MATH 4720 / MSCS 5720

Instructor: Mehdi Maadooliat

Chapter 5 (Part A)

Department of Mathematics, Statistics and Computer Science

CHAPTER 5 (PART A)

- Confidence Interval (CI)
- CI for μ, when σ is known
- Choosing Sample Size for Estimating μ
- A Statistical Test for μ
- Type I and Type II Errors
- Power of a Test
- Choosing Sample Size for Testing μ
- Level of Significance
- P-value
- Inference about μ, when σ is unknown

CONFIDENCE INTERVALS

- In statistics, when we cannot get information from the entire population, we take a sample.
- However, as we have seen before statistics calculated from samples vary from sample to sample.
- When we obtain a statistic from a sample, we do not expect it to be the same as the corresponding parameter.
- It would be desirable to have a range of plausible values which take into account the sampling distribution of the statistic. A range of values which will capture the value of the parameter of interest with some level of confidence.

- This is known as a confidence interval.

CONFIDENCE INTERVALS

- **A confidence interval is for a parameter, not a statistic.**
  - For example, we use the sample mean to form a confidence interval for the population mean.
  - We use the sample proportion to form a confidence interval for the population proportion.
- We never say, “The confidence interval of the sample mean is …”
- We say, “A confidence interval for the true population mean, μ, is …”

MAKING DECISIONS WITH CONFIDENCE INTERVALS:

- If a value is NOT covered by a confidence interval (it's not included in the range), then it's NOT a plausible value for the parameter in question and should be rejected as a plausible value for the population parameter.

- In general, a confidence interval has the form

$$estimate \pm margin\ of\ \ error\ (E)$$

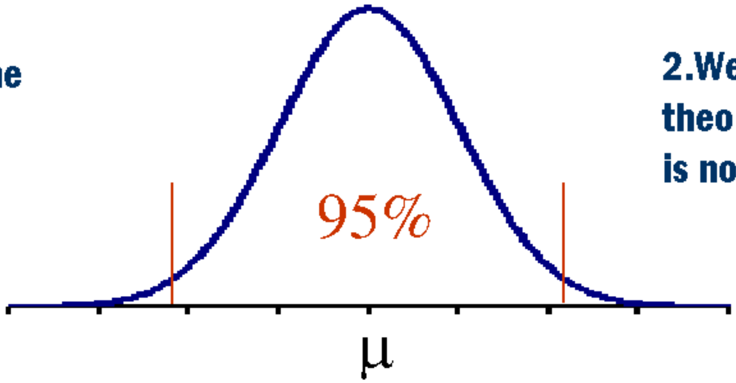
- We can find confidence intervals for any parameter of interest, however in this chapter we are primarily focus on the CI for the Population mean, μ

CONFIDENCE INTERVAL FOR POPULATION MEAN

Here we make use of the sampling distribution of the sample mean in the following way to develop a confidence interval for the population mean, μ, from the sample mean:

1. The unknown true mean of the sampling distribution is μ.

2. We know from the central limit theorem that the sampling distribution is normally distributed.



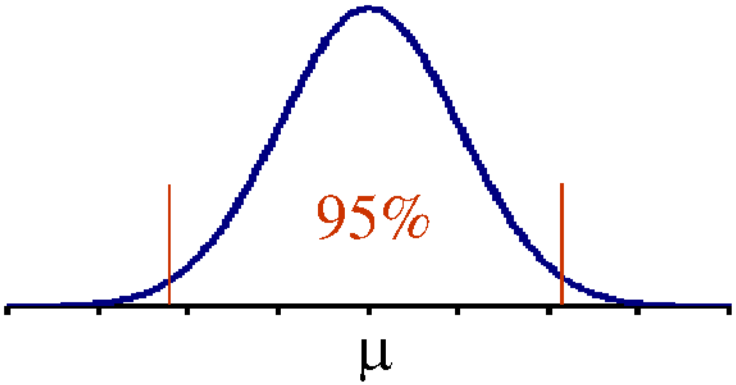
3. We know, from our study of normal distributions, the proportion of the values between two values (for example, two standard deviations).

4. We can thus say that we are 95% confident that a sample mean we find is within this interval.

CONFIDENCE INTERVAL FOR POPULATION MEAN

5. This is the same as saying that if we took many, many samples and found their means, 95% of them would fall within two standard deviations of the true mean.

6. If we took a hundred samples, we would expect that about 95 sample means would be within this interval.

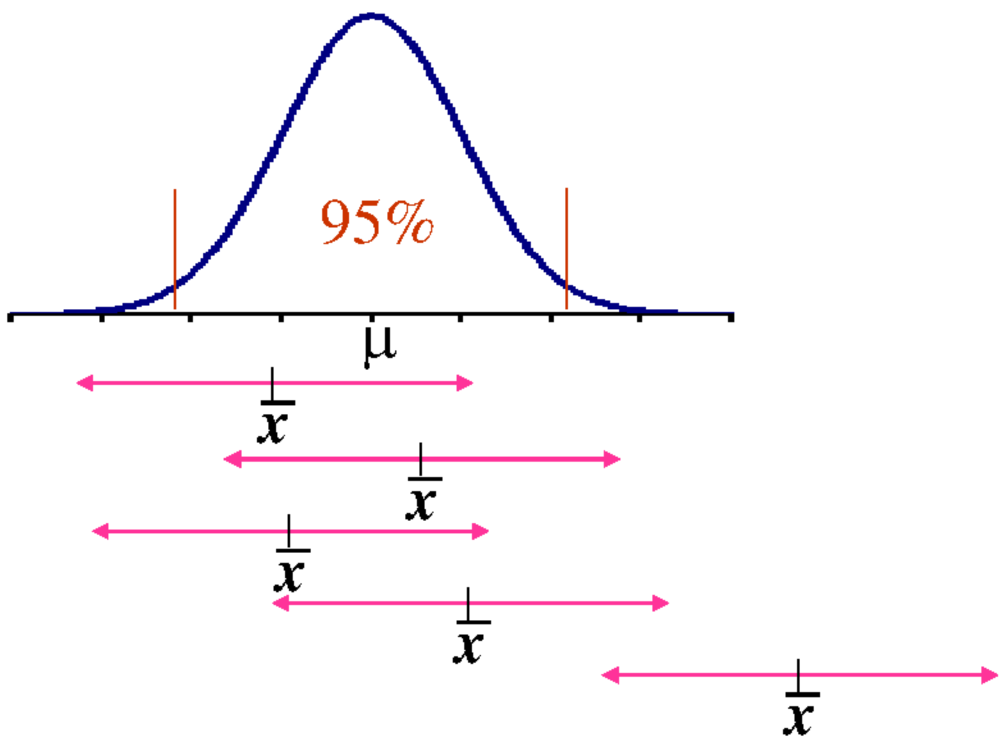


[Confidence Interval Applet]

CONFIDENCE INTERVAL FOR POPULATION MEAN

- In the previous slides we said we were confident that the sample mean was within a certain interval around the population mean.
- When we take a sample, we use the same principle to say we are confident that the true population mean will be in an interval around the sample mean.

That is, saying "we are 95% confident that the sample mean is in the interval around μ" is the same as saying "we are 95% confident that μ is in the interval around the sample mean."



CONFIDENCE INTERVAL FOR POPULATION MEAN

- The width of the confidence interval depends upon the **level of confidence** we wish to achieve.
- The **confidence level** (C) gives the probability that the method we are using will give a correct answer.
- Common confidence levels are $C = 90\%$, $C = 95\%$, and $C = 99\%$. The 95% confidence interval is the most common.
- The level of confidence directly affects the width of the interval.
  - Higher confidence yields wider intervals.
  - Lower confidence yields narrower intervals.
- The formula for a confidence interval for a population mean (when the population standard deviation σ is known) is

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

where $z_{\alpha/2}$ is the value on the standard normal curve with the confidence level between $-z_{\alpha/2}$ and $z_{\alpha/2}$

CONFIDENCE INTERVAL FOR POPULATION MEAN

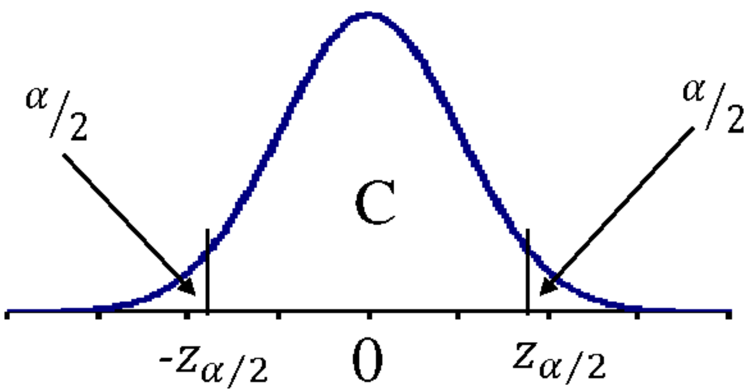
- The $z_{\alpha/2}$ for each of the three most common confidence levels are as follows:

  99%: $z_{\alpha/2} = 2.576$

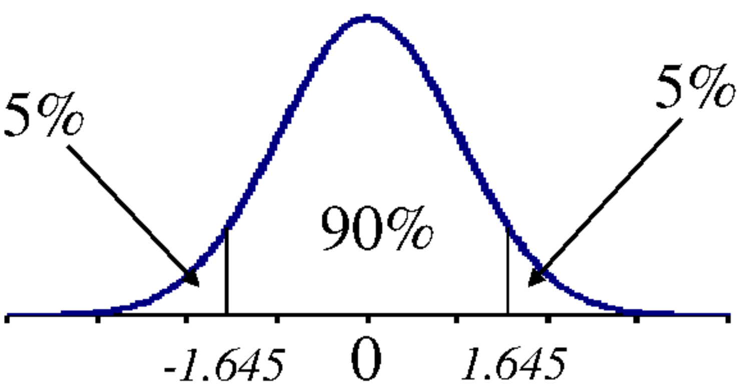
  95%: $z_{\alpha/2} = 1.960$

  90%: $z_{\alpha/2} = 1.645$

- A visual idea of $z_{\alpha/2}$ is



For a 90% confidence interval:



EXAMPLE 1

MARQUETTE UNIVERSITY
Be The Difference.

- Suppose we want to know the average systolic blood pressure of a healthy population of a age group $25 - 35$. Assume that the population distribution is normal with the standard deviation of 5 mm.
- We have a sample of 16 subjects of this population with $\bar{y} = 121.5$

(a) Estimate the average SBP with a 95% confidence interval.

- Formula: $\bar{y} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ (Note, pop. Dist. Is normal)
- $\alpha = 0.05,\ z_{\alpha/2} = 1.96$

EXAMPLE 1

- Suppose we want to know the average systolic blood pressure of a healthy population of a age group $25 - 35$. Assume that the population distribution is normal with the standard deviation of 5 mm.
- We have a sample of 16 subjects of this population with $\bar{y} = 121.5$

(a) Estimate the average SBP with a 95% confidence interval.

- Formula: $\bar{y} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ (Note, pop. Dist. Is normal)
- $\alpha = 0.05, \; z_{\alpha/2} = 1.96$

95% confidence interval of μ : $121.5 \pm 1.96 * \frac{5}{\sqrt{16}} = 121.5 \pm 2.45$

$121.5 - 2.45 \le \mu \le 121.5 + 2.45, \;\; i.e$

$119.050 \le \mu \le 123.950$

margin of error

(b) Estimate the average SBP with 99% CI

- $\alpha = 0.01, \;\; z_{\alpha/2} = invnorm(0.995, 0, 1) = 2.58$

99% confidence interval of μ : $121.5 \pm 2.58 * \frac{5}{\sqrt{16}} = 121.5 \pm 3.225$

$118.275 \le \mu \le 124.725$

Note:
There is a large margin of error in 99% CI estimate compared to the 95% CI estimate.
And the 99% CI is wider than the 95% CI.

EXAMPLE 2

- **Objective is to estimate the mean household income, μ, of Wisconsin households. Suppose the population st.dev. is $\$10{,}000$. If a sample of 100 households yield $\bar{y} = 51{,}500$, estimate μ with a 95% CI, and with $99\% \; CI$.**

- **Formula:**

$\bar{y} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ **(Note, pop. dist. Is not known, but $n \geq 30$)**

EXAMPLE 2

- Objective is to estimate the mean household income, μ, of Wisconsin households. Suppose the population st.dev. is $\$10{,}000$. If a sample of 100 households yield $\bar{y} = 51{,}500$, estimate μ with a 95% CI, and with $99\%\ CI$.

- Formula:

  $\bar{y} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ (Note, pop. dist. Is not known, but $n \geq 30$)

- 95% CI: $51{,}500 \pm 1.96 * \frac{10000}{\sqrt{100}}$

  $z_{\alpha/2} = 1.96$ $51{,}500 \pm \mathbf{1{,}960}$ margin of error

  $\$49{,}540 \leq \mu \leq \$53{,}460$

- 99% CI: $51{,}500 \pm 2.58 * \frac{10000}{\sqrt{100}}$

  $z_{\alpha/2} = 2.58$ $51{,}500 \pm \mathbf{2{,}580}$ margin of error

  $\$48{,}920 \leq \mu \leq \$54{,}080$

EXAMPLE 2 (CONT'D)

- Suppose, the state is planning to levy an average of 6% tax rate. What is the expected tax revenue per household?

- 99% confidence interval of μ: $\$48{,}920 \leq \mu \leq \$54{,}080$

EXAMPLE 2 (CONT'D)

- Suppose, the state is planning to levy an average of 6% tax rate. What is the expected tax revenue per household?

- 99% confidence interval of μ: $\$48{,}920 \leq \mu \leq \$54{,}080$

- Expected tax revenue per household is expected to be between $0.06 * 48{,}920$ and $0.06 * \$54{,}080$, or between

$$\$2{,}935.20 \ and \ \$3{,}244.80.$$

- One might say that this is a big range. In other words, the margin of error is too high.

EXAMPLE 2 (CONT'D)

- The only way, you can reduce the margin of error is to sample more households.

- Sample 400 households. $\bar{y} = \$51{,}125$

- 99% CI of μ: $51{,}125 \pm 2.58 * \frac{10000}{\sqrt{400}}$

$$51{,}125 \pm 1290$$

$$\$49{,}835 \leq \mu \leq \$52{,}415$$

- So, with 6% tax rate, the tax revenue per household is between

$$\$2{,}990 \; and \; \$3{,}145$$

CONFIDENCE INTERVAL FOR POPULATION MEAN

- **Recall that confidence intervals have the form**

$$estimate \pm margin\ of\ error\ (E)$$

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

- **There are three ways to reduce the margin of error (E):**
  - **Reduce σ**
  - **Increase n**
  - **Reduce $z_{\alpha/2}$**
    - $z_{\alpha/2}$ **can only be reduced by changing the confidence level C.**
    - $z_{\alpha/2}$ **is reduced by lowering the confidence level**
      - **Example:** $z_{\alpha/2}$ **for** $C = 95\%$ **is** 1.960 **while** $z_{\alpha/2}$ **for** 90% **is** 1.645.

CHOOSING SAMPLE SIZE FOR ESTIMATING μ

- The most common way to change the margin of error (E) is to change the sample size n.
- To get a desired margin of error (E) by adjusting the sample size n we use the following:

- Determine the desired margin of error (E).
- Use the following formula:

$$n = \left(\frac{z_{\alpha/2}\sigma}{E} \right)^2$$

EXAMPLE 1:

- State tax advisory board wants to estimate the mean household income μ within a margin of error of \$1,000 with 99% confidence. How many households they need to sample? Assume that the population st. dev. is \$10,000.

$$n = \frac{z_{\alpha/2}^2 \sigma^2}{E^2} = \frac{2.58^2 10000^2}{1000^2} = 656.7$$

$$n = 657$$

EXAMPLE 5.4 OF THE BOOK

- **A federal agency wants to investigate the average weight of a cereal box of a particular brand. How many boxes they need to sample to estimate the mean weight μ to within a margin of error Of 0.25 oz with 99% confidence. Assume** $\sigma = 0.75$

- $n = \frac{z_{\alpha/2}^2 \sigma^2}{E^2} = \frac{2.58^2 0.75^2}{0.25^2} = \mathbf{59.91}$

- $n = 60\ boxes.$