ISE 315: Engineering Statistics

Lecture 4: Multi-Population Estimators

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

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

Based on Montgomery & Runger, Applied Statistics and Probability for Engineers, 6th Ed.

Lecture 4

Multi-Population Estimators

Lecture 4 Outline

- Announcements:
  - HW 2 will be posted EOD today, due 2 weeks from today before class
  - Quiz #1 today, for those who want to get done early
  - You may take it next week if you want (no penalty, but different questions)
- Multi-Population Estimators
- Next class: Start Chapter 8 (Statistics Interval for Single Population)

Population and Sampling Distributions

- Given random sample $X_1, X_2, \ldots, X_n$ from population p with mean μ
- $\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$ converges to a Normal Distribution (by CLT)
  - Mean: $\mu_{\bar{X}} = \mu$
  - Variance:
    - If σ^2 is known: $\sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$
    - If σ^2 is unknown: $\sigma^2_{\bar{X}} = \frac{s^2}{n}$, where

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

Difference of Means $\mu_1 - \mu_2$

- Goal: estimate the difference in means between two populations with
  - μ_1 and μ_2 are the means
  - σ_1^2 and σ_2^2 are the variances
  - n_1 and n_2 are the sample sizes from the two populations
- We can use the difference of sample means $\bar{X}_1 - \bar{X}_2$ as an estimator
- The sampling distribution of $\bar{X}_1 - \bar{X}_2$ converges to a Normal Distribution:
  - Mean: $\mu_{\bar{X}_1 - \bar{X}_2} = \mathbb{E}[\bar{X}_1 - \bar{X}_2] = \mu_1 - \mu_2$
  - Variance: $\sigma^2_{\bar{X}_1 - \bar{X}_2} = \mathrm{Var}(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$
  - Standard Error: $\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Similar to Single Population Estimator

- Goal: estimate unknown parameter $\Delta\mu = \mu_1 - \mu_2$
  - μ_1 and μ_2 are the population means
  - σ_1^2 and σ_2^2 are the population variances
  - n_1 and n_2 are the sample sizes from the two populations
- Define estimator $\Delta\hat{\mu} = \Delta\bar{X} = \bar{X}_1 - \bar{X}_2$
- The sampling distribution of $\Delta\bar{X}$ converges to a Normal Distribution:
  - Mean: $\mu_{\Delta\bar{X}} = \mathbb{E}[\Delta\bar{X}] = \Delta\mu = \mu_1 - \mu_2$
  - Variance: $\sigma^2_{\Delta\bar{X}} = \text{Var}(\Delta\bar{X}) = \text{Var}(\bar{X}_1 - \bar{X}_2) = \text{Var}(\bar{X}_1) + \text{Var}(\bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$
  - Standard Error: $\sigma_{\Delta\bar{X}} = \sqrt{\text{Var}(\Delta\bar{X})} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

A refinery compares fuel efficiency of two catalyst types.
Type A: $n_1 = 50$ batches, $\sigma_1 = 4$ L/100km.
Type B: $n_2 = 45$ batches, $\sigma_2 = 5$ L/100km.
What is the probability that $\bar{X}_1 - \bar{X}_2$ differs from $\mu_1 - \mu_2$ by less than 1.5 L/100km?

Problem 1: Fuel Efficiency Comparison

What do we need to find first?

Step 1: Population distribution parameters:

- $\sigma_1 = 4,\ n_1 = 50,\ \sigma_2 = 5,\ n_2 = 45$ (μ_1, μ_2 are unknown)

Step 2: Sampling distribution parameters:

- $\mu_{\Delta\bar{X}} = \mu_1 - \mu_2$ (unknown)
- $\sigma^2_{\Delta\bar{X}} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \frac{16}{50} + \frac{25}{45} = 0.32 + 0.556 = 0.876$
- $\sigma_{\Delta\bar{X}} = \sqrt{\sigma^2_{\Delta\bar{X}}} = \sqrt{0.876} = 0.936$

Step 3: Standardize and find probability

$$P\left(|(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)| < 1.5\right) = P\left(|Z| < \frac{1.5}{0.936}\right) = P(|Z| < 1.60)$$

$$= \Phi(1.60) - \Phi(-1.60) = 0.9452 - 0.0548 = 0.8904$$

Two production lines manufacture electronic components. Line 1: $n_1 = 200$ units, true defect rate $p_1 = 0.08$. Line 2: $n_2 = 250$ units, true defect rate $p_2 = 0.05$. Find the probability that the difference in sample proportions $\hat{P}_1 - \hat{P}_2$ is within 0.03 of the true difference.

Problem 2: Defect Rate Comparison

Step 1: Population distribution parameters:

- $p_1 = 0.08$
- $p_2 = 0.05$
- $n_1 = 200$
- $n_2 = 250$

Step 2: Sampling distribution parameters (specifically standard error/SE)

$$
\begin{aligned}
\mathrm{SE} &= \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \\
&= \sqrt{\frac{0.08(0.92)}{200} + \frac{0.05(0.95)}{250}} \\
&= \sqrt{0.000368 + 0.00019} \\
&= \sqrt{0.000558} \\
&= 0.0236
\end{aligned}
$$

Step 3: Find the probability

$$P\left(|(\hat{P}_1 - \hat{P}_2) - (p_1 - p_2)| \leq 0.03\right) = P\left(|Z| \leq \frac{0.03}{0.0236}\right)$$
$$= P(|Z| \leq 1.27)$$
$$= \Phi(1.27) - \Phi(-1.27)$$
$$= 0.8980 - 0.1020 = \boxed{0.7960}$$

Summary: Sampling Distributions for Multi-Population Estimators

	Mean Difference $(\bar{X}_1 - \bar{X}_2)$	**Proportion Difference** $(\hat{P}_1 - \hat{P}_2)$
Mean	$\mu_1 - \mu_2$	$p_1 - p_2$
Variance	$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$	$\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$

Remember: For proportions, check $np \geq 10$ and $n(1-p) \geq 10$ before using Normal approximation.

Quiz #1 (available on Blackboard)

- 5 questions about estimators 7.1-7.3, 10-15 minutes
- Please stay if you are taking it today
- If not, you may leave early and see you next week!