# STA260: PROBABILITY AND STATISTICS II
# SPRING 2021

## TUTORIAL 6 (TUT9101)

**MAY 21, 2021** MAY 26, 2021

**YUEHUAN HE**

UNIVERSITY OF TORONTO

# TERM TEST 2 COMING

Date and time: 27th may, Thursday 8am – 9am

Some tips:

- Do the problems instead of reading them
- Use piazza and other resources
- **Keep an eye on updates on course outlines**, e.g. range, zoom link etc. on Quercus

**8.61** A small amount of the trace element selenium, from 50 to 200 micrograms ($\mu$g) per day, is considered essential to good health. Suppose that independent random samples of $n_1 = n_2 = 30$ adults were selected from two regions of the United States, and a day's intake of selenium, from both liquids and solids, was recorded for each person. The mean and standard deviation of the selenium daily intakes for the 30 adults from region 1 were $\overline{y}_1 = 167.1$ $\mu$g and $s_1 = 24.3$ $\mu$g, respectively. The corresponding statistics for the 30 adults from region 2 were $\overline{y}_2 = 140.9$ $\mu$g and $s_2 = 17.6$ $\mu$g. Find a 95% confidence interval for the difference in the mean selenium intake for the two regions.

## Problem 1

P430 [3]

**8.61** A small amount of the trace element selenium, from 50 to 200 micrograms ($\mu$g) per day, is considered essential to good health. Suppose that independent random samples of $n_1 = n_2 = 30$ adults were selected from two regions of the United States, and a day's intake of selenium, from both liquids and solids, was recorded for each person. The mean and standard deviation of the selenium daily intakes for the 30 adults from region 1 were $\overline{y}_1 = 167.1$ $\mu$g and $s_1 = 24.3$ $\mu$g, respectively. The corresponding statistics for the 30 adults from region 2 were $\overline{y}_2 = 140.9$ $\mu$g and $s_2 = 17.6$ $\mu$g. Find a 95% confidence interval for the difference in the mean selenium intake for the two regions.

$(n_1+n_2-2) > 30$

large sample

ss

small sample

$$\theta \in \left[\hat\theta - Z_{\frac{\alpha}{2}} \sigma_{\hat\theta},\ \hat\theta + Z_{\frac{\alpha}{2}} \sigma_{\hat\theta}\right]$$

$\alpha = 5\%$

$\frac{\alpha}{2} = 0.025$ $\quad Z_{\frac{\alpha}{2}} = 1.96$

$(15.463, 36.937).$

$\hat\theta = \overline{y}_1 - \overline{y}_2$

$$\sigma_{\hat\theta} \approx \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

**8.92** Solid copper produced by sintering (heating without melting) a powder under specified environmental conditions is then measured for porosity (the volume fraction due to voids) in a laboratory. A sample of $n_1 = 4$ independent porosity measurements have mean $\overline{y}_1 = .22$ and variance $s_1^2 = .0010$. A second laboratory repeats the same process on solid copper formed from an identical powder and gets $n_2 = 5$ independent porosity measurements with $\overline{y}_2 = .17$ and $s_2^2 = .0020$. Estimate the true difference between the population means $(\mu_1 - \mu_2)$ for these two laboratories, with confidence coefficient .95.

**8.92** Solid copper produced by sintering (heating without melting) a powder under specified environmental conditions is then measured for porosity (the volume fraction due to voids) in a laboratory. A sample of $n_1 = 4$ independent porosity measurements have mean $\overline{y}_1 = .22$ and variance $s_1^2 = .0010$. A second laboratory repeats the same process on solid copper formed from an identical powder and gets $n_2 = 5$ independent porosity measurements with $\overline{y}_2 = .17$ and $s_2^2 = .0020$. Estimate the true difference between the population means $(\mu_1 - \mu_2)$ for these two laboratories, with confidence coefficient .95.

Problem 2

$n_1+n_2-2$
$=7$

small sample

$t_{n_1+n_2-2,\frac{\alpha}{2}} = 2.365$ $\qquad \alpha = 0.05$ , $\frac{\alpha}{2} = 0.025$

prob. 8.4.17.

$$\mu_1-\mu_2 \in \left(\overline{Y}_1 - \overline{Y}_2 - t_{\frac{\alpha}{2}} S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}},\ \overline{Y}_1 - \overline{Y}_2 + t_{\frac{\alpha}{2}} S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$$

$\overline{Y}_1 = 0.22 \quad \overline{Y}_2 = 0.17, \quad S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = 0.0016$

$(-0.013, 0.113)$

UNIVERSITY OF TORONTO

**8.97** 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 $\sigma^2$.

**b** lower confidence bound for $\sigma^2$.

**8.99** In Exercise 8.97, you derived upper and lower confidence bounds, each with confidence coefficient $1-\alpha$, for $\sigma^2$. How would you construct a $100(1-\alpha)\%$

**a** upper confidence bound for $\sigma$?

**b** lower confidence bound for $\sigma$?

**8.97** 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 $\sigma^2$.

**b** lower confidence bound for $\sigma^2$.

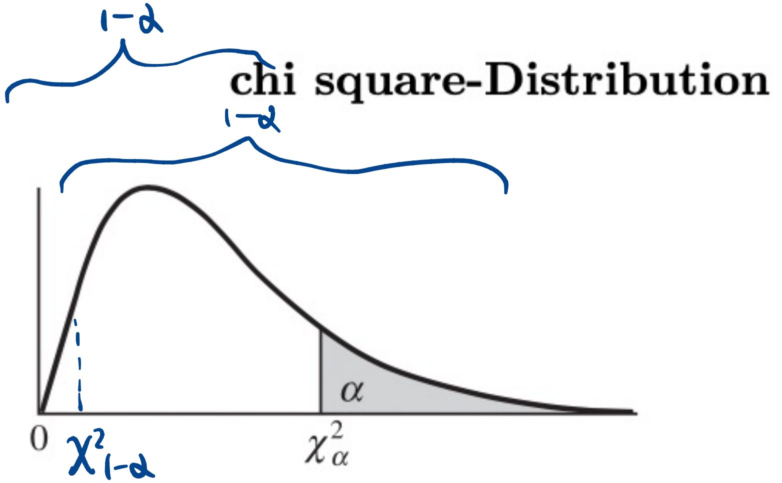
Problem 3

**8.99** In Exercise 8.97, you derived upper and lower confidence boun ficient $1-\alpha$, for $\sigma^2$. How would you construct a $100(1-\alpha)\%$

**a** upper confidence bound for $\sigma$?

**b** lower confidence bound for $\sigma$?

chi square-Distribution




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

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

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

$$\left(-\infty, \frac{(n-1)S^2}{\chi^2_{1-\alpha}}\right)$$

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

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

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha}}, +\infty\right)$$

UNIVERSITY OF TORONTO

**8.97** 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 $\sigma^2$.

**b** lower confidence bound for $\sigma^2$.

Problem 3

**8.99** In Exercise 8.97, you derived upper and lower confidence bounds, each with confidence coefficient $1-\alpha$, for $\sigma^2$. How would you construct a $100(1-\alpha)\%$

**a** upper confidence bound for $\sigma$?

**b** lower confidence bound for $\sigma$?

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

(b) $\sqrt{\frac{(n-1)S^2}{\chi^2_{\alpha}}}$

$\therefore P\left(\sigma \in \left(-\infty, \boxed{\sqrt{\frac{(n-1)S^2}{\chi^2_{1-\alpha}}}}\right)\right) = 1-\alpha$

Problem 4

**8.129** We noted in Section 8.3 that if

$$S'^2 = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{n} \quad \text{and} \quad S^2 = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{n-1},$$

then $S'^2$ is a biased estimator of $\sigma^2$, but $S^2$ is an unbiased estimator of the same parameter. If we sample from a normal population,

**a** find $V(S'^2)$.

**b** show that $V(S^2) > V(S'^2)$.

**8.129** We noted in Section 8.3 that if

$$S'^2 = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{n} \quad \text{and} \quad S^2 = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{n-1},$$

then $S'^2$ is a biased estimator of $\sigma^2$, but $S^2$ is an unbiased estimator of the same parameter. If we sample from a normal population,

**a** find $V(S'^2)$.

**b** show that $V(S^2) > V(S'^2)$.

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

$$\therefore V\left(\frac{(n-1)}{\sigma^2} S^2\right) = 2(n-1)$$

$$V(S^2) = \frac{\sigma^4}{(n-1)^2} 2(n-1)$$

$$= \frac{2\sigma^4}{n-1}$$

**8.129** We noted in Section 8.3 that if

## Problem 4

$$S'^2 = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{n} \quad \text{and} \quad S^2 = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{n-1},$$

then $S'^2$ is a biased estimator of $\sigma^2$, but $S^2$ is an unbiased estimator of the same parameter. If we sample from a normal population,

**a** find $V(S'^2)$.

**b** show that $V(S^2) > V(S'^2)$.

a. $V(S^2) = \frac{2\sigma^4}{(n-1)}$

$V(S'^2) = V\left(\frac{(n-1)}{n} S^2\right) = \frac{(n-1)^2}{n^2} V(S^2) = \frac{2(n-1)}{n^2}\sigma^4$

b. $\text{eff}(S^2, S'^2) = \frac{V(S'^2)}{V(S^2)} = \frac{\frac{2(n-1)}{n^2}\sigma^4}{\frac{2}{n-1}\sigma^4} = \frac{(n-1)^2}{n^2} < 1$

$\therefore V(S'^2) < V(S^2)$

**9.7** Suppose that $Y_1, Y_2, \ldots, Y_n$ denote a random sample of size $n$ from an exponential distribution with density function given by

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

In Exercise 8.19, we determined that $\hat{\theta}_1 = nY_{(1)}$ is an unbiased estimator of $\theta$ with $\text{MSE}(\hat{\theta}_1) = \theta^2$. Consider the estimator $\hat{\theta}_2 = \overline{Y}$ and find the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$.

**9.7** Suppose that $Y_1, Y_2, \ldots, Y_n$ denote a random sample of size $n$ from an exponential distribution with density function given by

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

In Exercise 8.19, we determined that $\hat{\theta}_1 = nY_{(1)}$ is an unbiased estimator of $\theta$ with $\text{MSE}(\hat{\theta}_1) = \theta^2$. Consider the estimator $\hat{\theta}_2 = \overline{Y}$ and find the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$.

$$B(\hat{\theta}_1) = 0 \Rightarrow MSE(\hat{\theta}_1) = V(\hat{\theta}_1)$$

$$V(\hat{\theta}_1) = \theta^2$$

$$E(\hat{\theta}_2) = \frac{1}{n}\sum E(Y_i) = \frac{1}{n} n \cdot \theta = \theta \qquad \therefore \text{unbiased}$$

$$V(\hat{\theta}_2) = \frac{1}{n^2}\sum V(Y_i) = \frac{1}{n^2} n \cdot \theta^2 = \frac{\theta^2}{n}$$

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1)} = \frac{\frac{\theta^2}{n}}{\theta^2} = \frac{1}{n}$$

**9.35** Let $Y_1, Y_2, \ldots$ be a sequence of random variables with $E(Y_i) = \mu$ and $V(Y_i) = \sigma_i^2$. Notice that the $\sigma_i^2$'s are not all equal.

**a** What is $E(\overline{Y}_n)$?

**b** What is $V(\overline{Y}_n)$?

**c** Under what condition (on the $\sigma_i^2$'s) can Theorem 9.1 be applied to show that $\overline{Y}_n$ is a consistent estimator for $\mu$?

# Consistency

**9.35** Let $Y_1, Y_2, \ldots$ be a sequence of random variables with $E(Y_i) = \mu$ and $V(Y_i) = \sigma_i^2$. Notice that the $\sigma_i^2$'s are not all equal.

**a** What is $E(\overline{Y}_n)$?

**b** What is $V(\overline{Y}_n)$?

**c** Under what condition (on the $\sigma_i^2$'s) can Theorem 9.1 be applied to show that $\overline{Y}_n$ is a consistent estimator for $\mu$?

a $E(\overline{Y}_n) = E(\frac{1}{n}\sum Y_i) = \frac{1}{n}\sum E(Y_i) = \frac{1}{n}\cdot n \cdot \mu = \mu$

b $V(\overline{Y}_n) = V(\frac{1}{n}\sum Y_i) = \frac{1}{n^2}\sum V(Y_i) = \frac{1}{n^2}(\sum_{i=1}^{n} \sigma_i^2)$

c. $\overline{Y}_n$ is consistent $\Leftrightarrow$ $V(\overline{Y}_n) \to 0,\ n \to \infty$

$\sigma_i^2 < \infty$ for all $i$. i.e. $\exists\, \sigma^2$ s.t. $\sigma_i^2 < \sigma^2\ \forall i$.

then use prob 9.2.3 from notes.

**9.53** Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a population with density function

$$f(y \mid \theta) = \begin{cases} \dfrac{2\theta^2}{y^3}, & \theta < y < \infty, \\ 0, & \text{elsewhere.} \end{cases}$$

Show that $Y_{(1)} = \min(Y_1, Y_2, \ldots, Y_n)$ is sufficient for $\theta$.

**9.53** Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a population with density function

$$f(y \mid \theta) = \begin{cases} \dfrac{2\theta^2}{y^3}, & \theta < y < \infty, \\ 0, & \text{elsewhere.} \end{cases} \quad \Rightarrow f(y|\theta) = \frac{2\theta^2}{y^3} I_{(\theta,\infty)} y$$

Show that $Y_{(1)} = \min(Y_1, Y_2, \ldots, Y_n)$ is sufficient for $\theta$.

$$\therefore \mathcal{L}(\theta) = \mathcal{L}(y_1, \cdots, y_n | \theta) = f(y_1|\theta)\, f(y_2|\theta) \cdots f(y_n|\theta)$$

$$= \left(\frac{2\theta^2}{y_1^3} I_{(\theta,\infty)} y_1\right)\left(\frac{2\theta^2}{y_2^3} I_{(\theta,\infty)} y_2\right) \cdots \left(\frac{2\theta^2}{y_n^3} I_{(\theta,\infty)} y_n\right)$$

$$= \frac{2^n \theta^{2n}}{\prod_{i=1}^n y_i^3} \left(\prod_{i=1}^n I_{(\theta,\infty)} y_i\right)$$

$$= \frac{2^n \theta^{2n}}{\prod_{i=1}^n y_i^3} I_{(\theta,\infty)} y_{(1)}$$

UNIVERSITY OF TORONTO

**9.53** Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a population with density function

$$f(y \mid \theta) = \begin{cases} \dfrac{2\theta^2}{y^3}, & \theta < y < \infty, \\ 0, & \text{elsewhere.} \end{cases}$$

Show that $Y_{(1)} = \min(Y_1, Y_2, \ldots, Y_n)$ is sufficient for $\theta$.

$$\mathcal{L}(\theta) = \frac{2^n \theta^{2n}}{\prod_{i=1}^{n} y_i^3} I_{(\theta,\infty)}\, y_{(1)}$$

$$g(y_{(1)}, \theta) = \theta^{2n} I_{(\theta,\infty)}\, y_{(1)} \qquad h(y_1 \cdots y_n) = \frac{2^n}{\prod_{i=1}^{n} y_i^3}$$

by factorization theorem, $Y_{(1)}$ is sufficient for $\theta$.

**9.63** Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a population with density function

$$f(y \mid \theta) = \begin{cases} \dfrac{3y^2}{\theta^3}, & 0 \le y \le \theta, \\ 0, & \text{elsewhere.} \end{cases}$$

In Exercise 9.52 you showed that $Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n)$ is sufficient for $\theta$.

**a** Show that $Y_{(n)}$ has probability density function

$$f_{(n)}(y \mid \theta) = \begin{cases} \dfrac{3ny^{3n-1}}{\theta^{3n}}, & 0 \le y \le \theta, \\ 0, & \text{elsewhere.} \end{cases}$$

**b** Find the MVUE of $\theta$.

9.52 show $Y_{(n)}$ is sufficient for $\theta$

**9.63** Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a population with density function

$$f(y \mid \theta) = \begin{cases} \dfrac{3y^2}{\theta^3}, & 0 \le y \le \theta, \\ 0, & \text{elsewhere.} \end{cases}$$

In Exercise 9.52 you showed that $Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n)$ is sufficient for $\theta$.

**a** Show that $Y_{(n)}$ has probability density function

$$f_{(n)}(y \mid \theta) = \begin{cases} \dfrac{3ny^{3n-1}}{\theta^{3n}}, & 0 \le y \le \theta, \\ 0, & \text{elsewhere.} \end{cases}$$

**b** Find the MVUE of $\theta$.

by factorization theorem

QED .

$$\mathcal{L}(\theta) = \frac{3^n \prod_{i=1}^{n} y_i^2}{\theta^{3n}} \left( \prod_{i=1}^{n} I_{(0,\theta)}\, y_i \right) = \frac{3^n \prod_{i=1}^{n} y_i^2}{\theta^{3n}} I_{(0,\theta)}\, y_{(n)} \left( \prod_{i=1}^{n} I_{(0,\infty)}\, y_i \right)$$

$$g(y_{(n)}, \theta) = \frac{1}{\theta^{3n}} I_{(0,\theta)}\, y_{(n)} \qquad h(y_1 \cdots y_n) = 3^n \prod_{i=1}^{n} y_i^2 \left( \prod_{i=1}^{n} I_{(0,\infty)}\, y_i \right).$$

**9.63** Let $Y_1, Y_2, \dots, Y_n$ be a random sample from a population with density function

$$f(y \mid \theta) = \begin{cases} \dfrac{3y^2}{\theta^3}, & 0 \le y \le \theta, \\ 0, & \text{elsewhere.} \end{cases}$$

In Exercise 9.52 you showed that $Y_{(n)} = \max(Y_1, Y_2, \dots, Y_n)$ is sufficient for $\theta$.

**a** Show that $Y_{(n)}$ has probability density function

$$f_{(n)}(y \mid \theta) = \begin{cases} \dfrac{3ny^{3n-1}}{\theta^{3n}}, & 0 \le y \le \theta, \\ 0, & \text{elsewhere.} \end{cases}$$

**b** Find the MVUE of $\theta$.

a. $$F(y \mid \theta) = \begin{cases} \int_0^y \dfrac{3t^2}{\theta^3}\, dt = \dfrac{y^3}{\theta^3} & 0 \le y \le \theta \\ 0 & y < 0 \\ 1 & y > \theta \end{cases}$$

$$f_{(n)}(y \mid \theta) = n[F(y)]^{n-1} f(y)$$

$$= \begin{cases} \dfrac{3n\, y^{3n-1}}{\theta^{3n}}, & 0 \le y \le \theta \\ 0 & \text{else} \end{cases}$$

**9.63** Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a population with density function

$$f(y \mid \theta) = \begin{cases} \dfrac{3y^2}{\theta^3}, & 0 \le y \le \theta, \\ 0, & \text{elsewhere.} \end{cases}$$

In Exercise 9.52 you showed that $Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n)$ is sufficient for $\theta$.

**a** Show that $Y_{(n)}$ has probability density function

$$f_{(n)}(y \mid \theta) = \begin{cases} \dfrac{3ny^{3n-1}}{\theta^{3n}}, & 0 \le y \le \theta, \\ 0, & \text{elsewhere.} \end{cases}$$

**b** Find the MVUE of $\theta$.

b. $E(Y_{(n)}) = \int_0^\theta \frac{3n\,t^{3n-1}}{\theta^{3n}} \cdot t \, dt = \frac{3n}{3n+1} \cdot \theta$.

$E\left(\frac{3n+1}{3n} Y_{(n)}\right) = \theta$. MVUE: $\frac{3n+1}{3n} \cdot Y_{(n)}$.

# THANK YOU FOR ATTENDING TODAY'S TUTORIAL

Feedbacks: yuehuan.he@utoronto.ca