ECON10005 Quantitative Methods 1

Tutorial in Week 7

Zheng Fan

The University of Melbourne

Introduction

Zheng Fan

- Ph.D student in Economics at Unimelb. Research interest in Bayesian Econometrics
- Personal website: *zhengfan.site* for some details

Don't be shy if you need help

- Discuss on Ed Discussion Board
- Attend consultation sessions: see Canvas for time and location
- Consult Stop 1, in case of special considerations,
- Contact QM-1@unimelb.edu.au for admin issues
- Send me an email: fan.z@unimelb.edu.au

Agenda

Today's focus

- Understanding sampling distributions
- Central Limit Theorem (CLT)
- Practice with normal probability calculations

Agenda

- **Next week** (Mid-semester break):
  - **No tutorial**
- **Following week:** We will spend time discussing your group report.
  - Overall, a strong start—well done!
  - Remember: this is your **first report**, and the goal is to **learn and improve**.
  - Losing marks here is expected and useful—it highlights where you can focus your attention for the next report.

PART A Questions

Answer the following questions.

1. Define a sampling distribution in your own words. How does it differ from a population distribution?
2. Explain how the mean and standard deviation of a sampling distribution are related to the population parameters. What happens to the standard deviation of the sampling distribution as the sample size increases?
3. State the Central Limit Theorem (CLT). Why is the CLT important for inferential statistics, especially in business decision-making?

PART A Question 1

Can anyone tell me what a sampling distribution is?

PART A Question 1

Can anyone tell me what a sampling distribution is?

- A sampling distribution is the probability distribution of a sample statistic (e.g., sample mean $\bar{X}$) based on repeated random samples from the population.

PART A Question 1

Can anyone tell me what a sampling distribution is?

- A sampling distribution is the probability distribution of a sample statistic (e.g., sample mean $\bar{X}$) based on repeated random samples from the population.
- An example of a statistic is the sample mean $\bar{X}$ or the sample proportion $\hat{p}$

PART A Question 1

Can anyone tell me what a sampling distribution is?

- A sampling distribution is the probability distribution of a sample statistic (e.g., sample mean $\bar{X}$) based on repeated random samples from the population.
- An example of a statistic is the sample mean $\bar{X}$ or the sample proportion $\hat{p}$
- The distribution shows all possible statistics values that could be obtained from all possible samples of a certain size taken from the population.

PART A Question 1

Example: Can anyone tell me what $\hat{p}$ is?

PART A Question 1

Example: Can anyone tell me what $\hat{p}$ is?

- $\hat{p}$ is the total number of successes in the sample divided by the sample size.

PART A Question 1

Example: Can anyone tell me what $\hat{p}$ is?

- $\hat{p}$ is the total number of successes in the sample divided by the sample size.
- It actually equals to the sample mean, think about CBD or not in your draft.

PART A Question 1

Example: Can anyone tell me what $\hat{p}$ is?

- $\hat{p}$ is the total number of successes in the sample divided by the sample size.
- It actually equals to the sample mean, think about CBD or not in your draft.

What values can $\hat{p}$ take?

PART A Question 1

Example: Can anyone tell me what $\hat{p}$ is?

- $\hat{p}$ is the total number of successes in the sample divided by the sample size.
- It actually equals to the sample mean, think about CBD or not in your draft.

What values can $\hat{p}$ take?

- $\hat{p}$ can be $0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1$ where n is the sample size.

PART A Question 1

How does this differ from a population distribution?

PART A Question 1

How does this differ from a population distribution?

- It differs from a population distribution because the sampling distribution is based on statistics derived from samples, whereas a population distribution describes the entire population.

PART A Question 1

How does this differ from a population distribution?

- It differs from a population distribution because the sampling distribution is based on statistics derived from samples, whereas a population distribution describes the entire population.
- An individual i in the sample is associated with a random variable X_i, which follows the population distribution.

PART A Question 1

How does this differ from a population distribution?

- It differs from a population distribution because the sampling distribution is based on statistics derived from samples, whereas a population distribution describes the entire population.
- An individual i in the sample is associated with a random variable X_i, which follows the population distribution.
- i is the index for each sample, $1, 2, \ldots, n$.

PART A Question 1

Example: Consider the income of individuals in a city.

PART A Question 1

Example: Consider the income of individuals in a city.

- A population distribution shows the income distribution of all individuals in a city

PART A Question 1

Example: Consider the income of individuals in a city.

- A population distribution shows the income distribution of all individuals in a city
- A sampling distribution shows how sample statistics (like the mean income $\bar{X}$ of randomly selected groups) would distribute across many samples X_i

PART A Question 1

Example: Consider the income of individuals in a city.

- A population distribution shows the income distribution of all individuals in a city
- A sampling distribution shows how sample statistics (like the mean income $\bar{X}$ of randomly selected groups) would distribute across many samples X_i
- Recall that $\bar{X} = \dfrac{X_1 + \cdots + X_n}{n}$

PART A Question 2

Explain how the mean and standard deviation of a sampling distribution are related to the population parameters. What happens to the standard deviation of the sampling distribution as the sample size increases?

- The population parameters relate to the population distribution.

PART A Question 2

Explain how the mean and standard deviation of a sampling distribution are related to the population parameters. What happens to the standard deviation of the sampling distribution as the sample size increases?

- The population parameters relate to the population distribution.
- $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$

PART A Question 2

- The population parameters relate to the population distribution.
- $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$

Consider sampling distribution of $\bar{X}$:

- $E(\bar{X}) = E(X_i) = \mu$

PART A Question 2

- The population parameters relate to the population distribution.
- $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$

Consider sampling distribution of $\bar{X}$:

- $E(\bar{X}) = E(X_i) = \mu$
- $Var(\bar{X}) = Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$

PART A Question 2

- The population parameters relate to the population distribution.
- $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$

Consider sampling distribution of $\bar{X}$:

- $E(\bar{X}) = E(X_i) = \mu$
- $Var(\bar{X}) = Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$
- The standard deviation of the sampling distribution, also called standard error (SE) =

$$\sqrt{Var(\bar{X})} = \frac{\sigma}{\sqrt{n}}$$

PART A Question 2

- The population parameters relate to the population distribution.
- $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$

Consider sampling distribution of $\bar{X}$:

- $E(\bar{X}) = E(X_i) = \mu$
- $Var(\bar{X}) = Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$
- The standard deviation of the sampling distribution, also called standard error (SE) =

$$\sqrt{Var(\bar{X})} = \frac{\sigma}{\sqrt{n}}$$

- As sample size n increases, SE decreases $\rightarrow$ more precise estimates

PART A Question 3

Can anyone tell me what the Central Limit Theorem (CLT) is?

PART A Question 3

Can anyone tell me what the Central Limit Theorem (CLT) is?

- For sufficiently large n, the sampling distribution of the sample mean is approximately normal, even if the population distribution is not.

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \overset{a}{\sim} N(0, 1)$$

PART A Question 3

Can anyone tell me what the Central Limit Theorem (CLT) is?

- For sufficiently large n, the sampling distribution of the sample mean is approximately normal, even if the population distribution is not.

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \overset{a}{\sim} N(0, 1)$$

- Allows use of normal distribution for inference.

PART A Question 3

Can anyone tell me what the Central Limit Theorem (CLT) is?

- For sufficiently large n, the sampling distribution of the sample mean is approximately normal, even if the population distribution is not.

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \overset{a}{\sim} N(0, 1)$$

- Allows use of normal distribution for inference.
- Critical for applications in business where underlying data is not normal.

PART B Questions

A business school announces that their graduates are among the top earners in the industry and earn an impressive **average** salary of \$2,000 per week just one year after graduation, with a **standard deviation** of \$100 per week.
A group of statistics students decided to investigate this claim using a random sample of **25 graduates** from this business school a year after they completed their degree. Assuming that the claims of the business school about the average salary is true, answer the following questions.

1. Find the probability that the students will find a sample mean of the weekly earnings:
   a. less than \$2,000.
   b. greater than \$2,000.
   c. greater than \$2,200. How do you interpret this probability?
   d. between \$1980 and \$2010.

Business School Salary Case

- True mean $\mu = 2000$, standard deviation $\sigma = 100$,

Business School Salary Case

- True mean $\mu = 2000$, standard deviation $\sigma = 100$,
- Sample size $n = 25$

Business School Salary Case

- True mean $\mu = 2000$, standard deviation $\sigma = 100$,
- Sample size $n = 25$
- Sample mean $\bar{X}$ from these $n = 25$ sample data

Business School Salary Case

- True mean $\mu = 2000$, standard deviation $\sigma = 100$,
- Sample size $n = 25$
- Sample mean $\bar{X}$ from these $n = 25$ sample data
- Standard deviation of $\bar{X}$: $\sigma_{\bar{X}} =?$
  - How to calculate standard deviation of a data?
  - We collect a bunch of $\bar{X}_1$, $\bar{X}_2$ ... and $\bar{X}_n$.
  - Then calculate the standard deviation via some formula
  - However, we only have one set of sample data, and thus only $\bar{X}_1$

Business School Salary Case

- True mean $\mu = 2000$, standard deviation $\sigma = 100$,
- Sample size $n = 25$
- Sample mean $\bar{X}$ from these $n = 25$ sample data
- Standard deviation of $\bar{X}$: $\sigma_{\bar{X}} = ?$
  - How to calculate standard deviation of a data?
  - We collect a bunch of $\bar{X}_1$, $\bar{X}_2$ ... and $\bar{X}_n$.
  - Then calculate the standard deviation via some formula
  - However, we only have one set of sample data, and thus only $\bar{X}_1$
- Standard error $= \frac{100}{\sqrt{25}} = 20$, which is $\sigma_{\bar{X}}$.

Q1 a b: Probabilities at Mean

- $P(\bar{X} < 2000) = 0.5$
- $P(\bar{X} > 2000) = 0.5$
- Follows from symmetry of normal distribution

Q1 c: $P(\bar{X} > 2200)$

- $Z = \dfrac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \dfrac{2200 - 2000}{20} = 10$
- $P(Z > 10) \approx 0$
- Sample mean of \$2200 is extremely unlikely

Q1 d: $P(1980 < \bar{X} < 2010)$

- $Z_1 = \dfrac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \dfrac{1980 - 2000}{20} = -1$
- $Z_2 = \dfrac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \dfrac{2010 - 2000}{20} = 0.5$
- $P(-1 < Z < 0.5) = P(Z < 0.5) - P(Z < -1) = 0.5328$

PART B Questions

A business school announces that their graduates are among the top earners in the industry and earn an impressive average salary of \$2,000 per week just one year after graduation, with a standard deviation of \$100 per week.
A group of statistics students decided to investigate this claim using a random sample of 25 graduates from this business school a year after they completed their degree. Assuming that the claims of the business school about the average salary is true, answer the following questions.

2. Based on your understanding of random variables and income distributions, you hypothesize that the weekly earnings of business graduates might not follow a normal distribution. How would this affect your previous findings? Use the Central Limit Theorem to support your explanation.

Q2: What If Population is Not Normal?

- CLT holds for large n (rule of thumb: $n \geq 30$)
- For $n = 25$, approximation still reasonable unless heavy skew/outliers
- For extreme cases, larger n needed to ensure (approximately distributed as) normality.

Some Discussion and Thoughts

Why is CLT powerful in real-world data?

- It allows inference even when data is not normally distributed.
- Applicable to various fields: economics, finance, health, etc.

When might the approximation break down?

- When sample size is small ($n < 30$) and population is skewed.
- Presence of extreme outliers or heavy tails.

Examples where sample size matters:

- Income distributions (often skewed)
- Customer spending behavior
- Hospital wait times

Any final questions?

Thanks for your attention! 😄

Let me know if you have any questions.

No tutorial next week!