HE2003 Econometrics I

Quiz Practice 2 (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. This paper is slightly harder than Practice 1: it includes an estimator-comparison-with-weights derivation, a Least Squares Assumption 1 counterexample, a $TSS/SSR/R^2$ computation, and a sampling-distribution question about the difference of two independent sample means.
- **Format:** 5 questions, 50 marks total, approximately 60 minutes.
- **Question mix:** Q1 derivation (efficient combination of two independent unbiased estimators, with weights); Q2 regression-output interpretation (prediction, change in prediction, and $TSS/SSR/R^2$); Q3 sampling distribution of a difference of two independent sample means (mean, variance, a standardised probability, and the effect of quadrupling n on the standard deviation); Q4 binary regressor (group means, and a common-slope/different-intercepts extension with a second regressor); Q5 conceptual (selection bias, a Least Squares Assumption 1 counterexample, Type I/II trade-off).
- **Standard normal table** ($\Phi(z) = P(Z \leq z)$ for $Z \sim N(0,1)$), needed in Question 3:

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, $\Phi(-z) = 1 - \Phi(z)$.

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)

A government agency wants to estimate the true population mean μ of weekly working hours in a sector. It combines two **independent** unbiased surveys: survey 1 has $\hat{\mu}_1$ with $\text{Var}(\hat{\mu}_1) = v_1$, and survey 2 has $\hat{\mu}_2$ with $\text{Var}(\hat{\mu}_2) = v_2$, where both are unbiased for μ, i.e. $E[\hat{\mu}_1] = E[\hat{\mu}_2] = \mu$. For a weight $w \in [0, 1]$, consider the combined estimator

$$\hat{\mu}_w = w\hat{\mu}_1 + (1 - w)\hat{\mu}_2.$$

(a) (3 marks) Show that $\hat{\mu}_w$ is unbiased for μ for *every* choice of $w \in [0, 1]$.

(b) (4 marks) Using independence of $\hat{\mu}_1$ and $\hat{\mu}_2$, derive $\text{Var}(\hat{\mu}_w)$ as a function of w, v_1, and v_2.

(c) (3 marks) Find the weight w^* that minimizes $\text{Var}(\hat{\mu}_w)$, and show that it can be written as $w^* = \dfrac{v_2}{v_1 + v_2}$ (i.e. the more precise survey, with the smaller variance, receives the larger weight).

Unofficial QRS@NTU – qrsntu.org

Question 2 (10 Marks)

PUB (Singapore's national water agency) commissions a study of $n = 200$ households and regresses monthly water consumption (in cubic metres, m^3) on household size (number of residents). The estimated regression is

$$\widehat{\text{Water}}_i = 3.8 + 5.4\,\text{Size}_i, \qquad \text{SE}(\hat{\beta}_1) = 1.2.$$

(a) (3 marks) What is the regression's predicted monthly water consumption for a household of 5 residents?

(b) (2 marks) A household grows from 3 to 6 residents. What is the regression's predicted change in monthly water consumption?

(c) (3 marks) For this regression, TSS = 5000 and SSR = 1500 (in units of $(\text{m}^3)^2$). Compute ESS and R^2, and interpret R^2 in words.

(d) (2 marks) Interpret the slope coefficient with units, and comment on whether the intercept is economically meaningful here.

Unofficial – QRS@NTU – qrs.ntu.edu.sg

Question 3 (10 Marks)

Two independent surveys estimate average weekly online-shopping spending (in SGD) in two regions of a large Southeast Asian city. Region A: sample size $n_A = 100$, sample mean $\bar{Y}_A$, with $\text{Var}(\bar{Y}_A) = 16$. Region B: sample size $n_B = 150$, sample mean $\bar{Y}_B$, with $\text{Var}(\bar{Y}_B) = 9$. The two samples are drawn independently, and the true population means satisfy $\mu_A - \mu_B = 10$.

(a) (3 marks) Derive $E[\bar{Y}_A - \bar{Y}_B]$ and $\text{Var}(\bar{Y}_A - \bar{Y}_B)$, using the independence of the two samples.

(b) (4 marks) In large samples, $\bar{Y}_A - \bar{Y}_B$ is approximately normally distributed. Standardise and use the standard normal table in the Overview to compute $P(\bar{Y}_A - \bar{Y}_B > 17.5)$.

(c) (3 marks) Suppose both sample sizes are quadrupled ($n'_A = 400$, $n'_B = 600$) while the population variances of individual spending stay the same. Explain what happens to $\text{Var}(\bar{Y}_A)$ and $\text{Var}(\bar{Y}_B)$, and hence find the new standard deviation of $\bar{Y}_A - \bar{Y}_B$.

Unofficial QRS@NTU practice

Question 4 (10 Marks)

A study of $n = 160$ hawker stalls in Singapore regresses average daily revenue (in SGD) on D_i, a dummy variable equal to 1 if the stall accepts e-payment and 0 if it is cash-only ($n_0 = 60$ cash-only, $n_1 = 100$ e-payment). The estimated regression is

$$\widehat{\text{Revenue}}_i = 520 + 85\, D_i, \qquad \text{SE}(\hat{\beta}_1) = 28.$$

(a) (3 marks) State the average daily revenue predicted by the regression for cash-only stalls and for e-payment stalls.

(b) (2 marks) Give an economic interpretation of $\hat{\beta}_1 = 85$.

(c) (3 marks) The platform's data team re-estimates the model to control for the number of hours each stall operates per day, Hours_i, assuming a *common slope* on Hours_i for both groups (i.e. the effect of an extra hour of operation is the same for cash-only and e-payment stalls):

$$\widehat{\text{Revenue}}_i = 150 + 85\, D_i + 45\, \text{Hours}_i.$$

Write the separate fitted lines for cash-only stalls ($D = 0$) and e-payment stalls ($D = 1$) as functions of Hours_i. State which group is the *base group*, and give an economic interpretation of the coefficient 85 on D_i in this extended model.

(d) (2 marks) Explain the *dummy variable trap*: why can the regression in part (c) *not* also include a second dummy $C_i = 1 - D_i$ (an indicator for cash-only) alongside D_i and the intercept? State the general rule for how many dummy variables to include for a categorical variable with g mutually exclusive groups.

Unofficial QRS Practice

Question 5 (10 Marks)

(a) (4 marks) A ride-hailing platform in a large Southeast Asian city regresses a driver's weekly earnings on the number of hours the driver chose to be online (available for trips), using data from drivers who freely set their own hours. The platform's data team finds a strong positive coefficient and concludes that staying online longer *causes* higher weekly earnings on a per-hour basis. Explain, using Least Squares Assumption 1, why this conclusion may not be valid, giving a specific unobserved factor and the likely direction of the resulting bias.

(b) (3 marks) **Least Squares Assumption 1 counterexample.** Let $X \in \{0, 1\}$ be a binary regressor with $P(X = 0) = P(X = 1) = 0.5$, and let u be an error term with $E[u \mid X = 0] = -2$ and $E[u \mid X = 1] = 2$.

(i) Show that the *unconditional* mean $E[u] = 0$.

(ii) Explain why Least Squares Assumption 1 is nonetheless violated, and briefly describe (in words or with a simple sketch) what this would look like: two groups whose regression errors are centred at different, non-zero values even though the overall average error is zero.

(c) (3 marks) A public-health agency screens travellers for a disease outbreak using a rapid test, classifying each traveller as "positive" (quarantine) or "negative" (release) based on a biomarker threshold. Define, in this context, a Type I error and a Type II error, and explain the trade-off the agency faces if it lowers the biomarker threshold required to be classified "positive".

Unofficial QRS@NTU