HE2003 Econometrics I

Quiz Practice 1 (Solutions)

AY 2026/2027, Semester 1

Quantitative Research Society @NTU

October 12, 2026

Overview

- **Coverage:** Lectures 1–5 and Lecture 6 slides 1–3 (large-sample distribution of $\hat{\beta}_1$ and its standard error); no t-tests, confidence intervals, p-values or Stata.
- **Format:** 5 questions, 50 marks total, approximately 60 minutes.
- **Standard normal table** ($\Phi(z) = P(Z \leq z)$ for $Z \sim N(0,1)$), needed only where a probability is asked:

z	0.50	1.00	1.28	1.50	1.645	1.96	2.00	2.50	2.58	3.00
$\Phi(z)$	0.6915	0.8413	0.8997	0.9332	0.9500	0.9750	0.9772	0.9938	0.9951	0.9987

  Use $P(Z > z) = 1 - \Phi(z)$ and, by symmetry, $\Phi(-z) = 1 - \Phi(z)$.
- **Question mix:** Q1 derivation (OLS FOCs and residual identities); Q2 regression-output interpretation (prediction, change in prediction, units); Q3 large-sample sampling distribution of $\hat{\beta}_1$ (standardising, probability calculations, the role of $\text{SE}(\hat{\beta}_1)$); Q4 binary regressor (group means, an OLS-with-a-dummy proof, Least Squares Assumption 1 in this setting); Q5 conceptual (causality, Least Squares Assumption 1, selection bias, Type I/II trade-off).

Instructions

Answer **all** questions. Show all working clearly; answers without justification may receive zero marks.

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.

Contents

Question 1 **(10 Marks)** **3**
Solution . . . 3
Method 1: Calculus first-order conditions and substitution . . . 3
Method 2: Perturbation of the minimiser and the closed-form estimators . . . 5
Method 3: Orthogonal projection onto the span of the constant and the regressor 7
Marking scheme (10 marks) . . . 7

Question 2 **(10 Marks)** **9**
Solution . . . 9
Method 1: Substitution into the fitted line and slope times change . . . 9
Method 2: Difference of two fitted values (parts (b) and (c)) . . . 11
Marking scheme (10 marks) . . . 11

Question 3 **(10 Marks)** **12**
Solution . . . 12
Method 1: Standardisation with the normal table and the sampling-variance formula . . . 12
Method 2: Slope decomposition, symmetry about β_1 and variance via linearity . 15
Marking scheme (10 marks) . . . 16

Question 4 **(10 Marks)** **18**
Solution . . . 18
Method 1: Substitution and the OLS normal equations . . . 18
Method 2: Group-wise minimisation of S (part (c)) . . . 21
Method 3: Closed-form OLS slope with a binary regressor (part (c)) . . . 22
Method 4: Bias from the conditional means of the error (part (d)) . . . 23
Marking scheme (10 marks) . . . 23

Question 5 **(10 Marks)** **25**
Solution . . . 25
Method 1: Least Squares Assumption 1, randomised assignment and error definitions . . . 25
Method 2: Slope decomposition, difference in means under random assignment and score distributions . . . 27
Marking scheme (10 marks) . . . 28

Question 1 (10 Marks)

Consider the simple linear regression model

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \qquad i = 1, \ldots, n.$$

For candidate values b_0, b_1, define the sum of squared errors

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

The OLS estimators $\hat{\beta}_0, \hat{\beta}_1$ are the values of b_0, b_1 that minimize $S(b_0, b_1)$.

(a) (4 marks) Derive the two first-order conditions for this minimization problem, and show that they can be written as

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

(b) (3 marks) Using part (a), show that the OLS residuals $\hat{u}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$ satisfy

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

(c) (3 marks) Using part (b), show that $\sum_{i=1}^{n} \hat{u}_i \hat{Y}_i = 0$, where $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ is the OLS predicted value.

Solution

Method 1: Calculus first-order conditions and substitution

(a) $S(b_0, b_1)$ is a sum of n squared terms, each a polynomial in (b_0, b_1), so S is differentiable on $\mathbb{R}^2$. Differentiate the i-th term with respect to b_0 by the chain rule:

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

Summing over $i = 1, \ldots, n$,

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

In the same way, with respect to b_1,

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

so

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

At an interior minimiser of a differentiable function both partial derivatives are zero. The minimiser is $(b_0, b_1) = (\hat{\beta}_0, \hat{\beta}_1)$, so

$$-2 \sum_{i=1}^{n} \big(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\big) = 0,$$

$$-2\sum_{i=1}^{n}\big(Y_i-\hat\beta_0-\hat\beta_1X_i\big)X_i=0.$$

These conditions describe a minimum and not a maximum or saddle. The second derivatives are

$$\frac{\partial^2 S}{\partial b_0^2}=2n,\qquad \frac{\partial^2 S}{\partial b_1^2}=2\sum_{i=1}^{n}X_i^2,\qquad \frac{\partial^2 S}{\partial b_0\,\partial b_1}=2\sum_{i=1}^{n}X_i,$$

and the Hessian determinant is

$$(2n)\Big(2\sum_{i=1}^{n}X_i^2\Big)-\Big(2\sum_{i=1}^{n}X_i\Big)^2=4\Big[n\sum_{i=1}^{n}X_i^2-n^2\bar X^2\Big]=4n\sum_{i=1}^{n}(X_i-\bar X)^2>0$$

whenever the X_i are not all equal, using $\sum_{i=1}^{n}(X_i-\bar X)^2=\sum_{i=1}^{n}X_i^2-2\bar X\sum_{i=1}^{n}X_i+n\bar X^2=\sum_{i=1}^{n}X_i^2-n\bar X^2$. With $2n>0$ as well, S is strictly convex, so the solution of the two conditions is the unique minimiser. Dividing each first-order condition by $-2\neq 0$ gives

$$\boxed{\sum_{i=1}^{n}\big(Y_i-\hat\beta_0-\hat\beta_1X_i\big)=0,\qquad \sum_{i=1}^{n}\big(Y_i-\hat\beta_0-\hat\beta_1X_i\big)X_i=0.}$$

(b) By definition, for every i,

$$\hat u_i=Y_i-\hat\beta_0-\hat\beta_1X_i.$$

Replace the bracket in the first condition of part (a) by $\hat u_i$:

$$0=\sum_{i=1}^{n}\big(Y_i-\hat\beta_0-\hat\beta_1X_i\big)=\sum_{i=1}^{n}\hat u_i.$$

Replace the bracket in the second condition of part (a) by $\hat u_i$:

$$0=\sum_{i=1}^{n}\big(Y_i-\hat\beta_0-\hat\beta_1X_i\big)X_i=\sum_{i=1}^{n}\hat u_iX_i.$$

Therefore

$$\boxed{\sum_{i=1}^{n}\hat u_i=0,\qquad \sum_{i=1}^{n}\hat u_iX_i=0.}$$

(c) Substitute $\hat Y_i=\hat\beta_0+\hat\beta_1X_i$ and expand the product inside the sum:

$$\sum_{i=1}^{n}\hat u_i\hat Y_i=\sum_{i=1}^{n}\hat u_i\big(\hat\beta_0+\hat\beta_1X_i\big)=\sum_{i=1}^{n}\big(\hat\beta_0\hat u_i+\hat\beta_1\hat u_iX_i\big).$$

Split the sum and take the constants $\hat\beta_0,\hat\beta_1$ (they do not depend on i) outside:

$$\sum_{i=1}^{n}\big(\hat\beta_0\hat u_i+\hat\beta_1\hat u_iX_i\big)=\hat\beta_0\sum_{i=1}^{n}\hat u_i+\hat\beta_1\sum_{i=1}^{n}\hat u_iX_i.$$

By part (b), $\sum_{i=1}^{n}\hat u_i=0$ and $\sum_{i=1}^{n}\hat u_iX_i=0$, so

$$\hat\beta_0\sum_{i=1}^{n}\hat u_i+\hat\beta_1\sum_{i=1}^{n}\hat u_iX_i=\hat\beta_0\cdot 0+\hat\beta_1\cdot 0=0.$$

Hence

$$\sum_{i=1}^{n}\hat u_i\hat Y_i=\boxed{0}$$

Method 2: Perturbation of the minimiser and the closed-form estimators

(a) This route uses no derivatives. Write $r_i := Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$ for the bracket evaluated at the minimiser.

Perturb b_0. For any real t, $Y_i - (\hat{\beta}_0 + t) - \hat{\beta}_1 X_i = r_i - t$, so

$$S(\hat{\beta}_0 + t, \hat{\beta}_1) = \sum_{i=1}^{n} (r_i - t)^2 = \sum_{i=1}^{n} \left(r_i^2 - 2tr_i + t^2 \right) = S(\hat{\beta}_0, \hat{\beta}_1) - 2t \sum_{i=1}^{n} r_i + nt^2.$$

Because $(\hat{\beta}_0, \hat{\beta}_1)$ minimises S, $S(\hat{\beta}_0 + t, \hat{\beta}_1) - S(\hat{\beta}_0, \hat{\beta}_1) \geq 0$ for every t, that is,

$$g(t) := nt^2 - 2t \sum_{i=1}^{n} r_i \ \geq \ 0 \qquad \text{for all } t \in \mathbb{R}.$$

Suppose $\sum_{i=1}^{n} r_i \neq 0$ and take $t = \frac{1}{n} \sum_{i=1}^{n} r_i$. Then

$$g(t) = n \cdot \frac{\left(\sum_{i=1}^{n} r_i \right)^2}{n^2} - \frac{2\left(\sum_{i=1}^{n} r_i \right)^2}{n} = -\frac{\left(\sum_{i=1}^{n} r_i \right)^2}{n} < 0,$$

which contradicts $g(t) \geq 0$. Therefore $\sum_{i=1}^{n} r_i = 0$.

Perturb b_1. For any real t, $Y_i - \hat{\beta}_0 - (\hat{\beta}_1 + t) X_i = r_i - tX_i$, so

$$S(\hat{\beta}_0, \hat{\beta}_1 + t) = \sum_{i=1}^{n} (r_i - tX_i)^2 = S(\hat{\beta}_0, \hat{\beta}_1) - 2t \sum_{i=1}^{n} r_i X_i + t^2 \sum_{i=1}^{n} X_i^2,$$

and minimality gives

$$h(t) := t^2 \sum_{i=1}^{n} X_i^2 - 2t \sum_{i=1}^{n} r_i X_i \ \geq \ 0 \qquad \text{for all } t \in \mathbb{R}.$$

If $\sum_{i=1}^{n} X_i^2 = 0$, every $X_i = 0$ and $\sum_{i=1}^{n} r_i X_i = 0$ holds directly. If $\sum_{i=1}^{n} X_i^2 > 0$ and $\sum_{i=1}^{n} r_i X_i \neq 0$, take $t = \sum_{i=1}^{n} r_i X_i / \sum_{i=1}^{n} X_i^2$:

$$h(t) = \frac{\left(\sum_{i=1}^{n} r_i X_i \right)^2}{\sum_{i=1}^{n} X_i^2} - \frac{2\left(\sum_{i=1}^{n} r_i X_i \right)^2}{\sum_{i=1}^{n} X_i^2} = -\frac{\left(\sum_{i=1}^{n} r_i X_i \right)^2}{\sum_{i=1}^{n} X_i^2} < 0,$$

a contradiction. Therefore $\sum_{i=1}^{n} r_i X_i = 0$. Substituting $r_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$ back,

$$\boxed{\sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i \right) = 0, \qquad \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i \right) X_i = 0.}$$

(b) When the X_i are not all equal, so that $\sum_{i=1}^{n} (X_i - \bar{X})^2 > 0$, solving the two conditions of part (a) gives the closed-form estimators of Lecture 4,

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

Substitute $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$ into the residual:

$$\hat{u}_i = Y_i - \left(\bar{Y} - \hat{\beta}_1 \bar{X} \right) - \hat{\beta}_1 X_i = (Y_i - \bar{Y}) - \hat{\beta}_1 (X_i - \bar{X}).$$

Deviations from a sample mean sum to zero:

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

Hence

$$\sum_{i=1}^{n}\hat{u}_i = \sum_{i=1}^{n}(Y_i - \bar{Y}) - \hat{\beta}_1 \sum_{i=1}^{n}(X_i - \bar{X}) = 0 - \hat{\beta}_1 \cdot 0 = 0.$$

For the second identity, write $X_i = (X_i - \bar{X}) + \bar{X}$ and split:

$$\sum_{i=1}^{n}\hat{u}_i X_i = \sum_{i=1}^{n}\hat{u}_i(X_i - \bar{X}) + \bar{X}\sum_{i=1}^{n}\hat{u}_i = \sum_{i=1}^{n}\hat{u}_i(X_i - \bar{X}) + \bar{X} \cdot 0.$$

Now expand $\hat{u}_i(X_i - \bar{X})$ using the centred form of $\hat{u}_i$:

$$\begin{aligned}\sum_{i=1}^{n}\hat{u}_i(X_i - \bar{X}) &= \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) - \hat{\beta}_1 \sum_{i=1}^{n}(X_i - \bar{X})^2 \\ &= \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) - \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \cdot \sum_{i=1}^{n}(X_i - \bar{X})^2 \\ &= \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) - \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) = 0.\end{aligned}$$

Therefore

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

(c) Substitute $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}$ into the fitted value to get its centred form:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i = \bar{Y} - \hat{\beta}_1\bar{X} + \hat{\beta}_1 X_i = \bar{Y} + \hat{\beta}_1(X_i - \bar{X}).$$

Then

$$\sum_{i=1}^{n}\hat{u}_i\hat{Y}_i = \sum_{i=1}^{n}\hat{u}_i\big(\bar{Y} + \hat{\beta}_1(X_i - \bar{X})\big) = \bar{Y}\sum_{i=1}^{n}\hat{u}_i + \hat{\beta}_1\sum_{i=1}^{n}\hat{u}_i(X_i - \bar{X}).$$

From part (b), $\sum_{i=1}^{n}\hat{u}_i = 0$ and $\sum_{i=1}^{n}\hat{u}_i(X_i - \bar{X}) = 0$ (the last display of part (b)), so

$$\bar{Y}\sum_{i=1}^{n}\hat{u}_i + \hat{\beta}_1\sum_{i=1}^{n}\hat{u}_i(X_i - \bar{X}) = \bar{Y} \cdot 0 + \hat{\beta}_1 \cdot 0 = 0.$$

Hence

$$\sum_{i=1}^{n}\hat{u}_i\hat{Y}_i = \boxed{0}$$

Method 3: Orthogonal projection onto the span of the constant and the regressor

This method should be used only if the course allows this interpretation (it uses vectors in $\mathbb{R}^n$ and the projection theorem from linear algebra).

Stack the data as vectors in $\mathbb{R}^n$:

$$\mathbf{y} = (Y_1, \dots, Y_n)', \quad \mathbf{1} = (1, \dots, 1)', \quad \mathbf{x} = (X_1, \dots, X_n)', \quad \hat{\mathbf{u}} = (\hat{u}_1, \dots, \hat{u}_n)',$$

and write $\langle \mathbf{a}, \mathbf{b} \rangle = \sum_{i=1}^n a_i b_i$ for the dot product. Then

$$S(b_0, b_1) = \sum_{i=1}^n (Y_i - b_0 - b_1 X_i)^2 = \big\| \mathbf{y} - b_0 \mathbf{1} - b_1 \mathbf{x} \big\|^2,$$

the squared distance from $\mathbf{y}$ to the point $b_0\mathbf{1} + b_1\mathbf{x}$ of the plane $V = \operatorname{span}\{\mathbf{1}, \mathbf{x}\}$.

(a) By the projection theorem, the point of the subspace V closest to $\mathbf{y}$ is the unique $\hat{\mathbf{y}} \in V$ whose error $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to every vector of V, equivalently to the two spanning vectors $\mathbf{1}$ and $\mathbf{x}$. The minimiser of S gives $\hat{\mathbf{y}} = \hat{\beta}_0\mathbf{1} + \hat{\beta}_1\mathbf{x}$, so

$$\langle \mathbf{y} - \hat{\beta}_0\mathbf{1} - \hat{\beta}_1\mathbf{x},\ \mathbf{1} \rangle = 0, \qquad \langle \mathbf{y} - \hat{\beta}_0\mathbf{1} - \hat{\beta}_1\mathbf{x},\ \mathbf{x} \rangle = 0.$$

Writing the dot products out coordinate by coordinate,

$$\boxed{\sum_{i=1}^n \big(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\big) = 0, \qquad \sum_{i=1}^n \big(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\big) X_i = 0.}$$

(b) The residual vector is exactly the projection error, $\hat{\mathbf{u}} = \mathbf{y} - \hat{\beta}_0\mathbf{1} - \hat{\beta}_1\mathbf{x}$, so the two orthogonality conditions read

$$\langle \hat{\mathbf{u}}, \mathbf{1} \rangle = \sum_{i=1}^n \hat{u}_i \cdot 1 = 0, \qquad \langle \hat{\mathbf{u}}, \mathbf{x} \rangle = \sum_{i=1}^n \hat{u}_i X_i = 0,$$

that is,

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

(c) The fitted-value vector $\hat{\mathbf{y}} = \hat{\beta}_0\mathbf{1} + \hat{\beta}_1\mathbf{x}$ lies in V, and $\hat{\mathbf{u}}$ is orthogonal to all of V. By linearity of the dot product,

$$\sum_{i=1}^n \hat{u}_i \hat{Y}_i = \langle \hat{\mathbf{u}}, \hat{\mathbf{y}} \rangle = \hat{\beta}_0 \langle \hat{\mathbf{u}}, \mathbf{1} \rangle + \hat{\beta}_1 \langle \hat{\mathbf{u}}, \mathbf{x} \rangle = \hat{\beta}_0 \cdot 0 + \hat{\beta}_1 \cdot 0.$$

Hence

$$\sum_{i=1}^n \hat{u}_i \hat{Y}_i = \boxed{0}$$

Marking scheme (10 marks)

(a) **(4 marks)** First-order conditions

- 2 marks: correct partial derivative of S with respect to b_0, set to zero.

Unofficial QRS@NTU Solutions

- 2 marks: correct partial derivative of S with respect to b_1, set to zero, and both simplified (divided by -2) to the stated form.

(b) **(3 marks)** Residual identities

- 1 mark: correctly identifies $\hat{u}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$ as the object appearing in the FOCs.
- 2 marks: correctly substitutes to obtain both $\sum_{i=1}^{n} \hat{u}_i = 0$ and $\sum_{i=1}^{n} \hat{u}_i X_i = 0$ (1 mark each).

(c) **(3 marks)** $\sum \hat{u}_i \hat{Y}_i = 0$

- 1 mark: correctly expands $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ and splits the sum.
- 1 mark: correctly invokes both identities from part (b).
- 1 mark: correct final conclusion $\sum_{i=1}^{n} \hat{u}_i \hat{Y}_i = 0$.

Unofficial - QRS@NTU - qrs.ntu.org

Question 2 (10 Marks)

A researcher studying the Singapore private rental market collects data on $n = 150$ HDB flat rental listings and regresses monthly rent (in SGD) on floor area (in square metres, sqm). The estimated OLS regression line is

$$\widehat{\text{Rent}}_i = 380 + 24.5\,\text{Area}_i.$$

(a) (3 marks) A 90 sqm flat is listed for rent. What is the regression's predicted monthly rent for this flat?

(b) (3 marks) A tenant moves from a 70 sqm flat to a 95 sqm flat. What is the regression's predicted change in monthly rent?

(c) (4 marks) Interpret the slope coefficient, including units. Is the intercept economically meaningful in this context? Explain briefly.

Solution

Method 1: Substitution into the fitted line and slope times change

The estimated coefficients are $\hat{\beta}_0 = 380$ (SGD per month) and $\hat{\beta}_1 = 24.5$ (SGD per month per sqm), so a floor area in sqm multiplied by $\hat{\beta}_1$ is in SGD per month, the same unit as $\hat{\beta}_0$.

(a) Substitute Area = 90 into the fitted line:

$$\widehat{\text{Rent}} = 380 + 24.5 \times 90.$$

Compute the product:

$$24.5 \times 90 = 24 \times 90 + 0.5 \times 90 = 2160 + 45 = 2205.$$

Add the intercept:

$$380 + 2205 = 2585.$$

Therefore

$$\boxed{\widehat{\text{Rent}} = \text{SGD } 2{,}585 \text{ per month}}$$

(b) The fitted line is linear in Area, so a change ΔArea in floor area changes the prediction by

$$\Delta\widehat{\text{Rent}} = \hat{\beta}_1\,\Delta\text{Area}.$$

The change in floor area is

$$\Delta\text{Area} = 95 - 70 = 25 \text{ sqm}.$$

Hence

$$\Delta\widehat{\text{Rent}} = 24.5 \times 25 = 24 \times 25 + 0.5 \times 25 = 600 + 12.5 = 612.5.$$

The change is positive because $\hat{\beta}_1 > 0$ and $\Delta\text{Area} > 0$. Therefore

$$\boxed{\Delta\widehat{\text{Rent}} = \text{SGD } 612.50 \text{ (an increase)}}$$

(c)
- The slope $\hat{\beta}_1 = 24.5$ is measured in SGD per month per sqm. Comparing two listings whose floor areas differ by one square metre, the regression predicts that the larger flat rents for SGD 24.50 more per month.

- For example, a flat that is 10 sqm larger is predicted to rent for $24.5 \times 10 = 245$ SGD more per month.
- $\hat{\beta}_1$ is a sample estimate, from these $n = 150$ listings, of the population slope β_1; it describes how average rent varies with floor area across listings, and it reads as the effect of floor area only if the other determinants of rent in u_i (location, floor level, age of the block) are unrelated to Area.
- The intercept $\hat{\beta}_0 = 380$ is the predicted monthly rent of a flat with Area = 0 sqm. No HDB flat has zero floor area, and the listings in the sample all lie far above zero, so this prediction is an extrapolation outside the range of the data.
- The intercept therefore has no economic meaning on its own; its role is to position the fitted line so that it passes through $(\overline{\text{Area}}, \overline{\text{Rent}})$ over the observed range of flat sizes.

Slope: +SGD 24.50/month per extra sqm; intercept not economically meaningful.

Unofficial QRS@NTU practice material

Method 2: Difference of two fitted values (parts (b) and (c))

(b) Predicted rent of the 70 sqm flat:

$$\widehat{\text{Rent}}(70) = 380 + 24.5 \times 70 = 380 + 1715 = 2095.$$

Predicted rent of the 95 sqm flat:

$$\widehat{\text{Rent}}(95) = 380 + 24.5 \times 95 = 380 + 2327.5 = 2707.5.$$

The predicted change is the difference:

$$\Delta\widehat{\text{Rent}} = \widehat{\text{Rent}}(95) - \widehat{\text{Rent}}(70) = 2707.5 - 2095 = 612.5.$$

The intercept 380 appears in both predictions and cancels, which is why only the slope matters for a change. Therefore

$$\boxed{\Delta\widehat{\text{Rent}} = \text{SGD } 612.50 \text{ (an increase)}}$$

(c) For any floor area A, the difference between the predictions at $A+1$ and A is

$$\widehat{\text{Rent}}(A+1) - \widehat{\text{Rent}}(A) = \big[380 + 24.5(A+1)\big] - \big[380 + 24.5A\big] = 24.5.$$

This difference is the same at every A, so one extra sqm raises predicted rent by SGD 24.50 per month whatever the size of the flat; for instance $\widehat{\text{Rent}}(91) - \widehat{\text{Rent}}(90) = 2609.5 - 2585 = 24.5$.

The intercept is the prediction at $A = 0$:

$$\widehat{\text{Rent}}(0) = 380 + 24.5 \times 0 = 380.$$

A flat with zero floor area does not exist and lies far outside the range of the 150 listings, so SGD 380 is not the rent of any flat that could be observed. Hence

$$\boxed{\text{Slope: +SGD 24.50/month per extra sqm; intercept not economically meaningful.}}$$

Marking scheme (10 marks)

(a) **(3 marks)** Predicted rent

- 1 mark: correct substitution Area = 90 into the regression line.
- 1 mark: correct arithmetic.
- 1 mark: final answer stated with correct units (SGD/month).

(b) **(3 marks)** Predicted change

- 1 mark: correct method (slope × change in Area, or difference of two predictions).
- 1 mark: correct arithmetic (ΔArea = 25, product = 612.5).
- 1 mark: correctly stated with units and direction (increase).

(c) **(4 marks)** Interpretation

- 2 marks: correct slope interpretation with units and correct sign/direction.
- 2 marks: correct discussion that the intercept is not economically meaningful (Area= 0 does not occur in the data), with brief justification.

Question 3 (10 Marks)

A team of researchers studies $n = 250$ Singapore tech start-ups, regressing monthly revenue growth (%) on monthly R&D spending (SGD '000s). Based on the large sample, they treat the OLS estimator $\hat{\beta}_1$ as following its large-sample sampling distribution

$$\hat{\beta}_1 \;\sim\; N\big(\beta_1,\; \sigma^2_{\hat{\beta}_1}\big),$$

with population slope $\beta_1 = 2.0$ and standard deviation of the sampling distribution $\sigma_{\hat{\beta}_1} = \mathrm{SE}(\hat{\beta}_1) = 0.5$.

(a) (2 marks) Explain, in words, why $\hat{\beta}_1$ – computed from this one sample of 250 start-ups – is itself a random variable, even though the population slope β_1 is a fixed (if unknown) number.

(b) (3 marks) Using the large-sample result $\dfrac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}} \sim N(0,1)$, standardise and use the standard normal table in the Overview to compute $P(\hat{\beta}_1 > 2.75)$.

(c) (3 marks) Compute $P\big(|\hat{\beta}_1 - \beta_1| < 0.98\big)$, i.e. the probability that the estimator falls within 0.98 of the true population slope. Use the standard normal table in the Overview.

(d) (2 marks) Explain what $\mathrm{SE}(\hat{\beta}_1)$ measures, and give two reasons – referring to the formula for $\sigma^2_{\hat{\beta}_1}$ – why it becomes *smaller* as (i) the sample size n increases, and (ii) the sample spread of the regressor X increases.

Solution

Method 1: Standardisation with the normal table and the sampling-variance formula

(a)

- The population slope $\beta_1 = 2.0$ describes the relation between revenue growth and R&D spending for the whole population of Singapore tech start-ups: it is a single fixed number, measured in percentage points of monthly revenue growth per SGD 1,000 of monthly R&D.
- The estimator is computed from the data by the formula $\hat{\beta}_1 = \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) \big/ \sum_{i=1}^{n}(X_i - \bar{X})^2$, so its value depends on which 250 start-ups happen to be drawn.
- Under Least Squares Assumption 2 the pairs (X_i, Y_i), $i = 1, \ldots, 250$, are i.i.d. random draws from the population, so $\hat{\beta}_1$ is a function of random variables and is therefore itself a random variable.
- A second random sample of 250 start-ups would contain different firms and give a different number for $\hat{\beta}_1$; the distribution of these values over all possible samples is the sampling distribution $N(\beta_1, \sigma^2_{\hat{\beta}_1}) = N(2.0,\, 0.5^2)$, while β_1 stays at 2.0.
- For example, one sample could give $\hat{\beta}_1 = 1.7$ and another $\hat{\beta}_1 = 2.4$; both are ordinary draws, each less than one standard deviation (0.5) away from $\beta_1 = 2.0$.

> $\hat{\beta}_1$ is a function of the random sample $(Y_i, X_i)_{i=1}^n$, so it varies across samples and is itself random.

(b) Let $Z = (\hat{\beta}_1 - \beta_1)/\sigma_{\hat{\beta}_1} = (\hat{\beta}_1 - 2.0)/0.5$, so $Z \sim N(0,1)$ by the large-sample result. Subtracting 2.0 and dividing by the positive number 0.5 preserves the direction of the inequality:

$$P(\hat{\beta}_1 > 2.75) = P\left(\frac{\hat{\beta}_1 - 2.0}{0.5} > \frac{2.75 - 2.0}{0.5}\right).$$

The standardised cutoff is

$$\frac{2.75 - 2.0}{0.5} = \frac{0.75}{0.5} = 1.5.$$

Hence, by the complement rule $P(Z > z) = 1 - \Phi(z)$ and the table value $\Phi(1.5) = 0.9332$,

$$P(\hat{\beta}_1 > 2.75) = P(Z > 1.5) = 1 - \Phi(1.5) = 1 - 0.9332 = 0.0668.$$

Therefore

$$\boxed{P(\hat{\beta}_1 > 2.75) = 0.0668}$$

(c) Divide both sides of the inequality inside the absolute value by $\sigma_{\hat{\beta}_1} = 0.5 > 0$:

$$P\big(|\hat{\beta}_1 - 2.0| < 0.98\big) = P\left(\left|\frac{\hat{\beta}_1 - 2.0}{0.5}\right| < \frac{0.98}{0.5}\right) = P(|Z| < 1.96),$$

since $0.98/0.5 = 1.96$. The complement of $\{|Z| < 1.96\}$ is the union of the two tails $\{Z \geq 1.96\}$ and $\{Z \leq -1.96\}$, which have equal probability because the $N(0,1)$ density is symmetric about 0:

$$P(Z \leq -1.96) = \Phi(-1.96) = 1 - \Phi(1.96) = P(Z \geq 1.96).$$

Therefore

$$\begin{aligned} P(|Z| < 1.96) &= 1 - 2P(Z \geq 1.96) = 1 - 2\big(1 - \Phi(1.96)\big) = 2\Phi(1.96) - 1 \\ &= 2(0.9750) - 1 = 1.9500 - 1 = 0.9500. \end{aligned}$$

Hence

$$\boxed{P\big(|\hat{\beta}_1 - \beta_1| < 0.98\big) = 0.9500}$$

(d) The variance of the large-sample sampling distribution of $\hat{\beta}_1$ is, with $\mu_X = E[X_i]$ and u_i the population errors,

$$\sigma^2_{\hat{\beta}_1} = \frac{1}{n} \cdot \frac{\mathrm{Var}\left[(X_i - \mu_X)u_i\right]}{\big[\mathrm{Var}(X_i)\big]^2},$$

and it is estimated by replacing each population moment with its sample analogue ($\bar{X}$ for μ_X, the residual $\hat{u}_i$ for u_i, sample averages for the variances):

$$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{1}{n} \cdot \frac{\frac{1}{n-2}\sum_{i=1}^n (X_i - \bar{X})^2 \hat{u}_i^2}{\left[\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2\right]^2}, \qquad \mathrm{SE}(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2_{\hat{\beta}_1}}.$$

- $\sigma_{\hat{\beta}_1}$ is the standard deviation of the sampling distribution of $\hat{\beta}_1$, a population object: it measures how far $\hat{\beta}_1$ typically falls from β_1 across repeated samples. $\mathrm{SE}(\hat{\beta}_1)$ is its estimate from the one sample at hand, in the units of the slope, here percentage points per SGD 1,000; a smaller $\mathrm{SE}(\hat{\beta}_1)$ means a more precise estimator.

- (i) Sample size. The population moments $\text{Var}[(X_i - \mu_X)u_i]$ and $\text{Var}(X_i)$ are fixed numbers that do not depend on n (and their sample analogues in $\hat{\sigma}^2_{\hat{\beta}_1}$ settle down to them as n grows). The leading factor $1/n$ therefore makes $\sigma^2_{\hat{\beta}_1}$ shrink in proportion to $1/n$, and $\sigma_{\hat{\beta}_1}$ in proportion to $1/\sqrt{n}$.
- Numerical illustration of (i): quadrupling the sample from $n = 250$ to $n = 1000$ multiplies $\sigma_{\hat{\beta}_1}$ by $\sqrt{250/1000} = \frac{1}{2}$, from 0.5 to 0.25.
- (ii) Spread of X. Replace every deviation $X_i - \mu_X$ by $c(X_i - \mu_X)$ with $c > 1$, holding the errors u_i fixed. Since $\text{Var}(aW) = a^2\,\text{Var}(W)$, the numerator becomes

$$\text{Var}\left[c(X_i - \mu_X)u_i\right] = c^2\ \text{Var}\left[(X_i - \mu_X)u_i\right],$$

and, because $\text{Var}(X_i) = E[(X_i - \mu_X)^2]$ becomes $E[c^2(X_i - \mu_X)^2] = c^2\,\text{Var}(X_i)$, the denominator becomes

$$\left[c^2\ \text{Var}(X_i)\right]^2 = c^4\left[\,\text{Var}(X_i)\right]^2.$$

So $\sigma^2_{\hat{\beta}_1}$ is multiplied by $c^2/c^4 = 1/c^2 < 1$. With twice the spread in R&D spending ($c = 2$), $\sigma_{\hat{\beta}_1}$ halves. More variation in X gives more leverage to trace out how Y moves with X.

$\text{SE}(\hat{\beta}_1)$ measures the spread of $\hat{\beta}_1$'s sampling distribution; it falls as n rises and as the spread of X rises.

Unofficial QRS@NTU Material

Method 2: Slope decomposition, symmetry about β_1 and variance via linearity

(a) Average the population model $Y_i = \beta_0 + \beta_1 X_i + u_i$ over i to get $\bar{Y} = \beta_0 + \beta_1 \bar{X} + \bar{u}$, and subtract:

$$Y_i - \bar{Y} = \beta_1 (X_i - \bar{X}) + (u_i - \bar{u}).$$

Multiply by $(X_i - \bar{X})$ and sum:

$$\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}) = \beta_1 \sum_{i=1}^{n} (X_i - \bar{X})^2 + \sum_{i=1}^{n} (X_i - \bar{X}) u_i - \bar{u} \sum_{i=1}^{n} (X_i - \bar{X}).$$

The last sum is $\sum_{i=1}^{n} (X_i - \bar{X}) = n\bar{X} - n\bar{X} = 0$. Dividing by $\sum_{i=1}^{n} (X_i - \bar{X})^2$,

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n} (X_i - \bar{X}) u_i}{\sum_{i=1}^{n} (X_i - \bar{X})^2}.$$

The first term is the fixed population slope 2.0. The second term is built from the regressors and the unobserved errors u_i of the particular 250 start-ups drawn, which are random under Least Squares Assumption 2, so it changes from sample to sample. Under Least Squares Assumption 1 it averages to zero across samples, which is why the sampling distribution is centred at β_1. Hence

$$\boxed{\begin{array}{c}\hat{\beta}_1 \text{ is a function of the random sample } (Y_i, X_i)_{i=1}^{n}, \\ \text{so it varies across samples and is itself random.}\end{array}}$$

(b) The $N(2.0, 0.5^2)$ density is symmetric about its mean $\beta_1 = 2.0$. The point 2.75 lies 0.75 above the mean; its mirror image lies 0.75 below, at $2.0 - 0.75 = 1.25$. By symmetry,

$$P(\hat{\beta}_1 > 2.75) = P(\hat{\beta}_1 > 2.0 + 0.75) = P(\hat{\beta}_1 < 2.0 - 0.75) = P(\hat{\beta}_1 < 1.25).$$

Standardise the lower tail:

$$P(\hat{\beta}_1 < 1.25) = P\left(Z < \frac{1.25 - 2.0}{0.5}\right) = P\left(Z < \frac{-0.75}{0.5}\right) = \Phi(-1.5).$$

Using $\Phi(-z) = 1 - \Phi(z)$ and the table,

$$\Phi(-1.5) = 1 - \Phi(1.5) = 1 - 0.9332 = 0.0668.$$

Therefore

$$\boxed{P(\hat{\beta}_1 > 2.75) = 0.0668}$$

(c) Work in the original units first. The event is an interval around β_1:

$$|\hat{\beta}_1 - 2.0| < 0.98 \iff -0.98 < \hat{\beta}_1 - 2.0 < 0.98 \iff 1.02 < \hat{\beta}_1 < 2.98.$$

The probability of an interval is the difference of the CDF at its endpoints:

$$P(1.02 < \hat{\beta}_1 < 2.98) = \Phi\left(\frac{2.98 - 2.0}{0.5}\right) - \Phi\left(\frac{1.02 - 2.0}{0.5}\right) = \Phi(1.96) - \Phi(-1.96).$$

From the table, $\Phi(1.96) = 0.9750$ and $\Phi(-1.96) = 1 - 0.9750 = 0.0250$, so

$$\Phi(1.96) - \Phi(-1.96) = 0.9750 - 0.0250 = 0.9500.$$

Hence

$$\boxed{P\big(|\hat{\beta}_1 - \beta_1| < 0.98\big) = 0.9500}$$

(d) This method should be used only if the course allows this interpretation (it adds homoskedasticity, $\mathrm{Var}(u_i \mid X_i) = \sigma_u^2$, from the Lecture 5 Gauss–Markov discussion).

From part (a), conditional on the regressors,

$$\hat{\beta}_1 - \beta_1 = \sum_{i=1}^{n} w_i u_i, \qquad w_i = \frac{X_i - \bar{X}}{\sum_{i=1}^{n}(X_j - \bar{X})^2},$$

a linear combination of the u_i with weights that are fixed once the X's are fixed (the weights are defined because the X_i are not all equal). By Least Squares Assumption 2 the pairs (X_i, u_i) are independent across i, so all covariance terms vanish and the variance of the sum is the sum of the variances:

$$\mathrm{Var}(\hat{\beta}_1 \mid X) = \sum_{i=1}^{n} w_i^2 \,\mathrm{Var}(u_i \mid X_i) = \sigma_u^2 \sum_{i=1}^{n} w_i^2.$$

Compute the sum of squared weights:

$$\sum_{i=1}^{n} w_i^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\left[\sum_{i=1}^{n}(X_j - \bar{X})^2\right]^2} = \frac{1}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{1}{n\, s_X^2}, \qquad s_X^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Therefore

$$\mathrm{Var}(\hat{\beta}_1 \mid X) = \frac{\sigma_u^2}{n\, s_X^2}.$$

- $\mathrm{SE}(\hat{\beta}_1)$ is the estimate of the square root of this variance: the typical distance between $\hat{\beta}_1$ and β_1 across repeated samples.
- (i) n sits in the denominator, so a larger sample lowers the variance in proportion to $1/n$; with n four times larger, $\mathrm{SE}(\hat{\beta}_1)$ halves, from 0.5 to 0.25.
- (ii) s_X^2 also sits in the denominator, so a larger spread of R&D spending lowers the variance; doubling every deviation $X_i - \bar{X}$ multiplies s_X^2 by 4 and halves $\mathrm{SE}(\hat{\beta}_1)$.

$\mathrm{SE}(\hat{\beta}_1)$ measures the spread of $\hat{\beta}_1$'s sampling distribution; it falls as n rises and as the spread of X rises.

Marking scheme (10 marks)

(a) **(2 marks)** Why $\hat{\beta}_1$ is random

- 1 mark: recognizes $\hat{\beta}_1$ is a function of the random sample (Y_i, X_i).
- 1 mark: correctly explains that a different sample gives a different numerical value, hence a sampling distribution, while β_1 stays fixed.

(b) **(3 marks)** $P(\hat{\beta}_1 > 2.75)$

- 1 mark: correct standardisation, $z = 1.5$.
- 1 mark: correctly reads $\Phi(1.5) = 0.9332$ from the table.
- 1 mark: correct final probability 0.0668.

(c) **(3 marks)** $P(|\hat{\beta}_1 - \beta_1| < 0.98)$

- 1 mark: correct standardisation to $P(|Z| < 1.96)$.

- 1 mark: correct use of symmetry, $2\Phi(1.96) - 1$.
- 1 mark: correct final probability 0.9500.

(d) **(2 marks)** Meaning of $\text{SE}(\hat{\beta}_1)$ and why it shrinks

- 1 mark: correctly states $\text{SE}(\hat{\beta}_1)$ measures the spread/dispersion of the sampling distribution of $\hat{\beta}_1$ (precision of the estimator).
- 1 mark: valid explanation of *both* why it falls with larger n and with more spread in X.

Unofficial - QRS@NTU - qrs.ntu.org

Question 4 (10 Marks)

A study of $n = 200$ food-delivery riders in Singapore regresses monthly income (in SGD) on D_i, a dummy variable equal to 1 if the rider owns their vehicle and 0 if the rider rents it ($n_0 = 120$ renters, $n_1 = 80$ owners). The estimated regression is

$$\widehat{\text{Income}}_i = 1450 + 320\, D_i, \qquad \text{SE}(\hat{\beta}_1) = 95.$$

(a) (3 marks) State the average monthly income predicted by the regression for renters and for owners.

(b) (2 marks) Give an economic interpretation of $\hat{\beta}_1 = 320$.

(c) (3 marks) Prove, using the OLS normal equations from Question 1 applied to this binary regressor $D_i \in \{0, 1\}$, that in general $\hat{\beta}_0 = \bar{Y}_0$ (the sample mean of income among renters) and $\hat{\beta}_1 = \bar{Y}_1 - \bar{Y}_0$ (the difference in sample mean income between owners and renters), where $\bar{Y}_0, \bar{Y}_1$ are the group sample means and n_0, n_1 are the group sizes.

(d) (2 marks) State what Least Squares Assumption 1 ($E[u_i \mid D_i] = 0$) requires in this dummy-regressor setting, give one specific unobserved factor that could plausibly make it fail here, and state the likely direction of the resulting bias in $\hat{\beta}_1$.

Solution

Method 1: Substitution and the OLS normal equations

(a) For a renter, $D_i = 0$:

$$\widehat{\text{Income}} = 1450 + 320 \times 0 = 1450.$$

For an owner, $D_i = 1$:

$$\widehat{\text{Income}} = 1450 + 320 \times 1 = 1450 + 320 = 1770.$$

Both are in SGD per month, the unit of the dependent variable. Therefore

Renters: SGD 1,450/month; Owners: SGD 1,770/month.

(b) The slope equals the difference between the two predictions of part (a):

$$\hat{\beta}_1 = \widehat{\text{Income}}(D = 1) - \widehat{\text{Income}}(D = 0) = 1770 - 1450 = 320.$$

- $\hat{\beta}_1 = 320$ is measured in SGD per month: it is the gap in predicted monthly income between a rider who owns the vehicle and a rider who rents.
- In this sample of $n = 200$ riders, the 80 owners earned on average SGD 320 per month more than the 120 renters; relative to the renters' average of SGD 1,450 this is $320/1450 \approx 22.1\%$ more.
- $\hat{\beta}_1$ is a sample estimate of the population difference in mean income between owners and renters,

  $$E[\text{Income}_i \mid D_i = 1] - E[\text{Income}_i \mid D_i = 0];$$

  this difference equals the causal effect β_1 of owning only if Least Squares Assumption 1 holds (part (d)).

> Riders who own their vehicle earn, on average, SGD 320/month more than those who rent.

(c) With $X_i = D_i$ and $Y_i = \text{Income}_i$, the two normal equations of Question 1 are

$$\text{(NE1)} \quad \sum_{i=1}^{n} \big(Y_i - \hat{\beta}_0 - \hat{\beta}_1 D_i\big) = 0,$$

$$\text{(NE2)} \quad \sum_{i=1}^{n} \big(Y_i - \hat{\beta}_0 - \hat{\beta}_1 D_i\big) D_i = 0.$$

Split each sum term by term:

$$\text{(NE1)} \quad \sum_{i=1}^{n} Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} D_i,$$

$$\text{(NE2)} \quad \sum_{i=1}^{n} Y_i D_i = \hat{\beta}_0 \sum_{i=1}^{n} D_i + \hat{\beta}_1 \sum_{i=1}^{n} D_i^2.$$

Four facts about the dummy evaluate these sums. Since $0^2 = 0$ and $1^2 = 1$,

$$D_i^2 = D_i \quad \text{for every } i.$$

$D_i = 1$ exactly for the n_1 owners and 0 otherwise, so

$$\sum_{i=1}^{n} D_i = n_1, \qquad \sum_{i=1}^{n} D_i^2 = \sum_{i=1}^{n} D_i = n_1.$$

$Y_i D_i$ equals Y_i for owners and 0 for renters, so

$$\sum_{i=1}^{n} Y_i D_i = \sum_{D_i=1} Y_i = n_1 \bar{Y}_1.$$

Splitting the full sum into the two groups,

$$\sum_{i=1}^{n} Y_i = \sum_{D_i=0} Y_i + \sum_{D_i=1} Y_i = n_0 \bar{Y}_0 + n_1 \bar{Y}_1, \qquad n = n_0 + n_1.$$

Substituting these into the normal equations:

$$\text{(NE1)} \quad n_0 \bar{Y}_0 + n_1 \bar{Y}_1 = (n_0 + n_1)\hat{\beta}_0 + n_1 \hat{\beta}_1,$$

$$\text{(NE2)} \quad n_1 \bar{Y}_1 = n_1 \hat{\beta}_0 + n_1 \hat{\beta}_1.$$

Both groups are non-empty ($n_1 > 0$; here $n_1 = 80$), so divide (NE2) by n_1:

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

Subtract (NE2) from (NE1); the terms $n_1 \bar{Y}_1$, $n_1 \hat{\beta}_0$ and $n_1 \hat{\beta}_1$ cancel:

$$n_0 \bar{Y}_0 = n_0 \hat{\beta}_0.$$

Since $n_0 > 0$ (here $n_0 = 120$), divide by n_0:

$$\hat{\beta}_0 = \bar{Y}_0.$$

Substitute into $\bar{Y}_1 = \hat{\beta}_0 + \hat{\beta}_1$:

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

Therefore

$$\boxed{\hat{\beta}_0 = \bar{Y}_0 \text{ (renter mean)}, \qquad \hat{\beta}_1 = \bar{Y}_1 - \bar{Y}_0 \text{ (owner mean} - \text{renter mean).}}$$

(d)

- In the population model $\text{Income}_i = \beta_0 + \beta_1 D_i + u_i$, the error u_i collects every determinant of income other than vehicle ownership. D_i takes only two values, so Least Squares Assumption 1, $E[u_i \mid D_i] = 0$, means two separate conditions: $E[u_i \mid D_i = 0] = 0$ and $E[u_i \mid D_i = 1] = 0$.
- In words, the unobserved determinants of income must have the same average (zero) among owners and among renters; then $E[\text{Income}_i \mid D_i = 1] - E[\text{Income}_i \mid D_i = 0] = \beta_1$ is the causal effect of owning.
- A plausible failure is rider experience and commitment (full-time versus part-time riding). Riders who plan to ride full-time for years are more likely to buy a vehicle, and they also work more hours and know the busy zones, which raises income directly. Then $E[u_i \mid D_i = 1] > E[u_i \mid D_i = 0]$.
- The omitted factor is positively related to ownership and positively related to income, so part of the SGD 320 gap reflects who chooses to own rather than ownership itself: $\hat{\beta}_1$ is biased upward and overstates β_1.

$E[u \mid D] \neq 0$ (self-selection into ownership by experience/commitment)
$\Rightarrow$ $\hat{\beta}_1$ is upward biased.

Unofficial QRS@NTU practice quiz

Method 2: Group-wise minimisation of S (part (c))

Reparametrise with

$$c_0 := b_0, \qquad c_1 := b_0 + b_1.$$

This is a one-to-one map of $\mathbb{R}^2$ onto itself, with inverse $b_0 = c_0$, $b_1 = c_1 - c_0$, so minimising over (b_0, b_1) is the same as minimising over (c_0, c_1). For a renter ($D_i = 0$) the i-th term of S is

$$Y_i - b_0 - b_1 \cdot 0 = Y_i - c_0,$$

and for an owner ($D_i = 1$) it is

$$Y_i - b_0 - b_1 \cdot 1 = Y_i - (b_0 + b_1) = Y_i - c_1.$$

Splitting the sum by group,

$$S = \sum_{D_i=0} (Y_i - c_0)^2 + \sum_{D_i=1} (Y_i - c_1)^2.$$

The first sum depends only on c_0 and the second only on c_1, so each can be minimised separately.

For a group $g \in \{0, 1\}$ with n_g members and mean $\bar{Y}_g$, write $Y_i - c = (Y_i - \bar{Y}_g) + (\bar{Y}_g - c)$ and square:

$$\sum_{D_i=g} (Y_i - c)^2 = \sum_{D_i=g} (Y_i - \bar{Y}_g)^2 + 2(\bar{Y}_g - c) \sum_{D_i=g} (Y_i - \bar{Y}_g) + n_g(\bar{Y}_g - c)^2.$$

The cross term vanishes because

$$\sum_{D_i=g} (Y_i - \bar{Y}_g) = n_g\bar{Y}_g - n_g\bar{Y}_g = 0,$$

so

$$\sum_{D_i=g} (Y_i - c)^2 = \sum_{D_i=g} (Y_i - \bar{Y}_g)^2 + n_g(\bar{Y}_g - c)^2.$$

The first term does not involve c, and the second is ≥ 0 with equality only at $c = \bar{Y}_g$ (as $n_g > 0$). Hence the unique minimisers are

$$\hat{c}_0 = \bar{Y}_0, \qquad \hat{c}_1 = \bar{Y}_1.$$

Map back with $b_0 = c_0$, $b_1 = c_1 - c_0$:

$$\hat{\beta}_0 = \hat{c}_0 = \bar{Y}_0, \qquad \hat{\beta}_1 = \hat{c}_1 - \hat{c}_0 = \bar{Y}_1 - \bar{Y}_0.$$

Therefore

$$\boxed{\hat{\beta}_0 = \bar{Y}_0 \text{ (renter mean)}, \qquad \hat{\beta}_1 = \bar{Y}_1 - \bar{Y}_0 \text{ (owner mean} - \text{renter mean).}}$$

Method 3: Closed-form OLS slope with a binary regressor (part (c))

Start from the closed-form estimators (which solve the normal equations of Question 1) with $X_i = D_i$:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(D_i - \bar{D})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(D_i - \bar{D})^2}, \qquad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{D}.$$

The sample means are

$$\bar{D} = \frac{1}{n}\sum_{i=1}^{n} D_i = \frac{n_1}{n}, \qquad \bar{Y} = \frac{n_0 \bar{Y}_0 + n_1 \bar{Y}_1}{n}.$$

Denominator. Expand the square and use $D_i^2 = D_i$:

$$\sum_{i=1}^{n}(D_i - \bar{D})^2 = \sum_{i=1}^{n} D_i^2 - 2\bar{D}\sum_{i=1}^{n} D_i + n\bar{D}^2 = n_1 - 2\frac{n_1}{n}n_1 + n\frac{n_1^2}{n^2} = n_1 - \frac{n_1^2}{n} = \frac{n_1(n-n_1)}{n} = \frac{n_0 n_1}{n}.$$

Numerator. Since $\sum_{i=1}^{n}(D_i - \bar{D}) = n_1 - n \cdot \frac{n_1}{n} = 0$, the term in $\bar{Y}$ drops out:

$$\sum_{i=1}^{n}(D_i - \bar{D})(Y_i - \bar{Y}) = \sum_{i=1}^{n}(D_i - \bar{D})Y_i - \bar{Y}\sum_{i=1}^{n}(D_i - \bar{D}) = \sum_{i=1}^{n} D_i Y_i - \bar{D}\sum_{i=1}^{n} Y_i.$$

Substitute the group sums:

$$\begin{aligned}\sum_{i=1}^{n} D_i Y_i - \bar{D}\sum_{i=1}^{n} Y_i &= n_1\bar{Y}_1 - \frac{n_1}{n}\big(n_0\bar{Y}_0 + n_1\bar{Y}_1\big) = \frac{n_1}{n}\Big[(n_0+n_1)\bar{Y}_1 - n_0\bar{Y}_0 - n_1\bar{Y}_1\Big] \\ &= \frac{n_1}{n}\cdot n_0\big(\bar{Y}_1 - \bar{Y}_0\big) = \frac{n_0 n_1}{n}\big(\bar{Y}_1 - \bar{Y}_0\big).\end{aligned}$$

Slope. Both groups are non-empty ($n_0 > 0$ and $n_1 > 0$), so the denominator $n_0 n_1/n$ is positive. Divide numerator by denominator:

$$\hat{\beta}_1 = \frac{\frac{n_0 n_1}{n}\big(\bar{Y}_1 - \bar{Y}_0\big)}{\frac{n_0 n_1}{n}} = \bar{Y}_1 - \bar{Y}_0.$$

Intercept.

$$\begin{aligned}\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{D} &= \frac{n_0\bar{Y}_0 + n_1\bar{Y}_1}{n} - \frac{n_1}{n}\big(\bar{Y}_1 - \bar{Y}_0\big) = \frac{n_0\bar{Y}_0 + n_1\bar{Y}_1 - n_1\bar{Y}_1 + n_1\bar{Y}_0}{n} \\ &= \frac{(n_0+n_1)\bar{Y}_0}{n} = \bar{Y}_0.\end{aligned}$$

With the sample figures, $\bar{Y} = (120 \times 1450 + 80 \times 1770)/200 = (174000 + 141600)/200 = 1578$ and $\bar{D} = 80/200 = 0.4$, so $\hat{\beta}_0 = 1578 - 320 \times 0.4 = 1578 - 128 = 1450 = \bar{Y}_0$. Therefore

$$\boxed{\hat{\beta}_0 = \bar{Y}_0 \text{ (renter mean)}, \qquad \hat{\beta}_1 = \bar{Y}_1 - \bar{Y}_0 \text{ (owner mean} - \text{renter mean).}}$$

Method 4: Bias from the conditional means of the error (part (d))

Average the population model $\text{Income}_i = \beta_0 + \beta_1 D_i + u_i$ within each group. For owners ($D_i = 1$),

$$\bar{Y}_1 = \beta_0 + \beta_1 + \bar{u}_1, \qquad \bar{u}_1 = \frac{1}{n_1} \sum_{D_i=1} u_i,$$

and for renters ($D_i = 0$),

$$\bar{Y}_0 = \beta_0 + \bar{u}_0, \qquad \bar{u}_0 = \frac{1}{n_0} \sum_{D_i=0} u_i.$$

By part (c), $\hat{\beta}_1 = \bar{Y}_1 - \bar{Y}_0$, so

$$\hat{\beta}_1 = \big(\beta_0 + \beta_1 + \bar{u}_1\big) - \big(\beta_0 + \bar{u}_0\big) = \beta_1 + \big(\bar{u}_1 - \bar{u}_0\big).$$

Take expectations conditional on the dummies; with i.i.d. sampling, $E[\bar{u}_1 \mid D] = E[u_i \mid D_i = 1]$ and $E[\bar{u}_0 \mid D] = E[u_i \mid D_i = 0]$:

$$E[\hat{\beta}_1 \mid D] = \beta_1 + \Big(E[u_i \mid D_i = 1] - E[u_i \mid D_i = 0]\Big).$$

- Least Squares Assumption 1 sets both conditional means to zero, so the bracket is zero and $E[\hat{\beta}_1 \mid D] = \beta_1$.
- If more experienced, full-time riders select into ownership, their unobserved income determinants are higher: $E[u_i \mid D_i = 1] > E[u_i \mid D_i = 0]$, the bracket is positive and $E[\hat{\beta}_1 \mid D] > \beta_1$.
- As an illustration with hypothetical numbers: if the experience gap adds SGD 200 per month to owners' average u_i relative to renters, then on average $\hat{\beta}_1$ exceeds β_1 by 200, and only about SGD 120 of the observed SGD 320 gap would be due to ownership.

$E[u \mid D] \neq 0$ (self-selection into ownership by experience/commitment) $\Rightarrow$ $\hat{\beta}_1$ is upward biased.

Marking scheme (10 marks)

(a) **(3 marks)** Group means

- 1 mark: correct method ($D = 0$ and $D = 1$ substituted into the regression line).
- 1 mark: correct renter mean (1450).
- 1 mark: correct owner mean (1770).

(b) **(2 marks)** Interpretation of $\hat{\beta}_1$

- 1 mark: recognizes $\hat{\beta}_1$ as the difference in group means.
- 1 mark: correct direction/magnitude with units (owners earn SGD 320 more).

(c) **(3 marks)** Proof that $\hat{\beta}_0 = \bar{Y}_0$, $\hat{\beta}_1 = \bar{Y}_1 - \bar{Y}_0$

- 1 mark: correctly uses the first normal equation (splitting $\sum_{i=1}^{n} Y_i$ by group and using $\sum_{i=1}^{n} D_i = n_1$).
- 1 mark: correctly uses the second normal equation, including $D_i^2 = D_i$, to obtain $\bar{Y}_1 = \hat{\beta}_0 + \hat{\beta}_1$.
- 1 mark: correct algebra combining the two to reach $\hat{\beta}_0 = \bar{Y}_0$ and $\hat{\beta}_1 = \bar{Y}_1 - \bar{Y}_0$.

(d) **(2 marks)** Least Squares Assumption 1

- 1 mark: correctly states Assumption 1 as requiring $E[u_i \mid D_i = 0] = 0$ and $E[u_i \mid D_i = 1] = 0$ separately.
- 1 mark: valid confounding factor with correct direction of bias.

Unofficial - QRS@NTU - qrsntu.org

Question 5 (10 Marks)

(a) (4 marks) A researcher regresses PSLE (Primary School Leaving Examination) score on weekly hours of private tuition, using observational data on Singapore primary-six pupils. She finds a strong positive coefficient on tuition hours and concludes that private tuition *causes* higher PSLE scores. Using Least Squares Assumption 1 ($E[u_i|X_i] = 0$) and the concept of selection bias, explain why this conclusion may not be valid. In your answer, state a specific unobserved factor that could be correlated with tuition hours, and predict the likely direction of the resulting bias.

(b) (3 marks) Suggest one research design that would let the researcher obtain a more credible estimate of the causal effect of tuition on PSLE score, and briefly explain why it addresses the problem identified in part (a).

(c) (3 marks) A bank's automated loan-approval system flags each application as "high risk" (reject) or "safe" (approve) based on a credit-score cutoff. Define, in the context of this bank's decision, what a Type I error and a Type II error would mean, and explain the trade-off the bank faces if it raises the credit-score cutoff required for approval.

Solution

Method 1: Least Squares Assumption 1, randomised assignment and error definitions

(a)

- The population model is $\text{PSLE}_i = \beta_0 + \beta_1 \,\text{Tuition}_i + u_i$, where β_1 is the causal effect of one more weekly hour of tuition on the PSLE score and u_i collects every other determinant of the score (ability, family income, parental education, motivation, school quality).
- The OLS slope $\hat{\beta}_1$ is an unbiased estimate of this causal β_1 only under Least Squares Assumption 1, $E[u_i \,|\, \text{Tuition}_i] = 0$: the average of all other determinants must be the same at every level of tuition hours.
- Tuition hours are chosen by families, not assigned at random. Family income (and the parental involvement that goes with it) is a specific unobserved factor: higher-income families buy more tuition hours, and the same families also raise PSLE scores directly through books, study space and help at home. This is selection bias: pupils with many tuition hours differ systematically from pupils with few, so $E[u_i \,|\, \text{Tuition}_i]$ rises with Tuition_i and Assumption 1 fails.
- The omitted factor is positively correlated with tuition hours and has a positive effect on PSLE scores, so the regression attributes part of the family-income effect to tuition: $\hat{\beta}_1$ is biased upward, and the true causal effect of tuition may be much smaller than the strong positive coefficient suggests, possibly zero.

$$\boxed{\begin{array}{c} E[u \mid \text{Tuition}] \neq 0 \text{ (selection on family resources and ability)} \\ \Rightarrow \;\; \hat{\beta}_1 \text{ is upward biased.} \end{array}}$$

(b) A randomised controlled trial: among a pool of eligible pupils, allocate free tuition places (a fixed number of weekly hours) by lottery, so that tuition hours are decided by a random draw rather than by the family.

- Because the lottery ignores every pupil characteristic, tuition hours are statistically independent of family income, motivation and ability, that is, of u_i:

$$\text{Tuition}_i \perp u_i.$$

- Independence implies that conditioning on tuition does not change the mean of u_i, and the intercept β_0 absorbs any non-zero mean of the error, so

$$E[u_i \mid \text{Tuition}_i] = E[u_i] = 0,$$

  which is Least Squares Assumption 1.

- With Assumption 1 restored, the selection channel of part (a) is cut: high-income and low-income families are equally represented at every tuition level, so the OLS slope estimates the causal effect β_1 without bias.

$$\boxed{\begin{array}{c}\text{Random assignment of tuition} \Rightarrow E[u \mid \text{Tuition}] = 0 \\ \Rightarrow \text{ causal interpretation valid.}\end{array}}$$

(c) Frame each application as a test of

$$H_0: \text{ the applicant is safe} \qquad \text{vs} \qquad H_1: \text{ the applicant is high risk,}$$

with the decision rule "reject H_0 (flag high risk and reject the loan) if the credit score is below the cutoff c".

- Type I error (rejecting a true H_0): a truly safe applicant scores below c, is flagged "high risk" and is rejected. The bank loses the interest income it would have earned and the customer goes elsewhere.
- Type II error (failing to reject a false H_0): a truly high-risk applicant scores at or above c, is flagged "safe" and is approved. The bank later suffers the loss from a default.
- Raising c enlarges the rejection region $\{\text{score} < c\}$. More safe applicants now fall below the cutoff, so the probability of a Type I error rises; fewer high-risk applicants clear the cutoff, so the probability of a Type II error falls.
- Moving the cutoff cannot lower both error probabilities at once; the bank chooses c by weighing the cost of a default against the profit lost on a rejected good customer.

$$\boxed{\begin{array}{c}\text{Higher cutoff: } \downarrow \text{Type II error (fewer bad loans),} \\ \uparrow \text{Type I error (more good applicants rejected).}\end{array}}$$

Method 2: Slope decomposition, difference in means under random assignment and score distributions

(a) This method should be used only if the course allows this interpretation (it uses the law of large numbers on the slope decomposition, the omitted-variable-bias argument).

Write $X_i = \text{Tuition}_i$. The OLS slope decomposes as (average the model, subtract, multiply by $X_i - \bar{X}$ and sum, as in the unbiasedness proof)

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

In a large i.i.d. sample the two averages approach their population counterparts, so

$$\hat{\beta}_1 \;\longrightarrow\; \beta_1 + \frac{\text{Cov}(X_i, u_i)}{\text{Var}(X_i)}.$$

Under Least Squares Assumption 1, $\text{Cov}(X_i, u_i) = E[(X_i - \mu_X)\,E[u_i \,|\, X_i]] = 0$ and $\hat{\beta}_1$ is centred on β_1. Now let family income W_i be part of the error, $u_i = \gamma W_i + v_i$ with $\gamma > 0$ (income raises PSLE) and v_i unrelated to X_i. Then

$$\text{Cov}(X_i, u_i) = \gamma\ \text{Cov}(X_i, W_i) + \text{Cov}(X_i, v_i) = \gamma\ \text{Cov}(X_i, W_i) + 0 > 0,$$

because higher-income families buy more tuition ($\text{Cov}(X_i, W_i) > 0$). The bias term $\text{Cov}(X_i, u_i)/\,\text{Var}(X_i)$ is positive, so $E[u_i \,|\, X_i]$ is not zero and $\hat{\beta}_1$ overstates β_1.

$E[u \mid \text{Tuition}] \neq 0$ (selection on family resources and ability) $\Rightarrow$ $\hat{\beta}_1$ is upward biased.

(b) Run the lottery of Method 1 with a binary treatment, $T_i = 1$ if the pupil wins a tuition place and $T_i = 0$ otherwise, and regress $\text{PSLE}_i = \beta_0 + \beta_1 T_i + u_i$. By the binary-regressor result of Question 4(c),

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

the difference between the mean PSLE score of winners and of losers. Averaging the model within each group, as in Question 4 Method 4,

$$\hat{\beta}_1 = \beta_1 + \big(\bar{u}_1 - \bar{u}_0\big), \qquad E[\hat{\beta}_1 \,|\, T] = \beta_1 + \Big(E[u_i \,|\, T_i = 1] - E[u_i \,|\, T_i = 0]\Big).$$

The lottery makes T_i independent of u_i, so winners and losers have the same average family income, motivation and ability:

$$E[u_i \,|\, T_i = 1] = E[u_i \,|\, T_i = 0] = E[u_i] = 0.$$

Hence $E[\hat{\beta}_1 \,|\, T] = \beta_1$: the difference in mean PSLE scores between the two lottery groups estimates the causal effect of tuition, with no selection term left.

Random assignment of tuition $\Rightarrow$ $E[u \mid \text{Tuition}] = 0$ $\Rightarrow$ causal interpretation valid.

(c) Model the credit scores of the two kinds of applicant with hypothetical distributions (for illustration only):

$$\text{safe: } N(700,\, 50^2), \qquad \text{high risk: } N(600,\, 50^2),$$

and approve an application when the score is at least c. Then

$$P(\text{Type I}) = P(\text{score} < c \,|\, \text{safe}) = \Phi\left(\frac{c-700}{50}\right),$$
$$P(\text{Type II}) = P(\text{score} \geq c \,|\, \text{high risk}) = 1 - \Phi\left(\frac{c-600}{50}\right).$$

At $c = 650$:

$$P(\text{Type I}) = \Phi(-1.00) = 1 - 0.8413 = 0.1587,$$
$$P(\text{Type II}) = 1 - \Phi(1.00) = 1 - 0.8413 = 0.1587.$$

Raise the cutoff to $c = 675$:

$$P(\text{Type I}) = \Phi(-0.50) = 1 - 0.6915 = 0.3085,$$
$$P(\text{Type II}) = 1 - \Phi(1.50) = 1 - 0.9332 = 0.0668.$$

In general Φ is increasing, so $\Phi((c-700)/50)$ rises with c and $1 - \Phi((c-600)/50)$ falls with c. Raising the cutoff from 650 to 675 lowers the chance of approving a high-risk applicant (Type II, defaults) from 15.87% to 6.68%, and raises the chance of rejecting a safe applicant (Type I, lost business) from 15.87% to 30.85%.

Higher cutoff: $\downarrow$ Type II error (fewer bad loans), $\uparrow$ Type I error (more good applicants rejected).

Marking scheme (10 marks)

(a) **(4 marks)** Selection bias

- 1 mark: correctly states Assumption 1 ($E[u \mid X] = 0$) as the relevant condition.
- 1 mark: identifies a specific, plausible confounding factor (e.g. family income, motivation, or ability).
- 1 mark: explains why the confounder is correlated with tuition hours (self-selection).
- 1 mark: correct direction of bias (upward) with valid justification.

(b) **(3 marks)** Alternative design

- 1 mark: proposes a valid design (RCT, natural experiment, or instrumental variable).
- 1 mark: explains that random/quasi-random assignment breaks the correlation between tuition and u.
- 1 mark: correctly links this back to Assumption 1 being restored / causal interpretation being valid.

(c) **(3 marks)** Type I/II trade-off

- 1 mark: correct Type I error definition in context.
- 1 mark: correct Type II error definition in context.
- 1 mark: correct description of the trade-off when the cutoff is raised.