ISE 315: Engineering Statistics

Lecture 3: Practice on Estimators

Instructor: Mansur M. Arief, PhD
Industrial and Systems Engineering, KFUPM

Office: 22-219 — Email: mansur.arief@kfupm.edu.sa

Course Handbook: Engineering Statistics — A Handbook, Chapter 2.

Texbook: Montgomery & Runger, Applied Statistics and Probability for Engineers, 6th Ed, Chapter 7.

Lecture 3

Practice on Estimators

Lecture 3 Outline

- **In-Class Practice #1** — attempt in teams, then we solve it together
- Q & A on homework or mini quizzes

Problem 1: A resin 3D-printing cell has a per-layer cure time that is normally distributed with mean $\mu = 60$ seconds and standard deviation $\sigma = 15$ seconds. A sample of $n = 25$ layers is timed and the sample mean $\bar{X}$ is recorded.

- Find the sampling distribution of $\bar{X}$: give $\mu_{\bar{X}}$ and $\sigma_{\bar{X}}$
- Find $P(56.4 \le \bar{X} \le 64.5)$
- The rule of thumb for CLT is $n \ge 30$, but $n = 25$. Does that make (a) and (b) wrong? Why?

Problem 1(a)

- Given: $\mu = 60$, $\sigma = 15$, $n = 25$
- Sampling distribution of $\bar{X}$ is Normal, with (2.2) and (2.5):

$$\mu_{\bar{X}=}\mu = 60 \text{ s} \quad \text{and} \quad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{25}} = 3 \text{ s}$$

- Careful: single layers vary by 15 s, but *averages of 25 layers* vary by only 3 s

Problem 1(b)

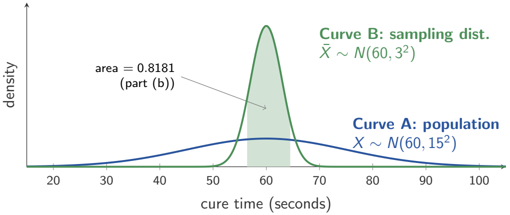
- Standardizing with (2.8):

$$
\begin{aligned}
P(56.4 \leq \bar{X} \leq 64.5) &= P\left(\frac{56.4-60}{3} \leq Z \leq \frac{64.5-60}{3}\right) \\
&= P(-1.20 \leq Z \leq 1.50) \\
&= \Phi(1.50) - \Phi(-1.20) \\
&= \underbrace{0.9332}_{\text{from table}} - \underbrace{0.1151}_{\text{from table}} \\
&= \boxed{0.8181}
\end{aligned}
$$

Problem 1(c): Does $n = 25 < 30$ break it?

- **No.** The $n \geq 30$ rule of thumb is only needed when the *population* is **not** normal and we lean on the CLT (2.7) to make $\bar{X}$ *approximately* normal
- Here the cure time is **stated to be normal**, so $\bar{X} \sim N(\mu, \sigma^2/n)$ exactly, for every n
- Same formula, two different justifications:
  - Normal population $\Rightarrow \bar{X}$ exactly normal (any n)
  - Non-normal population + large $n \Rightarrow \bar{X}$ approximately normal (CLT)

One curve is the *population* distribution, the other is the *sampling* distribution of $\bar{X}$ ($n = 25$). Which is which? Why are both at 60?



- **(d)** A is the **population distribution**; B the **sampling distribution**: narrower and taller
- Both at 60, $\bar{X}$ is **unbiased**: averaging changes the *precision*, not the *center*
- **(e)** At $n = 100$: A does not move; B narrows to $\sigma_{\bar{X}} = \sigma/\sqrt{n} = 15/\sqrt{100} = 1.5$

Problem 2: A bottling line records daily unplanned downtime, modelled as $N(\mu, \sigma^2)$ with $\sigma = 8$ minutes. From a sample of $n = 16$ days the maintenance team considers two estimators of the mean daily downtime μ: $\hat{\theta}_1 = \bar{X}$ and $\hat{\theta}_2 = 0.8\,\bar{X}$.

- Bias, variance, and MSE of $\hat{\theta}_1$
- Bias, variance, and MSE of $\hat{\theta}_2$
- Which has the smaller MSE when $\mu = 5$? When $\mu = 12$?

(a) For $\hat{\theta}_1 = \bar{X}$:

- $\text{Bias}(\hat{\theta}_1) = \mathbb{E}[\bar{X}] - \mu = 0$ (unbiased, by (2.2))
- $\text{Var}(\hat{\theta}_1) = \text{Var}(\bar{X}) = \dfrac{\sigma^2}{n} = \dfrac{8^2}{16} = 4 \quad (2.4)$
- $MSE(\hat{\theta}_1) = \text{Var}(\hat{\theta}_1) + \text{Bias}(\hat{\theta}_1)^2 = 4 + 0^2 = \boxed{4 \text{ min}^2}$

(b) For $\hat{\theta}_2 = 0.8\,\bar{X}$:

- $\text{Bias}(\hat{\theta}_2) = 0.8\,\mathbb{E}[\bar{X}] - \mu = 0.8\mu - \mu = -0.2\mu$ (biased low)
- $\text{Var}(\hat{\theta}_2) = (0.8)^2\text{Var}(\bar{X}) = 0.64 \cdot 4 = 2.56$
- $\text{MSE}(\hat{\theta}_2) = 2.56 + (-0.2\mu)^2 = \boxed{2.56 + 0.04\mu^2}$

Shrinking cut the variance from 4 to 2.56, but add a bias that grows with μ.

Problem 2: Solution (c)

- For $\mu = 5$ min:

$$\text{MSE}(\hat{\theta}_2) = 2.56 + 0.04(25) = 3.56 < 4 = \text{MSE}(\hat{\theta}_1) \quad \Rightarrow \hat{\theta}_2 \textbf{ wins}$$

- For $\mu = 12$ min:

$$\text{MSE}(\hat{\theta}_2) = 2.56 + 0.04(144) = 8.32 > 4 = \text{MSE}(\hat{\theta}_1) \quad \Rightarrow \hat{\theta}_1 \textbf{ wins}$$

- **The lesson:** unbiased $\neq$ accurate — which estimator is better depends on μ

Problem 3: A logistics firm audits $n = 400$ randomly selected parcels. For each parcel, $X_i = 1$ if the delivery was late and $X_i = 0$ otherwise. The sample late rate is

$$\hat{P} = \frac{X_1 + X_2 + \cdots + X_{400}}{400}$$

- Population distribution of X_i? Sampling distribution of $\hat{P}$? Is $\hat{P}$ unbiased?
- We want $\mathrm{SE}(\hat{P}) \leq 0.01$. How many parcels must be checked ($p \approx 0.10$)?

Problem 3(a)

Sampling distribution of $\hat{P} = \frac{X_1+\cdots+X_n}{n}$ (2.10) — it is a *sample average*:

- $\mathbb{E}[\hat{P}] = \mathbb{E}\left[\bar{X}\right] = \mathbb{E}[X] = p$ (unbiased!)
- $\mathrm{Var}(\hat{P}) = \mathrm{Var}\left(\bar{X}\right) = \frac{\mathrm{Var}(X)}{n} = \frac{p(1-p)}{n}$, $\mathrm{SE}(\hat{P}) = \sqrt{\frac{p(1-p)}{n}}$ (2.11)

By CLT (since $n = 400$ **is large), (2.12):** $\boxed{\hat{P} \overset{\text{approx}}{\sim} N\left(p, \frac{p(1-p)}{n}\right)}$

Problem 3: Solution (b) — how many parcels?

- **Given:** target $\mathrm{SE}(\hat{P}) \leq 0.01$, planning value $p \approx 0.10$
- **Set up** with (2.11): $\sqrt{\dfrac{p(1-p)}{n}} \leq 0.01$
- **Solve for** n ((2.9)):

$$n \geq \frac{p(1-p)}{(0.01)^2} = \frac{(0.10)(0.90)}{0.0001} = \boxed{900 \text{ parcels}}$$

- At the current $n = 400$: $\mathrm{SE} = \sqrt{0.09/400} = 0.015$
- Cutting the SE by a third costs $900/400 = 2.25\times$ the audit effort — precision improves only with $\sqrt{n}$

Class Announcement Lecture 3

- No homework yet, but start reading Handbook Chapter 3 and textbook Chapter 7
- Homework 1 will be posted next week
- Next class: Multiple population estimators