HE2003 Econometrics I

Tutorial 5 (Week 6) – Problems & Solutions

Academic Year 2026/2027, Semester 1

Quantitative Research Society @NTU

October 12, 2026

Overview

- **Lecture coverage:** Lecture 5 — *unbiasedness and efficiency of $\bar{Y}$, the Gauss–Markov theorem for the OLS estimators, and the general condition for a linear estimator $\sum a_i Y_i$ to be unbiased*; plus Lecture 4's decomposition TSS = ESS + SSR, used in Q2 to show $R^2 \leq 1$.
- **Questions:** 3 questions from Tutorial Question Set No. 5.
- **Notation:** as in lectures — u_i population error, $\hat{u}_i$ OLS residual; ESS explained, SSR residual, TSS = ESS + SSR; "Least Squares Assumption 1/2/3" for the OLS assumptions (never "SLR.1–5"); critical value 1.96 for a two-sided 5% test.
- **Question themes:** Q1 – unbiasedness, efficiency and BLUE for $\bar{Y}$ and for OLS (Gauss–Markov); Q2 – the decomposition TSS = ESS + SSR and $R^2 \leq 1$; Q3 – mean and variance of a difference of two independent sample means.

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 (Unbiasedness, efficiency, Gauss–Markov, and BLUE)

Problem

(a) Explain the notion of unbiasedness and efficiency by using the sample average $\bar{Y}$ as an example.

(b) Explain the Gauss–Markov Theorem for the OLS estimators.

(c) Find out the condition that guarantees the linear estimator $\sum_{i=1}^{n} a_i Y_i$ is an unbiased estimator of the population mean μ.

(d) Explain the example regarding the efficiency of $\bar{Y}$ on slides 6–8 of Lecture 5.

(Tutorial Question Set No. 5, Q1.)

Solution

Part (a): Unbiasedness and efficiency, via $\bar{Y}$

Method 1: Definitions, then direct expectation and variance of each estimator

Let $Y_1, \ldots, Y_n$ be i.i.d. draws from a population with mean $E[Y_i] = \mu$ and variance $\text{Var}(Y_i) = \sigma^2 > 0$. The numbers μ and σ^2 are population quantities: fixed and unknown. An estimator of μ is a sample statistic, a function of $Y_1, \ldots, Y_n$, so it is a random variable with its own sampling distribution.

- An estimator $\hat{\mu}$ of μ is *unbiased* if $E[\hat{\mu}] = \mu$. Averaged over all possible samples it neither over- nor under-states the population mean, so its sampling distribution is centred on the true value.
- If $\hat{\mu}$ and $\check{\mu}$ are both unbiased for μ, then $\hat{\mu}$ is *more efficient* than $\check{\mu}$ when $\text{Var}(\hat{\mu}) < \text{Var}(\check{\mu})$. Its sampling distribution is more tightly concentrated around μ, so a single estimate is typically closer to the truth.
- Efficiency is compared only among unbiased estimators. The constant guess $\hat{\mu} = 40$ has variance 0, but it is biased whenever $\mu \neq 40$, so it is not a competitor.
- If Y_i is an employee's weekly working hours, then μ, $\bar{Y}$ and every estimator below are measured in hours, and their variances are measured in hours^2.

To illustrate, compare three estimators of μ: the sample average $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$, the two-point average $\tilde{Y} = \frac{1}{2}(Y_1 + Y_n)$, and the single observation Y_1. Assume $n \geq 2$, so that Y_1 and Y_n are different observations.

Unbiasedness. The expectation is linear and $E[Y_i] = \mu$ for every i, so

$$E[\bar{Y}] = E\Big[\frac{1}{n}\sum_{i=1}^{n} Y_i\Big] = \frac{1}{n}\sum_{i=1}^{n} E[Y_i] = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{n}\cdot n\mu = \mu,$$
$$E[\tilde{Y}] = E\Big[\tfrac{1}{2}(Y_1 + Y_n)\Big] = \tfrac{1}{2}\big(E[Y_1] + E[Y_n]\big) = \tfrac{1}{2}(\mu + \mu) = \mu,$$
$$E[Y_1] = \mu.$$

Hence all three estimators are unbiased for μ.

Variances. For constants $c_1, \ldots, c_n$ the variance of a weighted sum is

$$\operatorname{Var}\Big(\sum_{i=1}^{n} c_i Y_i\Big) = \sum_{i=1}^{n} c_i^2 \operatorname{Var}(Y_i) + 2\sum_{i<j} c_i c_j \operatorname{Cov}(Y_i, Y_j).$$

The Y_i are independent, so $\operatorname{Cov}(Y_i, Y_j) = 0$ for $i \neq j$ and every cross term is zero. Therefore

$$\operatorname{Var}(\bar{Y}) = \operatorname{Var}\Big(\frac{1}{n}\sum_{i=1}^{n} Y_i\Big) = \frac{1}{n^2}\sum_{i=1}^{n} \operatorname{Var}(Y_i) = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n},$$
$$\operatorname{Var}(\tilde{Y}) = \tfrac{1}{4}\big[\operatorname{Var}(Y_1) + \operatorname{Var}(Y_n) + 2\operatorname{Cov}(Y_1, Y_n)\big] = \tfrac{1}{4}(\sigma^2 + \sigma^2 + 0) = \frac{\sigma^2}{2},$$
$$\operatorname{Var}(Y_1) = \sigma^2.$$

Ranking. Since $\sigma^2 > 0$,

$$\frac{\sigma^2}{n} < \frac{\sigma^2}{2} \iff \frac{1}{n} < \frac{1}{2} \iff n > 2, \qquad \text{and} \qquad \frac{\sigma^2}{2} < \sigma^2.$$

For example, with $n = 10$ and $\sigma^2 = 1$ hour^2, the three variances are 0.1, 0.5 and 1 hour^2, with standard deviations 0.32, 0.71 and 1 hour: the sampling error of $\bar{Y}$ is typically about one third of the sampling error of a single observation.

$$\boxed{\operatorname{Var}(\bar{Y}) = \frac{\sigma^2}{n} \;<\; \operatorname{Var}(\tilde{Y}) = \frac{\sigma^2}{2} \quad \text{for } n > 2,}$$

Adding the single observation Y_1, which is also unbiased, with $\operatorname{Var}(Y_1) = \sigma^2$, the full ranking for $n > 2$ is

$$\boxed{\operatorname{Var}(\bar{Y}) = \frac{\sigma^2}{n} \;<\; \operatorname{Var}(\tilde{Y}) = \frac{\sigma^2}{2} \;<\; \operatorname{Var}(Y_1) = \sigma^2.}$$

Method 2: Weights representation $\hat{\mu} = \sum_{i=1}^{n} a_i Y_i$

Each of the three estimators is a linear estimator $\hat{\mu} = \sum_{i=1}^{n} a_i Y_i$ with constant weights:

$$\begin{aligned} \bar{Y}: &\quad a_i = \tfrac{1}{n} \ \text{ for } i = 1, \dots, n, \\ \tilde{Y}: &\quad a_1 = a_n = \tfrac{1}{2}, \ \ a_i = 0 \ \text{ otherwise}, \\ Y_1: &\quad a_1 = 1, \ \ a_i = 0 \ \text{ otherwise}. \end{aligned}$$

For general weights, linearity of the expectation gives

$$E\Big[\sum_{i=1}^{n} a_i Y_i\Big] = \sum_{i=1}^{n} a_i E[Y_i] = \sum_{i=1}^{n} a_i \mu = \mu \sum_{i=1}^{n} a_i,$$

and independence (all covariances between different observations are zero) gives

$$\operatorname{Var}\Big(\sum_{i=1}^{n} a_i Y_i\Big) = \sum_{i=1}^{n} a_i^2 \operatorname{Var}(Y_i) + 2\sum_{i<j} a_i a_j \cdot 0 = \sigma^2 \sum_{i=1}^{n} a_i^2.$$

So an estimator of this form is unbiased exactly when its weights sum to one, and its variance is σ^2 times the sum of squared weights.

Weight sums.

$$\begin{aligned} \bar{Y}: &\quad \sum_{i=1}^{n} a_i = n \cdot \tfrac{1}{n} = 1, \\ \tilde{Y}: &\quad \sum_{i=1}^{n} a_i = \tfrac{1}{2} + \tfrac{1}{2} = 1, \\ Y_1: &\quad \sum_{i=1}^{n} a_i = 1. \end{aligned}$$

Hence $E[\bar{Y}] = E[\tilde{Y}] = E[Y_1] = \mu \cdot 1 = \mu$: all three are unbiased.

Sums of squared weights.

$$\begin{aligned} \bar{Y}: &\quad \sum_{i=1}^{n} a_i^2 = n \cdot \frac{1}{n^2} = \frac{1}{n} && \Longrightarrow \ \operatorname{Var}(\bar{Y}) = \frac{\sigma^2}{n}, \\ \tilde{Y}: &\quad \sum_{i=1}^{n} a_i^2 = \tfrac{1}{4} + \tfrac{1}{4} = \tfrac{1}{2} && \Longrightarrow \ \operatorname{Var}(\tilde{Y}) = \frac{\sigma^2}{2}, \\ Y_1: &\quad \sum_{i=1}^{n} a_i^2 = 1^2 = 1 && \Longrightarrow \ \operatorname{Var}(Y_1) = \sigma^2. \end{aligned}$$

All three spend a total weight of 1; spreading that weight evenly over more observations lowers $\sum_{i=1}^{n} a_i^2$, and with it the variance. Comparing,

$$\frac{1}{n} < \frac{1}{2} \iff n > 2, \qquad \frac{1}{2} < 1.$$

$$\boxed{\operatorname{Var}(\bar{Y}) = \frac{\sigma^2}{n} \ < \ \operatorname{Var}(\tilde{Y}) = \frac{\sigma^2}{2} \quad \text{for } n > 2,}$$

and, for $n > 2$, the full ranking is

$$\boxed{\operatorname{Var}(\bar{Y}) = \frac{\sigma^2}{n} \ < \ \operatorname{Var}(\tilde{Y}) = \frac{\sigma^2}{2} \ < \ \operatorname{Var}(Y_1) = \sigma^2.}$$

Part (b): Gauss–Markov Theorem for the OLS estimators

Method 1: Statement, assumptions and meaning

The population regression model is $Y_i = \beta_0 + \beta_1 X_i + u_i$, where β_0, β_1 are population parameters and u_i is the error. The OLS estimators are sample statistics:

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

Since $\sum_{i=1}^n (X_i - \bar{X})\bar{Y} = \bar{Y} \sum_{i=1}^n (X_i - \bar{X}) = \bar{Y}\,(n\bar{X} - n\bar{X}) = 0$, the numerator equals $\sum_{i=1}^n (X_i - \bar{X}) Y_i$, so

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

a weighted sum of the Y_i with weights that depend only on $X_1, \ldots, X_n$.

- *Assumptions.* Least Squares Assumption 1 ($E[u_i \mid X_i] = 0$), Least Squares Assumption 2 ((X_i, Y_i), $i = 1, \ldots, n$, are i.i.d.), Least Squares Assumption 3 (large outliers are unlikely: finite fourth moments), plus homoskedasticity, $\mathrm{Var}(u_i \mid X_i) = \sigma_u^2$ for every i (the "technical assumption" of the slides).

- *Linear* means an estimator of the form $\tilde{\beta} = \sum_{i=1}^n c_i Y_i$, where the weights c_i may depend on $X_1, \ldots, X_n$ but not on $Y_1, \ldots, Y_n$; the display above shows that $\hat{\beta}_1$ (and hence $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$) is linear. *Unbiased* means $E[\tilde{\beta}_1] = \beta_1$ (respectively $E[\tilde{\beta}_0] = \beta_0$) whatever the true population values of β_0, β_1.

- *Best* means smallest variance: for every linear unbiased $\tilde{\beta}_0, \tilde{\beta}_1$, $\mathrm{Var}(\hat{\beta}_0) \le \mathrm{Var}(\tilde{\beta}_0)$ and $\mathrm{Var}(\hat{\beta}_1) \le \mathrm{Var}(\tilde{\beta}_1)$. So OLS is the Best Linear Unbiased Estimator (BLUE): no other linear unbiased rule gives estimates of the population slope that are more tightly concentrated around β_1.

- *Scope.* The theorem ranks only linear unbiased estimators (a biased estimator can have a smaller variance); it does not need normal errors; and without homoskedasticity OLS is still unbiased (Least Squares Assumption 1 suffices) but need not be best, since weighted least squares can then have a smaller variance.

- *Illustration (hypothetical numbers).* Let $X = (1, 2, 4)$ be hours of tutoring, Y a test score in points, and $\sigma_u = 10$ points. Then $\bar{X} = 7/3$, $X_i - \bar{X} = (-\frac{4}{3}, -\frac{1}{3}, \frac{5}{3})$, $\sum_{i=1}^n (X_i - \bar{X})^2 = \frac{16+1+25}{9} = \frac{14}{3}$, and $\mathrm{Var}(\hat{\beta}_1 \mid X) = \sigma_u^2 / \frac{14}{3} = \frac{300}{14} = 21.43$ (points per hour)^2, a standard deviation of 4.63 points per hour. The endpoint slope $\tilde{\beta}_1 = (Y_3 - Y_1)/(4 - 1)$ has weights $(-\frac{1}{3}, 0, \frac{1}{3})$ with $\sum c_i = 0$ and $\sum c_i X_i = -\frac{1}{3} + \frac{4}{3} = 1$, so it is linear and unbiased, but $\mathrm{Var}(\tilde{\beta}_1 \mid X) = 100\big(\frac{1}{9} + \frac{1}{9}\big) = 22.22$ (points per hour)^2, a standard deviation of 4.71 points per hour: larger than for OLS.

Least Squares Assumptions 1–3 + homoskedasticity
$\Longrightarrow$ $\mathrm{Var}(\hat{\beta}_0) \le \mathrm{Var}(\tilde{\beta}_0)$ and $\mathrm{Var}(\hat{\beta}_1) \le \mathrm{Var}(\tilde{\beta}_1)$
for all linear unbiased $\tilde{\beta}_0, \tilde{\beta}_1$: $\hat{\beta}_0, \hat{\beta}_1$ are BLUE.

Method 2: Proof through the OLS weights and a variance decomposition

(Non-examinable proof; beyond the Lecture 5 slides, which state the theorem without proof.)
All expectations and variances below are conditional on $X = (X_1, \ldots, X_n)$; write $S_{XX} = \sum_{i=1}^n (X_i - \bar{X})^2 > 0$.

Step 1: the OLS weights. From Method 1, $\hat{\beta}_1 = \sum_{i=1}^n w_i Y_i$ with $w_i = (X_i - \bar{X})/S_{XX}$. Also

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = \sum_{i=1}^n \frac{1}{n} Y_i - \bar{X} \sum_{i=1}^n w_i Y_i = \sum_{i=1}^n v_i Y_i, \qquad v_i = \frac{1}{n} - \bar{X} w_i.$$

The weights satisfy four identities. Using $\sum_{i=1}^n X_i = n\bar{X}$,

$$\begin{aligned}
\sum_{i=1}^n w_i &= \frac{1}{S_{XX}} \sum_{i=1}^n (X_i - \bar{X}) = \frac{1}{S_{XX}} (n\bar{X} - n\bar{X}) = 0, \\
\sum_{i=1}^n w_i X_i &= \frac{1}{S_{XX}} \sum_{i=1}^n (X_i - \bar{X}) X_i = \frac{1}{S_{XX}} \Big[\sum_{i=1}^n (X_i - \bar{X})^2 + \bar{X} \sum_{i=1}^n (X_i - \bar{X}) \Big] \\
&= \frac{S_{XX} + \bar{X} \cdot 0}{S_{XX}} = 1, \\
\sum_{i=1}^n v_i &= n \cdot \frac{1}{n} - \bar{X} \sum_{i=1}^n w_i = 1 - \bar{X} \cdot 0 = 1, \\
\sum_{i=1}^n v_i X_i &= \frac{1}{n} \sum_{i=1}^n X_i - \bar{X} \sum_{i=1}^n w_i X_i = \bar{X} - \bar{X} \cdot 1 = 0.
\end{aligned}$$

Step 2: what unbiasedness requires. Let $\tilde{\beta} = \sum_{i=1}^n c_i Y_i$ be any linear estimator. Substituting $Y_i = \beta_0 + \beta_1 X_i + u_i$,

$$\tilde{\beta} = \beta_0 \sum_{i=1}^n c_i + \beta_1 \sum_{i=1}^n c_i X_i + \sum_{i=1}^n c_i u_i.$$

By Least Squares Assumptions 1 and 2, $E[u_i \,|\, X] = E[u_i \,|\, X_i] = 0$, so

$$E[\tilde{\beta} \,|\, X] = \beta_0 \sum_{i=1}^n c_i + \beta_1 \sum_{i=1}^n c_i X_i.$$

For this to equal β_1 for every value of (β_0, β_1) we need $\sum_{i=1}^n c_i = 0$ and $\sum_{i=1}^n c_i X_i = 1$; for it to equal β_0 for every value we need $\sum_{i=1}^n c_i = 1$ and $\sum_{i=1}^n c_i X_i = 0$. By Step 1 the OLS weights w_i and v_i satisfy these, so $\hat{\beta}_1$ and $\hat{\beta}_0$ are unbiased.

Step 3: variance of a linear estimator. Given X, the u_i are independent (Least Squares Assumption 2) with $\mathrm{Var}(u_i \,|\, X) = \sigma_u^2$ (homoskedasticity), so all conditional covariances between different u_i vanish and

$$\mathrm{Var}(\tilde{\beta} \,|\, X) = \mathrm{Var}\Big(\sum_{i=1}^n c_i u_i \,\Big|\, X \Big) = \sum_{i=1}^n c_i^2 \sigma_u^2 = \sigma_u^2 \sum_{i=1}^n c_i^2.$$

Step 4: decomposition for β_1. Let $\tilde{\beta}_1 = \sum_{i=1}^n c_i Y_i$ be linear and unbiased for β_1, and put $d_i = c_i - w_i$. By Steps 1–2,

$$\sum_{i=1}^n d_i = \sum_{i=1}^n c_i - \sum_{i=1}^n w_i = 0 - 0 = 0, \qquad \sum_{i=1}^n d_i X_i = \sum_{i=1}^n c_i X_i - \sum_{i=1}^n w_i X_i = 1 - 1 = 0.$$

The cross term between w and d is zero:

$$\sum_{i=1}^n w_i d_i = \frac{1}{S_{XX}} \sum_{i=1}^n (X_i - \bar{X}) d_i = \frac{1}{S_{XX}} \Big[\sum_{i=1}^n X_i d_i - \bar{X} \sum_{i=1}^n d_i \Big] = \frac{1}{S_{XX}} (0 - 0) = 0.$$

Expanding the square,

$$\sum_{i=1}^n c_i^2 = \sum_{i=1}^n (w_i + d_i)^2 = \sum_{i=1}^n w_i^2 + 2 \sum_{i=1}^n w_i d_i + \sum_{i=1}^n d_i^2 = \sum_{i=1}^n w_i^2 + \sum_{i=1}^n d_i^2.$$

Multiplying by σ_u^2 and using Step 3,

$$\mathrm{Var}(\tilde{\beta}_1 \,|\, X) = \mathrm{Var}(\hat{\beta}_1 \,|\, X) + \sigma_u^2 \sum_{i=1}^n d_i^2 \;\geq\; \mathrm{Var}(\hat{\beta}_1 \,|\, X),$$

with equality if and only if every $d_i = 0$, that is, $\tilde{\beta}_1 = \hat{\beta}_1$. In the illustration of Method 1, $d = (-\frac{1}{3} + \frac{2}{7},\ \frac{1}{14},\ \frac{1}{3} - \frac{5}{14}) = (-\frac{1}{21}, \frac{1}{14}, -\frac{1}{42})$, $\sum d_i^2 = \frac{4+9+1}{1764} = \frac{1}{126}$, and $21.43 + 100 \cdot \frac{1}{126} = 21.43 + 0.79 = 22.22$.

Step 5: decomposition for β_0. Let $\tilde{\beta}_0 = \sum_{i=1}^n c_i Y_i$ be linear and unbiased for β_0, and put $d_i = c_i - v_i$. Then

$$\sum_{i=1}^n d_i = 1 - 1 = 0, \qquad \sum_{i=1}^n d_i X_i = 0 - 0 = 0,$$

so the computation of Step 4 again gives $\sum_{i=1}^n w_i d_i = 0$, and therefore

$$\sum_{i=1}^n v_i d_i = \frac{1}{n} \sum_{i=1}^n d_i - \bar{X} \sum_{i=1}^n w_i d_i = 0 - \bar{X} \cdot 0 = 0.$$

Expanding $\sum_{i=1}^n c_i^2 = \sum_{i=1}^n (v_i + d_i)^2 = \sum_{i=1}^n v_i^2 + \sum_{i=1}^n d_i^2$ and multiplying by σ_u^2,

$$\mathrm{Var}(\tilde{\beta}_0 \,|\, X) = \mathrm{Var}(\hat{\beta}_0 \,|\, X) + \sigma_u^2 \sum_{i=1}^n d_i^2 \;\geq\; \mathrm{Var}(\hat{\beta}_0 \,|\, X).$$

Least Squares Assumptions 1–3 + homoskedasticity
$\Longrightarrow$ $\mathrm{Var}(\hat{\beta}_0) \leq \mathrm{Var}(\tilde{\beta}_0)$ and $\mathrm{Var}(\hat{\beta}_1) \leq \mathrm{Var}(\tilde{\beta}_1)$
for all linear unbiased $\tilde{\beta}_0, \tilde{\beta}_1$: $\hat{\beta}_0, \hat{\beta}_1$ are BLUE.

Part (c): unbiasedness condition for $\sum a_i Y_i$

Method 1: Direct expectation, then completing the square in the weights

Let $E[Y_i] = \mu$ for every i, let $a_1, \ldots, a_n$ be constants (non-random, fixed before the data are seen), and let $\hat{\mu} = \sum_{i=1}^{n} a_i Y_i$.

Unbiasedness. Constants pass through the expectation, and the expectation of a sum is the sum of the expectations:

$$E[\hat{\mu}] = E\Big[\sum_{i=1}^{n} a_i Y_i\Big] = \sum_{i=1}^{n} E[a_i Y_i] = \sum_{i=1}^{n} a_i E[Y_i] = \sum_{i=1}^{n} a_i \mu = \mu \sum_{i=1}^{n} a_i.$$

If $\sum_{i=1}^{n} a_i = 1$, then $E[\hat{\mu}] = \mu \cdot 1 = \mu$, so $\hat{\mu}$ is unbiased. Conversely, unbiasedness must hold whatever the value of the unknown population mean μ, because the weights are chosen before μ is known. Hence

$$\mu \sum_{i=1}^{n} a_i = \mu \ \text{ for every } \mu \iff \mu\Big(\sum_{i=1}^{n} a_i - 1\Big) = 0 \ \text{ for every } \mu.$$

Taking $\mu = 1$ gives $\sum_{i=1}^{n} a_i - 1 = 0$, that is, $\sum_{i=1}^{n} a_i = 1$. For example, with $n = 3$ the weights $(0.5, 0.3, 0.2)$ sum to 1 and give an unbiased estimator, while $a_i = \frac{1}{n-1}$ for all i gives $E[\hat{\mu}] = \frac{n}{n-1}\mu$, a bias of $\frac{n}{n-1}\mu - \mu = \frac{\mu}{n-1}$ (upward when $\mu > 0$).

$$\boxed{\hat{\mu} = \sum_{i=1}^{n} a_i Y_i \text{ is unbiased for } \mu \iff \sum_{i=1}^{n} a_i = 1.}$$

Variance. Now let the Y_i be i.i.d. with $\mathrm{Var}(Y_i) = \sigma^2$. Independence makes every covariance between different observations zero, so

$$\mathrm{Var}(\hat{\mu}) = \sum_{i=1}^{n} a_i^2 \,\mathrm{Var}(Y_i) + 2\sum_{i<j} a_i a_j \,\mathrm{Cov}(Y_i, Y_j) = \sum_{i=1}^{n} a_i^2 \sigma^2 + 0 = \sigma^2 \sum_{i=1}^{n} a_i^2.$$

Minimising $\sum_{i=1}^{n} a_i^2$ subject to $\sum_{i=1}^{n} a_i = 1$. Put $b_i = a_i - \frac{1}{n}$, so $a_i = \frac{1}{n} + b_i$. Under unbiasedness,

$$\sum_{i=1}^{n} b_i = \sum_{i=1}^{n} a_i - n \cdot \frac{1}{n} = 1 - 1 = 0.$$

Expanding each square and summing,

$$\begin{aligned}\sum_{i=1}^{n} a_i^2 &= \sum_{i=1}^{n} \Big(\frac{1}{n} + b_i\Big)^2 = \sum_{i=1}^{n} \Big(\frac{1}{n^2} + \frac{2}{n} b_i + b_i^2\Big) = n \cdot \frac{1}{n^2} + \frac{2}{n}\sum_{i=1}^{n} b_i + \sum_{i=1}^{n} b_i^2 \\ &= \frac{1}{n} + \frac{2}{n} \cdot 0 + \sum_{i=1}^{n} b_i^2 = \frac{1}{n} + \sum_{i=1}^{n} \Big(a_i - \frac{1}{n}\Big)^2.\end{aligned}$$

$$\boxed{\sum_{i=1}^{n} a_i^2 = \sum_{i=1}^{n} \Big(a_i - \frac{1}{n}\Big)^2 + \frac{1}{n} \ \geq \ \frac{1}{n},}$$

A sum of squares is zero only when every term is zero, so equality holds if and only if $a_i = \frac{1}{n}$ for every i, that is, $\hat{\mu} = \bar{Y}$. Multiplying by σ^2,

$$\text{Var}(\hat{\mu}) = \sigma^2 \sum_{i=1}^{n} a_i^2 \;\geq\; \frac{\sigma^2}{n} = \text{Var}(\bar{Y})$$

for every linear unbiased $\hat{\mu}$, with equality only at $\hat{\mu} = \bar{Y}$.

$\bar{Y}$ is the Best Linear Unbiased Estimator (BLUE) of μ.

Unofficial_QRS@NTU_unofficial.org

Method 2: Error representation and Cauchy–Schwarz

Write each observation as its population mean plus a deviation,

$$Y_i = \mu + \varepsilon_i, \qquad \varepsilon_i = Y_i - \mu, \qquad E[\varepsilon_i] = 0, \qquad \mathrm{Var}(\varepsilon_i) = \sigma^2,$$

with $\varepsilon_1, \ldots, \varepsilon_n$ independent. This is the regression model $Y_i = \beta_0 + u_i$ with $\beta_0 = \mu$, and the step below mirrors $\hat{\beta}_1 - \beta_1 = \sum_{i=1}^n w_i u_i$.

Unbiasedness. Substituting,

$$\hat{\mu} = \sum_{i=1}^n a_i(\mu + \varepsilon_i) = \mu \sum_{i=1}^n a_i + \sum_{i=1}^n a_i \varepsilon_i, \qquad \text{so} \qquad \hat{\mu} - \mu = \mu\Big(\sum_{i=1}^n a_i - 1\Big) + \sum_{i=1}^n a_i \varepsilon_i.$$

Taking expectations and using $E[\varepsilon_i] = 0$, the bias is

$$E[\hat{\mu}] - \mu = \mu\Big(\sum_{i=1}^n a_i - 1\Big) + \sum_{i=1}^n a_i \cdot 0 = \mu\Big(\sum_{i=1}^n a_i - 1\Big).$$

The bias must be zero for every value of μ, in particular for $\mu = 1$, which forces $\sum_{i=1}^n a_i - 1 = 0$; conversely $\sum_{i=1}^n a_i = 1$ makes the bias zero for every μ.

$$\boxed{\hat{\mu} = \sum_{i=1}^n a_i Y_i \text{ is unbiased for } \mu \iff \sum_{i=1}^n a_i = 1.}$$

Variance. When $\sum_{i=1}^n a_i = 1$, $\hat{\mu} - \mu = \sum_{i=1}^n a_i \varepsilon_i$. For $i \neq j$, independence gives $E[\varepsilon_i \varepsilon_j] = E[\varepsilon_i]E[\varepsilon_j] = 0$, so

$$\mathrm{Var}(\hat{\mu}) = E\Big[\Big(\sum_{i=1}^n a_i \varepsilon_i\Big)^2\Big] = \sum_{i=1}^n \sum_{j=1}^n a_i a_j E[\varepsilon_i \varepsilon_j] = \sum_{i=1}^n a_i^2 E[\varepsilon_i^2] = \sigma^2 \sum_{i=1}^n a_i^2.$$

Lower bound. Recall the Cauchy–Schwarz inequality for real numbers, $\big(\sum_{i=1}^n x_i y_i\big)^2 \leq \big(\sum_{i=1}^n x_i^2\big)\big(\sum_{i=1}^n y_i^2\big)$, with equality if and only if (x_i) is proportional to (y_i). With $x_i = a_i$ and $y_i = 1$,

$$1 = \Big(\sum_{i=1}^n a_i\Big)^2 = \Big(\sum_{i=1}^n a_i \cdot 1\Big)^2 \leq \Big(\sum_{i=1}^n a_i^2\Big)\Big(\sum_{i=1}^n 1^2\Big) = n \sum_{i=1}^n a_i^2, \qquad \text{so} \qquad \sum_{i=1}^n a_i^2 \geq \frac{1}{n}.$$

To measure the gap in this inequality, expand the squared deviations of the weights from $\frac{1}{n}$:

$$\sum_{i=1}^n \Big(a_i - \frac{1}{n}\Big)^2 = \sum_{i=1}^n a_i^2 - \frac{2}{n}\sum_{i=1}^n a_i + n \cdot \frac{1}{n^2} = \sum_{i=1}^n a_i^2 - \frac{2}{n} + \frac{1}{n} = \sum_{i=1}^n a_i^2 - \frac{1}{n}.$$

$$\boxed{\sum_{i=1}^n a_i^2 = \sum_{i=1}^n \Big(a_i - \frac{1}{n}\Big)^2 + \frac{1}{n} \;\geq\; \frac{1}{n},}$$

Equality in Cauchy–Schwarz holds if and only if $a_i = k$ for all i; then $\sum_{i=1}^n a_i = nk = 1$ gives $k = \frac{1}{n}$, so the unique minimiser is $\hat{\mu} = \bar{Y}$ with $\mathrm{Var}(\bar{Y}) = \sigma^2/n \leq \mathrm{Var}(\hat{\mu})$.

$$\boxed{\bar{Y} \text{ is the Best Linear Unbiased Estimator (BLUE) of } \mu.}$$

Method 3: $\hat{\mu}$ as $\bar{Y}$ plus a perturbation, with an orthogonal variance decomposition

Put $b_i = a_i - \frac{1}{n}$ (no restriction yet), so that

$$\hat{\mu} = \sum_{i=1}^{n} \Big(\frac{1}{n} + b_i\Big) Y_i = \bar{Y} + Z, \qquad Z = \sum_{i=1}^{n} b_i Y_i.$$

Unbiasedness. By linearity, $E[\bar{Y}] = \mu$ and $E[Z] = \sum_{i=1}^{n} b_i \mu = \mu \sum_{i=1}^{n} b_i$, so

$$E[\hat{\mu}] = \mu + \mu \sum_{i=1}^{n} b_i = \mu\Big(1 + \sum_{i=1}^{n} b_i\Big).$$

This equals μ for every μ if and only if $\sum_{i=1}^{n} b_i = 0$, and

$$\sum_{i=1}^{n} b_i = \sum_{i=1}^{n} a_i - n \cdot \frac{1}{n} = \sum_{i=1}^{n} a_i - 1,$$

so the condition is $\sum_{i=1}^{n} a_i = 1$.

$$\boxed{\hat{\mu} = \sum_{i=1}^{n} a_i Y_i \text{ is unbiased for } \mu \iff \sum_{i=1}^{n} a_i = 1.}$$

Variance. Let the Y_i be i.i.d. with $\mathrm{Var}(Y_i) = \sigma^2$, and let $\hat{\mu}$ be unbiased, so $\sum_{i=1}^{n} b_i = 0$. By bilinearity of the covariance, and since $\mathrm{Cov}(Y_i, Y_j) = 0$ for $i \neq j$ and $\mathrm{Cov}(Y_j, Y_j) = \sigma^2$,

$$\begin{aligned}\mathrm{Cov}(\bar{Y}, Z) &= \mathrm{Cov}\Big(\frac{1}{n}\sum_{i=1}^{n} Y_i,\ \sum_{j=1}^{n} b_j Y_j\Big) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} b_j \,\mathrm{Cov}(Y_i, Y_j) \\ &= \frac{1}{n}\sum_{j=1}^{n} b_j \sigma^2 = \frac{\sigma^2}{n}\sum_{j=1}^{n} b_j = \frac{\sigma^2}{n} \cdot 0 = 0.\end{aligned}$$

So the perturbation Z is uncorrelated with $\bar{Y}$, and

$$\mathrm{Var}(\hat{\mu}) = \mathrm{Var}(\bar{Y}) + \mathrm{Var}(Z) + 2\,\mathrm{Cov}(\bar{Y}, Z) = \frac{\sigma^2}{n} + \sigma^2 \sum_{i=1}^{n} b_i^2 + 0,$$

where $\mathrm{Var}(Z) = \sigma^2 \sum_{i=1}^{n} b_i^2$ and $\mathrm{Var}(\hat{\mu}) = \sigma^2 \sum_{i=1}^{n} a_i^2$ both follow from independence (all cross covariances vanish). Dividing by $\sigma^2 > 0$:

$$\sum_{i=1}^{n} a_i^2 = \frac{1}{n} + \sum_{i=1}^{n} b_i^2 = \frac{1}{n} + \sum_{i=1}^{n} \Big(a_i - \frac{1}{n}\Big)^2.$$

$$\boxed{\sum_{i=1}^{n} a_i^2 = \sum_{i=1}^{n} \Big(a_i - \frac{1}{n}\Big)^2 + \frac{1}{n} \ \geq\ \frac{1}{n},}$$

Hence $\mathrm{Var}(\hat{\mu}) - \mathrm{Var}(\bar{Y}) = \mathrm{Var}(Z) = \sigma^2 \sum_{i=1}^{n} b_i^2 \geq 0$, which is zero only when every $b_i = 0$, that is, $\hat{\mu} = \bar{Y}$. Any other linear unbiased estimator is $\bar{Y}$ plus noise that is uncorrelated with $\bar{Y}$, and that noise only adds variance (the same structure as Method 2 of Part (b)).

$$\boxed{\bar{Y} \text{ is the Best Linear Unbiased Estimator (BLUE) of } \mu.}$$

Part (d): the Lecture 5 slides 6–8 simulation

Method 1: Reading the simulation output

The slides draw 20 independent samples, each of size $n = 10$, from the population $N(2, 1)$, so the population mean is $\mu = 2$ and the population variance is $\sigma^2 = 1$. For each sample two estimates of μ are computed: the first observation Y_1 and the sample average $\bar{Y} = \frac{1}{10}(Y_1 + \cdots + Y_{10})$. The 20 values of each estimator are 20 draws from its sampling distribution.

- *Centre (unbiasedness).* Across the 20 samples the average of Y_1 is about 1.89 and the average of $\bar{Y}$ is about 1.96. Both are close to $\mu = 2$; the gaps -0.11 and -0.04 are the size that chance alone produces with 20 replications, which is what unbiasedness ($E[Y_1] = E[\bar{Y}] = 2$) predicts.
- *Spread (efficiency).* Y_1 ranges over $[-0.64, 4.27]$, a width of $4.27 - (-0.64) = 4.91$, while $\bar{Y}$ ranges over $[1.16, 2.58]$, a width of $2.58 - 1.16 = 1.42$. The sampling distribution of $\bar{Y}$ is about $4.91/1.42 = 3.46$ times narrower.
- *Closeness.* $\bar{Y}$ lands closer to $\mu = 2$ than Y_1 does in 16 of the 20 samples (80%).
- *Theory.* By Part (a), $\mathrm{Var}(Y_1) = \sigma^2 = 1$ and $\mathrm{Var}(\bar{Y}) = \sigma^2/n = 1/10 = 0.1$, so the standard deviations are 1 and $\sqrt{0.1} = 0.32$, a predicted spread ratio of $\sqrt{10} = 3.16$, close to the observed 3.46. Both estimators are unbiased, and $\bar{Y}$, which uses all ten observations, is ten times more efficient in variance terms.

$$\boxed{\mathrm{Var}(Y_1) = 1, \quad \mathrm{Var}(\bar{Y}) = 0.1; \quad \bar{Y} \text{ is unbiased and far more efficient than } Y_1.}$$

Unofficial QRS@NTU

Method 2: Predicting the simulation from the sampling distributions

Under $Y_i \sim N(2,1)$ i.i.d., a linear combination of independent normal variables is normal, and Part (a) supplies its mean and variance:

$$Y_1 \sim N(\mu, \sigma^2) = N(2,1),$$
$$\bar{Y} \sim N\Big(\mu, \frac{\sigma^2}{n}\Big) = N\Big(2, \frac{1}{10}\Big) = N(2,\, 0.1), \qquad \sqrt{0.1} = 0.3162.$$

Spread. The central 95% ranges of the two sampling distributions are

$$Y_1: \quad 2 \pm 1.96 \times 1 = [0.04,\ 3.96],$$
$$\bar{Y}: \quad 2 \pm 1.96 \times 0.3162 = 2 \pm 0.620 = [1.38,\ 2.62].$$

Their widths are 3.92 and 1.24, in the ratio $1/\sqrt{0.1} = \sqrt{10} = 3.16$. The observed ranges $[-0.64, 4.27]$ and $[1.16, 2.58]$ have widths 4.91 and 1.42 (ratio 3.46). With 20 draws the minimum and maximum typically reach slightly beyond the central 95% range, since about one draw in twenty falls outside it, so the observed spreads match the predicted ones.

Centre. The average of 20 independent copies of an estimator with variance V has variance $V/20$. Hence the average of the 20 values of Y_1 has standard deviation $\sqrt{1/20} = 0.224$, and the average of the 20 values of $\bar{Y}$ has standard deviation $\sqrt{0.1/20} = \sqrt{0.005} = 0.0707$. The observed averages deviate from $\mu = 2$ by

$$\frac{1.89 - 2}{0.224} = -0.49 \text{ standard deviations}, \qquad \frac{1.96 - 2}{0.0707} = -0.57 \text{ standard deviations},$$

both well inside ± 1.96, so the simulation is consistent with $E[Y_1] = E[\bar{Y}] = 2$.

$$\boxed{\text{Var}(Y_1) = 1, \quad \text{Var}(\bar{Y}) = 0.1; \quad \bar{Y} \text{ is unbiased and far more efficient than } Y_1.}$$

Method 3: Probability that $\bar{Y}$ is closer to μ than Y_1

(Uses the bivariate normal orthant probability; beyond the Lecture 5 slides.)
Let $A = \bar{Y} - 2$ and $B = Y_1 - 2$. They are jointly normal with mean 0, $\mathrm{Var}(A) = \mathrm{Var}(\bar{Y}) = 0.1$, $\mathrm{Var}(B) = \mathrm{Var}(Y_1) = 1$, and, because Y_1 is uncorrelated with $Y_2, \ldots, Y_{10}$,

$$\mathrm{Cov}(A, B) = \mathrm{Cov}\left(\frac{1}{10}\sum_{i=1}^{10} Y_i,\ Y_1\right) = \frac{1}{10}\mathrm{Var}(Y_1) + \frac{1}{10}\sum_{i=2}^{10}\mathrm{Cov}(Y_i, Y_1) = 0.1 + 0 = 0.1.$$

$\bar{Y}$ is closer to 2 exactly when $|A| < |B|$, that is,

$$A^2 < B^2 \iff B^2 - A^2 > 0 \iff (B - A)(B + A) > 0 \iff UV > 0,$$

where $U = B - A$ and $V = B + A$. The moments of (U, V) are

$$\begin{aligned}
\mathrm{Var}(U) &= \mathrm{Var}(B) + \mathrm{Var}(A) - 2\,\mathrm{Cov}(A, B) = 1 + 0.1 - 0.2 = 0.9,\\
\mathrm{Var}(V) &= \mathrm{Var}(B) + \mathrm{Var}(A) + 2\,\mathrm{Cov}(A, B) = 1 + 0.1 + 0.2 = 1.3,\\
\mathrm{Cov}(U, V) &= \mathrm{Var}(B) - \mathrm{Var}(A) = 1 - 0.1 = 0.9,
\end{aligned}$$

so their correlation is

$$\rho = \frac{0.9}{\sqrt{0.9 \times 1.3}} = \sqrt{\frac{0.9}{1.3}} = \sqrt{\frac{9}{13}} = 0.8321.$$

For a zero-mean bivariate normal pair with correlation ρ, $P(UV > 0) = \frac{1}{2} + \frac{1}{\pi}\arcsin\rho$. Therefore

$$P\big(|\bar{Y} - 2| < |Y_1 - 2|\big) = \frac{1}{2} + \frac{\arcsin(0.8321)}{\pi} = \frac{1}{2} + \frac{0.9828}{\pi} = 0.5 + 0.3128 = 0.813.$$

In 20 samples the expected number of samples with $\bar{Y}$ closer is $20 \times 0.813 = 16.3$, against the observed 16.

$\mathrm{Var}(Y_1) = 1, \quad \mathrm{Var}(\bar{Y}) = 0.1; \quad \bar{Y}$ is unbiased and far more efficient than Y_1.

Question 2 (Why R^2 cannot exceed 1)

Problem

In this question we show why R^2 cannot be larger than 1.

(a) Show that $\text{TSS} = \sum_{i=1}^{n}[\hat{u}_i + (\hat{Y}_i - \bar{Y})]^2$.

(b) Show that TSS = ESS + SSR, where TSS and ESS are as defined in the lecture slides and $\text{SSR} = \sum_{i=1}^{n} \hat{u}_i^2$. [Hint: use $\sum_{i=1}^{n} \hat{u}_i \hat{Y}_i = 0$ and $\sum_{i=1}^{n} \hat{u}_i = 0$.]

(c) Show that $R^2 \leq 1$.

(Tutorial Question Set No. 5, Q2.)

Solution

Throughout, $\hat{\beta}_0, \hat{\beta}_1$ minimise $\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i)^2$ over (b_0, b_1); the fitted values are $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ and the residuals are $\hat{u}_i = Y_i - \hat{Y}_i$. The sums of squares are

$$\text{TSS} = \sum_{i=1}^{n}(Y_i - \bar{Y})^2, \qquad \text{ESS} = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2, \qquad \text{SSR} = \sum_{i=1}^{n} \hat{u}_i^2,$$

and we assume the Y_i are not all equal, so $\text{TSS} > 0$ and $R^2 = \text{ESS}/\text{TSS}$ is defined.

Method 1: Algebraic decomposition with the OLS first-order conditions

The first-order conditions. Setting the two partial derivatives of $\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i)^2$ to zero at $(\hat{\beta}_0, \hat{\beta}_1)$:

$$\frac{\partial}{\partial b_0}: \quad -2\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0 \implies \sum_{i=1}^{n} \hat{u}_i = 0,$$

$$\frac{\partial}{\partial b_1}: \quad -2\sum_{i=1}^{n} X_i(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0 \implies \sum_{i=1}^{n} X_i \hat{u}_i = 0.$$

Part (a). Since $Y_i = \hat{Y}_i + \hat{u}_i$, subtracting $\bar{Y}$ from both sides gives

$$Y_i - \bar{Y} = \hat{u}_i + (\hat{Y}_i - \bar{Y}) \qquad \text{for every } i.$$

Squaring both sides and summing over i,

$$\boxed{\text{TSS} = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}\left[\hat{u}_i + (\hat{Y}_i - \bar{Y})\right]^2.}$$

Part (b). Expand the square inside the sum of Part (a):

$$\left[\hat{u}_i + (\hat{Y}_i - \bar{Y})\right]^2 = \hat{u}_i^2 + 2\hat{u}_i(\hat{Y}_i - \bar{Y}) + (\hat{Y}_i - \bar{Y})^2.$$

Summing over i,

$$\text{TSS} = \sum_{i=1}^{n} \hat{u}_i^2 + 2\sum_{i=1}^{n} \hat{u}_i(\hat{Y}_i - \bar{Y}) + \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 = \text{SSR} + 2\sum_{i=1}^{n} \hat{u}_i(\hat{Y}_i - \bar{Y}) + \text{ESS}.$$

Split the cross term:

$$\sum_{i=1}^{n} \hat{u}_i(\hat{Y}_i - \bar{Y}) = \sum_{i=1}^{n} \hat{u}_i\hat{Y}_i - \bar{Y}\sum_{i=1}^{n} \hat{u}_i.$$

Substituting $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ and using both first-order conditions,

$$\sum_{i=1}^{n} \hat{u}_i\hat{Y}_i = \sum_{i=1}^{n} \hat{u}_i(\hat{\beta}_0 + \hat{\beta}_1 X_i) = \hat{\beta}_0 \sum_{i=1}^{n} \hat{u}_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i\hat{u}_i = \hat{\beta}_0 \cdot 0 + \hat{\beta}_1 \cdot 0 = 0,$$

and by the first condition

$$\bar{Y}\sum_{i=1}^{n} \hat{u}_i = \bar{Y} \cdot 0 = 0.$$

Hence the cross term is $0 - 0 = 0$, and $\text{TSS} = \text{SSR} + 0 + \text{ESS}$.

$$\boxed{\text{TSS} = \text{SSR} + \text{ESS} = \text{ESS} + \text{SSR}.}$$

Part (c). By Part (b), $\text{ESS} = \text{TSS} - \text{SSR}$, so

$$R^2 = \frac{\text{ESS}}{\text{TSS}} = \frac{\text{TSS} - \text{SSR}}{\text{TSS}} = 1 - \frac{\text{SSR}}{\text{TSS}}.$$

Since $\text{SSR} = \sum_{i=1}^{n} \hat{u}_i^2 \geq 0$ and $\text{TSS} > 0$, $\text{SSR}/\text{TSS} \geq 0$ and therefore $R^2 \leq 1$. Since $\text{ESS} = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 \geq 0$, $R^2 = \text{ESS}/\text{TSS} \geq 0$. For example, the hypothetical values $\text{TSS} = 200$ and $\text{SSR} = 50$ give $R^2 = 1 - 50/200 = 0.75$: the regression accounts for 75% of the sample variation of Y around $\bar{Y}$, and R^2 could reach 1 only if every residual were 0.

$$\boxed{0 \leq R^2 = 1 - \frac{\text{SSR}}{\text{TSS}} \leq 1.}$$

Method 2: Orthogonality of the residual vector (Pythagoras)

Stack the data as vectors in $\mathbb{R}^n$: $\mathbf{y} = (Y_1, \ldots, Y_n)'$, $\mathbf{x} = (X_1, \ldots, X_n)'$, $\mathbf{1} = (1, \ldots, 1)'$, $\hat{\mathbf{y}} = (\hat{Y}_1, \ldots, \hat{Y}_n)'$, $\hat{\mathbf{u}} = (\hat{u}_1, \ldots, \hat{u}_n)'$, and define $\mathbf{v} = \hat{\mathbf{y}} - \bar{Y}\mathbf{1}$, so $v_i = \hat{Y}_i - \bar{Y}$. Write $\|\mathbf{z}\|^2 = \mathbf{z}'\mathbf{z} = \sum_{i=1}^n z_i^2$.

Part (a). From $\mathbf{y} = \hat{\mathbf{y}} + \hat{\mathbf{u}}$,

$$\mathbf{y} - \bar{Y}\mathbf{1} = \hat{\mathbf{u}} + (\hat{\mathbf{y}} - \bar{Y}\mathbf{1}) = \hat{\mathbf{u}} + \mathbf{v},$$

so

$$\boxed{\text{TSS} = \|\mathbf{y} - \bar{Y}\mathbf{1}\|^2 = \|\hat{\mathbf{u}} + \mathbf{v}\|^2 = \sum_{i=1}^n \left[\hat{u}_i + (\hat{Y}_i - \bar{Y})\right]^2.}$$

Part (b). The OLS objective is $\|\mathbf{y} - b_0\mathbf{1} - b_1\mathbf{x}\|^2$. Its gradient at $(\hat{\beta}_0, \hat{\beta}_1)$ is zero:

$$-2\,\mathbf{1}'(\mathbf{y} - \hat{\beta}_0\mathbf{1} - \hat{\beta}_1\mathbf{x}) = -2\,\mathbf{1}'\hat{\mathbf{u}} = 0, \qquad -2\,\mathbf{x}'(\mathbf{y} - \hat{\beta}_0\mathbf{1} - \hat{\beta}_1\mathbf{x}) = -2\,\mathbf{x}'\hat{\mathbf{u}} = 0.$$

So the residual vector is orthogonal to both $\mathbf{1}$ and $\mathbf{x}$. The vector $\mathbf{v}$ is a linear combination of these two,

$$\mathbf{v} = \hat{\beta}_0\mathbf{1} + \hat{\beta}_1\mathbf{x} - \bar{Y}\mathbf{1} = (\hat{\beta}_0 - \bar{Y})\mathbf{1} + \hat{\beta}_1\mathbf{x},$$

hence

$$\hat{\mathbf{u}}'\mathbf{v} = (\hat{\beta}_0 - \bar{Y})\,\mathbf{1}'\hat{\mathbf{u}} + \hat{\beta}_1\,\mathbf{x}'\hat{\mathbf{u}} = (\hat{\beta}_0 - \bar{Y}) \cdot 0 + \hat{\beta}_1 \cdot 0 = 0.$$

Expanding the squared length,

$$\|\hat{\mathbf{u}} + \mathbf{v}\|^2 = (\hat{\mathbf{u}} + \mathbf{v})'(\hat{\mathbf{u}} + \mathbf{v}) = \hat{\mathbf{u}}'\hat{\mathbf{u}} + 2\,\hat{\mathbf{u}}'\mathbf{v} + \mathbf{v}'\mathbf{v} = \|\hat{\mathbf{u}}\|^2 + 0 + \|\mathbf{v}\|^2,$$

which is Pythagoras' theorem for the right angle between $\hat{\mathbf{u}}$ and $\mathbf{v}$. Since $\|\hat{\mathbf{u}}\|^2 = \text{SSR}$ and $\|\mathbf{v}\|^2 = \text{ESS}$,

$$\boxed{\text{TSS} = \text{SSR} + \text{ESS} = \text{ESS} + \text{SSR}.}$$

Part (c). Let θ be the angle between $\mathbf{y} - \bar{Y}\mathbf{1} = \hat{\mathbf{u}} + \mathbf{v}$ and $\mathbf{v}$. Using $\hat{\mathbf{u}}'\mathbf{v} = 0$,

$$\cos\theta = \frac{\mathbf{v}'(\hat{\mathbf{u}} + \mathbf{v})}{\|\mathbf{v}\|\,\|\hat{\mathbf{u}} + \mathbf{v}\|} = \frac{\mathbf{v}'\hat{\mathbf{u}} + \mathbf{v}'\mathbf{v}}{\|\mathbf{v}\|\,\|\hat{\mathbf{u}} + \mathbf{v}\|} = \frac{\|\mathbf{v}\|^2}{\|\mathbf{v}\|\,\|\hat{\mathbf{u}} + \mathbf{v}\|} = \frac{\|\mathbf{v}\|}{\|\hat{\mathbf{u}} + \mathbf{v}\|},$$

so

$$R^2 = \frac{\text{ESS}}{\text{TSS}} = \frac{\|\mathbf{v}\|^2}{\|\hat{\mathbf{u}} + \mathbf{v}\|^2} = \cos^2\theta.$$

(If $\mathbf{v} = \mathbf{0}$, then ESS $= 0$ and $R^2 = 0$ directly.) A squared cosine lies in $[0, 1]$. Equivalently, the hypotenuse is at least as long as either leg: $\|\mathbf{v}\|^2 = \|\hat{\mathbf{u}} + \mathbf{v}\|^2 - \|\hat{\mathbf{u}}\|^2 \le \|\hat{\mathbf{u}} + \mathbf{v}\|^2$, and $R^2 = 1 - \|\hat{\mathbf{u}}\|^2/\|\hat{\mathbf{u}} + \mathbf{v}\|^2 = 1 - \text{SSR/TSS}$.

$$\boxed{0 \le R^2 = 1 - \frac{\text{SSR}}{\text{TSS}} \le 1.}$$

Method 3: Deviation form of the fitted line and the squared sample correlation

Use the closed-form OLS estimators

$$\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}}, \qquad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X},$$

where $S_{XX} = \sum_{i=1}^{n}(X_i - \bar{X})^2 > 0$ (the X_i are not all equal), $S_{XY} = \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$ and $S_{YY} = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \text{TSS}$.

Part (a). Substituting $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$,

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

so the residual is

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

Adding the last two displays, $\hat{u}_i + (\hat{Y}_i - \bar{Y}) = Y_i - \bar{Y}$; squaring and summing,

$$\boxed{\text{TSS} = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}\left[\hat{u}_i + (\hat{Y}_i - \bar{Y})\right]^2.}$$

Part (b). From the deviation form,

$$\text{ESS} = \sum_{i=1}^{n} \hat{\beta}_1^2 (X_i - \bar{X})^2 = \hat{\beta}_1^2 S_{XX} = \frac{S_{XY}^2}{S_{XX}^2} S_{XX} = \frac{S_{XY}^2}{S_{XX}}.$$

Expanding the squared residual and summing,

$$\begin{aligned}
\text{SSR} &= \sum_{i=1}^{n}\left[(Y_i - \bar{Y}) - \hat{\beta}_1 (X_i - \bar{X})\right]^2 \\
&= \sum_{i=1}^{n}(Y_i - \bar{Y})^2 - 2\hat{\beta}_1 \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) + \hat{\beta}_1^2 \sum_{i=1}^{n}(X_i - \bar{X})^2 \\
&= S_{YY} - 2\hat{\beta}_1 S_{XY} + \hat{\beta}_1^2 S_{XX} = S_{YY} - 2\frac{S_{XY}^2}{S_{XX}} + \frac{S_{XY}^2}{S_{XX}} = S_{YY} - \frac{S_{XY}^2}{S_{XX}} = \text{TSS} - \text{ESS}.
\end{aligned}$$

$$\boxed{\text{TSS} = \text{SSR} + \text{ESS} = \text{ESS} + \text{SSR}.}$$

Part (c). Dividing the two expressions,

$$R^2 = \frac{\text{ESS}}{\text{TSS}} = \frac{S_{XY}^2}{S_{XX} S_{YY}} = \left(\frac{S_{XY}}{\sqrt{S_{XX} S_{YY}}}\right)^2 = r_{XY}^2,$$

the square of the sample correlation between X and Y, so $R^2 \geq 0$. For the upper bound, every real t gives a sum of squares that is non-negative:

$$0 \leq \sum_{i=1}^{n}\left[(Y_i - \bar{Y}) - t(X_i - \bar{X})\right]^2 = S_{YY} - 2tS_{XY} + t^2 S_{XX}.$$

Choosing $t = S_{XY}/S_{XX}$,

$$0 \le S_{YY} - 2\frac{S_{XY}^2}{S_{XX}} + \frac{S_{XY}^2}{S_{XX}} = S_{YY} - \frac{S_{XY}^2}{S_{XX}} \implies S_{XY}^2 \le S_{XX}S_{YY} \implies r_{XY}^2 \le 1,$$

which is the Cauchy–Schwarz inequality. Finally, from Part (b), $1 - \text{SSR/TSS} = 1 - (S_{YY} - S_{XY}^2/S_{XX})/S_{YY} = S_{XY}^2/(S_{XX}S_{YY}) = R^2$.

$$\boxed{0 \le R^2 = 1 - \frac{\text{SSR}}{\text{TSS}} \le 1.}$$

Final answer

$$\boxed{\text{TSS} = \sum_{i=1}^{n}[\hat{u}_i + (\hat{Y}_i - \bar{Y})]^2 = \text{ESS} + \text{SSR}, \qquad 0 \le R^2 = \frac{\text{ESS}}{\text{TSS}} = 1 - \frac{\text{SSR}}{\text{TSS}} \le 1.}$$

Unofficial QRS@NTU question bank

Question 3 (Comparing weekly working hours: SG versus HK)

Problem

Suppose that the population mean and variance of employees' working hours per week in Singapore (SG) are μ_1 and σ_1^2, respectively, and N_1 i.i.d. sample observations were drawn from this population; the population mean and variance of employees' working hours per week in Hong Kong (HK) are μ_2 and σ_2^2, respectively, and N_2 i.i.d. sample observations were drawn from this population.

The SG and HK random samples are totally independent. We are interested in $\mu_1 - \mu_2$, the difference between the two population means, and use $\hat{d} = \bar{Y}_1 - \bar{Y}_2$ as its estimator, where $\bar{Y}_1$ is the sample average of the N_1 SG observations and $\bar{Y}_2$ is the sample average of the N_2 HK observations.

**The formulas of means, variances, and covariances of sums of random variables can be found in Key Concept 2.3 (p.74) of "Introduction to Econometrics" or in the Supplement "Expectation, Variance, and Covariance".*

(a) Compute $E[\bar{Y}_1]$, $E[\bar{Y}_2]$, $\text{Var}[\bar{Y}_1]$, and $\text{Var}[\bar{Y}_2]$.

(b) Compute $E[\hat{d}]$, $\text{Cov}[\bar{Y}_1, \bar{Y}_2]$, and $\text{Var}[\hat{d}]$.

(Tutorial Question Set No. 5, Q3.)

Solution

Write the SG observations as $Y_{1,1}, \ldots, Y_{1,N_1}$, i.i.d. with $E[Y_{1,i}] = \mu_1$ and $\text{Var}(Y_{1,i}) = \sigma_1^2$, and the HK observations as $Y_{2,1}, \ldots, Y_{2,N_2}$, i.i.d. with $E[Y_{2,j}] = \mu_2$ and $\text{Var}(Y_{2,j}) = \sigma_2^2$; every SG observation is independent of every HK observation. The population quantities μ_1, μ_2 are in hours per week and σ_1^2, σ_2^2 in hours^2; the sample averages $\bar{Y}_1 = \frac{1}{N_1}\sum_{i=1}^{N_1} Y_{1,i}$, $\bar{Y}_2 = \frac{1}{N_2}\sum_{j=1}^{N_2} Y_{2,j}$ and $\hat{d} = \bar{Y}_1 - \bar{Y}_2$ are random variables measured in hours.

Method 1: Direct expectation and variance algebra

Part (a). Take a country $g \in \{1, 2\}$. By linearity of the expectation,

$$E[\bar{Y}_g] = E\Big[\frac{1}{N_g}\sum_{i=1}^{N_g} Y_{g,i}\Big] = \frac{1}{N_g}\sum_{i=1}^{N_g} E[Y_{g,i}] = \frac{1}{N_g} \cdot N_g\mu_g = \mu_g.$$

For the variance, pull out the constant $1/N_g$ as $1/N_g^2$ and expand the variance of the sum:

$$\text{Var}(\bar{Y}_g) = \frac{1}{N_g^2}\,\text{Var}\Big(\sum_{i=1}^{N_g} Y_{g,i}\Big) = \frac{1}{N_g^2}\Big[\sum_{i=1}^{N_g}\text{Var}(Y_{g,i}) + 2\sum_{i<k}\text{Cov}(Y_{g,i}, Y_{g,k})\Big].$$

Within a sample the observations are independent, so every $\text{Cov}(Y_{g,i}, Y_{g,k})$ with $i \neq k$ is zero, and

$$\text{Var}(\bar{Y}_g) = \frac{1}{N_g^2}\big[N_g\sigma_g^2 + 0\big] = \frac{\sigma_g^2}{N_g}.$$

Setting $g = 1$ (SG) and $g = 2$ (HK):

$$\boxed{E[\bar{Y}_1] = \mu_1, \quad E[\bar{Y}_2] = \mu_2, \quad \mathrm{Var}[\bar{Y}_1] = \frac{\sigma_1^2}{N_1}, \quad \mathrm{Var}[\bar{Y}_2] = \frac{\sigma_2^2}{N_2}.}$$

Part (b). By linearity and Part (a),

$$E[\hat{d}] = E[\bar{Y}_1 - \bar{Y}_2] = E[\bar{Y}_1] - E[\bar{Y}_2] = \mu_1 - \mu_2,$$

so $\hat{d}$ is an unbiased estimator of the population gap $\mu_1 - \mu_2$. The covariance is bilinear, so

$$\mathrm{Cov}[\bar{Y}_1, \bar{Y}_2] = \mathrm{Cov}\left(\frac{1}{N_1}\sum_{i=1}^{N_1} Y_{1,i},\ \frac{1}{N_2}\sum_{j=1}^{N_2} Y_{2,j}\right) = \frac{1}{N_1 N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} \mathrm{Cov}(Y_{1,i}, Y_{2,j}).$$

Each $Y_{1,i}$ is independent of each $Y_{2,j}$, so $E[Y_{1,i}Y_{2,j}] = E[Y_{1,i}]E[Y_{2,j}] = \mu_1\mu_2$ and

$$\mathrm{Cov}(Y_{1,i}, Y_{2,j}) = E[Y_{1,i}Y_{2,j}] - E[Y_{1,i}]E[Y_{2,j}] = \mu_1\mu_2 - \mu_1\mu_2 = 0.$$

Every term of the double sum is zero.

$$\boxed{\mathrm{Cov}[\bar{Y}_1, \bar{Y}_2] = 0.}$$

For the variance of $\hat{d} = 1 \cdot \bar{Y}_1 + (-1) \cdot \bar{Y}_2$, use $\mathrm{Var}(aU + bV) = a^2\,\mathrm{Var}(U) + b^2\,\mathrm{Var}(V) + 2ab\,\mathrm{Cov}(U, V)$ with $a = 1$, $b = -1$:

$$\mathrm{Var}[\hat{d}] = 1^2\,\mathrm{Var}[\bar{Y}_1] + (-1)^2\,\mathrm{Var}[\bar{Y}_2] + 2(1)(-1)\,\mathrm{Cov}[\bar{Y}_1, \bar{Y}_2] = \frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2} - 2 \cdot 0.$$

The two sampling errors are uncorrelated, so neither cancels the other and their variances add; $\mathrm{Var}[\hat{d}]$ falls when either sample grows.

$$\boxed{\mathrm{Var}[\hat{d}] = \frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}.}$$

Standard error and 95% CI. Replace σ_1^2, σ_2^2 by the sample variances s_1^2, s_2^2 to get $\mathrm{SE}(\hat{d}) = \sqrt{s_1^2/N_1 + s_2^2/N_2}$. For large N_1, N_2 the central limit theorem gives $\big(\hat{d} - (\mu_1 - \mu_2)\big)/\,\mathrm{SE}(\hat{d}) \approx N(0,1)$, and $P(-1.96 \le Z \le 1.96) = 0.95$ for $Z \sim N(0,1)$. Hence

$$\begin{aligned} 0.95 &\approx P\Big(-1.96 \le \frac{\hat{d} - (\mu_1 - \mu_2)}{\mathrm{SE}(\hat{d})} \le 1.96\Big) \\ &= P\Big(-1.96\ \mathrm{SE}(\hat{d}) \le \hat{d} - (\mu_1 - \mu_2) \le 1.96\ \mathrm{SE}(\hat{d})\Big) \\ &= P\Big(\hat{d} - 1.96\ \mathrm{SE}(\hat{d}) \le \mu_1 - \mu_2 \le \hat{d} + 1.96\ \mathrm{SE}(\hat{d})\Big). \end{aligned}$$

For example, with the hypothetical data $N_1 = 400$, $s_1 = 10$ hours, $\bar{Y}_1 = 45.0$ hours and $N_2 = 225$, $s_2 = 9$ hours, $\bar{Y}_2 = 42.5$ hours: $\hat{d} = 2.5$ hours, $\mathrm{SE}(\hat{d}) = \sqrt{100/400 + 81/225} = \sqrt{0.25 + 0.36} = \sqrt{0.61} = 0.781$ hours, and the interval is $2.5 \pm 1.96(0.781) = 2.5 \pm 1.531 = [0.97,\ 4.03]$ hours.

$$\boxed{\mathrm{CI}_{95\%}(\mu_1 - \mu_2) = (\bar{Y}_1 - \bar{Y}_2) \pm 1.96\ \mathrm{SE}(\hat{d}) = (\bar{Y}_1 - \bar{Y}_2) \pm 1.96\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}.}$$

Method 2: Pooled weights representation

Stack all $N = N_1 + N_2$ observations into one list $W_1, \ldots, W_N$: $W_k = Y_{1,k}$ for $k \le N_1$ (SG) and $W_k = Y_{2,k-N_1}$ for $k > N_1$ (HK). The N variables are mutually independent (i.i.d. within each sample, independent across samples), with $E[W_k] = \mu_1$, $\mathrm{Var}(W_k) = \sigma_1^2$ for SG and $E[W_k] = \mu_2$, $\mathrm{Var}(W_k) = \sigma_2^2$ for HK.

Rules for linear combinations. For constants a_k, b_k, let $L = \sum_{k=1}^N a_k W_k$ and $M = \sum_{k=1}^N b_k W_k$. Linearity gives $E[L] = \sum_k a_k E[W_k]$. Bilinearity gives

$$\mathrm{Cov}(L, M) = \sum_{k=1}^{N} \sum_{l=1}^{N} a_k b_l \,\mathrm{Cov}(W_k, W_l),$$

and independence makes every term with $k \neq l$ zero, so

$$\mathrm{Cov}(L, M) = \sum_{k=1}^{N} a_k b_k \,\mathrm{Var}(W_k), \qquad \mathrm{Var}(L) = \mathrm{Cov}(L, L) = \sum_{k=1}^{N} a_k^2 \,\mathrm{Var}(W_k).$$

The weights. The three estimators are linear combinations of the pooled list:

$$\begin{aligned} \bar{Y}_1: &\quad a_k = \tfrac{1}{N_1} \text{ (SG)}, \qquad a_k = 0 \text{ (HK)}, \\ \bar{Y}_2: &\quad b_k = 0 \text{ (SG)}, \qquad b_k = \tfrac{1}{N_2} \text{ (HK)}, \\ \hat{d} = \bar{Y}_1 - \bar{Y}_2: &\quad c_k = a_k - b_k = \tfrac{1}{N_1} \text{ (SG)}, \qquad c_k = -\tfrac{1}{N_2} \text{ (HK)}. \end{aligned}$$

Part (a). Summing over the N_1 SG positions and the N_2 HK positions separately,

$$E[\bar{Y}_1] = N_1 \cdot \frac{1}{N_1}\mu_1 + N_2 \cdot 0 \cdot \mu_2 = \mu_1, \qquad \mathrm{Var}[\bar{Y}_1] = N_1 \cdot \frac{1}{N_1^2}\sigma_1^2 + N_2 \cdot 0^2 \cdot \sigma_2^2 = \frac{\sigma_1^2}{N_1},$$
$$E[\bar{Y}_2] = N_1 \cdot 0 \cdot \mu_1 + N_2 \cdot \frac{1}{N_2}\mu_2 = \mu_2, \qquad \mathrm{Var}[\bar{Y}_2] = N_1 \cdot 0^2 \cdot \sigma_1^2 + N_2 \cdot \frac{1}{N_2^2}\sigma_2^2 = \frac{\sigma_2^2}{N_2}.$$

The SG weights of $\bar{Y}_1$ sum to $N_1 \cdot \frac{1}{N_1} = 1$, which is the unbiasedness condition of Q1(c) applied to the SG sample.

$$\boxed{E[\bar{Y}_1] = \mu_1, \quad E[\bar{Y}_2] = \mu_2, \quad \mathrm{Var}[\bar{Y}_1] = \frac{\sigma_1^2}{N_1}, \quad \mathrm{Var}[\bar{Y}_2] = \frac{\sigma_2^2}{N_2}.}$$

Part (b). For the mean,

$$E[\hat{d}] = N_1 \cdot \frac{1}{N_1}\mu_1 + N_2 \cdot \Big(-\frac{1}{N_2}\Big)\mu_2 = \mu_1 - \mu_2.$$

For the covariance, the product of weights is zero at every position, because $\bar{Y}_1$ and $\bar{Y}_2$ put weight on disjoint sets of observations:

$$a_k b_k = \frac{1}{N_1} \cdot 0 = 0 \;\; \text{(SG)}, \qquad a_k b_k = 0 \cdot \frac{1}{N_2} = 0 \;\; \text{(HK)},$$

so

$$\mathrm{Cov}[\bar{Y}_1, \bar{Y}_2] = \sum_{k=1}^{N} a_k b_k \,\mathrm{Var}(W_k) = 0.$$

$$\boxed{\text{Cov}[\bar{Y}_1, \bar{Y}_2] = 0.}$$

For the variance,

$$\text{Var}[\hat{d}] = \sum_{k=1}^{N} c_k^2 \,\text{Var}(W_k) = N_1 \cdot \Big(\frac{1}{N_1}\Big)^2 \sigma_1^2 + N_2 \cdot \Big(-\frac{1}{N_2}\Big)^2 \sigma_2^2 = \frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}.$$

$$\boxed{\text{Var}[\hat{d}] = \frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}.}$$

Unofficial QRS@NTU - qrsntu.org

Method 3: Binary-regressor OLS on the pooled sample

Use the pooled list $W_1, \ldots, W_N$ of Method 2 and the binary regressor $D_k = 1$ for an SG observation, $D_k = 0$ for an HK observation. The sample sizes are fixed by design, so the D_k are non-random. Define

$$\beta_0 = \mu_2, \qquad \beta_1 = \mu_1 - \mu_2, \qquad u_k = W_k - \beta_0 - \beta_1 D_k,$$

so that $W_k = \beta_0 + \beta_1 D_k + u_k$ with $E[u_k] = 0$ in both groups ($E[W_k] = \mu_2 + (\mu_1 - \mu_2) = \mu_1$ when $D_k = 1$, $E[W_k] = \mu_2$ when $D_k = 0$), $\mathrm{Var}(u_k) = \sigma_1^2$ for SG and σ_2^2 for HK, and the u_k independent across k.

OLS equals the group means. Let $\bar{D} = \frac{1}{N}\sum_k D_k = \frac{N_1}{N}$ and $\bar{W} = \frac{1}{N}\sum_k W_k = \frac{N_1\bar{Y}_1 + N_2\bar{Y}_2}{N}$. The deviations of the regressor are

$$D_k - \bar{D} = 1 - \frac{N_1}{N} = \frac{N_2}{N} \ \text{(SG)}, \qquad D_k - \bar{D} = 0 - \frac{N_1}{N} = -\frac{N_1}{N} \ \text{(HK)},$$

and they sum to zero: $\sum_k (D_k - \bar{D}) = N_1 \frac{N_2}{N} - N_2 \frac{N_1}{N} = 0$. Hence

$$S_{DD} = \sum_k (D_k - \bar{D})^2 = N_1 \frac{N_2^2}{N^2} + N_2 \frac{N_1^2}{N^2} = \frac{N_1 N_2 (N_2 + N_1)}{N^2} = \frac{N_1 N_2}{N}$$

and

$$\begin{aligned} \sum_k (D_k - \bar{D})(W_k - \bar{W}) &= \sum_k (D_k - \bar{D}) W_k - \bar{W} \sum_k (D_k - \bar{D}) \\ &= \frac{N_2}{N} \sum_{\text{SG}} W_k - \frac{N_1}{N} \sum_{\text{HK}} W_k - 0 \\ &= \frac{N_2}{N} N_1 \bar{Y}_1 - \frac{N_1}{N} N_2 \bar{Y}_2 = \frac{N_1 N_2}{N} \big(\bar{Y}_1 - \bar{Y}_2\big). \end{aligned}$$

Therefore

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum_k (D_k - \bar{D})(W_k - \bar{W})}{S_{DD}} = \bar{Y}_1 - \bar{Y}_2 = \hat{d}, \\ \hat{\beta}_0 &= \bar{W} - \hat{\beta}_1 \bar{D} = \frac{N_1\bar{Y}_1 + N_2\bar{Y}_2 - N_1\bar{Y}_1 + N_1\bar{Y}_2}{N} = \frac{(N_2 + N_1)\bar{Y}_2}{N} = \bar{Y}_2, \end{aligned}$$

so $\bar{Y}_2 = \hat{\beta}_0$ and $\bar{Y}_1 = \hat{\beta}_0 + \hat{\beta}_1$.

Weights and sampling errors. In weight form, $\hat{\beta}_1 = \sum_k w_k W_k$ and $\hat{\beta}_0 = \sum_k v_k W_k$ with $w_k = (D_k - \bar{D})/S_{DD}$ and $v_k = \frac{1}{N} - \bar{D} w_k$:

$$\begin{aligned} \text{SG:} \quad w_k &= \frac{N_2/N}{N_1 N_2/N} = \frac{1}{N_1}, & v_k &= \frac{1}{N} - \frac{N_1}{N} \cdot \frac{1}{N_1} = 0, \\ \text{HK:} \quad w_k &= \frac{-N_1/N}{N_1 N_2/N} = -\frac{1}{N_2}, & v_k &= \frac{1}{N} + \frac{N_1}{N} \cdot \frac{1}{N_2} = \frac{N_2 + N_1}{N N_2} = \frac{1}{N_2}. \end{aligned}$$

These satisfy $\sum_k w_k = N_1\frac{1}{N_1} - N_2\frac{1}{N_2} = 0$, $\sum_k w_k D_k = N_1\frac{1}{N_1} = 1$, $\sum_k v_k = N_2\frac{1}{N_2} = 1$ and $\sum_k v_k D_k = 0$. Substituting $W_k = \beta_0 + \beta_1 D_k + u_k$,

$$\hat{\beta}_1 = \beta_0 \sum_k w_k + \beta_1 \sum_k w_k D_k + \sum_k w_k u_k = \beta_1 + \sum_k w_k u_k,$$
$$\hat{\beta}_0 = \beta_0 \sum_k v_k + \beta_1 \sum_k v_k D_k + \sum_k v_k u_k = \beta_0 + \sum_k v_k u_k,$$
$$\hat{\beta}_0 + \hat{\beta}_1 = \beta_0 + \beta_1 + \sum_k (v_k + w_k) u_k, \qquad v_k + w_k = \frac{1}{N_1} \text{ (SG)}, \quad 0 \text{ (HK)}.$$

Part (a). With $E[u_k] = 0$ and independent u_k,

$$E[\bar{Y}_1] = E[\hat{\beta}_0 + \hat{\beta}_1] = \beta_0 + \beta_1 = \mu_2 + (\mu_1 - \mu_2) = \mu_1,$$
$$E[\bar{Y}_2] = E[\hat{\beta}_0] = \beta_0 = \mu_2,$$
$$\mathrm{Var}[\bar{Y}_1] = \sum_k (v_k + w_k)^2 \,\mathrm{Var}(u_k) = N_1 \cdot \frac{1}{N_1^2}\sigma_1^2 + N_2 \cdot 0^2 \cdot \sigma_2^2 = \frac{\sigma_1^2}{N_1},$$
$$\mathrm{Var}[\bar{Y}_2] = \sum_k v_k^2 \,\mathrm{Var}(u_k) = N_1 \cdot 0^2 \cdot \sigma_1^2 + N_2 \cdot \frac{1}{N_2^2}\sigma_2^2 = \frac{\sigma_2^2}{N_2}.$$

$$\boxed{E[\bar{Y}_1] = \mu_1, \quad E[\bar{Y}_2] = \mu_2, \quad \mathrm{Var}[\bar{Y}_1] = \frac{\sigma_1^2}{N_1}, \quad \mathrm{Var}[\bar{Y}_2] = \frac{\sigma_2^2}{N_2}.}$$

Part (b). The slope is unbiased for the population gap:

$$E[\hat{d}] = E[\hat{\beta}_1] = \beta_1 + \sum_k w_k E[u_k] = \beta_1 = \mu_1 - \mu_2.$$

For the covariance, only $k = l$ terms survive, and $(v_k + w_k)v_k = \frac{1}{N_1} \cdot 0 = 0$ (SG), $0 \cdot \frac{1}{N_2} = 0$ (HK):

$$\mathrm{Cov}[\bar{Y}_1, \bar{Y}_2] = \mathrm{Cov}\Big(\sum_k (v_k + w_k)u_k, \ \sum_l v_l u_l\Big) = \sum_k (v_k + w_k) v_k \,\mathrm{Var}(u_k) = 0.$$

$$\boxed{\mathrm{Cov}[\bar{Y}_1, \bar{Y}_2] = 0.}$$

For the variance of the slope,

$$\mathrm{Var}[\hat{d}] = \mathrm{Var}(\hat{\beta}_1) = \sum_k w_k^2 \,\mathrm{Var}(u_k) = N_1\Big(\frac{1}{N_1}\Big)^2\sigma_1^2 + N_2\Big(-\frac{1}{N_2}\Big)^2\sigma_2^2$$
$$= \frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}.$$

The error variance differs between the two groups unless $\sigma_1^2 = \sigma_2^2$, so this regression is heteroskedastic; the homoskedasticity-only formula $\sigma_u^2/S_{DD} = \sigma_u^2\big(\frac{1}{N_1} + \frac{1}{N_2}\big)$ agrees with the line above only when $\sigma_1^2 = \sigma_2^2 = \sigma_u^2$.

$$\boxed{\mathrm{Var}[\hat{d}] = \frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}.}$$

Final answer

$$\boxed{E[\bar{Y}_1] = \mu_1, \quad E[\bar{Y}_2] = \mu_2, \quad \text{Var}[\bar{Y}_1] = \frac{\sigma_1^2}{N_1}, \quad \text{Var}[\bar{Y}_2] = \frac{\sigma_2^2}{N_2}.}$$

$$\boxed{E[\hat{d}] = \mu_1 - \mu_2, \quad \text{Cov}[\bar{Y}_1, \bar{Y}_2] = 0, \quad \text{Var}[\hat{d}] = \frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}.}$$

Unofficial QRS@NTU - qrsntu.org

Final Answers

Q1(a) $\mathrm{Var}(\bar{Y}) = \sigma^2/n < \mathrm{Var}(\tilde{Y}) = \sigma^2/2 < \mathrm{Var}(Y_1) = \sigma^2$ for $n > 2$; all unbiased.

Q1(b) LSA1–3 + homoskedasticity $\Rightarrow \hat{\beta}_0, \hat{\beta}_1$ BLUE (smallest variance among linear unbiased).

Q1(c) $\sum_{i=1}^n a_i Y_i$ unbiased $\iff \sum_{i=1}^n a_i = 1$; minimum variance at $a_i = 1/n$, so $\bar{Y}$ is BLUE.

Q1(d) $\mathrm{Var}(Y_1) = 1$, $\mathrm{Var}(\bar{Y}) = 0.1$; both unbiased, $\bar{Y}$ far more efficient.

Q2 $\mathrm{TSS} = \sum[\hat{u}_i + (\hat{Y}_i - \bar{Y})]^2 = \mathrm{ESS} + \mathrm{SSR}$; $0 \le R^2 = \mathrm{ESS}/\mathrm{TSS} = 1 - \mathrm{SSR}/\mathrm{TSS} \le 1$.

Q3 $E[\bar{Y}_1] = \mu_1$, $E[\bar{Y}_2] = \mu_2$, $\mathrm{Var}[\bar{Y}_1] = \sigma_1^2/N_1$, $\mathrm{Var}[\bar{Y}_2] = \sigma_2^2/N_2$; $E[\hat{d}] = \mu_1 - \mu_2$, $\mathrm{Cov}[\bar{Y}_1, \bar{Y}_2] = 0$, $\mathrm{Var}[\hat{d}] = \sigma_1^2/N_1 + \sigma_2^2/N_2$.

Unofficial QRS@NTU – qrsntu.org