FUNDAMENTALS OF STATISTICS (4ECON013C)

TUTORIAL 6

Point and Interval estimations

Learning outcomes:

- Point estimation for μ and π;
- Confidence Interval for population mean (μ);
- Confidence Interval for population proportion (π);
- Confidence Interval for population means difference (μ_1-μ_2);
- Confidence Interval for population proportions difference ($\pi_1 - \pi_2$).

Principle of confidence intervals

The **principle of confidence intervals** is based on the idea of using sample data to estimate an unknown population parameter (such as μ, π) with an associated level of certainty. Instead of providing a single point estimate, such as a sample mean or proportion, a confidence interval gives a **range of plausible values** that are likely to contain the true population value. The width of the interval reflects the **degree of uncertainty** in the estimate: narrower intervals indicate more precise estimates, while wider intervals reflect greater variability or smaller sample sizes. The **confidence level** (such as 90%, 95%, or 99%) represents the proportion of intervals that would contain the true parameter if the sampling process were repeated many times. In essence, a confidence interval combines information about the sample estimate, its variability, and the desired confidence level to make probabilistic statements about the population parameter.

100(1-α)% Confidence interval for a single mean (μ):

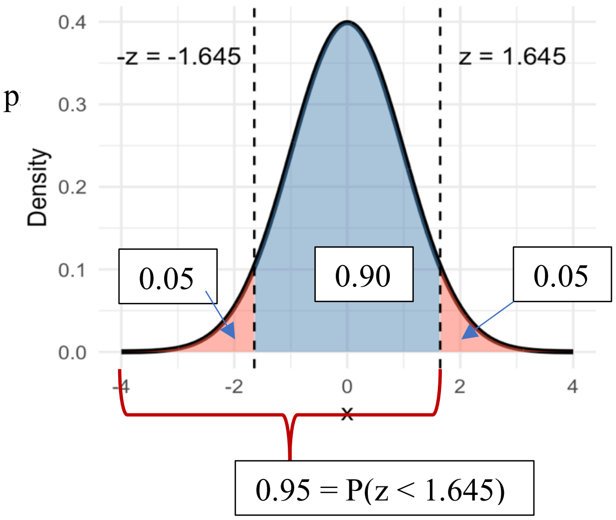
σ known	**σ unknown (given n ≥ 30)**
$\bar{x} \pm z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}$	$\bar{x} \pm z_{\alpha/2} * \frac{s}{\sqrt{n}}$

where $z_{\alpha/2}$ is the z-value which cuts off the $\alpha/2$*100% probability in the upper tail of the standard normal (z) distribution. Here, **Margin of Error (ME)** = $z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}$ or $z_{\alpha/2} * \frac{s}{\sqrt{n}}$

Example 6.1. A farmer wants to estimate the average weight of apples from his orchard. He randomly selects a sample of 40 apples, which have an average weight of 155 grams with a sample standard deviation of 12 grams. He wants to construct a 90% confidence interval for the true mean weight of all apples.

Solution.

$\alpha = 1 - 0.90 = 0.10$ for 90% confidence level
Then $z_{\alpha/2} = z_{0.05} = 1.645$ from the z table for cumulative p



100(1-α)% Confidence interval for a population proportion (π):

The sampling distribution of the sample proportion, P, (estimator of the population proportion, π) is: $P \sim N(\pi, \frac{\pi(1-\pi)}{n})$ approximately, as $n \to \infty$.

The confidence interval for a single proportion is given as follows:

$$\boxed{\mathbf{p} \pm \mathbf{z}_{\alpha/2} * \sqrt{\frac{\mathbf{p}(1-\mathbf{p})}{\mathbf{n}}}}$$

where $z_{\alpha/2}$ is the z-value which cuts off the α/2*100% probability in the upper tail of the standard normal (z) distribution.

Example 6.2. A random survey was conducted by Emerson College Polling to find out which candidate was favorite to win the US election in the state of Georgia. Out of 800 surveyed, 392 favored Kamala Harris, 384 favored Donald Trump and the rest indicated the independent candidate. If we denote the population proportion of voters who support Harris by π, and then what is the 95% confidence interval for population proportion?

Solution.

$p = 392/800 = 0.49$ and $z_{\alpha/2} = 1.96$ for 95% CI.

$$0.49 \pm 1.96 * \sqrt{\frac{0.49*0.51}{800}} = 0.49 \pm 0.0346$$

So, 95% CI for π: (0.455, 0.525) or (45.5%, 52.5%)

Choosing minimum sample size

If you want to limit your Margin of Error (ME) then how many observations (at minimum) do you need for your study? The sample size can be found using the following formulas:

For mean:	**For proportion*:**
$\mathbf{n} = \left(\frac{\sigma * z_{\alpha/2}}{ME}\right)^2$	$\mathbf{n} = p(1-p)\left(\frac{z_{\alpha/2}}{ME}\right)^2$ *if p is unknown, use p = 0.5

Confidence Interval for Two Population mean difference (μ_1 - μ_2):

When σ is known:

$$\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2} * \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

When σ is unknown, provided n_1, n_2 ≥ 30:

$$\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2} * \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Confidence Interval for Two Population proportion difference (π_1 - π_2)

$$p_1 - p_2 \pm z_{\frac{\alpha}{2}} * \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Example 6.3. A health researcher wants to compare COVID-19 vaccination rates between urban and rural residents.

- In a sample of 400 urban adults, 320 are vaccinated.
- In a sample of 500 rural adults, 350 are vaccinated.

We want a 98% confidence interval for the difference in population proportions, $\pi_U - \pi_R$

Solution. $p_U = 320/400 = 0.8$ and $p_R = 350/500 = 0.7$, $z_{0.01} = 2.33$ ($\frac{\alpha}{2} = 0.01$)

Then, 98% confidence interval for $\pi_U - \pi_R$:

$$0.8 - 0.7 \pm 2.33 * \sqrt{\frac{0.8*0.2}{400} + \frac{0.7*0.3}{500}}$$

$$0.1 \pm 0.067 \Rightarrow (0.033, 0.167)$$

TASKS

1. ***True/False*** questions. Explain your choice.

   a. A 95% confidence interval for the mean always contains the true population mean.

   b. Increasing the sample size will reduce the width of a confidence interval for the population mean.

   c. The confidence interval for a population proportion will be narrower if the sample proportion is close to 0.5 compared to when it is closer to 0 or 1.

   d. When constructing a confidence interval for the difference between two population means, we assume that both populations have similar variances.

   e. A 99% confidence interval is wider than a 95% confidence interval for the same dataset and variable.

   f. If the confidence interval for the difference between two proportions does not contain zero, we can conclude that the difference is statistically significant.

   g. If we construct a 95% confidence interval for a mean and later decide to increase the confidence level to 99% using the same data and want to keep the same margin of error, then we need to increase the sample size.

2. A simple random sample of 100 workers had weekly salaries with a mean of $315 and a standard deviation of $20. Calculate a 90% confidence interval for the mean weekly salary of all workers in the factory.

3. The following exit poll was provided by CNN on the US election day in the state of Michigan.

   Link: https://edition.cnn.com/election/2024/exit-polls/michigan/general/president/52

Age
1,791 total respondents

Filtered By: Election Day Voters

	18-29 18%	30-44 39%	45-64 33%	65 or older 10%
Harris	**51%**	42%	37%	45%
Trump	46%	**56%**	**62%**	**54%**

a. How many respondents supported Harris from 18-29 age group in this survey?

b. Based on your 99% confidence interval, is it plausible that more than half of all voters aged 18–29 support Harris?

4. A university researcher aims to determine the percentage of graduate students who believe that the skills and knowledge gained at the university will be useful in their future endeavors. Out of 164 students 130 students provided a positive answer. Compute a 95% confidence interval for the true population proportion (π) of students who find the skills/knowledge to be useful in the future.

5. An economist wants to estimate mean annual income from the first year of work for university graduates who have had the profound wisdom to take a Fundamentals of Statistics module. How many such incomes must be found if she wants to be 98% confident that the sample mean is within \$50 of the true population mean? Assume that a previous study has revealed that for such incomes, σ = \$300.

6. Suppose a president wants to estimate the proportion of the population that supports his current policy toward revisions in the health care system. The president wants the estimate to be within 4% of the true percentage. Assume a 90% level of confidence.

a. The president's political advisors found a similar survey from two years ago that reported that 60% of people supported health care revisions. How large of a sample is required?

b. If there was no previous study, then how large of a sample is required?

7. To compare customer satisfaction levels of two competing cable television companies, 174 customers of Company 1 and 355 customers of Company 2 were randomly selected and were asked to rate their cable companies on a five-point scale, with 1 being least satisfied and 5 most satisfied. The survey results are summarized in the following table:

Company 1	**Company 2**
$n_1 = 174$	$n_2 = 355$
$\overline{x_1} = 3.51$	$\overline{x_2} = 3.24$
$s_1 = 0.51$	$s_2 = 0.52$

Construct a 95% confidence interval for $\mu_1 - \mu_2$ and provide interpretations.

8. The use of cellular phones in automobiles has increased dramatically in the last few years. Of concern to traffic experts, as well as manufacturers of cellular phones, is the effect on accident rates. Is someone who is using a cellular phone more likely to be involved in a traffic accident? What is your conclusion from the following sample information? Use 90% confidence level.

	Had accident in the last year	Did not have an accident in the last year
Uses a cell phone	45	250
Does not use a cell phone	50	450

HOMEWORK

9. Construct the confidence interval for $\mu_1 - \mu_2$ for the level of confidence and the data from independent samples are given.

a. 97% Confidence,

$n_1 = 30$, $\bar{x}_1 = -112$, $s_1 = 9$

$n_2 = 40$, $\bar{x}_2 = -98$, $s_2 = 4$

b. 95% Confidence,

$n_1 = 45$, $\bar{x}_1 = 27$, $s_1 = 2$

$n_2 = 60$, $\bar{x}_2 = 22$, $s_2 = 3$

10. A small private university is planning to start a volunteer football program. A random sample of alumni is surveyed. It was found that 250 were in favor of this program, 75 were opposed, and 25 had no opinion.

a. Estimate the percent of alumni in favor of this program. Use $\alpha = 0.05$

b. Estimate the percent of alumni opposed to this volunteer football program with a 90% confidence level.

11. A personnel manager has found that historically the scores on aptitude tests given to applicants for entry level positions follow a normal distribution with a standard deviation of 32.4 points. A random sample of nine test scores from the current group of applicants had a mean score of 187.9 points.

a. Find an 80% confidence interval for the population mean score of the current group of applicants.

b. Based on these sample results, a statistician found for the population mean a confidence interval extending from 165.8 to 210.0 points. Find the confidence level of this interval.

12. In order to assess the impact of an advertising campaign, a restaurateur monitors her daily revenue before and after the campaign. The table below shows some sample statistics of 5daily sales calculated over a period of 60 days prior to the campaign, and 45 days after the campaign. Determine a 95% confidence interval for the increase in average daily sales due to the campaign. Is there strong evidence that the advertising campaign has increased sales? We do not assume the variances are equal.

	Before campaign	After Campaign
Number of days	60	45
Mean daily sales	$503	$559
Standard deviation	$21	$29

13. In the past decade, intensive antismoking campaigns have been sponsored by both federal and private agencies. Suppose the American Cancer Society randomly sampled 1,500 adults in 2010 and then sampled 1,750 adults in 2020 to determine whether there was evidence that the percentage of smokers had decreased. Provide the 96% confidence interval for the difference in proportions between 2010 and 2020.

Results of Smoking Survey	
2010	2020
$n_1 = 1500$	$n_2 = 1750$
$x_1 = 555$	$x_2 = 578$