ECO227Y5 Tutorial 18

William Hsu

Department of Economics
University of Toronto Mississauga

March 2nd, 2026

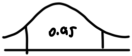
Question: Do SAT scores for high school students differ depending on the students' intended field of study? Fifteen students who intended to major in engineering were compared with 15 students who intended to major in language and literature. Given in the accompanying table are the means and standard deviations of the scores on the verbal and mathematics portion of the SAT for the two groups of students:

		Verbal	Math
Engineering	1, n	$\bar{y} = 446,\ s = 42$	$\bar{y} = 548,\ s = 57$
Language/literature	2, m	$\bar{y} = 534,\ s = 45$	$\bar{y} = 517,\ s = 52$

(a) Construct a 95% confidence interval for the difference in average verbal scores of students majoring in engineering and of those majoring in language/literature.

let $\bar{x}$ be $\bar{Y}_1 - \bar{Y}_2$ $n = m = 15$, t statistic, t_{n+m-2}

- How spread out your estimate would be



$\bar{x} = 446 - 534 = -88$ $S.E = \sqrt{\frac{42^2}{15} + \frac{45^2}{15}} = 15.89$, t_{28}

$P\left(-a \le \frac{-88 - (\mu_1 - \mu_2)}{15.89} \le a\right) = 0.95 \ni a = 2.048$

95% C.I for $\mu_1 - \mu_2$: $[-88 - 2.048 \cdot 15.89,\ -88 + 2.048 \cdot 15.89] = [-120.54,\ -55.46]$

Question 8.90

Question: Do SAT scores for high school students differ depending on the students' intended field of study? Fifteen students who intended to major in engineering were compared with 15 students who intended to major in language and literature. Given in the accompanying table are the means and standard deviations of the scores on the verbal and mathematics portion of the SAT for the two groups of students:

	Verbal	Math
Engineering	$\bar{y} = 446,\ s = 42$	$\bar{y}_1 = 548,\ s = 57$
Language/literature	$\bar{y} = 534,\ s = 45$	$\bar{y}_2 = 517,\ s = 52$

(b) Construct a 95% confidence interval for the difference in average math scores of students majoring in engineering and of those majoring in language/literature. t_{n+m-2}, t_{28} again

Let $\bar{x} = \bar{y}_1 - \bar{y}_2$. $\bar{x} = 548 - 517 = 31$ $SE = \sqrt{\frac{57^2}{15} + \frac{52^2}{15}} = 19.92$

$P\left(-a \le \frac{31 - (\mu_1 - \mu_2)}{19.92} \le a\right) \Rightarrow a = 2.048$

95% C.I for $\mu_1 - \mu_2$: $[\,31 - 2.048 \cdot 19.92,\ 31 + 2.048 \cdot 19.92\,] = [-9.79, 71.80]$

UNIVERSITY OF TORONTO

Question 8.90

Question: Do SAT scores for high school students differ depending on the students' intended field of study? Fifteen students who intended to major in engineering were compared with 15 students who intended to major in language and literature. Given in the accompanying table are the means and standard deviations of the scores on the verbal and mathematics portion of the SAT for the two groups of students:

	Verbal	Math
Engineering	$\bar{y} = 446,\ s = 42$	$\bar{y} = 548,\ s = 57$
Language/literature	$\bar{y} = 534,\ s = 45$	$\bar{y} = 517,\ s = 52$

(c) Interpret the results obtained in parts (a) and (b).

(d) What assumptions are necessary for the methods used previously to be valid?

c.) Since the 95% C.I for verbal is all negative, Language/literature students are likely to score higher.

For math, the interval contains 0, so there is a chance that there is no difference between scores.

d.) Independent samples

UNIVERSITY OF TORONTO

Question 8.93

Question: A factory operates with two machines of type A and one machine of type B. The weekly repair costs X for type A machines are normally distributed with mean μ_1 and variance σ^2. The weekly repair costs Y for machines of type B are also normally distributed but with mean μ_2 and variance $3\sigma^2$. The expected repair cost per week for the factory is thus $2\mu_1 + \mu_2$. If you are given a random sample $X_1, X_2, \ldots, X_n$ on costs of type A machines and an independent random sample $Y_1, Y_2, \ldots, Y_m$ on costs for type B machines, show how you would construct a 95% confidence interval for $2\mu_1 + \mu_2$

$\bar{X} \sim N(\mu_1, \frac{\sigma^2}{n})$ $\quad$ $\bar{Y} \sim N(\mu_2, \frac{3\sigma^2}{m})$

(a) if σ^2 is known.

(b) if σ^2 is not known.

a.) want C.I. for $2\mu_1+\mu_2$

$V(2\bar{X}+\bar{Y}) = 4\frac{\sigma^2}{n} + \frac{3\sigma^2}{m} \Rightarrow S.E. = \sigma\sqrt{\frac{4}{n}+\frac{3}{m}}$, can use Z

$P\left(-a \leq \frac{(2\bar{X}+\bar{Y})-(2\mu_1+\mu_2)}{\sigma\sqrt{\frac{4}{n}+\frac{3}{m}}} \leq a\right) = 0.95 \Rightarrow a = 1.96$

so 95% C.I for $2\mu_1+\mu_2$: $\left(2\bar{x}+\bar{y} - 1.96\cdot\sigma\sqrt{\frac{4}{n}+\frac{3}{n}},\ 2\bar{x}+\bar{y}+1.96\,\sigma\sqrt{\frac{4}{m}+\frac{3}{n}}\right]$

b.) If σ^2 not known, we can use pooled estimator instead:

$S_p^2 = \frac{(n-1)S_X^2 + (m-1)\frac{S_Y^2}{3}}{n+m-2}$, $\quad t_{n+m-2}$

$P\left(-a \leq \frac{(2\bar{x}+\bar{y})-(2\mu_1+\mu_2)}{S_p\sqrt{\frac{4}{n}+\frac{3}{m}}} \leq a\right) = 0.95 \Rightarrow a = 2.048$

$\Rightarrow$ 95% C.I for $2\mu_1+\mu_2$: $\left[2\bar{x}+\bar{y} - 2.048\cdot S_p\sqrt{\frac{4}{n}+\frac{3}{m}},\ 2\bar{x}+\bar{y}+2.048\cdot S_p\sqrt{\frac{4}{n}+\frac{3}{m}}\right]$

UNIVERSITY OF TORONTO

Question 8.97

Question: Suppose that S^2 is the sample variance based on a sample of size n from a normal population with unknown mean and variance. Derive a $100(1-\alpha)\%$

(a) upper confidence bound for σ^2.

(b) lower confidence bound for σ^2.

Recall: $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

a.) $P\left(\frac{(n-1)S^2}{\sigma^2} \geq \chi^2_{\alpha}\right) = 1-\alpha$

$\Rightarrow P\left(\frac{1}{\sigma^2} \geq \frac{\chi^2_{\alpha}}{(n-1)S^2}\right) = 1-\alpha$

$\Rightarrow P\left(\sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{\alpha}}\right) = 1-\alpha$ so $100(1-\alpha)\%$ upper C.I: $\left(0, \frac{(n-1)S^2}{\chi^2_{\alpha}}\right]$

b.) $P\left(\frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{1-\alpha}\right) = 1-\alpha$

$\Rightarrow P\left(\frac{1}{\sigma^2} \leq \frac{\chi^2_{1-\alpha}}{(n-1)S^2}\right) = 1-\alpha \Rightarrow P\left(\sigma^2 \geq \frac{(n-1)S^2}{\chi^2_{1-\alpha}}\right)$

$\Rightarrow 100(1-\alpha)\%$ lower bound C.I: $\left[\frac{(n-1)S^2}{\chi^2_{1-\alpha}}, \infty\right)$

UNIVERSITY OF TORONTO

Question 9.3

Question: Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the uniform distribution on the interval $(\theta, \theta+1)$. Let

$Y_i \overset{iid}{\sim} unif(\theta, \theta+1) \ni \bar{Y} = \theta + \frac{1}{2}$

$f(y_i) = \begin{cases} 1, & y_i \in (\theta, \theta+1) \\ 0, & \text{elsewhere} \end{cases}$

$$\hat{\theta}_1 = \bar{Y} - \frac{1}{2} \qquad \text{and} \qquad \hat{\theta}_2 = Y_{(n)} - \frac{n}{n+1}.$$

(a) Show that both $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased estimators of θ.

(b) Find the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$.

a.) $E(\hat{\theta}_1) = E(\bar{Y}) - \frac{1}{2} = \theta + \frac{1}{2} - \frac{1}{2} = \theta$ so unbiased $\quad F(y) = y - \theta$

$P(Y_{(n)} \le y) = P(Y_1 \le y, \ldots, Y_n \le y) = F(y)^n = (y-\theta)^n \ni f(y_{(n)}) = n(y-\theta)^{n-1}$

$y = \theta + (y - \theta)$

$E(Y_{(n)}) = \int_\theta^{\theta+1} n\, y (y-\theta)^{n-1} dy = n\int_\theta^{\theta+1} y\, (y-\theta)^{n-1} dy = n\int_\theta^{\theta+1} (\theta + y - \theta)(y-\theta)^{n-1} dy$

$= n\theta \int_\theta^{\theta+1} (y-\theta)^{n-1} dy + n\int_\theta^{\theta+1} (y-\theta)^n dy = \theta + \frac{n}{n+1} \ni E(\hat{\theta}_2) = \theta + \frac{n}{n+1} - \frac{n}{n+1} = \theta$

b.) $\dfrac{var(\hat{\theta}_2)}{var(\hat{\theta}_1)} =$

$var(\hat{\theta}_1) = var(\bar{Y})$. $\quad var(Y_i) = \frac{1}{12} \ni var(\bar{Y}) = \frac{1}{12n}$

$var(\hat{\theta}_2) = var(Y_{(n)}) = E(Y_{(n)}^2) - E(Y_{(n)})^2$

continued on next slide

UNIVERSITY OF TORONTO

Question 9.3

Question: Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the uniform distribution on the interval $(\theta, \theta + 1)$. Let

$$\hat{\theta}_1 = \bar{Y} - \frac{1}{2} \qquad \text{and} \qquad \hat{\theta}_2 = Y_{(n)} - \frac{n}{n+1}.$$

(a) Show that both $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased estimators of θ.

(b) Find the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$.

b.) $E(Y_{(n)}^2) = \int_{\theta}^{\theta+1} y^2 n (y-\theta)^{n-1} dy$ $\qquad u = y - \theta \;\therefore\; y = u + \theta$, $\quad du = dy$

$$= n \int_0^1 (\theta^2 + 2\theta u + u^2) u^{n-1} du$$

$$= n \int_0^1 \theta^2 u^{n-1} + 2\theta u^n + u^{n+1} \, du = \theta^2 + \frac{2n\theta}{n+1} + \frac{n}{n+2}$$

$$V(\hat{\theta}_2) = \left(\theta^2 + \frac{2n\theta}{n+1} + \frac{n}{n+2}\right) - \left(\theta + \frac{n}{n+1}\right)^2 = \left(\theta^2 + \frac{2n\theta}{n+1} + \frac{n}{n+2}\right) - \left(\theta^2 + \frac{2\theta n}{n+1} + \frac{n^2}{(n+1)^2}\right)$$

$$V(\hat{\theta}_2) = \frac{n}{n+2} - \frac{n^2}{(n+1)^2} = \frac{n}{(n+1)^2(n+2)} \quad \text{so} \quad \text{Eff}(\hat{\theta}_1 \text{ rel. } \hat{\theta}_2) = \frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1)} = \frac{\frac{n}{(n+1)^2(n+2)}}{\frac{1}{12n}} = \frac{12n^2}{(n+1)^2(n+2)}$$

notice as $n \to \infty$ Eff $\to 0$ so $\hat{\theta}_2$ is more efficient.

Question 9.15

Question: Refer to Exercise 9.3. Show that both $\hat{\theta}_1$ and $\hat{\theta}_2$ are consistent estimators for θ.

Recall: If $\hat{\theta}$ is unbiased, and if $\lim_{n\to\infty} V(\hat{\theta}) = 0$

Then $\hat{\theta} \xrightarrow{p} \theta$ (consistent)

faster than definition method

$\hat{\theta}_1 = \bar{Y} - \frac{1}{2}$, $V(\hat{\theta}_1) = \frac{1}{12n}$ $\lim_{n\to\infty} \frac{1}{12n} = 0$

degree 2

$\hat{\theta}_2 = Y_{(n)} - \frac{n}{n+1}$, $V(\hat{\theta}_2) = \frac{12n^2}{(n+1)^2(n+2)}$, $\lim_{n\to\infty} \frac{12n^2}{(n+1)^2(n+2)} = 0$

degree 3

Question 9.19

Question: Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the probability density function

$$f(y) = \begin{cases} \theta y^{\theta-1}, & 0 < y < 1, \\ 0, & \text{elsewhere,} \end{cases}$$

where $\theta > 0$. Show that $\bar{Y}$ is a consistent estimator of $\dfrac{\theta}{\theta+1}$.

$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n} \cdot n\left(\frac{\theta}{\theta+1}\right) = \frac{\theta}{\theta+1}$ unbiased

$E(Y_i^2) = \int_0^1 \theta\, y^{\theta+1}\, dy = \frac{\theta}{\theta+2}$

$E(Y_i) = \int_0^1 \theta y^{\theta}\, dy = \frac{\theta y^{\theta+1}}{\theta+1}\Big|_0^1 = \frac{\theta}{\theta+1}$

$V(\bar{Y}) = V\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n^2}\, n \cdot V(Y_i) = \frac{1}{n}\left(\frac{\theta}{\theta+2} - \frac{\theta^2}{(\theta+1)^2}\right)$ clearly as $n \to \infty$, $V(\bar{Y}) \to 0$

so consistent

$V(Y_i) = E(Y_i^2) - E(Y_i)^2 = \frac{\theta}{\theta+2} - \frac{\theta^2}{(\theta+1)^2}$

UNIVERSITY OF TORONTO

Question 9.20

Question: If Y has a binomial distribution with n trials and success probability p, show that Y/n is a consistent estimator of p.

$Y \sim Bin(n, p)$

$E(\frac{Y}{n}) = \frac{E(Y)}{n} = \frac{np}{n} = p$ unbiased

$V(\frac{Y}{n}) = \frac{V(Y)}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n} \to 0$ as $n \to \infty$ so consistent

Question: Refer to Exercise 9.21. Suppose that $Y_1, Y_2, \ldots, Y_n$ is a random sample of size n from a population for which the first four moments are finite. That is, $m_1 = E(Y_1) < \infty$, $m_2 = E(Y_1^2) < \infty$, $m_3 = E(Y_1^3) < \infty$, and $m_4 = E(Y_1^4) < \infty$. (Note: This assumption is valid for the normal and Poisson distributions in Exercises 9.21 and 9.22, respectively.) Again, assume that $n = 2k$ for some integer k. Consider

n is even

$$\hat{\sigma}^2 = \frac{1}{2k}\sum_{i=1}^{k}(Y_{2i} - Y_{2i-1})^2 .$$

(a) Show that $\hat{\sigma}^2$ is an unbiased estimator for σ^2.

$$E(\hat{\sigma}^2) = \frac{1}{2k}\sum_{i=1}^{k} E[(Y_{2i} - Y_{2i-1})^2] = \frac{1}{2k}\sum_{i=1}^{k} E(Y_{2i}^2 - 2Y_{2i}Y_{2i-1} + Y_{2i-1}^2)$$

– independent

Recall: $E(X^2) = V(X) + E(X)^2$

But $Y_j \overset{iid}{\sim}$, so $E(Y_{2i}) = E(Y_{2i-1})$ etc

$$= \frac{1}{2k}\sum_{i=1}^{k} E(Y_{2i}^2) - 2E(Y_{2i})E(Y_{2i-1}) + E(Y_{2i-1}^2)$$

$$= \frac{1}{2k}\sum_{i=1}^{k} V(Y_{2i}) + E(Y_{2i})^2 - 2E(Y_{2i})E(Y_{2i-1}) + V(Y_{2i-1}) + E(Y_{2i-1})^2$$

$$= \frac{1}{2k}\sum_{i=1}^{k} \sigma^2 + \mu^2 - 2\mu^2 + \sigma^2 + \mu^2 = \frac{1}{2k}\sum_{i=1}^{k} 2\sigma^2$$

$$= \frac{1}{k}\sum_{i=1}^{k}\sigma^2 = \frac{k\sigma^2}{k} = \sigma^2$$

UNIVERSITY OF TORONTO

Question 9.23

Question: Refer to Exercise 9.21. Suppose that $Y_1, Y_2, \ldots, Y_n$ is a random sample of size n from a population for which the first four moments are finite. That is, $m_1 = E(Y_1) < \infty$, $m_2 = E(Y_1^2) < \infty$, $m_3 = E(Y_1^3) < \infty$, and $m_4 = E(Y_1^4) < \infty$. (Note: This assumption is valid for the normal and Poisson distributions in Exercises 9.21 and 9.22, respectively.) Again, assume that $n = 2k$ for some integer k. Consider

$$\hat{\sigma}^2 = \frac{1}{2k} \sum_{i=1}^{k} (Y_{2i} - Y_{2i-1})^2 .$$

(b) Show that $\hat{\sigma}^2$ is a consistent estimator for σ^2.

(c) Why did you need the assumption that $m_4 = E(Y_1^4) < \infty$?

b.) Define $\hat{\sigma}^2 = \frac{1}{2k} \sum_{i=1}^{k} J_i$, $J_i = (Y_{2i} - Y_{2i-1})^2$. J_i iid and $E(J_i) = 2\sigma^2$. $var(\hat{\sigma}^2) = var(\frac{1}{2k} \sum_{i=1}^{k} J_i) = \frac{1}{4k^2} k \cdot var(J_i) = \frac{var(J_i)}{4k}$ to if $var(J_i) < \infty$

c.) To make limit of $var(\hat{\sigma}^2)$ as $n \to \infty$ become 0.

Question 9.24

Question: Let $Y_1, Y_2, Y_3, \ldots, Y_n$ be independent standard normal random variables.

(a) What is the distribution of $\sum_{i=1}^{n} Y_i^2$?

(b) Let $W_n = \frac{1}{n}\sum_{i=1}^{n} Y_i^2$. Does W_n converge in probability to some constant? If so, what is the value of the constant?

a.) $Y_i \overset{iid}{\sim} N(0,1)$

$\sum_{i=1}^{n} Y_i^2 \sim \chi_n^2$

χ_1^2

$V(Y_i) = 1 = E(Y_i^2) - \overset{=0}{E(Y_i)^2}$

$\Rightarrow E(Y_i^2) = 1$

b.) $E(W_n) = \frac{1}{n}\sum_{i=1}^{n} E(Y_i^2) = \frac{n \cdot 1}{n} = 1$

$V(W_n) = \frac{1}{n^2}\sum_{i=1}^{n} V(Y_i^2) = \frac{2}{n}$ so yes $W_n \xrightarrow{p} 1$ by law of large numbers

and consistency