_i - (b_0 + b_1 X_i)]^2$.
2. β: true unknown parameters; b: candidate arguments; $\hat{\beta}$: minimisers, random variables (estimates on one sample).
3. $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$, $\hat{\beta}_1 = \dfrac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$; Hessian positive definite.
4. $\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2$; $\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}$.
5. (a) 517.05; (b) -14.79 points; (c) 527.896 points.

Unofficial_QRS@NTU_answers