Confidence Intervals

CSE 407 - Class 3

LEC RAIYAN RAHMAN
Dept of CSE, MIST
raiyan@cse.mist.ac.bd

Your Company’s is getting Sued! What to do?
We could use the good old Hypothesis Testing to reject (hopefully)

Label the Weight as: (150g ± 2%)

Or We could be smarter!

But how to get that “magic range”?

Let’s Consider a Scenario

You are tasked to find the average price of a particular ceiling fan, after a market analysis.

How to do that?

1. You go and visit EVERY outlet in dhaka city that sells that fan.
2. You add up their price and make an average.
3. But, that's not the best way, because -
   a. Not realistically possible to visit every shop, specially in Dhaka traffic!
   b. There could be always that pone shop in Gulshan that sells in for triple the price.
   c. There could be one shop that sells it for half the avg price.
   d. And, an average doesn't really help. **Why….?**

4. Rather, if we could say that, I'm pretty sure if you go to buy this fan, take 4500-5500 tk with you, you'll be fine.

   OR, **statistically,** you could say that I'm **95% certain that the price is between 4500-5500** in Dhaka city.

And in Statistics, this is called
“confidence intervals”

Confidence Intervals: Defn

- A confidence interval, in statistics, refers to the probability that a population parameter will fall between a set of values for a certain proportion of times.

- Confidence intervals measure the degree of uncertainty or certainty in a sampling method. They can take any number of probability limits, with the most common being a 95% or 99% confidence level.

Confidence Intervals: Defn

- Confidence levels are expressed as a percentage (for example, a 95% confidence level).
- It means that should you repeat an experiment or survey over and over again, 95 percent of the time your results will match the results you get from a population (in other words, your statistics would be sound!).
- For example, you survey the shops in dhaka (the prev example) to see how the price of a fan varies based on location (and you also find the avg/confidence range).

Confidence Intervals: Some Notes

1. The higher shops you visit (the higher the sample size), the more precise (or smaller) you can be with you statistical range (the magic range!).

   So n ↑ , range ↓

2. Inversely, the higher level of confidence you want in your range, the wiser it is to give a wider range.

   So C ↑ , range ↑

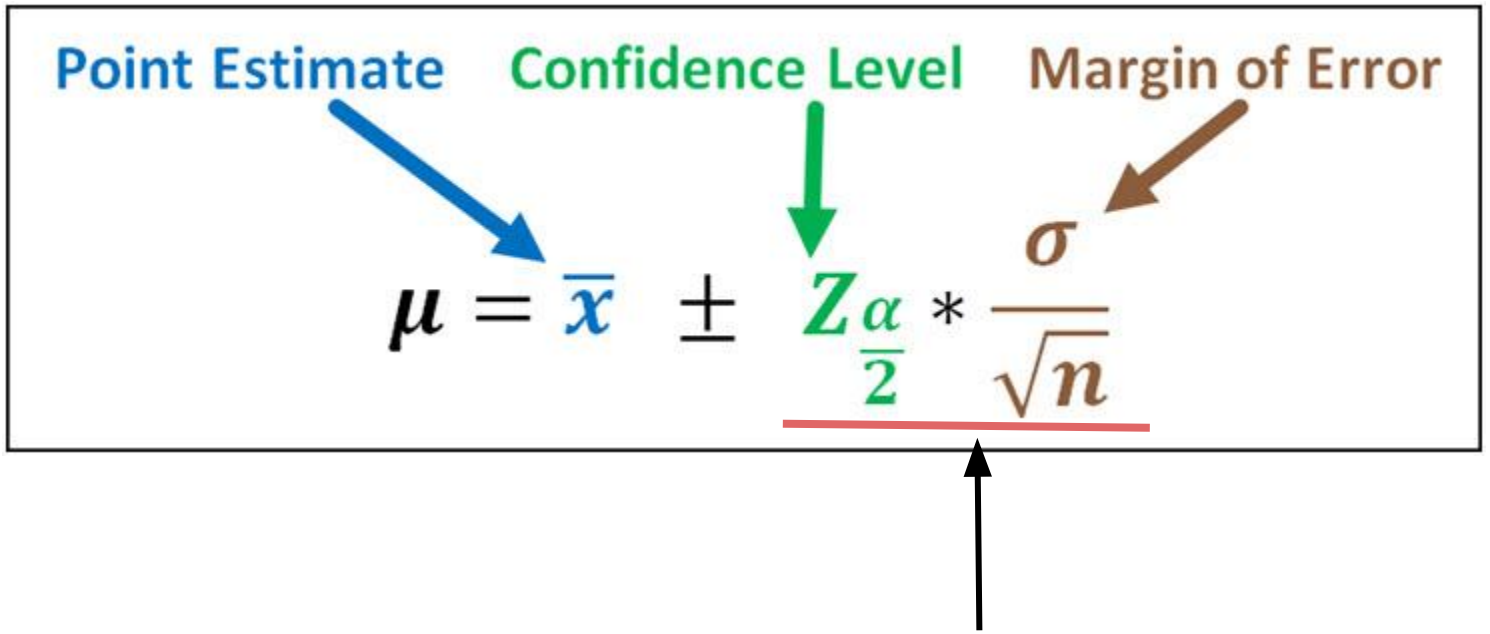
How to Calculate?

Decision Tree for selecting What Formula to use:

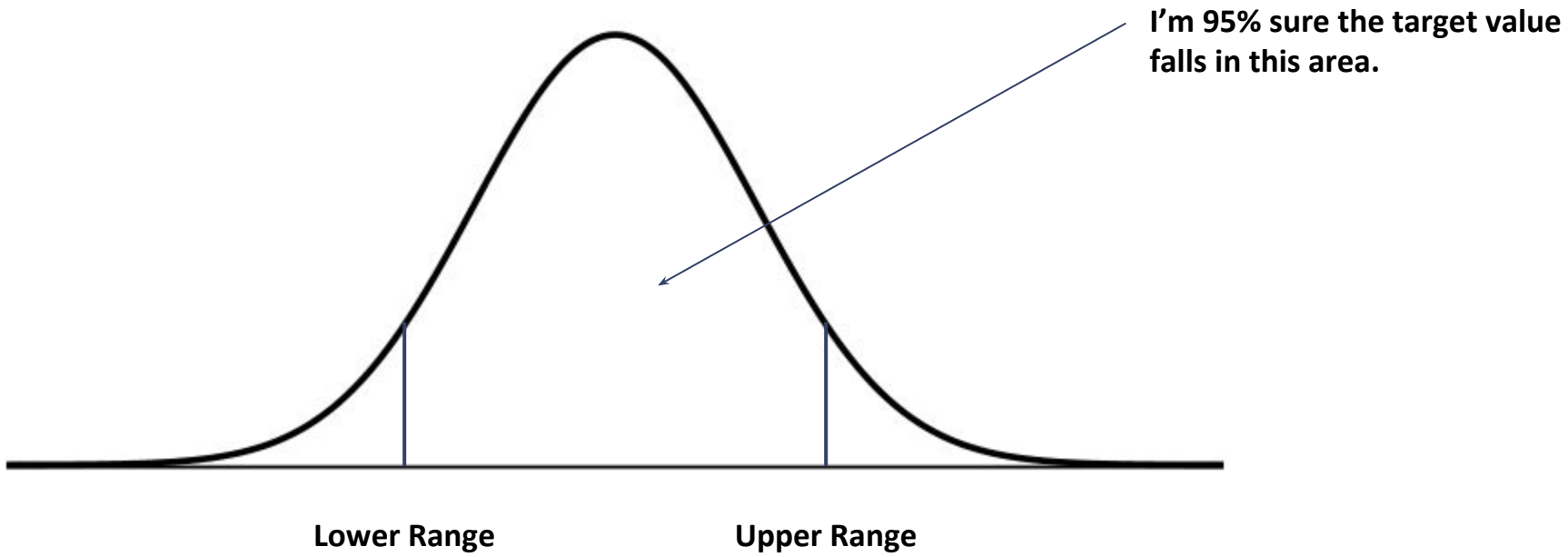
For Mean

with known variance & n>= 30	Use Z Distn Confidence Interval	$C.I. = \bar{X} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$
with unknown variance or n<30	Use t Dstn Confidence Interval	$C.I. = \bar{X} \pm t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$
For Variance	Use the χ^2 Distn Confidence Interval	$C.I. = \frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}$

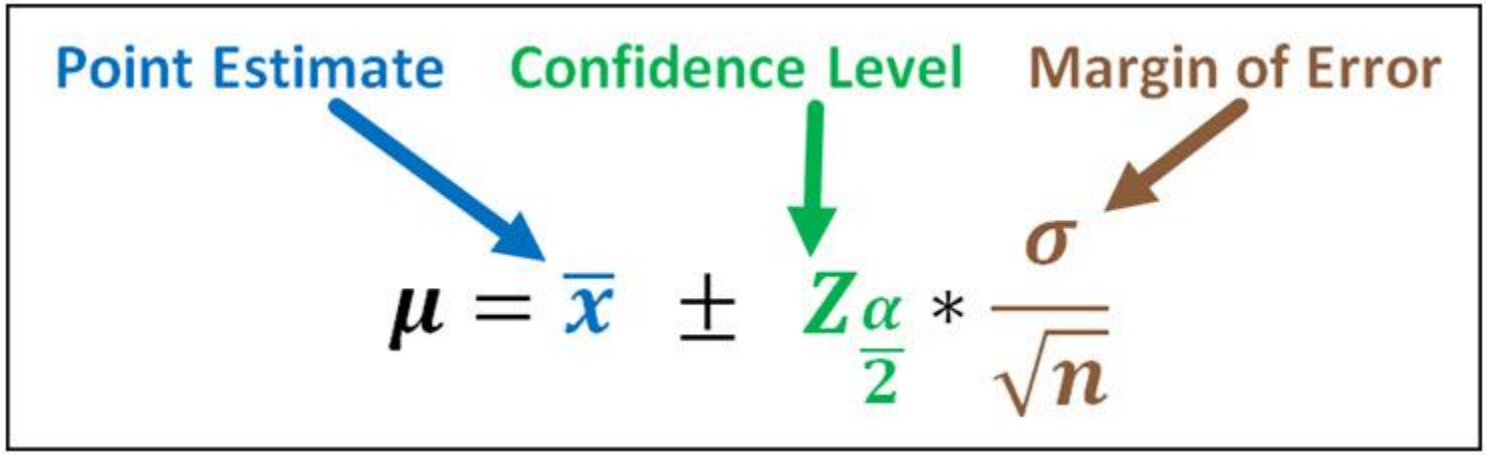
How do these formulae work?



What do these mean?



Relation between n, C/α and the interval



1. The higher the value of n, the narrower the C.I (confidence interval) i'll get. Form eqn, we see for **a quadruple sample size, we can half the width of the range.**

2. The more confident we want to be the wider the range we need.

Example Math 1

Suppose as a business analyst, you were tasked to find out the average price of a ceiling fan that you competitor recently launched.

So, you went to 5 sample shops from varied localities in Dhaka city. You assumed that the price would be normally distributed with mean μ and variance σ^2 unknown.

Now, you in the 5 shops you visited the price were 1500, 1600, 1650, 1550, 2000.

Now as the data analyst of the company, compute a 95% confidence interval for the price of your competitor's new fan. so that your company can market a even cheaper fan!

Also, if the company asked you to be 98% certain that the price you calculated is correct within +- 150, how large of a sample size would be needed?

Steps

1. Find out from the scenario which formula to use.
2. Then write down and calculate (if needed) the necessary value for the data set.
3. Find the critical value from given confidence interval. Use the z/t or chi-sq table appropriately.
4. Now use all the data to find the confidence interval.

Solution to Example 1

n = 5

$\bar{X}$ **= 1660**

S = 198.11

α = 1 - C.

Here, C = 0.95. So, α = 1 - 0.95 = 0.05. α/2 = 0.025.

So, $t_{.025,\,4} = 2.776$

Solution to Example 1

We'll use the second formula.

Putting values, we get (1414.04, 1905.95)

So, we're 95% certain that the price of these fans in retail is between, lets say, 1415 tk to 1906 tk.

Exercice Math 2

Suppose as a business analyst, you were tasked to find out the average price of a ceiling fan that you competitor recently launched.

So, you went to 6 sample shops from varied localities in Dhaka city. You assumed that the price would be normally distributed with mean μ and variance σ^2, which is unknown. Now, you in the 5 shops you visited the price were 1500, 1400, 1550, 1750, 1900, 2000.

Now as the data analyst of the company, compute a 90% confidence interval for the price of your competitor's new fan. so that your company can market a even cheaper fan!

Also, if the company asked you to be 97% certain that the price you calculated is correct within 200, how large of a sample size would be needed? Consider population std deviation to be 140 in this case.

Exercice Math 3

You work as the quality control engineer in the automation chain for a chip manufacturing company. Your company introduced a new automated chip manufacturing plant that can produce 10,000 chips an hour! However, the error rate of these chips suffers a bit due to this high speed.

You were tasked to measure the variation of error rates of these new chips to determine whether the new plant was worth it. To do so, you took 8 chips from the manufacturing line over 8 hours, one from a new batch of chips every hour. The error rate of these chips were 1.2, 2.4, 2.4, 1.1, 1.9, 2.9, 4.5, 5.6.

Now compute a 95% confidence interval of the variance of the error rates in these new chips.

Note: Use the formula for variance and use the chi square table in this case.

Thank You!