BAN 602: Quantitative Fundamentals

Lecture 3: Central Limit Theorem and Confidence Interval

Agenda

1. Big Picture of Chapter 7: Sampling Distribution and Central Limit Theorem
2. Big Picture of Chapter 8: Confidence Interval
3. Example 1 (population mean - quantitative data)
4. Example 2 (population proportion - categorical data)

Big Picture of Chapter 7: Sampling Distribution & CLT

Sampling and Point Estimation	Sample Distribution	Properties of Point Estimators

• Population (finite and infinite) vs. Sample
• Sampling
  ○ Simple random sampling
  ○ Stratified random sampling
  ○ Cluster sampling
  ○ Systematic sampling
  ○ Convenience sampling
  ○ Judgment sampling
• Sample statistic (a sample characteristic: sample mean; sample proportion; sample standard deviation)
• Point estimation
	• Sampling Distribution (prob. Dist. of a sample statistic)
	  ○ **Sampling distribution of sample mean**
	  ○ **Sampling distribution of sample proportion**
	  ○ Sampling distribution of sample variance or sd
		• Unbiased
		• Efficiency
		• Consistency

Big Picture of Chapter 8: Interval Estimation

Basics of Interval Estimation	Population Mean	Population Proportion

• Interval estimate
• Margin of error
• Interval estimate of a population mean
• Interval estimate of a population proportion
• Interval estimate of difference of two population means/proportions (ch. 10)
  ○ What if more than 2 population means (ch. 13)
  ○ What if more than 2 population proportions (ch. 12)
• Interval estimate of a population variance & ratio of two population variances (ch. 11)
	• **Confidence level and confidence interval**
	• **When σ is known (generally speaking, use normal distribution)**
	• **When σ is unknown (generally speaking, use t distribution)**
	• Given desired margin of error, determine the appropriate sample size
		• **Interval estimation of population proportion**
		• **Determine the sample size**

Example 1 (Population Mean)

We want to study the annual cost of auto insurance.

- What would be the population of our interest? Let X be the random variable that describes the population. How can we interpret this random variable?

Example 1 (Population Mean)

We want to study the annual cost of auto insurance.

- What probability distribution does X follow? Note that this probability distribution describe the population.

Example 1 (Population Mean)

- How can we know about the (population) mean μ and the standard deviation σ?

Example 1 (Population Mean)

- Now, let's assume the (population) mean μ = \$939 and the (population) standard deviation σ = \$245, and n = 25. Can you describe the sampling distribution of $\bar{X}$?

Example 1 (Population Mean)

- What is the probability that a simple random sample of automobile insurance policies will have a sample mean within $25 of the population mean?

Example 1 (Population Mean)

- In reality, we rarely know μ and σ. We instead apply the process reversely: use the sample mean to infer the population mean. And this process is called **statistical inference**. Now suppose we collect information on 25 auto insurance polices and their average annual cost is \$1000. For the time being, assume we still know σ = \$245 (We will handle the situation where σ is unknown later). What is the probability that the population mean is within \$50 of the sample mean?

Example 1 (Population Mean)

- To answer the previous question, let's consider the general case. We randomly sampled n auto policies and the mean price is $\bar{X}$. What is the probability that the population mean is within \$m (this is called margin of error) of the sample mean? This range of $\bar{X}$ $\pm$m is called a confidence interval and the probability $Pr(\bar{X} - m \leq \mu \leq \bar{X} + m)$ is called the confidence level.

Example 1 (Population Mean)

- In practice, we typically have desirable confidence levels, with 90%, 95%, and 99% being the most commonly used ones. We, instead, want to find the corresponding margin of error and the resulting confidence interval.
- Once again, assume σ, the population standard deviation is known. We randomly sampled n auto policies and the mean price is $\bar{X}$.
- Suppose the confidence level we want is 1-α. (α is called significance level, which we will use extensively later on. If the confidence level is, say 90%, then the significance level is 10%, vice versa.) What would be the margin of error that provides the confidence level of 1-α? And what would be the confidence interval that provides the confidence level of 1-α?.

Example 1 (Population Mean)

- Now suppose we collect information on 25 auto insurance polices and their average annual cost is $1000; σ = $245. What would be the margin of error that provides a 95% confidence level? And what would be a 95% confidence interval?. How can we interpret this confidence interval?

Example 1 (Population Mean)

- Sometimes we have desirable margin of error and confidence level. We want to find the corresponding sample size that can help us achieve the desirable m and 1-α. How?

Example 1 (Population Mean)

- Now suppose the population standard deviation of annual cost of auto insurance polices σ = \$245. What sample size can ensure that the margin of error of 99% confidence level is \$50?

Example 1 (Population Mean)

- What are the impacts of large sample size n?

$$m = F_{s.n.}^{-1}\left(1 - \frac{\alpha}{2}\right) * \sigma/\sqrt{n}$$

Example 1 (Population Mean)

- Now what if σ is unknow, which typically is the case?

Example 2 (Population Proportion)

The president of Doerman Distributors, Inc., wants to know the percentage of the firm’s orders that come from first-time customers. To that end, a random sample of 100 orders is collected among which 30% of the firm’s orders come from first-time customers.

- What is the population parameter of interest? What do we know about this population parameter?

Example 2 (Population Proportion)

- We use sample statistic, sample proportion $\bar{p}$, to study p. What do we know about $\bar{p}$?

Example 2 (Population Proportion)

The president of Doerman Distributors, Inc., wants to know the percentage of the firm's orders that come from first-time customers. To that end, a random sample of 100 orders is collected among which 30% of the firm's orders come from first-time customers.

- If the confidence level is 90%, what is the margin of error and what is a confidence interval for the population proportion?
- How can we interpret this confidence interval?

Example 2 (Population Proportion)

You are probably not happy with so wide a confidence interval. Suppose you want a confidence level of 95% ($1-\alpha$) and a margin of error 3% (m). What sample size will be needed?

Sample Size Requirements (Population Proportion)

	m = 1%	m = 3%	m = 5%
90%	6764	752	271
95%	9604	1068	385
99%	16588	1844	664

Quiz 3 Part 2

1. Describe the sampling distribution of sample mean or proportion.
2. Write down R code for computing margin of error.
3. Provide confidence interval and its interpretation.
4. Compute the sample size needed for desirable confidence level and margin of error.