HE2003 Econometrics I

Quiz Practice 1 (Questions)

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.

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

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

(b) (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.

Unofficial QRS material

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.

Unofficial – QRS@NTU – qrs.ntu.edu.sg

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.

Unofficial QRS@NTU

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 $\hat{\beta}_1$.

Unofficial QRS@NTU question

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.

Unofficial QRS@NTU practice set