Unit IV

Topics covered:

Role of Statistics -Estimation of Parameter and Sampling Distribution: Point Estimation - Sampling Distributions and the Central Limit Theorem. Statistical Intervals for a Single Sample: Confidence Interval on Mean

A. Role of statistics

Statistics Terminologies for Data Science:

Term	Definition	Example
Population	The entire group of interest in a study.	All patients in a hospital.
Sample	A subset selected from the population.	100 randomly selected patients.
Mean	The average of values in a dataset.	Average age of all patients.
Sample Mean ($\bar{X}$)	The mean of a sample, used to estimate the population mean.	Mean age of 100 selected patients.
Population Mean (μ)	The mean of all values in the entire population.	Average age of all patients in the hospital.
Sampling Distribution	The distribution of a statistic (e.g., sample mean) over many samples.	Distribution of mean ages from multiple samples of patients.

A . I. Terminologies:

1. Population Mean (μ:

The population mean represents the average of all values in a population.

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N}$$

- Xi: each individual value in the population.
- N: total number of values in the population.

2. Population Variance (σ^2):

The population variance measures how much the values in the population vary around the mean. (that is the average of the squared differences from the mean.)

$$\sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}$$

- μ: population mean.
- Xi: each individual value in the population.

3. Population Standard Deviation (σ):

The standard deviation is the square root of the variance.

$$\sigma = \sqrt{\sigma^2}$$

4. Sample Mean ($\bar{X}$):

The sample mean represents the average of values in a sample from the population.

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

- Xi: each individual value in the sample.
- n: total number of values in the sample.

5. Sample Variance (s^2):

The sample variance measures the spread of values in a sample. i.e. Variance measures **how spread out the data is** around the mean.

$$s^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}$$

- $\bar{X}$ is the sample mean.
- Xi: each individual value in the sample.
- n−1: **Bessel's correction**, used to correct the bias in the estimation of the population variance from a sample.

6. Sample Standard Deviation (s):

The standard deviation of the sample is the square root of the sample variance.

$$s = \sqrt{s^2}$$

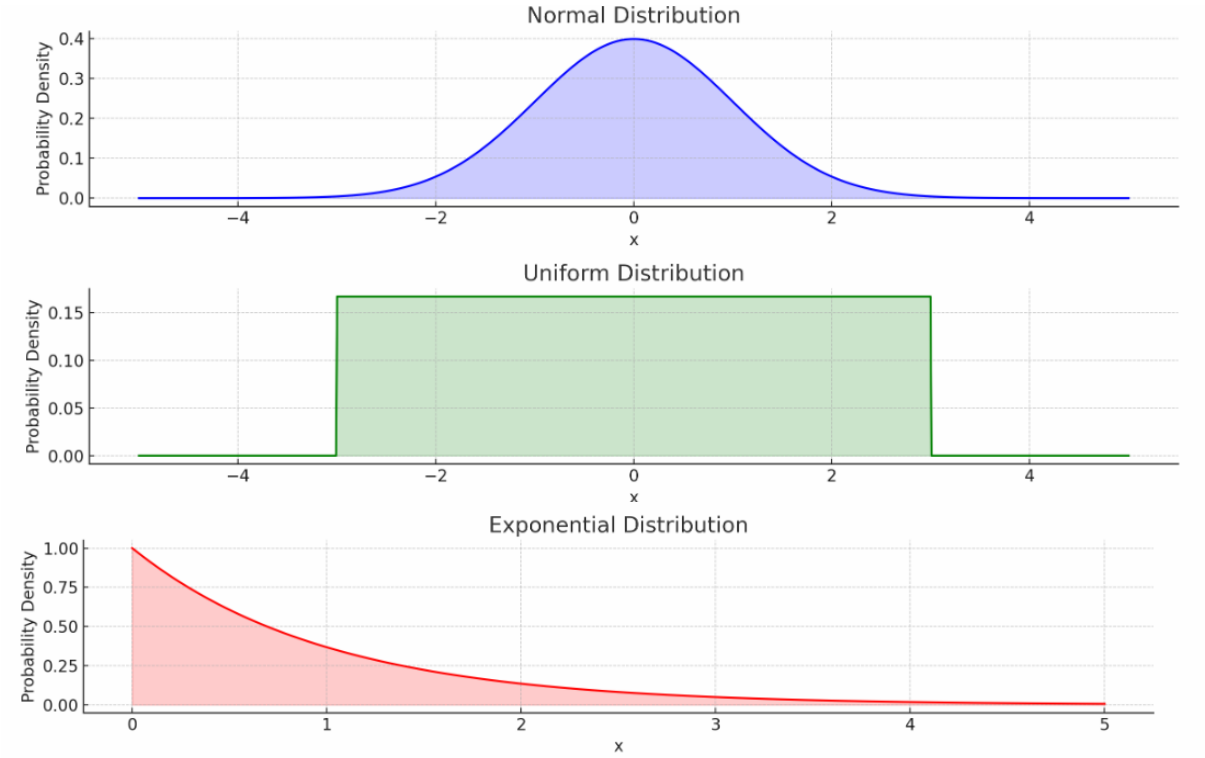
A . II. Probability Distributions:

List of distributions :

a. Normal (Gaussian) Distribution
b. Uniform Distribution
c. Exponential Distribution
d. Binomial Distribution
e. Poisson Distribution
f. Beta Distribution
g. Gamma Distribution
h. Geometric Distribution

i. Log-Normal Distribution

Illustrations :



We will use Normal and Uniform Distributions.

a. Normal Distribution :

to calculate Mean, Variance, and Standard Deviation for a Normal Distribution, formulas are like those for general populations: (**Exam scores in a large class** → if the test is fairly designed, most students score near the average, with a few high/low extremes.)

1. **Mean**:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

2. **Variance**:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} (x_i - \mu)^2$$

3. **Standard Deviation**:

$$\sigma = \sqrt{\sigma^2}$$

Example :

Consider the following population of 5 data points:

$$\text{Population} = \{2, 4, 4, 4, 6\}$$

1. Calculate the Mean (μ):

$$\mu = \frac{1}{5}(2 + 4 + 4 + 4 + 6) = \frac{20}{5} = 4$$

2. Calculate the Variance (σ^2):

First, compute the squared differences from the mean:

$$(x_i - \mu)^2 = (2-4)^2, (4-4)^2, (4-4)^2, (4-4)^2, (6-4)^2$$

$$= 4, 0, 0, 0, 4$$

Now, find the variance:

$$\sigma^2 = \frac{1}{5}(4 + 0 + 0 + 0 + 4) = \frac{8}{5} = 1.6$$

3. Calculate the Standard Deviation (σ):

$$\sigma = \sqrt{1.6} \approx 1.26$$

b. Uniform Distribution : To calculate Mean, Variance, and Standard Deviation for a uniform Distribution . (Rolling a fair dice→ each number (1 to 6) has equal probability (1/6).)

For a uniform distribution (where all values between a and b are equally likely):

- Mean: $\mu = \frac{a+b}{2}$
- Variance: $\sigma^2 = \frac{(b-a)^2}{12}$
- Standard Deviation: $\sigma = \sqrt{\frac{(b-a)^2}{12}} = \frac{b-a}{\sqrt{12}}$

Example :

Suppose the data is uniformly distributed between a=1and b=5, we can calculate as follows :

- Mean: $\mu = \frac{1+5}{2} = 3$
- Variance: $\sigma^2 = \frac{(5-1)^2}{12} = \frac{16}{12} = 1.333$
- Standard Deviation: $\sigma = \sqrt{1.333} \approx 1.154$

B. **Point Estimation** : **statistical methods** refer to the various techniques and procedures used to collect, analyze, interpret, and present data to makedecisions and draw conclusions about a population based on a sample.

These methods fall into two broad categories:

a. Descriptive Statistics

- **Definition**: These methods summarize and describe the characteristics of a dataset.
- **Examples**:
    - **Measures of Central Tendency**: Mean, median, mode (to describe the center of a data distribution).
    - **Measures of Dispersion**: Range, variance, standard deviation (to describe the spread of data).
    - **Graphical Representations**: Histograms, bar charts, pie charts, and box plots (to visualize data).

b. Inferential Statistics

- **Definition**: These methods use data from a sample to make inferences or predictions about a population. The main goal is to conclude the immediate data.
- **Examples**:
    - **Estimation**: Point estimation (like estimating the population mean) and confidence intervals (to provide a range of values within which the population parameter is likely to fall).
    - **Hypothesis Testing**: Methods like t-tests, chi-square tests, ANOVA, and regression analysis to determine if observed effects in the sample are statistically significant and can be generalized to the population.

2.1 The Parameters in statistical inference include population mean, proportion, variance, standard deviation, median, and regression coefficients. Statistical inference always focuses on drawing conclusions about one or more parameters ofa population.

Point estimation refers to the process of using sample data to calculate a single value (called a **point estimate**) as an approximation of an unknown population parameter.

Example 1: Point Estimation of the Population Mean (μ)

Definition: The sample mean Ẍ is used as a point estimate for the population mean (μ).

Formula:

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

.

Where:

- n = the number of observations in the sample
- x_i = each observation

Example: Suppose a sample of 5 students' scores in a test is: 80, 85, 90, 95, and 100.
The sample mean is:

$$\bar{x} = \frac{80 + 85 + 90 + 95 + 100}{5} = \frac{450}{5} = 90$$

Thus, the point estimate for the population mean score is 90.

Before the data are collected, the observations are considered to be random variables,say, X_1 ,X_2 ,X_n … . Therefore, any function of the observation, or any **statistic**, is alsoa random variable. For example, the sample mean and the sample variance are statisticsand random variables.

Because a statistic is a random variable, it has a probability distribution. The probabilitydistribution of a statistic is called as **sampling distribution**.

The symbol θ(theta) represents the parameter.The symbol θcan represent the mean μ, the variance σ^2, or any parameter. Theobjective of point estimation is to select a single number based on sample data that is the mostplausible value for θ. A numerical value of a sample statistic will be used as the point estimate.

A **point estimate** of some population parameter θis a single numerical $\hat{\theta}$ value of a $\hat{\Theta}$. statistic $\hat{\Theta}$.

The statistic(a formula or rule) is called the **point estimator**.

Example 1: Estimating Population Mean

- Suppose we want to know the **average height of all students in a university** (population mean, θ).
- Measuring everyone is impossible, so we take a **sample of 50 students**.
- From the sample, we calculate the **sample mean** $\hat{\theta}$

If the sample mean = **165 cm**, then:

- **Sample mean ($\bar{x}$)** is the **point estimator**.
- **165 cm** is the **point estimate** of the true population mean.

Example for Clarity:

Imagine you want to estimate the **mean height (population parameter)** (θ) of students in a school:

Step 1: Define the Point $\hat{\Theta}$. Estimator

A common way to estimate the **mean** of a population is to use the **sample mean**.

Let's say you collect a **random sample** of students' heights X1,X2,X3,X4,X5

The **sample mean** is calculated as,

$$\hat{\Theta} = \bar{X} = \frac{X_1 + X_2 + X_3 + X_4 + X_5}{5}$$

Here, $\hat{\Theta}$. is the point estimator because it is the **rule or formula** for how to calculate the average height

using a sample.

Step 2: Calculate a Point $\hat{\theta}$ Estimate

Suppose you collect a sample of **5 students' heights**: 150 cm, 155 cm, 160 cm, 165 cm, and 170 cm.

Using the formula of $\hat{\Theta}$.

$$\hat{\Theta} = \bar{X} = \frac{150 + 155 + 160 + 165 + 170}{5} = 160\,\text{cm}$$

160 cm is the **point $\hat{\theta}$ estimate** of the population mean height (θ).

Note:

$\hat{\Theta}$. is a **random variable** because the heights X1,X2,X3,X4,,X5 are random. Different samples will have different heights, and thus, different **sample means**.

For instance, if we take another random sample of 5 different students, the **average height** (the value of) could be $\hat{\Theta}$. different.

Point Estimator is the method (e.g., for estimating μ).

Point Estimate is the result of applying that method to a specific set of sample data (e.g., $\bar{X}$=160 cm).

C. Sampling Distributions and the Central Limit Theorem

Because a statistic is a random variable, it has a probability distribution. The probabilitydistribution of a statistic is called **sampling distribution**.

Each numerical value in the data is the observed value of a random variable. Furthermore, the random variables are usually assumed to be independent and identically distributed. These random variables are known as a random sample.

C.1 Random Sample :

The random variables $X_1, X_2 \ldots X_n$ are a random sample of size n if

(a) the X_i's areindependent random variables and

(b) every X_i has the same probability distribution.

The observed data are also referred to as a random sample.

The purpose of taking a random sample is to obtain information about the unknown population parameters.

C.2 Statistics

A statistic is any function of the observations in a random sample.

This means that a statistic is a numerical value calculated from a random sample of data points. It is used to summarize or infer information about a larger population. A statistic can be a simple function like the mean, median, variance, or more complex functions.

Example : The primary purpose of taking a random sample is to obtain information about the unknown population parameters. Suppose, for example, that we wish to reach a conclusion about theproportion of people in the United States who prefer a particular brand of soft drink. Let p represent the unknown value of this proportion. It is impractical to question every individual in thepopulation to determine the true value of p. To make an inference regarding the true proportionp, a more reasonable procedure would be to select a random sample (of an appropriate size)and use the observed proportion $\hat{p}$ of people in this sample favoring the brand of soft drink.The sample proportion, $\hat{p}$, is computed by dividing the number of individuals in the sample who prefer the brand of soft drink by the total sample size n. Thus, $\hat{p}$ is a function of theobserved values in the random sample. Because many random samples are possible from apopulation, the value of $\hat{p}$ will vary from sample to sample. That is, $\hat{p}$ is a random variable.Such a random variable is called a **statistic.**
For example, if $X_1, X_2, \ldots X_n$ is a random sample of size n, the sample mean $\bar{X}$, the sample variance S^2 , and the sample standard deviation S are statistics. Because a statistic is a random variable, it has a probability distribution.

C.3 Sampling Distribution
The probability distribution of a statistic is called a sampling distribution.
Example :

- A **population** has all the data (e.g., heights of all students in a university).
- A **sample** is a smaller group from that population (e.g., 50 students).
- A **statistic** is something you calculate from that sample (e.g., the sample mean height).
- If you take **many different samples** and compute the statistic each time, the pattern of those values forms a **sampling distribution**.

Example: The probability distribution of $\bar{X}$ is called the sampling distribution of the mean. The sampling distribution of a statistic depends on the distribution of the population, the size of the sample, and the method of sample selection.
How to calculate the sampling distribution of the sample mean $\bar{X}$:
Samplemean is calculated as :

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

The mean of the sampling distribution of the sample mean is equal to the population mean (μ)
This means, on average, the sample mean is an unbiased estimator of the population mean.

$$\mu_{\bar{X}} = \frac{\mu + \mu + \cdots + \mu}{n} = \mu$$

and variance (The variance of the sampling distribution is the **population variance divided by sample size**)

$$\sigma^2_{\bar{X}} = \frac{\sigma^2 + \sigma^2 + \cdots + \sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Note : Larger samples → smaller variance → sample means are closer to the population mean.
Smaller samples → larger variance → more spread out

Standard deviation of sampling distribution is called the **Standard Error (SE)**:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Note :
The **sampling distribution of the sample mean** shows how sample means vary if we take many random samples.
Its **average** is the true population mean (μ).
Its **spread (variance)** gets smaller as the sample size increases.
So, bigger samples give more **accurate estimates** of the population mean.

Explanation :

$$\mu_{\bar{X}} = \frac{\mu + \mu + \cdots + \mu}{n} = \mu$$

where μ_x refers**mean of the sampling distribution** of the sample mean $\bar{X}$

In the above, each μ represents that when taking a sample of size n from a population where each observation has a mean μ and the **mean of the sample mean** $\bar{X}$ is equal to the **population mean** (μ). That is each μ is the sample mean of a sample population. We generally take many sample populations and take average to find the sample mean. It means that even though each sample mean $\bar{X}$ might vary from one sample to another, the **average of all these sample means** would be equal to μ.

Example:

- Suppose we have a **population mean** μ=165 cm (average height of all students in a class).
- We take **random samples** of size n=10 from this population multiple times.
- For each sample, we calculate the **sample mean** $\bar{X}$
- Let's say the first sample of 10 students gives a mean of **162 cm**, the second gives **167 cm**, the third gives **163 cm**, and so on.
- If we were to continue this process and take an **infinite number of samples**, the **average of all these sample means** would equal the population mean μ=165 cm.
- Sampling distribution of sample mean ($\bar{X}$) is normal with meanμ_x=165 cm and variance of σ_x^2. (which is σ^2/n). From this variance, the standard deviation is $\sqrt{(\sigma_x^2)}$ that is (σ/), which is also known as standard error.

Simple example to calculate sampling distribution of sample mean :

Population

Consider the population: {2, 4, 6, 8}

Population size (N) = 4

Population mean (μ): (2+4+6+8)/4 = 20/4 = 5

Population variance (σ^2): $((2-5)^2 + (4-5)^2 + (6-5)^2 + (8-5)^2)/4 = (9+1+1+9)/4 = 20/4 = 5$

Step 1: Possible Samples of Size n=2

Sample (2, 4) → mean = 3

Sample (2, 6) → mean = 4

Sample (2, 8) → mean = 5

Sample (4, 6) → mean = 5

Sample (4, 8) → mean = 6

Sample (6, 8) → mean = 7

Step 2: Sampling Distribution of Sample Means

Sample means = {3, 4, 5, 5, 6, 7}

Mean of sample means ($\mu_{\bar{x}}$): (3+4+5+5+6+7)/6 = 30/6 = 5

~~This shows that the mean of the sample means equals the population mean (μ = 5)~~

Example : To calculate the sampling distribution of the sample mean, take a numerical example. When the population is normal, the sampling distribution of the sample mean is also normal, with mean μ (same as the population mean) and standard deviation $\sigma/\sqrt{n}$ (called the standard error of the mean), where σ is the population standard deviation and n is the sample size.

Given:
Population mean, μ=50
Population standard deviation, σ=10
Sample size, n=25
Find the sampling distribution of the sample mean.

1. **Mean of the Sampling Distribution of the Sample Mean ($\mu_{\bar{X}}$):**
   - The mean of the sampling distribution of the sample mean is the same as the population mean.

$$\mu_{\bar{X}} = \mu = 50$$

2. **Standard Error of the Mean ($\sigma_{\bar{X}}$):**
   - The standard deviation of the sampling distribution of the sample mean, also called the standard error, is calculated as:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{25}} = \frac{10}{5} = 2$$

3. **Resulting Sampling Distribution of the Sample Mean:** Since the population distribution is normal, the sampling distribution of the sample mean will also be normal, with the following parameters:
   - Mean = $\mu_{\bar{X}} = 50$
   - Standard deviation (standard error) = $\sigma_{\bar{X}} = 2$

C.4. Central limit theorem :

Generally, the probability distribution of a population will be normal distribution. Even if we are sampling from a population that has an unknown probability distribution, the sampling distribution of the sample mean will still be approximately normal with mean μ and variance σ^2 if the sample size n is large.

The **Central limit theorem is defined as follows :**

If $X_1, X_2, \ldots, X_n$ is a random sample of size n taken from a population (either finite or infinite) with mean μ and finite variance σ^2 and if $\bar{X}$ is the sample mean, the limiting form of the distribution of

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

as $n \to \infty$, is the standard normal distribution.

In the above equation, see (σ/) which is the standard error.

Z represents the **standardized version** of the sample mean $\bar{X}$. . It is a **standard normal variable**, which means it follows a **standard normal distribution** with a mean of 0 and a standard deviation of 1. When we use the above formula, we can then refer to the standard normal table (Z-table) to find probabilities or percentiles related to $\bar{X}$.

How variance is 1?

See below

🔹 Step 1: Recall variance scaling property

If $Y = aX$, then

$$Var(Y) = a^2 Var(X)$$

So when you multiply/divide a random variable by a constant, its variance scales by the **square** of that constant.

🔹 Step 2: Variance of $\bar{X}$ (sample mean)

From sampling theory:

$$Var(\bar{X}) = \frac{\sigma^2}{n}$$

This makes sense because bigger samples vary less than smaller ones.

🔹 Step 3: Variance of Z

Now,

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

The denominator $\sigma/\sqrt{n}$ is a **constant** (not random).

So,

$$Var(Z) = Var\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right)$$

Since subtracting μ (a constant) doesn't affect variance:

$$Var(Z) = Var\left(\frac{\bar{X}}{\sigma/\sqrt{n}}\right)$$

Now apply the scaling rule ($a = 1/(\sigma/\sqrt{n})$):

$$Var(Z) = \left(\frac{1}{\sigma/\sqrt{n}}\right)^2 Var(\bar{X})$$

- **Step 4: Substitute $Var(\bar{X})$**

$$Var(Z) = \left(\frac{\sqrt{n}}{\sigma}\right)^2 \cdot \frac{\sigma^2}{n}$$

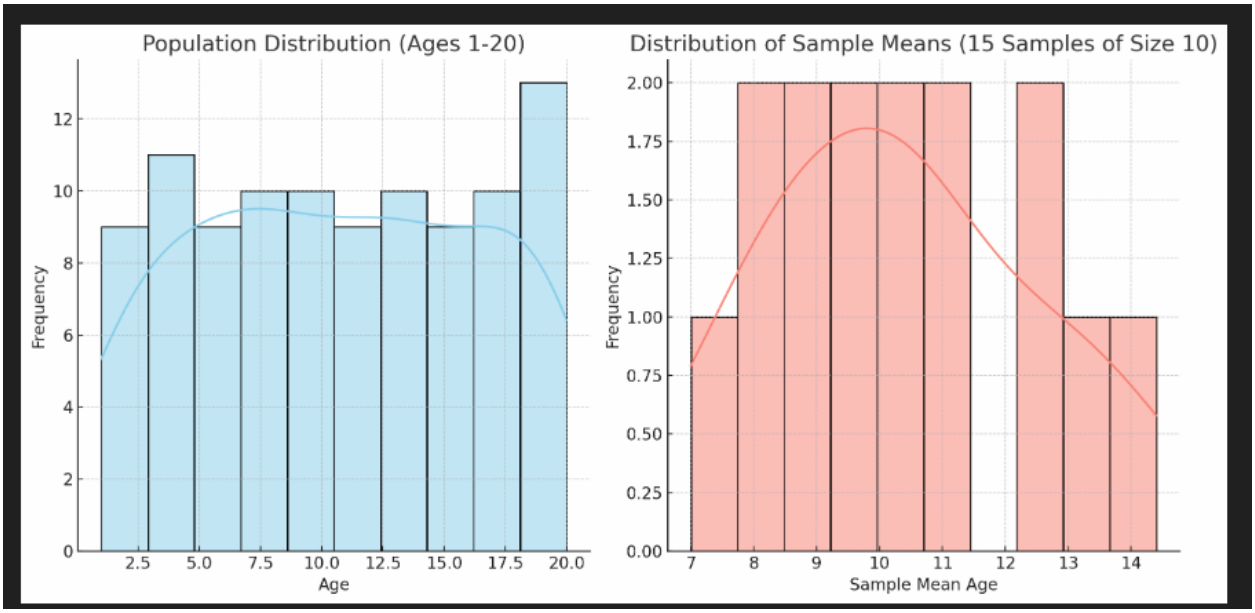
Simplify:

$$Var(Z) = \frac{n}{\sigma^2} \cdot \frac{\sigma^2}{n} = 1$$

Note:

- If the population is normally distributed, the sampling distribution of the statistic (e.g., sample mean) will also be normal. For large enough sample sizes, even if the population is not normal, the sampling distribution of the mean tends to be normally distributed due to the **Central Limit Theorem (CLT).**
- If the underlying distribution is symmetric and unimodal (not too far from normal), the central limit theorem will apply for small values of *n*, say 4 or 5. If the sampled population is very non-normal, larger samples will be required. As a general guideline, if *n* > 30, the central limit theorem will almost always apply.

Numerical Example for CLT 1:
Consider a population of 100 ages ranging from 1 to 20, representing students attending school and college. This population is not normally distributed. Now, draw 15 samples, each containing 10 ages, and calculate the mean for each sample. The resulting distribution of these sample means will tend to be normal, demonstrating the Central Limit Theorem.
The following visualization demonstrates the Central Limit Theorem with a population of 100 students' ages (ranging from 1 to 20):



The **left plot** shows the distribution of the ages in the population, which is not normal due to the ages being uniformly distributed between 1 and 20.
The **right plot** displays the distribution of the means of 15 samples, each containing 10 randomly selected ages. Despite the non-normality of the original age distribution, the distribution of these sample means approximates a normal shape.
This example confirms that the sampling distribution of the mean tends to be normal even when the original population distribution is not normal, as suggested by the CLT.

Example for the CLT theorem 2:
An electronics company manufactures resistors that have a mean resistance of 100 ohms and a standard deviation of 10 ohms. The distribution of resistance is normal.
a Find the probability that a random sample of n = 25 resistors will have an average (mean) resistance of fewer(lesser) than 95 ohms.
b. Also find the probability that a random sample of n = 25 resistors will have an average (mean) resistance of more than 105 ohms

Solution :

- Population size – NOT Given, Population mean (μ) – 100 ohms, Population standard deviation (σ) – 10 and distribution of resistance is NORMAL
- From CLT, we understand that sampling distribution of sample mean is normal with mean μ_x=100 ohms and standard deviation σ_x =σ/
- Sample Size (n) = 25, sample mean Ẍ=95 ohms

a. To find probability that a random sample of n = 25 resistors will have an average (mean) resistance of fewer(lesser) than 95 ohms,

And therefore
Use a **Z-table** or a calculator to find the probability corresponding to Z=−2.500.

Z-Table (Negative Z-Scores)

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084

P(Ẍ<95)=P(Z<-2.500)=0.0062

The Z-table is typically divided into rows and columns where the rows represent the integer and first decimal (e.g., -2.5), and the columns represent the second decimal place (e.g., 0.00).

Finding P(Z<-2.500)
Step 1: Use the Z-table to look up **-2.5** in the **row** and **0.00** in the **column**.
Step 3: The table value at the intersection of -2.5 **row** and **0.00 column** gives **0.0062**
This means P(Z<-2.5)=0.0062.
So, there is an **0.62% probability** that a standard normal variable is less than -2.5.

Hint to find the probability:
P(Z<z) → use Z table value directly
P(Z>z) → 1-P(Z<=z) subtract the Z table value from 1
P(z1<=Z<=z2) → subtract the Z table value of z1 from z2

b. To find probability that a random sample of n = 25 resistors will have an average (mean) resistance of more than 105 ohms,

Now, we need to find the probability that the z-score is greater than 2.5. Using the standard normal distribution table (z table), we find the probability corresponding to z=2.5
The area to the left of z=2.5 is approximately 0.9938. Therefore, the probability of the z-score being greater than 2.5 is:

P(Ẍ> 105)=1- P(Z<=z) =**1-P(Z<=2.500)=1-0.9938=0.0062 (by** P(Z>=z) → subtract the Z table value from 1)

Hence, sample of resistors with a sample mean less than 95 ohms or more than 105 ohms is a **rare event**. If this happens, it casts doubt as to whether the true mean is really 100 ohms or if the true standard deviation is really 10 ohms.

Example 3 :
Suppose that a random variable *X* has a continuous uniform distribution

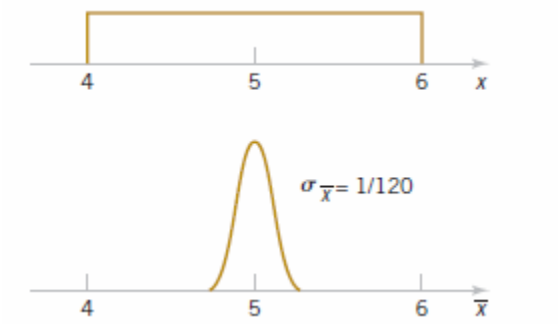
$$f(x)=\begin{cases}1/2, & 4\le x\le 6\\ 0, & \text{otherwise}\end{cases}$$

Find the distribution of the sample mean of a random sample of size $n = 40$.
Solution :

For uniform distribution,

- Mean: $\mu = \frac{a+b}{2}$
- Variance: $\sigma^2 = \frac{(b-a)^2}{12}$
- Standard Deviation: $\sigma = \sqrt{\frac{(b-a)^2}{12}} = \frac{b-a}{\sqrt{12}}$

The mean and variance of *X* are μ =(4+6)/2= 5 and σ^2 = (6-4)^2/12= 4/12=1/3.
The central limit theorem indicates that thedistribution of $\ddot{X}$is approximately normal with mean μ_x = 5 and variance σ_{x^2} = σ^2 / n =(1/3)/40 =1/120.
The distribution of the *$\ddot{X}$ and X is shown below:*



CLT with two independent populations:
Fundamental property in probability theory: linear combinations of independent normal random variables from two independent populations follow a normal distribution.
Consider the case in which we have two independent populations. Let the first population have mean μ_1 andvariance σ_1^2 and the second population have mean $\mu_2$and variance σ_2^2. Suppose that both populations are normallydistributed. Then,using the above property in probability theory, we can say that the sampling distribution of $\ddot{X}_1 - \ddot{X}_2$ is normal with

Mean= $$\mu_{\bar{X}_1-\bar{X}_2} = \mu_{\bar{X}_1} - \mu_{\bar{X}_2} = \mu_1 - \mu_2$$

(1)

And variance = $$\sigma^2_{\bar{X}_1-\bar{X}_2} = \sigma^2_{\bar{X}_1} + \sigma^2_{\bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

(2)

If the two populations are not normally distributed and if both sample sizes $n_1$and n_2 aremore than 30, then use the central limit theorem and assume that$\ddot{X}_1$ and $\ddot{X}_2$ follow approximatelyindependent normal distributions. Therefore, the sampling distribution of $\ddot{X}_1$- $\ddot{X}_2$ isapproximately normal with mean and variance given by the above two equations.

Approximate Sampling Distribution of a Difference in Sample Means:
If we have two independent populations with means μ_1 and μ_2 and variances σ_1^2 andσ_2^2 and if $\ddot{X}_1$ and $\ddot{X}_2$are the sample means of two independent random samples ofsizes n_1 and n_2 from these populations, then the sampling distribution of

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2 / n_1 + \sigma_2^2 / n_2}} \quad (3)$$

is approximately standard normal if the conditions of the central limit theorem apply.If the two populations are normal, the sampling distribution of *Z* is exactly standardnormal.

Example for above theorem :

The effective life of a component used in a jet-turbine aircraft engine is arandom variable with mean 5000 hours and standard deviation 40 hours. The distribution of effectivelife is fairly close to a normal distribution. The engine manufacturer introduces an improvement into the manufacturingprocess for this component that increases the mean life to 5050 hours and decreases the standard deviation to 30hours. Suppose that a random sample of n_1= 16 components is selected from the "old" process and a random sampleof n_2= 25 components is selected from the "improved" process. What is the probability that the difference in the twosamples means $\ddot{X}_1$ and $\ddot{X}_2$is at least 25 hours? Assume that the old and improved processes can be regarded as independentpopulations.

Solution :

Given :

Process 1 : μ_1=5000, σ_1= 40, n_1=16, Normal Distribution

Process 2 : μ_2=5000, σ_2=30, n_2=25, Norml distribution

P($\ddot{X}_1$-$\ddot{X}_2$>=25) =?

Since both sample means $\ddot{X}_1$ and $\ddot{X}_2$are normally distributed and the populations are independent, the difference in sample means $\ddot{X}_1$-$\ddot{X}_2$ will also follow a normal distribution.

The mean of the difference in the sample means is:

$\ddot{X}_1$-$\ddot{X}_2$ = μ2 −μ1 =5050−5000=50 hours (from (1))

The standard deviation (standard error) of the difference in sample means is:

From equation (2)

$$SE_{\bar{X}_2 - \bar{X}_1} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Substituting the values,

$$SE_{\bar{X}_2 - \bar{X}_1} = \sqrt{\frac{40^2}{16} + \frac{30^2}{25}} = \sqrt{\frac{1600}{16} + \frac{900}{25}} = \sqrt{100 + 36} = \sqrt{136} \approx 11.66$$

We are looking for the probability that $\bar{X}_2 - \bar{X}_1$ is at least 25 hours, i.e., $P(\bar{X}_2 - \bar{X}_1 \geq 25)$.

To standardize this, we calculate the z-score using the formula:

$$z = \frac{(\bar{X}_2 - \bar{X}_1) - (\mu_2 - \mu_1)}{SE_{\bar{X}_2 - \bar{X}_1}} = \frac{25 - 50}{11.66} = \frac{-25}{11.66} \approx -2.14$$

Now we need to find P(Z≥−2.14), where Z is a standard normal variable. Using Z Table,

From standard normal tables:

$$P(Z \geq -2.14) = 1 - P(Z < -2.14)$$

But $P(Z < -2.14) = 0.0162$, so:

$$P(Z \geq -2.14) = 1 - 0.0162 = 0.9838$$

$$\begin{aligned} P(\overline{X}_2 - \overline{X}_1 \geq 25) &= P\left(Z \geq -2.14\right) \\ &= 0.9838 \end{aligned}$$

Therefore, there is a high probability (0.9838) that the difference in sample means between the new and the oldprocess will be at least 25 hours if the sample sizes are $n1 = 16$ and $n2 = 25$.

Book Back problems:

1. PVC pipe is manufactured with a mean diameter of 1.01inch and a standard deviation of 0.003 inch. Find the probabilitythat a random sample of n = 9 sections of pipe will have a samplemean diameter greater than 1.009 inch and less than 1.012 inch.

2.Suppose that samples of size n = 25 are selected atrandom from a normal population with mean 100 and standarddeviation 10. What is the probability that the sample mean fallsin the interval from $\mu_X - 1.8\sigma_X$ to $\mu_X + 1.0\sigma_X$?

3. A synthetic fiber used in manufacturing carpet hastensile strength that is normally distributed with mean 75.5 psiand standard deviation 3.5 psi. Find the probability that a randomsample of n = 6 fiber specimens will have sample meantensile strength that exceeds 75.75 psi.

4. Consider the synthetic fiber in the previous exercise.How is the standard deviation of the sample mean changedwhen the sample size is increased from n = 6 to n = 49?

5. The compressive strength of concrete is normally distributedwith μ = 2500 psi and σ = 50 psi. Find the probabilitythat a random sample of n = 5 specimens will have a samplemean diameter that falls in the interval from 2499 psi to 2510 psi.

6. A normal population has mean 100 and variance 25.How large must the random sample be if you want the standarderror of the sample average to be 1.5?

7. Suppose that the random variable X has the continuousuniform distribution

$$f(x)=\begin{cases}1, & 0\le x\le 1\\ 0, & \text{otherwise}\end{cases}$$

Suppose that a random sample of n = 12 observations is selectedfrom this distribution. What is the approximate probability distributionof $X - 6$? Find the mean and variance of this quantity.

8. A random sample of size n_1 = 16 is selected froma normal population with a mean of 75 and a standard deviationof 8. A second random sample of size n_2= 9 is taken fromanother normal population with mean 70 and standard deviation12. Let X_1 and X_2 be the two sample means. Find:

(a) The probability that $X_1 - X_2$ exceeds 4

(b) The probability that $3.5 \le X_1 - X_2 \le 5.5$

Statistical Intervals for a Single Sample :

- When we analyze data, using just the **sample mean** provides a single estimate of the population mean. However, this single number does not give any information about the **uncertainty** or **variability** in our estimate.
- Because of natural sample variability (**sample variability** refers to the natural differences or **fluctuations** that occur in sample statistics like the sample mean, when we take different samples from the same population), we need a range within which we can reasonably expect the true population parameter to fall. This is where **statistical intervals**—like **confidence intervals, prediction intervals** and **tolerance intervals**—come in.
- Because of sampling variability, it is almost never the case that the true mean μ is exactly equal to the estimate Ẍ . The point estimate says nothing about how close Ẍ is to μ.
- Example : Is the population mean (μ) likely to be between 900 and 1100? Or is it likely to be between 990 and 1010?

Each of these intervals serves different purposes:

- **Confidence intervals** estimate the population parameter.
- **Prediction intervals** predict future individual values . A prediction interval provides a range within which a single future observation is expected to fall with a certain level of confidence, such as 95%.
- **Tolerance intervals** ensure that a specified proportion of the population lies within the range.A tolerance interval provides a range that is expected to contain a specified proportion of the population (for example, 90%) with a given confidence level (e.g., 95%).

1. **Confidence intervals:**

- Bounds that represent an interval of plausible values for a parameter are examples of an interval estimate.
- An interval estimate for a population parameter is called a **confidence interval**. Eg : confidence interval provides a range within which the population parameter (such as the mean or proportion) is expected to fall with a given level of confidence, such as 95%.
- Example : With 95% confidence, the mean weight of the population lies between 67.16 kg and 72.84 kg, where the sample mean (Ẍ) is 70 kg,
- Formula for a confidence interval for the population mean:

i. **If the population standard deviation is known**

$$\mathrm{CI} = \bar{X} \pm Z\frac{\sigma}{\sqrt{n}}$$

Where:

- Ẍ= sample mean
- Z = Z-score corresponding to the desired confidence level (e.g., 1.96 for 95% confidence)
- σ = population standard deviation

- n = sample size

ii. If the population standard deviation is unknown

$$\mathrm{CI} = \bar{X} \pm t\frac{s}{\sqrt{n}}$$

- Where t is the t-score based on degrees of freedom (n−1) for the desired confidence level.

1.a. Confidence Interval on the Mean of a Normal Distribution, Variance Known
Suppose that we have a normal population **with unknown mean μ and known variance σ^2**

1.a.i. Development of the Confidence Interval and its Basic Properties:
Suppose that $X_1, X_2 \ldots X_n$ is a random sample from a normal distribution with unknown meanμ and known variance σ^2. The sample mean Ẍ isnormally distributed with mean μ and variance σ^2/n. We may standardize Ẍ by subtractingthe mean and dividing by the standard deviation, which results in the variable :

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

The random variable Z has a standard normal distribution.

A confidence interval estimate for μ is an interval of the form $l \leq \mu \leq u$, where the end-points l and u are computed from the sample data. Because different samples will produce different values ofl and u, these end-points are values of random variables L and U, respectively. Suppose that we can determine values of L and U such that the following probability statement is true:

$$P\{L \leq \mu \leq U\} = 1 - \alpha$$

where $0 \leq \alpha \leq 1$.The value αis the **significance level**, which represents the probability of the confidence interval **not containing** the true mean μ. There is a probability of **$1 - \alpha$** of selecting a sample for which the CI will contain the true value of μ. Once we have selected the sample, so that $X_1, X_2 \ldots X_n$, , and computed l and u, the resulting confidence interval for μ isl $\leq \mu \leq$ u. The end-points or bounds l and u are called the lower- and upper-confidence limits (bounds), respectively, and $1 - \alpha$ is called the confidence coefficient.
We know that Z represents a **standardized score** (or Z-score) of the sample meanẌ when compared to the population mean μ:

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

The Z-score in this context represents the **standardized difference** between the sample mean Ẍ and the population mean μ.
The variable Z follows a **standard normal distribution** (with mean 0 and standard deviation 1). Z tells us how much standard errors Ẍ is away from the population mean μ. A higher |Z| value indicates that the sample mean Ẍ is farther from μ, while a Zclose to 0 suggests Ẍ is close to μ.

Since Z is **normally distributed**, we can say that:

$$P\left(-z_{\alpha/2} \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right) = 1-\alpha$$

Here, $z_{\alpha/2}$is the z-score that corresponds to the upper α/2 tail of the standard normal distribution. For a 95% confidence interval, α=0.05 (1-0.95). The z-score $z_{\alpha/2}$ represents the point on the **standard normal distribution** where the upper tail has an area of α/2. In other words:

- $z_{\alpha/2}$ is the value on the z-scale (standard normal scale) where 2.5% (or 0.025) of the distribution lies to the right, for a 95% confidence level.
- Similarly, $-z_{\alpha/2}$ is the point on the left where α/2 of the area is to its left.

Now by,(1) multiplying through by σ/√n, (2) subtracting X from each term, and (3) multiplying through by −1. This results in

1. **Multiplying through by $\frac{\sigma}{\sqrt{n}}$:**

   The goal here is to eliminate the denominator $\frac{\sigma}{\sqrt{n}}$ so that we can isolate terms involving μ. By multiplying each part of the inequality by $\frac{\sigma}{\sqrt{n}}$, we get:

   $$P\left\{-z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \bar{X}-\mu \leq z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right\} = 1-\alpha$$

   After this step, the inequality is:

   $$-z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \bar{X}-\mu \leq z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

2. **Subtracting $\bar{X}$ from each term:**

   The next step is to rearrange the inequality so that we can isolate μ by itself in the middle. To do this, we subtract $\bar{X}$ from each term in the inequality:

   $$-z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} - \bar{X} \leq -\mu \leq z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} - \bar{X}$$

   Now, the inequality is:

   $$-\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq -\mu \leq -\bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

3. **Multiplying through by -1:**

Finally, we multiply each term by -1 to switch the inequality direction and make μ positive. Remember that multiplying an inequality by a negative number reverses the inequality signs:

$$\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

This gives us:

$$P\left\{\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\} = 1 - \alpha$$

----(1)

Example :

To switch the inequality direction in an expression like $-10 \leq z < 12$, you need to multiply each part of the inequality by -1. However, when multiplying by a negative number, each inequality sign flips direction.

Here's the original inequality:

$$-10 \leq z < 12$$

When we multiply each part by -1, we get:

$$10 \geq -z > -12$$

Now, we can rewrite this in a more conventional order (smallest to largest):

$$-12 < -z \leq 10$$

So, after switching the direction, the inequality becomes:

$$-12 < -z < 10$$

This is a random interval because the end-points $\bar{X} \pm z_{\alpha/2}\sigma/\sqrt{n}$ involve the random variable X. The lower and upper end-points or limits of the inequalities in Equation 1 are the lower- and upper-confidence limits L and U, respectively. This leads to the following definition.

If $\bar{x}$ is the sample mean of a random sample of size n from a normal population with known variance σ^2, a $100(1-\alpha)\%$ CI on μ is given by

$$\bar{x} - z_{\alpha/2}\,\sigma/\sqrt{n} \leq \mu \leq \bar{x} + z_{\alpha/2}\,\sigma/\sqrt{n} \qquad (8\text{-}5)$$

where $z_{\alpha/2}$ is the upper $100\alpha/2$ percentage point of the standard normal distribution.

Example 1 : to find confidence level and range of population mean
ASTM Standard E23 defines standard test methods for notched bar impact testing of metallic materials. The Charpy V-notch (CVN) technique measures impact energy and is often used to determine whether or not a material experiences a ductile-to-brittle transition with decreasing temperature. Ten measurements of impact energy (J) on specimens of A238 steel cut at 60ºC are as follows: 64.1, 64.7, 64.5, 64.6, 64.5, 64.3, 64.6, 64.8, 64.2, and 64.3. Assume that impact energy is normally distributed with σ = 1J. Find a 95% CI for μ, the mean impact energy.
Solution:
Given:
Sample mean Ẍ=(64.1+64.7+ 64.5+ 64.6+ 64.5+ 64.3+ 64.6+ 64.8+64.2+ 64.3)/10 = **64.46**
Population standard deviation (σ) : 1

For 95%CI, the α is 0.05 (1-95% that 1-0.95), α/2 is 0.025
Since **95% CI is given**, the **standard Z-score** for **a 95% confidence level** is **1.96**.
The required quantities are $z_{\alpha/2}=z_{0.025}$ = 1.96, n=10, σ =1, and x = 64 46

The resulting 95% CI is found using the equation

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Therefore,

$$64.46 - 1.96\frac{1}{\sqrt{10}} \le \mu \le 64.46 + 1.96\frac{1}{\sqrt{10}}$$

$$63.84 \le \mu \le 65.08$$

Hence, based on the sample data, a range of highly plausible values for mean impact energy for A238 steel at 60°C is 63.84 ≤μ≤ 65.08 .

Note:
The problem is asking for a **95% confidence interval** for the population mean μ, given a sample mean Ẍ, sample size n=10, and known population standard deviation σ=1.

Finding 1.96:

To find the Z-score for the 95% confidence level, we look for the Z-score that leaves 2.5% in each tail (since 100% - 95% = 5%, and 5%/2 = 2.5% per tail).
The Z-score that corresponds to 97.5% of the distribution (from the left up to the central 95%) is approximately 1.96. This Z-score can be found using statistical tables

90% confidence → Z-score ≈ 1.645
99% confidence → Z-score ≈ 2.576

Example 2 : For a normal population with known variance σ^2 , answer the following questions:
(a) What is the confidence level for the interval Ẍ−2 14 σ/√n ≤μ≤ Ẍ+2 14 σ/√n ?

Solution :

- A **normal population** with a known variance σ^2.
- An interval around the sample mean $\bar{X}$ given as:

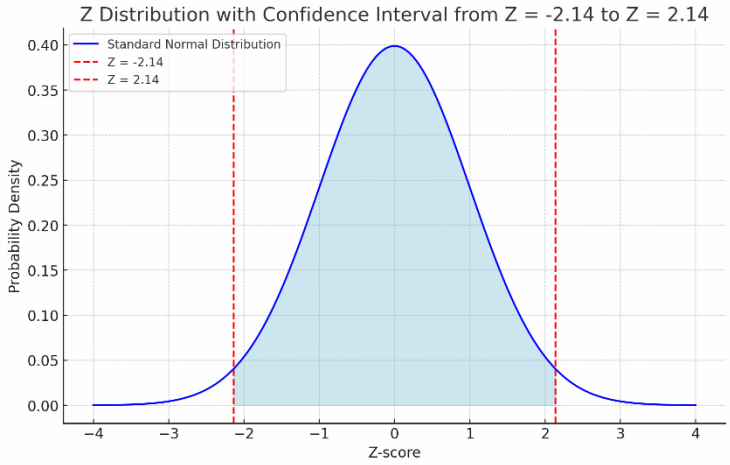
$$\bar{X} - 2.14\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 2.14\frac{\sigma}{\sqrt{n}}$$

where μ is the population mean.

We are asked to find the **confidence level** associated with this interval.

Given:

Here, 2.14 is the **Z-score**, which corresponds to the number of standard errors away from the sample mean X̄. That is, Z-score tells us how far away a value is from the mean in terms of standard deviations.



The shaded area is the range of μ with certain CI, that CI we need to find.

Calculate the Probability for $Z \leq 2.14$:

- We need to find the probability associated with the range $-2.14 \leq Z \leq 2.14$.
- Using a Z-table or standard normal distribution table:

$$P(Z \leq 2.14) \approx 0.9838$$

$$P(Z \leq -2.14) \approx 0.0162$$

(z-score always represents the area covered to the left of z-score value. From the above graph, it is clear that we need to find the CI for the shaded area)

Therefore, we need to P(Z<=-2.14) and P(Z<=2.14) and subtract : P(Z<=2.14)-P(Z<=-2.14)

- So, the probability that Z lies between -2.14 and 2.14 is:

$$0.9838 - 0.0162 = 0.9676$$

So, the confidence level associated with this interval is approximately **96.76%**.

1.a.ii. Confidence Level and Precision of Estimation

The **length of a confidence interval** is the distance (or width) between the lower and upper bounds of the interval. It quantifies the **range of values** that we are confident includes the population parameter (like the mean) based on our sample data.

The **length of the confidence interval** is the difference between the upper bound and the lower bound:

$$\text{Length} = \left(\bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) - \left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

Simplifying this, we get:

$$\text{Length} = 2 \times z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

The length of the 95% confidence interval is

$$2\left(1.96\sigma/\sqrt{n}\right) = 3.92\sigma/\sqrt{n}$$

whereas the length of the 99% CI is

$$2\left(2.58\sigma/\sqrt{n}\right) = 5.16\sigma/\sqrt{n}$$

Thus, the 99% CI is longer than the 95% CI. This is why we have a higher level of confidence in the 99% confidence interval.As the confidence level increases from 95% to 99%, the confidence interval becomes wider. This wider interval provides greater confidence that the interval contains the true population mean but **reduces precision** by providing a broader range of plausible values.

Example

Suppose we have:

- A sample mean $\bar{X} = 50$,
- Population standard deviation $\sigma = 10$,
- Sample size $n = 25$,
- Desired confidence level of 95% (with $z_{\alpha/2} = 1.96$).

The length of the confidence interval would be:

$$\text{Length} = 2 \times 1.96 \times \frac{10}{\sqrt{25}} = 2 \times 1.96 \times 2 = 7.84$$

So, the confidence interval is 7.84 units wide.

The length of a confidence interval is a measure of the **precision of estimation.**

A higher confidence level (like 99%) means we want to be more certain that the interval contains the true population parameter. To increase our certainty, we need to make the interval wider, capturing more potential values.
Conversely, **a lower confidence level (like 90%)** doesn't require as much certainty, so we can afford to have a narrower interval. This narrower interval is closer to the sample mean, offering more precision but with less certainty about containing the true mean.

Example :

a numerical example where a lower confidence level results in a narrower confidence interval, giving a more precise estimate of the population mean. However, the trade-off is that a lower confidence level means less certainty that the interval actually contains the true population mean.

Given Data

- Sample mean ($\ddot{X}$) = 75
- Population standard deviation (σ) = 10 (assumed known)
- Sample size (n) = 50

The confidence interval for the population mean, when the population standard deviation is known, is given by:

$$\text{Confidence Interval} = \bar{x} \pm Z \times \frac{\sigma}{\sqrt{n}}$$

Step 1: Calculate the Standard Error

$$\text{Standard Error} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{50}} \approx 1.414$$

Step 2: Calculate Confidence Intervals at Different Confidence Levels

1. **80% Confidence Level**

- For an 80% confidence level, the Z-score is approximately **1.282**.
- Confidence interval:

$$75 \pm 1.282 \times 1.414 = 75 \pm 1.813$$

So, the 80% confidence interval is:

$$(75 - 1.813, 75 + 1.813) = (73.187, 76.813)$$

2. **95% Confidence Level**

- For a 95% confidence level, the Z-score is approximately **1.96**.
- Confidence interval:

$$75 \pm 1.96 \times 1.414 = 75 \pm 2.772$$

So, the 95% confidence interval is:

$$(75 - 2.772, 75 + 2.772) = (72.228, 77.772)$$

- **Length of the 95% Confidence Interval**: $2 \times 2.772 = 5.544$

Interpretation

80% Confidence Interval: (73.187,76.813) with a length of **3.626**
95% Confidence Interval: (72.228,77.772) with a length of **5.544**
In this case:
The **80% confidence interval** is narrower, providing a more precise estimate of the population mean, but we are only 80% confident it contains the true mean.
The **95% confidence interval** is wider, giving a less precise estimate but with greater confidence that it includes the true mean.
In the context of confidence intervals, **precision** refers to how tightly the interval surrounds the sample mean as an estimate of the population mean. A **narrower confidence interval** means that we have a smaller range of values, which indicates a **more precise estimate** of where the true population mean is likely to fall.
In the example given:
For the **80% confidence interval**: the interval range is (73.187,76.813) with a length of **3.626**.
For the **95% confidence interval**: the interval range is (72.228,77.772) with a length of **5.544**.

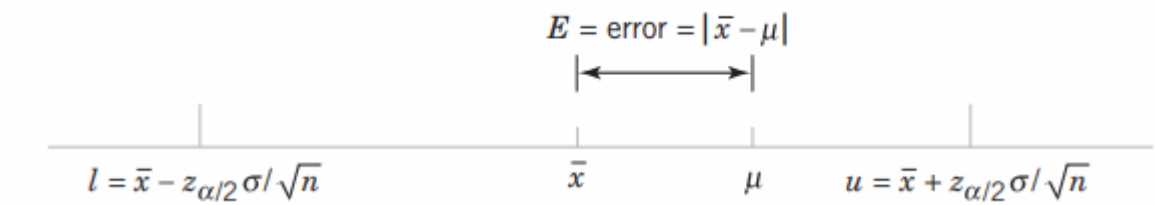
The **80% confidence interval is narrower** (3.626 vs. 5.544) because it has a lower confidence level. This narrower interval gives us a more focused range around the sample mean (75), suggesting a more **precise estimate** of where we believe the population mean to be.

However, the trade-off is that with an 80% confidence level, we are less certain that this narrower interval contains the true population mean compared to a 95% confidence level.

1.a.iii. Choice of Sample Size
The precision of the estimation is, that is the length of CI,

$$\text{Length} = 2 \times z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

This indicates that the difference between Ẍ and μ is E, that is E=|μ-Ẍ|. That is, E is <= to $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ with confidence 100 (1−α)% as shown below:



Therefore, Sample Size for Specified Error (E) on the Mean, when Variance is known , is given as follows:

If $\bar{x}$ is used as an estimate of μ, we can be $100(1-\alpha)\%$ confident that the error $|\bar{x}-\mu|$ will not exceed a specified amount E when the sample size is

$$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2 \qquad (8\text{-}6)$$

If the right-hand side of the above equation(8-6)is not an integer, it must be rounded up.

Example :
Suppose that we want to determine how many specimens must be tested to ensure that the 95% CI on μ for A238 steel cut at 60°C has a length of at most 1.0 J and standard deviation is 1. Because the bound on error in estimation E is one-half of the length of the CI. Determine the sample size, n.
The required sample size is

$$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2 = \left[\frac{(1.96)1}{0.5}\right]^2 = 15.37$$

and because n must be an integer, the required sample size is n = 16.

1.a.iv. 3 One-Sided Confidence Bounds :
The confidence interval in the following Equation

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

gives both a lower confidence bound, and an upper confidence bound for μ. Thus, it provides a two-sided CI. It is also possible to obtain onesided confidence bounds for μ by setting either the lower bound $l = -\infty$ or the upper bound $u = \infty$ and replacing $z_{\alpha/2}$ by z.

One-Sided Confidence Bounds on the Mean, Variance Known

A $100(1-\alpha)\%$ **upper-confidence bound** for μ is

$$\mu \le \bar{x} + z_\alpha \sigma/\sqrt{n}$$

and a $100(1-\alpha)\%$ **lower-confidence bound** for μ is

$$\bar{x} - z_\alpha \sigma/\sqrt{n} = l \le \mu$$

Example : to find confidence level and range of population mean
ASTM Standard E23 defines standard test methods for notched bar impact testing of metallic materials. The Charpy V-notch (CVN) technique measures impact energy and is often used to determine whether or not a material experiences a ductile-to-brittle transition with decreasing temperature. Ten measurements of impact energy (J) on specimens of A238 steel cut at 60ºC are as follows: 64.1, 64.7, 64.5, 64.6, 64.5, 64.3, 64.6, 64.8, 64.2, and 64.3. Assume that impact energy is normally distributed with σ = 1J. Construct a lower, one-sided 95% confidence interval for the mean, the mean impact energy. Also upper, one sided CI.

Solution:
Given:
Measurements Ẍ =[64.1,64.7,64.5,64.6,64.5,64.3,64.6,64.8,64.2,64.3]
Population standard deviation σ=1
Sample size n=10
Confidence level = 95%

For a normally distributed population, a one-sided confidence interval for the mean μ is given by
For the lower one-sided confidence interval:

$$\bar{X} - z_\alpha \frac{\sigma}{\sqrt{n}}$$

For the upper one-sided confidence interval:

$$\bar{X} + z_\alpha \frac{\sigma}{\sqrt{n}}$$

z_α is the z-score corresponding to the desired confidence level (for a 95% one-sided confidence interval, $z_{0.05} \approx 1.64$
for lower CI

$$\bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}} \le \mu$$

$$64.46 - 1.64 \frac{1}{\sqrt{10}} \le \mu$$

$$63.94 \le \mu$$

For upper CI,
64.46+1.64(0.316) = 64.98
Upper 95% confidence interval: 64.98

1.a.v. Large-Sample Confidence Interval for μ:
If the standard deviation σ is unknown and when n is large, replace σ by the sample standard deviation S. This has little effect on the distribution of Z.

When n is large, the quantity

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

has an approximate standard normal distribution. Consequently,

$$\bar{x} - z_{\alpha/2} \frac{s}{\sqrt{n}} \le \mu \le \bar{x} + z_{\alpha/2} \frac{s}{\sqrt{n}} \qquad (8\text{-}11)$$

is a **large-sample confidence interval** for μ, with confidence level of approximately $100(1 - \alpha)\%$.

Note; Generally, n should be at least 40 to use this result reliably. The central limit theorem generally holds for $n \geq 30$, but the larger sample size is recommended here because replacing s with S in Z results in additional variability.

Book Back problems:

1. PVC pipe is manufactured with a mean diameter of 1.01inch and a standard deviation of 0.003 inch. Find the probabilitythat a random sample of $n = 9$ sections of pipe will have a samplemean diameter greater than 1.009 inch and less than 1.012 inch.

Given: $\mu = 1.01$ $\sigma = 0.003$, $n = 9$, $\bar{x} \geq 1.009$ & less than 1.012

(ii) $1.009 \leq \bar{x} \leq 1.012$

(i) Std. error $= \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.003}{\sqrt{9}} = 0.001$

(ii) for $\bar{x}$ greater than 1.009

$$z = \frac{1.009 - 1.01}{0.001} = -1.0$$

for $\bar{x}$ less than 1.012

$$z = \frac{1.012 - 1.01}{.001} = 2.0$$

(iii) $P(z \leq -1.0) = 0.1587$

$P(z \leq 2) = 0.9772$

(iv) to find the prob. that sample mean is between 1.009 & 1.012

$P(1.009 \leq x \leq 1.012) = P(z \leq 2) - P(z \leq -1)$

$= 0.9772 - 0.1587$

$= 0.8185$

$= 81.7\%$

2. Suppose that samples of size $n = 25$ are selected atrandom from a normal population with mean 100 and standarddeviation 10. What is the probability that the sample mean fallsin the interval from $\mu X - 1.8\sigma X$ to $\mu X + 1.0\sigma X$?

Given $n = 25$, $\mu = 100$, $\sigma = 10$, Normal distribution.

(i) Std error $= \sigma / \sqrt{n} = 10/\sqrt{25} = 2 = \sigma_{\bar{x}}$

(ii) for $\bar{x} = \mu_{\bar{x}} - 1.8\sigma_{\bar{x}}$, $\mu_{\bar{x}} = \mu$ $\sigma_{\bar{x}} = 2$

$$z = \frac{100 - 1.8 * (2) - 100}{2} = \frac{96.4 - 100}{2} = -1.8$$

(iii) for $\bar{x} = \mu_{\bar{x}} + 1.0\sigma_{\bar{x}} = \mu_{\bar{x}} = \mu = 100$ $\sigma_{\bar{x}} = 2$

$$z = \frac{100 + 1.0(2) - 100}{2} = 2/2 = 1.0$$

$P(z <= -1.8) = 0.03593$, $P(z <= 1.0) = 0.8413$

(iv) Find the prob. that $\bar{x}$ falls within 96.4 & 102.0

$P(96.4 <= \bar{x} <= 102.0) = P(z <= 1.0) - P(z <= -1.8)$

$= 0.8413 - 0.3593 = 0.8054$

$= 80.5\%$

Explanation: Pb 1: For $\bar{x} = 1.009$, the z score was -1, so we need to find the cumulative probability upto this point i $P(z <= -1)$, Similarly $\bar{x} = 1.012 \Rightarrow P(z <= 2)$

The total prob. of the sample mean lying between 1.009 & 1.012 is $P(1.009 \leq x \leq 1.012) = P(z <= 2) - P(z <= -1)$

3. A synthetic fiber used in manufacturing carpet hastensile strength that is normally distributed with mean 75.5 psiand standard deviation 3.5 psi. Find the probability that a randomsample of $n = 6$ fiber specimens will have sample meantensile strength that exceeds 75.75 psi.

Given : $\mu = 75.5$ $\sigma = 3.5$ $n = 6$, Normal

(i) $SE = \sigma_{\bar{x}} = \sigma / \sqrt{n} = 3.5/\sqrt{6} = 1.428$

(ii) for sample mean $\bar{x} > 75.75$

$$Z = \frac{75.75 - 75.5}{1.428} = 0.175$$

(iii) Use Z table,

$P(\bar{x} > 75.75) = P(Z > 0.175)$

we find $P(Z <= 0.175) = 0.5695$

To find $P(Z > 0.175) = 1 - 0.5695$

$= 0.4305 = 43.5\%$

4. Consider the synthetic fiber in the previous exercise.How is the standard deviation of the sample mean changedwhen the sample size is increased from $n = 6$ to $n = 49$?

$\mu = 75.5$ $\sigma = 3.5$ Normal, $n = 49$

(i) $SE = \sigma / \sqrt{n} = \frac{3.5}{\sqrt{49}} = \frac{3.5}{7} = 1.43$ When $n = 49$

When $n = 6$ $SE = \sigma / \sqrt{n} = \frac{3.5}{\sqrt{6}} = 0.5$

When Sample Sz is increased from $n = 6$ to $n = 49$, the Sample mean reduced from 1.43 to 0.5 psi.

5. The compressive strength of concrete is normally distributedwith $\mu = 2500$ psi and $\sigma = 50$ psi. Find the probabilitythat a random sample of $n = 5$ specimens will have a samplemean diameter that falls in the interval from 2499 psi to 2510 psi.

(i) $\mu = 2500$ $\sigma = 50$, Normal distribution

$n = 5$

$P(2499 <= \bar{X} <= 2510) = ?$

(ii)

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

(iii) For $P(\bar{x} >= 2499) = \frac{2499 - 2500}{50/\sqrt{5}}$

$z = -0.0447$

(iv) for $P(\bar{x} <= 2510) = \frac{2510 - 2500}{50/\sqrt{5}} = 0.4472$

(v) $P(z < -0.0447) = 0.4822$

$P(z < 0.4472) = 0.6710$

(vi) To find prob. that $\bar{x}$ falls between 2499 & 2510

(vii) $P(2499 <= \bar{X} <= 2510)$

$= P(Z < 0.4472) - P(Z < -0.0047)$

$= 0.6710 - 0.4472$

$= 0.1888$

6. A normal population has mean 100 and variance 25.How large must the random sample be if you want the standarderror of the sample average to be 1.5?

$\mu = 100$, $\sigma^2 = 25 \Rightarrow \sigma = 5$

$SE = \sigma/\sqrt{n} = 1.5$

$\Rightarrow 25/\sqrt{n} = 1.5$

$25 = 1.5\sqrt{n}$

$(25)^2 = (1.5\sqrt{n})^2$

$625 = (1.5)^2 n$

$SE \Rightarrow 5/\sqrt{n} = 1.5$

$5 = 1.5\sqrt{n}$

$5^2 = (1.5\sqrt{n})^2$

$25 = 1.5^2 n$

$25 = 2.25n$

$n = \frac{25}{2.25} = 11.11$

$n \approx 12$ samples

7. Suppose that the random variable X has the continuousuniform distribution

$$f(x)=\begin{cases}1, & 0\le x\le 1\\ 0, & \text{otherwise}\end{cases}$$

Suppose that a random sample of n = 12 observations is selectedfrom this distribution. What is the approximate probability distributionof $X-6$? Find the mean and variance of this quantity.

(i) Mean $M=\frac{a+b}{2}=\frac{0+1}{2}=0.5$ $(a=0, b=1)$

$\sigma^2=\frac{(b-a)^2}{12}=\frac{(1-0)^2}{12}=0.0833$

Distribution of sample mean & variance

(ii) $n=12$. Due to CLT, the sample mean will follow normal distribution with $\bar{X}=M=0.5$

Variance$(s^2)=\sigma^2/n=\frac{0.0833}{12}=0.00694$

(iii) Prob. distribution of $\bar{X}-6$,

Mean $\bar{X}-6=0.5-6=-5.5$

Variance $\bar{X}-6$ $\Rightarrow$ The variance remain unchanged because subtracting a constant does not affect variance.

$\Rightarrow 0.00694$.

8. A random sample of size n_1 = 16 is selected froma normal population with a mean of 75 and a standard deviationof 8. A second random sample of size n_2= 9 is taken fromanother normal population with mean 70 and standard deviation12. Let X_1 and X_2 be the two sample means. Find:

(a) The probability that X_1-X_2exceeds 4

(b) The probability that 3.5 ≤X_1-X_2≤5.5

(i) $n_1=16, M_1=75, \sigma_1=8$

$n_2=9, M_2=70, \sigma_2=12$

(a) Prob. that $\bar{x}_1-\bar{x}_2$ exceeds 4,

$M_1-M_2=75-70=5$

$$Z=\frac{(\bar{x}_1-\bar{x}_2)-(M_1-M_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}$$

$$=\frac{4-5}{\sqrt{8^2/16+12^2/9}}$$

$$=\frac{-1}{\sqrt{\frac{64}{16}+\frac{144}{9}}}=\frac{-1}{\sqrt{4+16}}=\frac{-1}{\sqrt{20}}=-0.2237$$

$P(Z>-0.2237)=0.5892$

$\therefore$ The prob. that $\bar{x}_1-\bar{x}_2$ exceeds 4 is 58.92%.

(b) $P(\bar{x}_1-\bar{x}_2\ge 3.5)=P(Z<-0.3358)$

$=0.3694$

$P(\bar{x}_1-\bar{x}_2\le 5.5)=P(Z<0.1119)$

$=0.5448$

$\therefore P(3.5\le \bar{x}_1-\bar{x}_2\le 5.5)$

$=P(Z<-0.1119)-P(Z<-0.3358)$

$=0.5448-0.3694$

$=0.1754$

$\Rightarrow 17.5\%$.

Unit IV

Topics covered: Statistical Intervals for a Single Sample: Confidence Interval on Mean – variance and Standard Deviation - Guidelines - Bootstrap - Tolerance and Prediction Intervals

Statistical Intervals for a Single Sample

- Introduction
- Confidence interval on the mean of a normal distribution, variance known
- Confidence interval on the mean of a normal distribution, variance unknown
- Confidence interval on the variance and standard deviation of a normal distribution
- Large-sample confidence interval for a population proportion
- Tolerance and prediction intervals

Introduction

- In the previous chapter we illustrated how a parameter can be estimated from sample data.However, it is important to understand how good is the estimate obtained.
- Bounds that represent an interval of plausible values for a parameter are an example of an interval estimate.
  - Three types of intervals will be presented:
    - Confidence intervals
    - Prediction intervals
    - Tolerance intervals

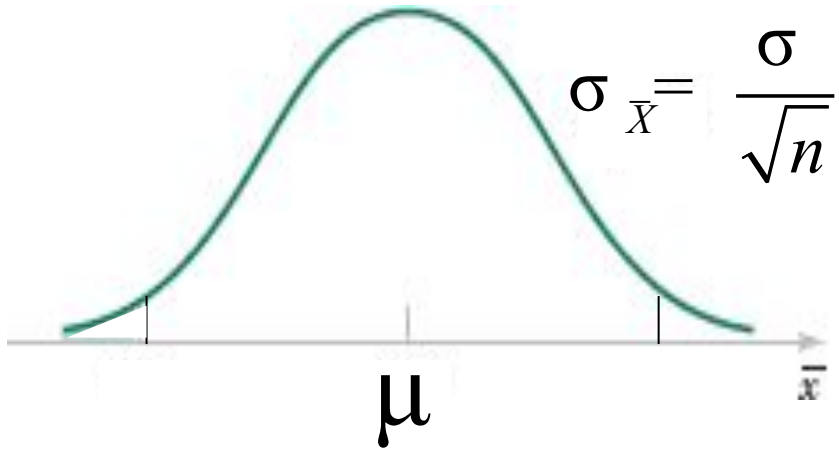
1. Confidence Interval on the Mean of a Normal Distribution, Variance Known

1.1 Development of the Confidence Interval and its Basic Properties

POPULATION
$N(\mu,\sigma^2)$
σ^2: known
μ : unknown

Sample $X_1, X_2, \ldots, X_n$

$$\bar{X} \sim N\left(\mu, \sigma^2/n\right)$$



$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$ **has a standard normal distribution**

1. Confidence Interval on the Mean of a Normal Distribution, Variance Known

1.1 Development of the Confidence Interval and its Basic Properties

A **confidence interval** estimate for μ is an interval of the form $l \leq \mu \leq u$, where the end-points l and u are computed from the sample data. Because different samples will produce different values of l and u, these end-points are values of random variables L and U, respectively. Suppose that we can determine values of L and U such that the following probability statement is true:

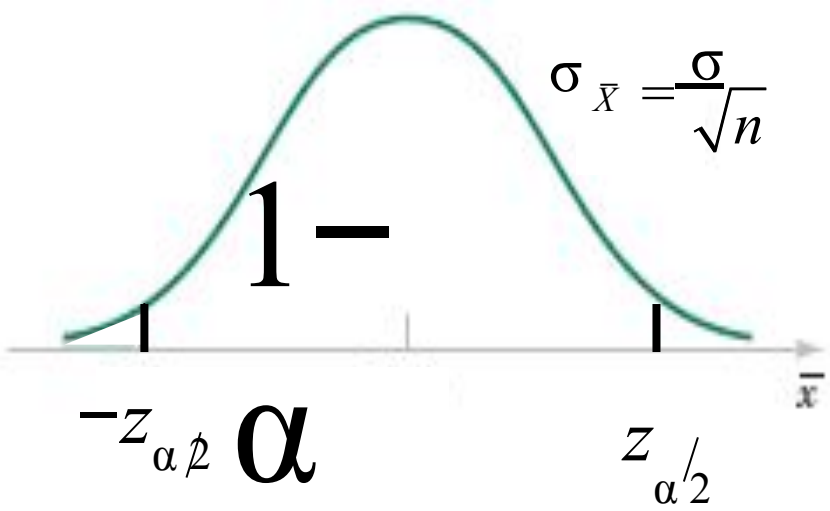
$$P\{L \leq \mu \leq U\} = 1 - \alpha \tag{8-4}$$

where $0 \leq \alpha \leq 1$. There is a probability of $1 - \alpha$ of selecting a sample for which the CI will contain the true value of μ. Once we have selected the sample, so that $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, and computed l and u, the resulting **confidence interval** for μ is

$$l \leq \mu \leq u \tag{8-5}$$

1. Confidence Interval on the Mean of a Normal Distribution, Variance Known

1.1 Development of the Confidence Interval and its Basic Properties



$$P\left\{-z_{\alpha/2} \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le z_{\alpha/2}\right\} = 1-\alpha$$

$$P\left\{\bar{X} - z_{\alpha/2}\sigma/\sqrt{n} \le \mu \le \bar{X} + z_{\alpha/2}\sigma/\sqrt{n}\right\} = 1-\alpha$$

1. Confidence Interval on the Mean of a Normal Distribution, Variance Known

1.1 Development of the Confidence Interval and its Basic Properties

Definition

If $\bar{x}$ is the sample mean of a random sample of size n from a normal population with known variance σ^2, a $100(1-\alpha)\%$ CI on μ is given by

$$\bar{x} - z_{\alpha/2}\sigma/\sqrt{n} \leq \mu \leq \bar{x} + z_{\alpha/2}\sigma/\sqrt{n} \qquad (8\text{-}7)$$

where $z_{\alpha/2}$ is the upper $100\alpha/2$ percentage point of the standard normal distribution.

1. Confidence Interval on the Mean of a Normal Distribution, Variance Known

Example 8-1

ASTM Standard E23 defines standard test methods for notched bar impact testing of metallic materials. The Charpy V-notch (CVN) technique measures impact energy and is often used to determine whether or not a material experiences a ductile-to-brittle transition with decreasing temperature. Ten measurements of impact energy (J) on specimens of A238 steel cut at 60°C are as follows: 64.1, 64.7, 64.5, 64.6, 64.5, 64.3, 64.6, 64.8, 64.2, and 64.3. Assume that impact energy is normally distributed with $\sigma = 1J$. We want to find a 95% CI for μ, the mean impact energy.

$n = 10$

$\sigma = 1$

$\bar{x} = 64.46$

$\alpha = 0.05$

$z_{\alpha/2} = z_{0.025} = 1.96$

1. Confidence Interval on the Mean of a Normal Distribution, Variance Known

Example 8-1

ASTM Standard E23 defines standard test methods for notched bar impact testing of metallic materials. The Charpy V-notch (CVN) technique measures impact energy and is often used to determine whether or not a material experiences a ductile-to-brittle transition with decreasing temperature. Ten measurements of impact energy (J) on specimens of A238 steel cut at 60°C are as follows: 64.1, 64.7, 64.5, 64.6, 64.5, 64.3, 64.6, 64.8, 64.2, and 64.3. Assume that impact energy is normally distributed with $\sigma = 1J$. We want to find a 95% CI for μ, the mean impact energy. The required quantities are $z_{\alpha/2} = z_{0.025} = 1.96$, $n = 10$, $\sigma = 1$, and $\bar{x} = 64.46$. The resulting 95% CI is found from Equation 8-7 as follows:

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$64.46 - 1.96\frac{1}{\sqrt{10}} \le \mu \le 64.46 + 1.96\frac{1}{\sqrt{10}}$$

$$63.84 \le \mu \le 65.08$$

That is, based on the sample data, a range of highly plausible vaules for mean impact energy for A238 steel at 60°C is $63.84J \le \mu \le 65.08J$.

1. Confidence Interval on the Mean of a Normal Distribution, Variance Known

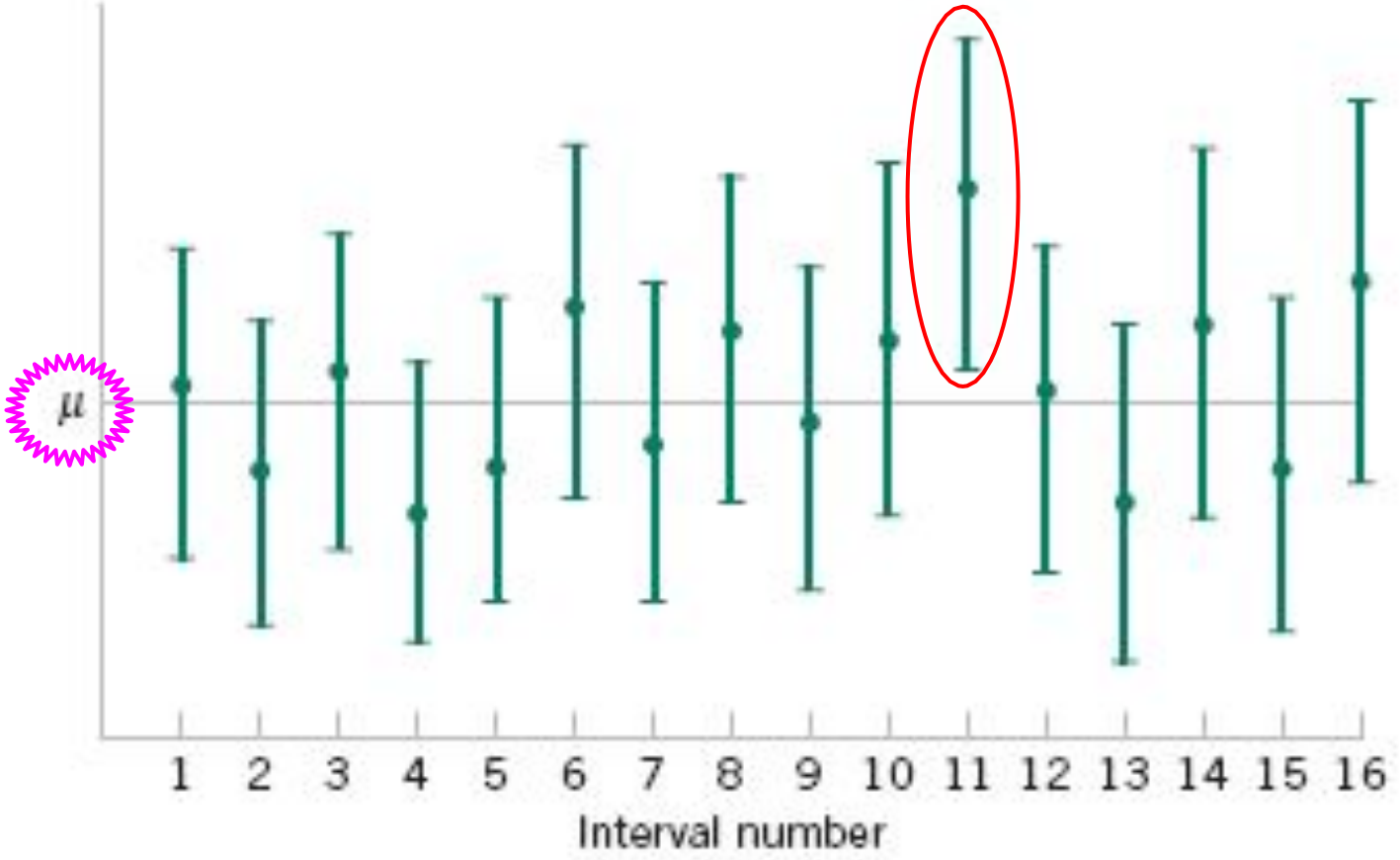
Interpreting a Confidence Interval

- The confidence interval is a random interval

- The appropriate interpretation of a confidence interval (for example on μ) is:
  The observed interval $[l, u]$ brackets the true value of μ, with confidence $100(1-\alpha)\%$.

- Examine the figure on the next slide.

1. Confidence Interval on the Mean of a Normal Distribution, Variance Known



Repeated construction of a confidence interval for μ.

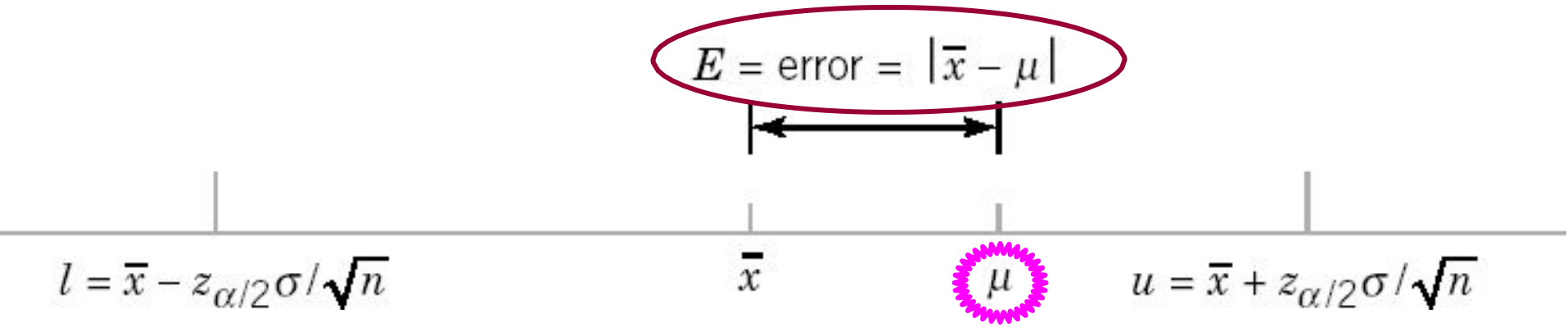
A 95% CI means in the long run only 5% of the intervals would fail to contain μ.

1. Confidence Interval on the Mean of a Normal Distribution, Variance Known

Confidence Level and Precision of Estimation

For fixed n and σ, the higher the confidence level, the longer the resulting confidence interval.

The length of a confidence interval is a measure of the precision of estimation.



By using $\bar{x}$ to estimate μ, the error E is less than or equal to $z_{\alpha/2}\,\sigma/\sqrt{n}$ with confidence 100(1-α)%.

1 Confidence Interval on the Mean of a Normal Distribution, Variance Known

1.2 Choice of Sample Size

Bound on error is $E \geq z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

Choose n to be large enough to assure a predetermined CI and precision !

If $\bar{x}$ is used as an estimate of μ, we can be $100(1-\alpha)\%$ confident that the error $|\bar{x} - \mu|$ will not exceed a specified amount E when the sample size is

$$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2 \qquad (8\text{-}8)$$

1 Confidence Interval on the Mean of a Normal Distribution, Variance Known

Example 8-2

To illustrate the use of this procedure, consider the CVN test described in Example 8-1, and suppose that we wanted to determine how many specimens must be tested to ensure that the 95% CI on μ for A238 steel cut at 60°C has a length of at most 1.0*J*.

1 Confidence Interval on the Mean of a Normal Distribution, Variance Known

Example 8-2

To illustrate the use of this procedure, consider the CVN test described in Example 8-1, and suppose that we wanted to determine how many specimens must be tested to ensure that the 95% CI on μ for A238 steel cut at 60°C has a length of at most 1.0*J*. Since the bound on error in estimation E is one-half of the length of the CI, to determine n we use Equation 8-8 with $E = 0.5$, $\sigma = 1$, and $z_{\alpha/2} = 0.025$. The required sample size is 16

$$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2 = \left[\frac{(1.96)1}{0.5}\right]^2 = 15.37$$

and because n must be an integer, the required sample size is $n = 16$.

1 Confidence Interval on the Mean of a Normal Distribution, Variance Known

1.3 One-Sided Confidence Bounds

Definition

A $100(1-\alpha)\%$ **upper-confidence bound** for μ is

$$\mu \leq u = \bar{x} + z_\alpha \sigma/\sqrt{n} \qquad (8\text{-}9)$$

and a $100(1-\alpha)\%$ **lower-confidence bound** for μ is

$$\bar{x} - z_\alpha \sigma/\sqrt{n} = l \leq \mu \qquad (8\text{-}10)$$

Attention!

1. Confidence Interval on the Mean of a Normal Distribution, Variance Known

Example 8-3

The same data for impact testing from example 8.1.

Construct a lower, one sided 95% CI for the mean impact energy

$$z_{\alpha} = z_{0.05} = 1.64$$

$$\bar{x} - z_{\alpha}\frac{\sigma}{\sqrt{n}} \leq \mu$$

$$64.46 - 1.64\frac{1}{\sqrt{10}} \leq \mu$$

$$63.94 \leq \mu$$

1. Confidence Interval on the Mean of a Normal Distribution, Variance Known

1.4 A Large-Sample Confidence Interval for μ

-Holds regardless of the population distribution
-Additional variability because of replacing σ by S
-So **n>=40**

Definition

When n is large, the quantity

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

has an approximate standard normal distribution. Consequently,

$$\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}} \qquad (8\text{-}13)$$

is a **large sample confidence interval** for μ, with confidence level of approximately $100(1-\alpha)\%$.

1. Confidence Interval on the Mean of a Normal Distribution, Variance Known

Example 8-4

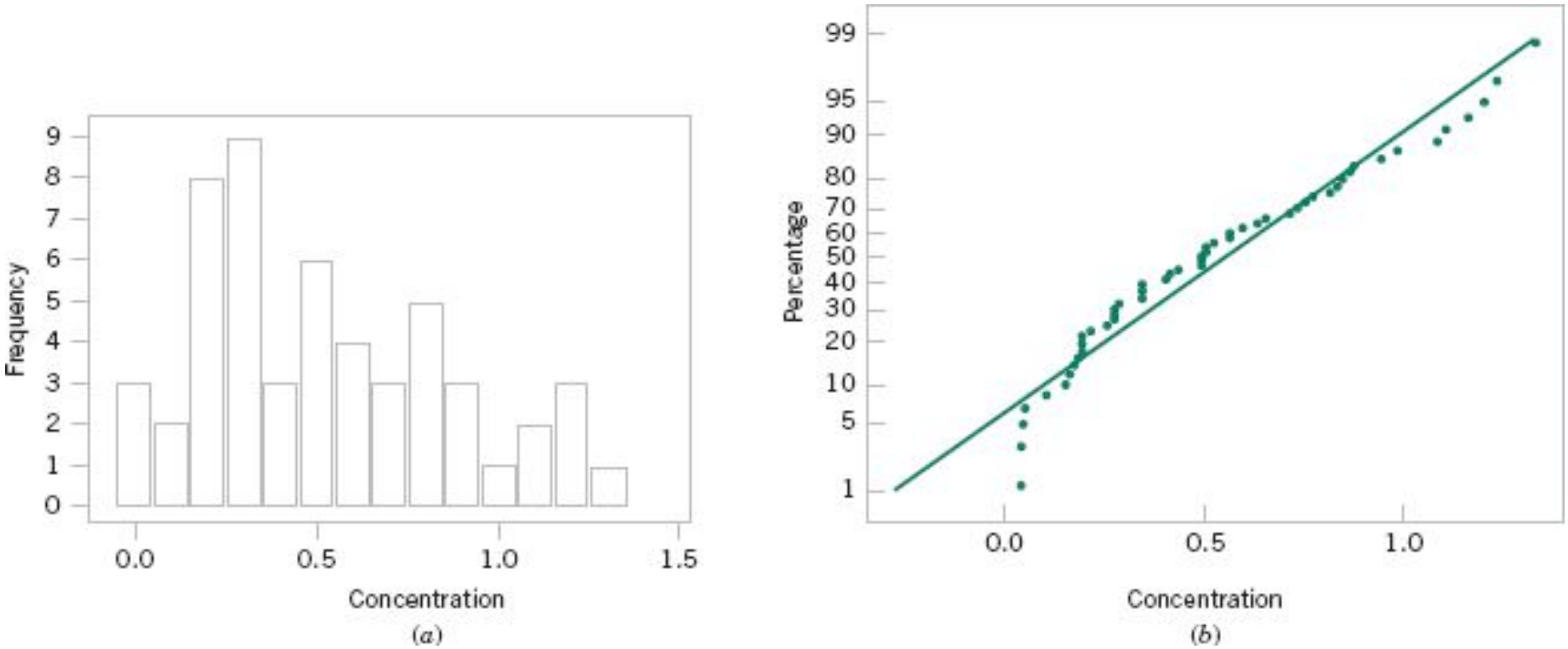
An article in the 1993 volume of the *Transactions of the American Fisheries Society* reports the results of a study to investigate the mercury contamination in largemouth bass. A sample of fish was selected from 53 Florida lakes and mercury concentration in the muscle tissue was measured (ppm). The mercury concentration values are

1.230	0.490	0.490	1.080	0.590	0.280	0.180	0.100	0.940
1.330	0.190	1.160	0.980	0.340	0.340	0.190	0.210	0.400
0.040	0.830	0.050	0.630	0.340	0.750	0.040	0.860	0.430
0.044	0.810	0.150	0.560	0.840	0.870	0.490	0.520	0.250
1.200	0.710	0.190	0.410	0.500	0.560	1.100	0.650	0.270
0.270	0.500	0.770	0.730	0.340	0.170	0.160	0.270	

Find an approximate 95% CI on μ.

1. Confidence Interval on the Mean of a Normal Distribution, Variance Known

Example 8-4 (continued)



Mercury concentration in largemouth bass

(a) Histogram. (b) Normal probability plot

1. Confidence Interval on the Mean of a Normal Distribution, Variance Known

Example 8-4 (continued)

The summary statistics from Minitab are displayed below:

Descriptive Statistics: Concentration

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Concentration	53	0.5250	0.4900	0.5094	0.3486	0.0479

Variable	Minimum	Maximum	Q1	Q3
Concentration	0.0400	1.3300	0.2300	0.7900

1. Confidence Interval on the Mean of a Normal Distribution, Variance Known

Example 8-4 (continued)

Figure 8-3(a) and (b) presents the histogram and normal probability plot of the mercury concentration data. Both plots indicate that the distribution of mercury concentration is not normal and is positively skewed. We want to find an approximate 95% CI on μ. Because $n > 40$, the assumption of normality is not necessary to use Equation 8-13. The required quantities are $n = 53$, $\bar{x} = 0.5250$, $s = 0.3486$, and $z_{0.025} = 1.96$. The approximate 95% CI on μ is

$$\bar{x} - z_{0.025}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + z_{0.025}\frac{s}{\sqrt{n}}$$

$$0.5250 - 1.96\frac{0.3486}{\sqrt{53}} \le \mu \le 0.5250 + 1.96\frac{0.3486}{\sqrt{53}}$$

$$0.4311 \le \mu \le 0.6189$$

This interval is fairly wide because there is a lot of variability in the mercury concentration measurements.

1. Confidence Interval on the Mean of a Normal Distribution, Variance Known

A General Large Sample Confidence Interval

If the estimator of a population parameter Θ
- has an approximate normal distribution
- is approximately unbiased for Θ
- has standard deviation that can be estimated from the sample data

Then the estimator has an approximate normal distribution and a large-sample approximate CI for Θ is as follows

$$\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\Theta}} \leq \theta \leq \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\Theta}} \qquad (8\text{-}14)$$

2. Confidence Interval on the Mean of a Normal Distribution, Variance Unknown

2.1 The *t* distribution

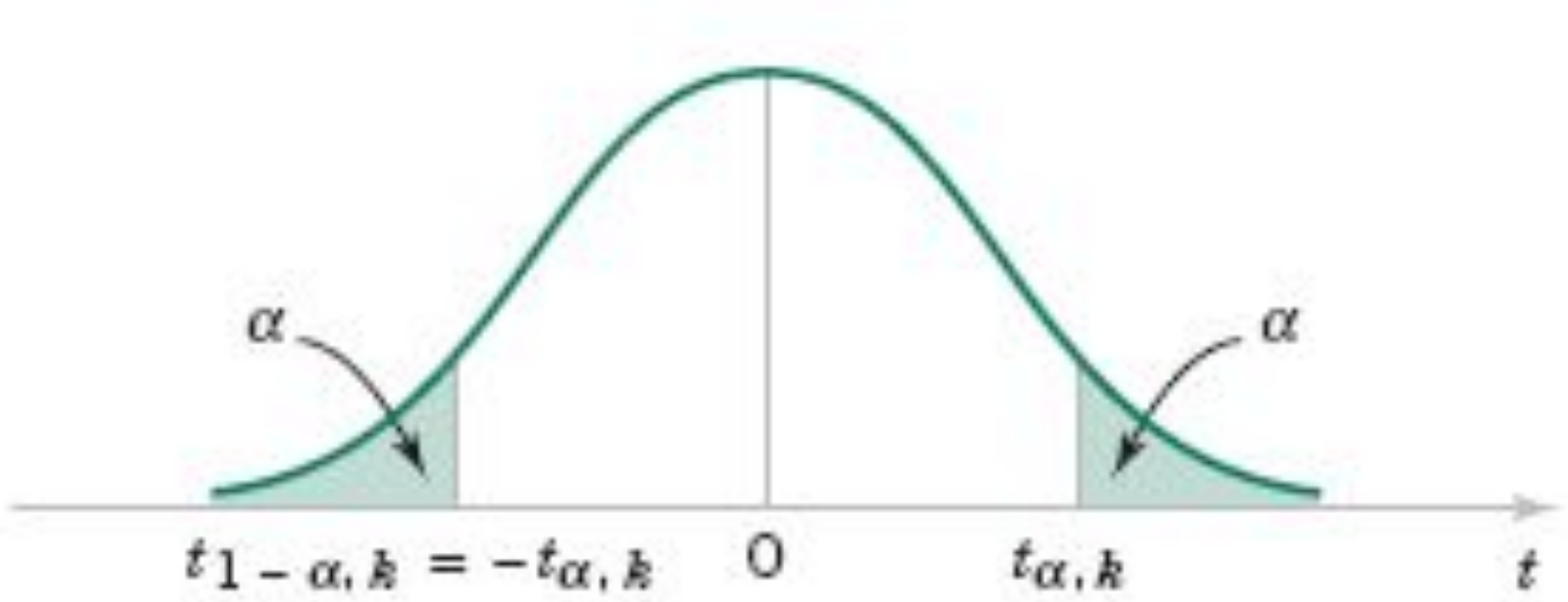
Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with unknown mean μ and unknown variance σ^2. The random variable

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \qquad (8\text{-}15)$$

has a t distribution with $n - 1$ degrees of freedom.

2. Confidence Interval on the Mean of a Normal Distribution, Variance Unknown

2.1 The *t* distribution



Percentage points of the *t* distribution.

$t_{1-\alpha,k} = -t_{\alpha,k}$

2. Confidence Interval on the Mean of a Normal Distribution, Variance Unknown

2.2 The *t* Confidence Interval on μ

If $\bar{x}$ and s are the mean and standard deviation of a random sample from a normal distribution with unknown variance σ^2, a **$100(1-\alpha)$ percent confidence interval on μ** is given by

$$\bar{x} - t_{\alpha/2,n-1}s/\sqrt{n} \le \mu \le \bar{x} + t_{\alpha/2,n-1}s/\sqrt{n} \qquad (8\text{-}18)$$

where $t_{\alpha/2,n-1}$ is the upper $100\alpha/2$ percentage point of the t distribution with $n-1$ degrees of freedom.

One-sided confidence bounds on the mean are found by replacing $t_{\alpha/2,n-1}$ in the equation with $t_{\alpha,n-1}$.

2, Confidence Interval on the Mean of a Normal Distribution, Variance Unknown

Example 8-5

An article in the journal *Materials Engineering* (1989, Vol. II, No. 4, pp. 275–281) describes the results of tensile adhesion tests on 22 U-700 alloy specimens. The load at specimen failure is as follows (in megapascals):

19.8	10.1	14.9	7.5	15.4	15.4
15.4	18.5	7.9	12.7	11.9	11.4
11.4	14.1	17.6	16.7	15.8	
19.5	8.8	13.6	11.9	11.4	

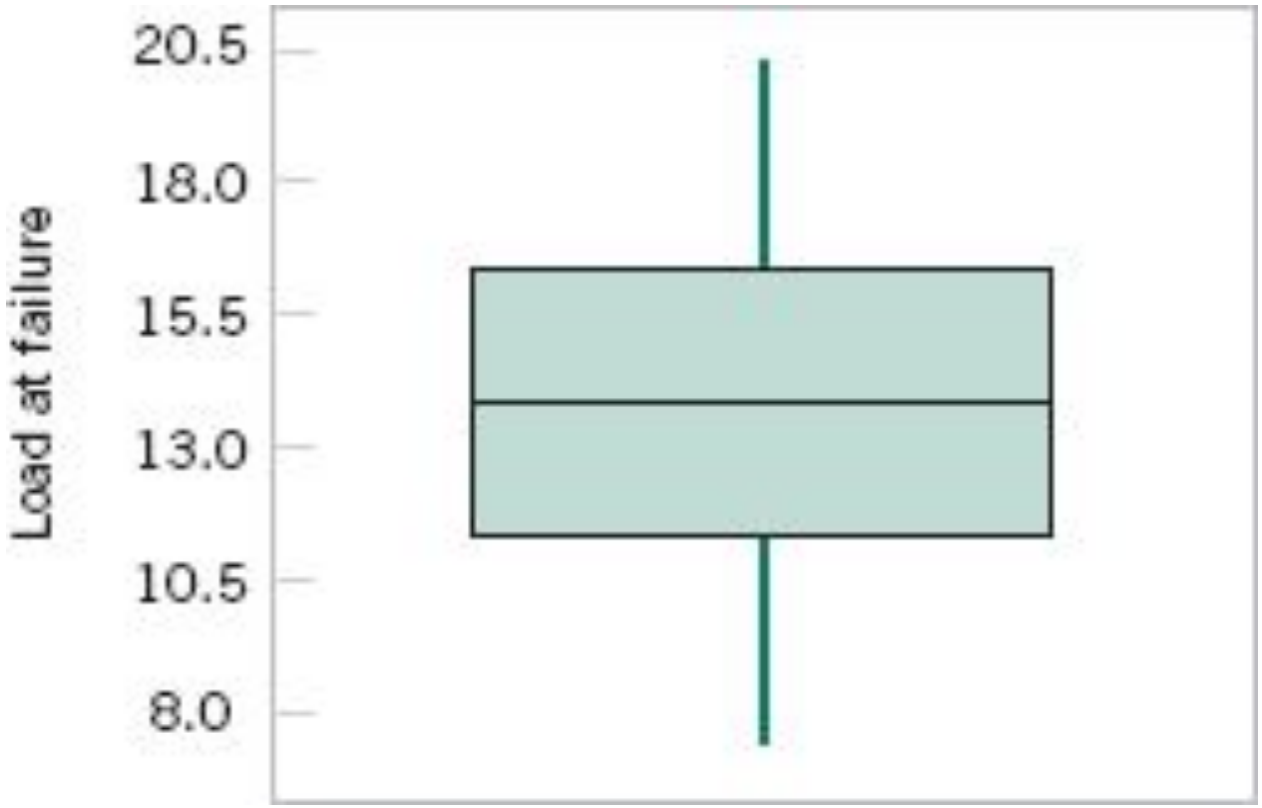
Find a 95% CI on μ.

$\bar{x} = 13.71 \quad s = 3.55$

Let's check whether we can assume that the population is normally distributed by

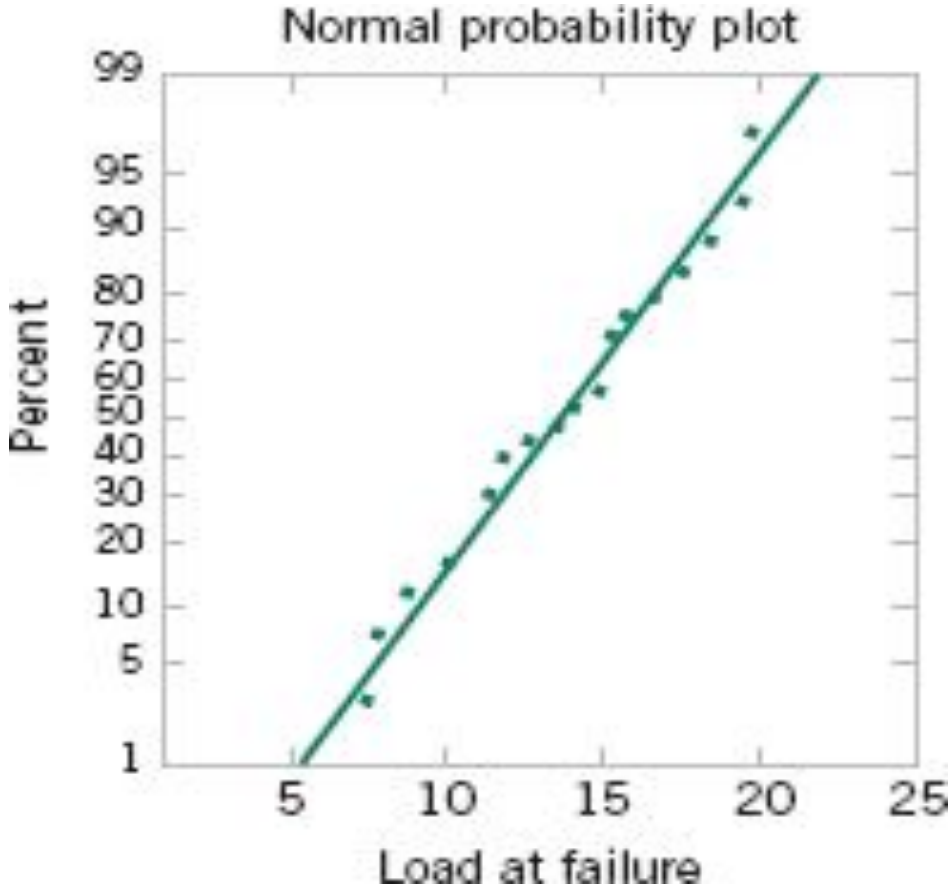
- box plot
- normal probability plot

2. Confidence Interval on the Mean of a Normal Distribution, Variance Unknown



Box and Whisker plot for the load at failure data in Example 8-5.

2. Confidence Interval on the Mean of a Normal Distribution, Variance Unknown



Normal probability plot of the load at failure data in Example 8-5.

2. Confidence Interval on the Mean of a Normal Distribution, Variance Unknown

Example 8-5 (continues)

- Box plot and normal probability plot provide good support for the assumption that the population is normally distributed.
  - Since $n = 22$, we have n-1 = 21 degrees of freedom for t.
  - $t_{0.025,21} = 2.080$
  - The resulting 95% CI is

$$\bar{x} - t_{\alpha/2,n-1}s/\sqrt{n} \le \mu \le \bar{x} + t_{\alpha/2,n-1}s/\sqrt{n}$$

$$13.71 - 2.080(3.55)/\sqrt{22} \le \mu \le 13.71 + 2.080(3.55)/\sqrt{22}$$

$$13.71 - 1.57 \le \mu \le 13.71 + 1.57$$

$$12.14 \le \mu \le 15.28$$

2. Confidence Interval on the Variance and Standard Deviation of a Normal Distribution

Definition

Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with mean μ and variance σ^2, and let S^2 be the sample variance. Then the random variable

$$X^2 = \frac{(n-1)S^2}{\sigma^2} \qquad (8\text{-}19)$$

has a chi-square (χ^2) distribution with $n-1$ degrees of freedom.

3. Confidence Interval on the Variance and Standard Deviation of a Normal Distribution

3.1 Definition

If s^2 is the sample variance from a random sample of n observations from a normal distribution with unknown variance σ^2, then **a $100(1-\alpha)\%$ confidence interval on σ^2 is**

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}} \qquad (8\text{-}21)$$

where $\chi^2_{\alpha/2,n-1}$ and $\chi^2_{1-\alpha/2,n-1}$ are the lower and upper $100\alpha/2$ percentage points of the chi-square distribution with $n-1$ degrees of freedom, respectively. A **confidence interval for σ** has lower and upper limits that are the square roots of the corresponding limits in Equation 8-21.

- Note: **Smaller chi-square value in the denominator gives larger bound (upper limit).**
- Larger chi-square value in the denominator gives **smaller bound (lower limit)**.

3. Confidence Interval on the Variance and Standard Deviation of a Normal Distribution

3.2 One-Sided Confidence Bounds

The $100(1-\alpha)\%$ lower and upper confidence bounds on σ^2 are

$$\frac{(n-1)s^2}{\chi^2_{\alpha,n-1}} \leq \sigma^2 \quad \text{and} \quad \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha,n-1}} \tag{8-22}$$

respectively.

3. Confidence Interval on the Variance and Standard Deviation of a Normal Distribution

Example 8-6

An automatic filling machine is used to fill bottles with liquid detergent. A random sample of 20 bottles results in a sample variance of fill volume of s^2=0.0153 (fluid ounces)^2. If the variance of fill volume is too large, an unacceptable proportion of bottles will be under- or overfilled. We will assume that the fill volume is approximately normally distributed. Compute a 95% upper-confidence interval for the variance.

$$\sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{0.95,19}}$$

$$\sigma^2 \leq \frac{(19)0.0153}{10.117} = 0.0287 \text{ (fluid ounce)}^2$$

$$\sigma \leq 0.17$$

4. Guidelines for Constructing Confidence Intervals

Two primary comments can help identify the analysis:
1. Determine the parameter (and the distribution of the data) that will be bounded by the confidence interval or tested by the hypothesis.
2. Check if other parameters are known or need to be estimated.

5. Bootstrap Confidence Interval

Bootstrapping Confidence Interval is a non-parametric method used to estimate the confidence interval for a population parameter (e.g., mean, median, variance) by resampling from the original data with replacement.

How Bootstrapping Works:

1. **Original Sample**: Start with an original dataset of size n.
2. **Resampling**: Randomly draw samples *with replacement* from the original dataset to create a large number of *bootstrap samples*, each of the same size n as the original sample.
3. **Statistic Calculation**: For each bootstrap sample, calculate the desired statistic (e.g., mean, median).
4. **Distribution of Bootstrap Statistics**: Construct the distribution of the statistic calculated from all the bootstrap samples.
5. **Confidence Interval Estimation**:
    1. Sort the calculated statistics and use the percentiles of this distribution to estimate the confidence interval.
    2. For example, a 95% confidence interval can be found by taking the 2.5th and 97.5th percentiles of the bootstrap distribution.

5. Bootstrap Confidence Interval – an example

Suppose we have a sample:

$$X = \{4, 5, 7, 9, 10\}$$

We want a **95% confidence interval for the mean**.

1. Original sample mean = 7.
2. Generate 1000 bootstrap resamples (each resample also has 5 values, sampled with replacement).
   i. Example bootstrap sample: {4, 7, 7, 9, 10} → mean = 7.4
   ii. Another bootstrap sample: {5, 5, 4, 7, 9} → mean = 6.0
3. Collect all 1000 bootstrap means → distribution of means.
4. Find the **2.5th percentile** = 5.9 and the **97.5th percentile** = 8.2.

So, the **bootstrap 95% CI for the mean = [5.9, 8.2]**.

6. Tolerance and Prediction Intervals

6.1 Prediction Interval for Future Observation

- Predicting the next future observation with a 100(1-α)% prediction interval
- A random sample of $X_1, X_2, \ldots, X_n$ from a normal population
- What will be X_{n+1} ?
- A point prediction of X_{n+1} is the sample mean X-bar.
- The prediction error is (X_{n+1} - X-bar). Since X_{n+1} and X-bar are independent, prediction error is normally distributed with

$$E(X_{n+1} - \bar{X}) = \mu - \mu = 0$$

$$V(X_{n+1} - \bar{X}) = \sigma^2 + \frac{\sigma^2}{n} = \sigma^2\left(1 + \frac{1}{n}\right)$$

$$Z = \frac{X_{n+1} - \bar{X}}{\sigma\sqrt{1 + \frac{1}{n}}} \qquad \text{replace } \sigma \text{ with } S \qquad T = \frac{X_{n+1} - \bar{X}}{S\sqrt{1 + \frac{1}{n}}}$$

6. Tolerance and Prediction Intervals

6.2 Prediction Interval for Future Observation

A $100(1-\alpha)\%$ prediction interval on a single future observation from a normal distribution is given by

$$\bar{x} - t_{\alpha/2,n-1}\, s\sqrt{1+\frac{1}{n}} \le X_{n+1} \le \bar{x} + t_{\alpha/2,n-1}\, s\sqrt{1+\frac{1}{n}} \qquad (8\text{-}29)$$

The prediction interval for X_{n+1} will always be longer than the confidence interval for μ.

6. Tolerance and Prediction Intervals

Example 8-9

Reconsider the tensile adhesion tests on specimens of U-700 alloy described in Example 8-5. The load at failure for n = 22 specimens was observed, and we found that x = 13.71 and s = 3.55. The 95% confidence interval on μ was $12.14 \leq \mu \leq 15.28$. We plan to test a 23rd specimen. A 95% prediction interval on the load at failure for this specimen is

$$\bar{x} - t_{\alpha/2,n-1} s\sqrt{1+\frac{1}{n}} \leq X_{n+1} \leq \bar{x} + t_{\alpha/2,n-1} s\sqrt{1+\frac{1}{n}}$$

$$13.71 - (2.080)3.55\sqrt{1+\frac{1}{22}} \leq X_{23} \leq 13.71 + (2.080)3.55\sqrt{1+\frac{1}{22}}$$

$$6.16 \leq X_{23} \leq 21.26$$

Notice that the prediction interval is considerably longer than the CI.

6. Tolerance and Prediction Intervals

6.3 Tolerance Interval for a Normal Distribution

Definition

If **μ and σ are unkown**, capturing a specific percentage of values of a population will contain less than this percentage (probably) because of sampling variability in x-bar and s

(covering)

A **tolerance interval** for capturing at least γ% of the values in a normal distribution with confidence level $100(1-\alpha)\%$ is

$$\bar{x} - ks, \qquad \bar{x} + ks$$

where k is a tolerance interval factor found in Appendix Table XII. Values are given for $\gamma = 90\%, 95\%$, and 99% and for 90%, 95%, and 99% confidence.

6. Tolerance and Prediction Intervals

Example 8-10

Consider the tensile adhesion tests originally described in Example 8-5. The load at failure for n =22 specimens was observed, and we found that $\bar{x}$ = 13.71 and s =3.55.Find a tolerance interval for the load at failure that includes 90% of the values in the population with 95% confidence.

$n = 22$, $\gamma = 0.90$, and 95% confidence ➡ k=2.264

$$(\bar{x} - ks, \bar{x} + ks) \Rightarrow [13.71 - (2.264)3.55 \, , \, 13.71 + (2.264)3.55]$$

$$(5.67 \, , \, 21.74)$$

Chapter 8

Book Back problems

Section 8.1 Problems on Confidence Interval on the Mean of a Normal Distribution, Variance Known

Formula :

1.

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

The random variable Z has a standard normal distribution.

2.

Confidence Interval on the Mean, Variance Known

If $\bar{x}$ is the sample mean of a random sample of size n from a normal population with known variance σ^2, a $100(1-\alpha)\%$ CI on μ is given by

$$\bar{x} - z_{\alpha/2}\,\sigma/\sqrt{n} \le \mu \le \bar{x} + z_{\alpha/2}\,\sigma/\sqrt{n} \qquad (8\text{-}5)$$

where $z_{\alpha/2}$ is the upper $100\alpha/2$ percentage point of the standard normal distribution.

Problems

1. For a normal population with known variance σ^2, answer the following questions:

(a) What is the confidence level for the interval $\ddot{x} - 2.14\sigma\, n \le \mu \le \ddot{x} + 2.14\sigma\, n$?

Explanation:

- The z-value is 2.14.
- We need to find the probability $P(-2.14 \le Z \le 2.14)$.

Solution: Using the standard normal distribution table:

- $P(Z \le 2.14) \approx 0.9838$
- $P(Z \le -2.14) \approx 0.0162$

So, the confidence level is:

$$P(-2.14 \le Z \le 2.14) = 0.9838 - 0.0162 = 0.9676 \text{ or } 96.76\%$$

(b) What is the confidence level for the interval $\ddot{x} - 2.49\sigma\, n \le \mu \le \ddot{x} + 2.49\sigma\, n$?

Explanation:

- The z-value is 2.49.
- We need to find the probability $P(-2.49 \leq Z \leq 2.49)$.

Solution: Using the standard normal distribution table:

- $P(Z \leq 2.49) \approx 0.9936$
- $P(Z \leq -2.49) \approx 0.0064$

So, the confidence level is:

$$P(-2.49 \leq Z \leq 2.49) = 0.9936 - 0.0064 = 0.9872 \text{ or } 98.72\%$$

(c) What is the confidence level for the interval ẍ −1.85σ n ≤ μ ≤ ẍ + 1.85σ n ?

Explanation:

- The z-value is 1.85.
- We need to find the probability $P(-1.85 \leq Z \leq 1.85)$.

Solution: Using the standard normal distribution table:

- $P(Z \leq 1.85) \approx 0.9678$
- $P(Z \leq -1.85) \approx 0.0322$

So, the confidence level is:

$$P(-1.85 \leq Z \leq 1.85) = 0.9678 - 0.0322 = 0.9356 \text{ or } 93.56\%$$

(d) What is the confidence level for the interval μ ≤ ẍ +2.00σ n ?

Explanation:

- We need to find the one-sided probability $P(Z \leq 2.00)$.

Solution: Using the standard normal distribution table:

- $P(Z \leq 2.00) \approx 0.9772$

So, the confidence level is:

$$P(\mu \leq \bar{x} + 2.00\frac{\sigma}{\sqrt{n}}) = 0.9772 \text{ or } 97.72\%$$

2. For a normal population with known variance σ^2:

(a) What value of $z_{\alpha/2}$ gives 98% confidence?

(a) 98% Confidence Level

- $\alpha = 1 - 0.98 = 0.02$
- $\alpha/2 = 0.01$

Solution: We need to find $z_{\alpha/2}$ such that $P(Z \leq z_{\alpha/2}) = 0.99$ (since 98% confidence means 1% in each tail).

From the standard normal distribution table:

- $z_{0.01} \approx 2.33$

Answer: $z_{\alpha/2} = 2.33$

(b) What value of $z_{\alpha/2}$ gives 80% confidence?

(b) 80% Confidence Level

- $\alpha = 1 - 0.80 = 0.20$
- $\alpha/2 = 0.10$

Solution: We need to find $z_{\alpha/2}$ such that $P(Z \leq z_{\alpha/2}) = 0.90$.

From the standard normal distribution table:

- $z_{0.10} \approx 1.28$

Answer: $z_{\alpha/2} = 1.28$

(c) What value of $z_{\alpha/2}$ gives 75% confidence?

(c) 75% Confidence Level

- $\alpha = 1 - 0.75 = 0.25$
- $\alpha/2 = 0.125$

Solution: We need to find $z_{\alpha/2}$ such that $P(Z \leq z_{\alpha/2}) = 0.875$.

From the standard normal distribution table:

- $z_{0.125} \approx 1.15$

Answer: $z_{\alpha/2} = 1.15$

3. Consider the one-sided confidence interval expressions for a mean of a normal population.
   (a) What value of z_{α} would result in a 90% CI?

(a) 90% One-Sided Confidence Interval

- $\alpha = 1 - 0.90 = 0.10$

Solution: We need z_{α} such that $P(Z \leq z_{\alpha}) = 0.90$.

From the standard normal distribution table:

- $z_{0.10} \approx 1.28$

Answer: $z_{\alpha} = 1.28$

   (b) What value of z_{α} would result in a 95% CI?

(b) 95% One-Sided Confidence Interval

- $\alpha = 1 - 0.95 = 0.05$

Solution: We need z_α such that $P(Z \leq z_\alpha) = 0.95$.

From the standard normal distribution table:

- $z_{0.05} \approx 1.645$

Answer: $z_\alpha = 1.645$

(c) What value of z_α would result in a 99% CI?

(c) 99% One-Sided Confidence Interval

- $\alpha = 1 - 0.99 = 0.01$

Solution: We need z_α such that $P(Z \leq z_\alpha) = 0.99$.

From the standard normal distribution table:

- $z_{0.01} \approx 2.33$

Answer: $z_\alpha = 2.33$

4. A confidence interval estimate is desired for the gain in a circuit on a semiconductor device. Assume that gain is normally distributed with standard deviation σ = 20.

To construct a confidence interval for the mean μ of a normal distribution with a known standard deviation σ, we use the following formula for the confidence interval:

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

where:

- $\bar{x}$ is the sample mean,
- σ is the known standard deviation,
- n is the sample size,
- $z_{\alpha/2}$ is the critical value from the standard normal distribution corresponding to the desired confidence level.

Given:

- $\sigma = 20$
- $\bar{x} = 1000$

(a) Find a 95% CI for μ when n = 10 and ẍ = 1000.

(a) 95% CI for μ when $n = 10$

Solution:

- $n = 10$
- $z_{\alpha/2}$ for 95% confidence = 1.96 (from standard normal distribution)

Calculate the margin of error:

$$\text{Margin of error} = 1.96 \cdot \frac{20}{\sqrt{10}} \approx 1.96 \cdot 6.32 \approx 12.39$$

Confidence interval:

$$1000 \pm 12.39 \implies (987.61, 1012.39)$$

Answer: The 95% CI for μ when $n = 10$ is approximately $(987.61, 1012.39)$.

(b) Find a 95% CI for μ when n = 25 and ẍ = 1000.

(b) 95% CI for μ when $n = 25$

Solution:

- $n = 25$
- $z_{\alpha/2}$ for 95% confidence = 1.96

Calculate the margin of error:

$$\text{Margin of error} = 1.96 \cdot \frac{20}{\sqrt{25}} \approx 1.96 \cdot 4 \approx 7.84$$

Confidence interval:

$$1000 \pm 7.84 \implies (992.16, 1007.84)$$

Answer: The 95% CI for μ when $n = 25$ is approximately $(992.16, 1007.84)$.

(c) Find a 99% CI for μ when n = 10 and ẍ = 1000.

(c) 99% CI for μ when $n = 10$

Solution:

- $n = 10$
- $z_{\alpha/2}$ for 99% confidence = 2.576

Calculate the margin of error:

$$\text{Margin of error} = 2.576 \cdot \frac{20}{\sqrt{10}} \approx 2.576 \cdot 6.32 \approx 16.29$$

Confidence interval:

$$1000 \pm 16.29 \implies (983.71, 1016.29)$$

Answer: The 99% CI for μ when $n = 10$ is approximately $(983.71, 1016.29)$.

(d) Find a 99% CI for v when n = 25 and ẍ = 1000.

(d) 99% CI for μ when $n = 25$

Solution:

- $n = 25$
- $z_{\alpha/2}$ for 99% confidence = 2.576

Calculate the margin of error:

$$\text{Margin of error} = 2.576 \cdot \frac{20}{\sqrt{25}} \approx 2.576 \cdot 4 \approx 10.304$$

Confidence interval:

$$1000 \pm 10.304 \implies (989.696, 1010.304)$$

Answer: The 99% CI for μ when $n = 25$ is approximately $(989.696, 1010.304)$.

5. A random sample has been taken from a normal distribution and the following confidence intervals were constructed using the same data: (38.02, 61.98) and (39.95, 60.05)

(a) What is the value of the sample mean?

(a) Value of the Sample Mean

A confidence interval is symmetric around the sample mean ($\bar{x}$). The sample mean can be calculated as the midpoint of the interval:

$$\bar{x} = \frac{\text{Lower bound} + \text{Upper bound}}{2}$$

1. For the interval $(38.02, 61.98)$:

$$\bar{x} = \frac{38.02 + 61.98}{2} = \frac{100}{2} = 50$$

2. For the interval $(39.95, 60.05)$:

$$\bar{x} = \frac{39.95 + 60.05}{2} = \frac{100}{2} = 50$$

(b) One of these intervals is a 95% CI and the other is a 90% CI. Which one is the 95% CI and why?

A wider confidence interval corresponds to a higher confidence level because it provides a larger range to ensure that the true mean is captured. A narrower interval corresponds to a lower confidence level.

1. **Interval** $(38.02, 61.98)$: Width = $61.98 - 38.02 = 23.96$
2. **Interval** $(39.95, 60.05)$: Width = $60.05 - 39.95 = 20.10$

Since the interval $(38.02, 61.98)$ is wider than $(39.95, 60.05)$, it corresponds to a higher confidence level.

6. A random sample has been taken from a normal distribution and the following confidence intervals were constructed using the same data: (37.53, 49.87) and (35.59, 51.81)

(a) What is the value of the sample mean?

(b) One of these intervals is a 99% CI and the other is a 95% CI. Which one is the 95% CI and why?

Do it yourself.

7. A confidence interval estimate is desired for the gain in a circuit on a semiconductor device. Assume that gain is normally distributed with standard deviation σ = 20.

(a) How large must n be if the length of the 95% CI is to be 40?

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

2E will be the length of the confidence interval.

From the above

$$E = z \cdot \frac{\sigma}{\sqrt{n}}$$

For a 95% confidence interval, the z-score is approximately 1.96.

Substituting into the margin of error formula:

$$20 = 1.96 \cdot \frac{20}{\sqrt{n}}$$

Solving for n:

$$\sqrt{n} = \frac{1.96 \cdot 20}{20}$$

$$\sqrt{n} = 1.96$$

$$n = (1.96)^2$$

$$n = 3.84$$

(b) How large must n be if the length of the 99% CI is to be 40?

Do it yourself.

2. Problems on Confidence Interval on the Mean of a Normal Distribution, Variance Unknown

***t* Distribution**

Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with unknown mean μ and unknown variance σ^2. The random variable

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \qquad (8\text{-}13)$$

has a t distribution with $n - 1$ degrees of freedom.

Confidence Interval on the Mean, Variance Unknown

If $\bar{x}$ and s are the mean and standard deviation of a random sample from a normal distribution with unknown variance σ^2, a **100(1 – α)% confidence interval on μ** is given by

$$\bar{x} - t_{\alpha/2,n-1}\, s/\sqrt{n} \le \mu \le \bar{x} + t_{\alpha/2,n-1}\, s/\sqrt{n} \qquad (8\text{-}16)$$

where $t_{\alpha/2,n-1}$ is the upper $100\alpha/2$ percentage point of the t distribution with $n - 1$ degrees of freedom.

1. Find the values of the following percentiles: $t_{0.025,15}$, $t_{0.05,10}$, $t_{0.10,20}$, $t_{0.005,25}$, and $t_{0.001,30}$.

1. $t_{0.025,15}$ (with 15 degrees of freedom, $\alpha = 0.025$):
   - $t_{0.025,15} \approx 2.131$
2. $t_{0.05,10}$ (with 10 degrees of freedom, $\alpha = 0.05$):
   - $t_{0.05,10} \approx 1.812$
3. $t_{0.10,20}$ (with 20 degrees of freedom, $\alpha = 0.10$):
   - $t_{0.10,20} \approx 1.325$
4. $t_{0.005,25}$ (with 25 degrees of freedom, $\alpha = 0.005$):
   - $t_{0.005,25} \approx 2.787$
5. $t_{0.001,30}$ (with 30 degrees of freedom, $\alpha = 0.001$):
   - $t_{0.001,30} \approx 3.646$

2. Determine the t-percentile that is required to construct each of the following two-sided confidence intervals:
   (a) Confidence level = 95%, degrees of freedom = 12
   (b) Confidence level = 95%, degrees of freedom = 24
   (c) Confidence level = 99%, degrees of freedom = 13
   (d) Confidence level = 99.9%, degrees of freedom = 15

$$\alpha = 1 - \text{Confidence level}$$

$$\alpha/2 = \frac{1 - \text{Confidence level}}{2}$$

(a) 95% Confidence Level, Degrees of Freedom = 12

- $\alpha = 1 - 0.95 = 0.05$
- $\alpha/2 = 0.025$

Critical value: $t_{0.025,12} \approx 2.179$

(b) 95% Confidence Level, Degrees of Freedom = 24

- $\alpha = 1 - 0.95 = 0.05$
- $\alpha/2 = 0.025$

Critical value: $t_{0.025,24} \approx 2.064$

(c) 99% Confidence Level, Degrees of Freedom = 13

- $\alpha = 1 - 0.99 = 0.01$
- $\alpha/2 = 0.005$

Critical value: $t_{0.005,13} \approx 3.012$

(d) 99.9% Confidence Level, Degrees of Freedom = 15

- $\alpha = 1 - 0.999 = 0.001$
- $\alpha/2 = 0.0005$

Critical value: $t_{0.0005,15} \approx 4.073$

3. Determine the *t*-percentile that is required to construct each of the following one-sided confidence intervals:

(a) Confidence level = 95%, degrees of freedom = 14

(b) Confidence level = 99%, degrees of freedom = 19

(c) Confidence level = 99.9%, degrees of freedom = 24

For one-sided confidence intervals, we need to find the t-percentile $t_{\alpha,\nu}$, where α represents the upper-tail probability. For a one-sided confidence interval, $\alpha = 1 - \text{confidence level}$.

(a) 95% Confidence Level, Degrees of Freedom = 14

- $\alpha = 1 - 0.95 = 0.05$

Critical value: $t_{0.05,14} \approx 1.761$

(b) 99% Confidence Level, Degrees of Freedom = 19

- $\alpha = 1 - 0.99 = 0.01$

Critical value: $t_{0.01,19} \approx 2.539$

(c) 99.9% Confidence Level, Degrees of Freedom = 24

- $\alpha = 1 - 0.999 = 0.001$

Critical value: $t_{0.001,24} \approx 3.746$

4. A random sample has been taken from a normal distribution.

Output from a software package follows:

Variable	N	Mean	SE Mean	StDev	Varianc	Sum
x	10	?	0.507	1.605	?	251.848

(a) Fill in the missing quantities.

(b)

Part (a): Fill in the Missing Quantities

1. **Mean**:

   The mean can be calculated using the sum of the sample values and the sample size.

$$\text{Mean} = \frac{\text{Sum}}{N} = \frac{251.848}{10} = 25.1848$$

2. **Variance**:

   The variance is the square of the standard deviation.

$$\text{Variance} = (\text{StDev})^2 = (1.605)^2 = 2.576025$$

So, the missing quantities are:

- Mean = 25.1848

(b) Find a 95% CI on the population mean.

Part (b): Find a 95% Confidence Interval for the Population Mean

To find the 95% confidence interval (CI) for the population mean, we can use the formula:

$$\text{CI} = \text{Mean} \pm t \cdot \text{SE Mean}$$

For a sample size of $N = 10$, the degrees of freedom are $N - 1 = 9$. For a 95% confidence level, the critical t-value (two-tailed) for 9 degrees of freedom is approximately 2.262.

Now, we can calculate the margin of error and the confidence interval:

1. **Margin of Error**:

$$\text{Margin of Error} = t \cdot \text{SE Mean} = 2.262 \times 0.507 = 1.146834$$

2. **Confidence Interval**:

$$\text{CI} = 25.1848 \pm 1.146834$$

$$\text{CI} = (25.1848 - 1.146834, 25.1848 + 1.146834)$$

$$\text{CI} = (24.038, 26.3316)$$

5. A random sample has been taken from a normal distribution.

Output from a software package follows:

Variable	N	Mean	SE Mean	StDev	Varianc	Sum
x	?	?	1.58	6.11	?	751.40

(a) Fill in the missing quantities.

We can calculate N using the formula for the standard error of the mean:

$$\text{SE Mean} = \frac{\text{StDev}}{\sqrt{N}}$$

Solving for N:

$$N = \left(\frac{\text{StDev}}{\text{SE Mean}}\right)^2$$

Substitute the given values:

$$N = \left(\frac{6.11}{1.58}\right)^2$$

$$N = (3.86582)^2$$

$$N \approx 14.94 \approx 15$$

(b) Find a 95% CI on the population mean

Do as in previous problem

5. A research engineer for a tire manufacturer is investigating tire life for a new rubber compound and has built 16 tires and tested them to end-of-life in a road test. The sample mean and standard deviation are 60,139.7 and 3645.94 kilometers. Find a 95% confidence interval on mean tire life.

To find the 95% confidence interval for the mean tire life, we will use the formula for the confidence interval when the sample size is small (i.e., $n < 30$) and the population standard deviation is unknown:

$$\text{Confidence Interval} = \bar{x} \pm t_{\alpha/2,\nu} \cdot \frac{s}{\sqrt{n}}$$

Where:

- $\bar{x} = 60,139.7$ (sample mean),
- $s = 3645.94$ (sample standard deviation),
- $n = 16$ (sample size),
- $t_{\alpha/2,\nu}$ is the critical value from the t-distribution with $\nu = n - 1 = 16 - 1 = 15$ degrees of freedom and $\alpha = 0.05$ (for a 95% confidence level. $\alpha/2 = 0.025$).

Step 1: Find the t-critical value

For a 95% confidence interval with 15 degrees of freedom, the t-critical value $t_{0.025,15}$ is approximately **2.131** (you can find this value in a t-distribution table or using statistical software).

Step 2: Calculate the standard error

The standard error of the mean is given by:

$$SE = \frac{s}{\sqrt{n}} = \frac{3645.94}{\sqrt{16}} = \frac{3645.94}{4} = 911.485$$

Step 3: Calculate the confidence interval

Now we can calculate the confidence interval:

$$\text{Confidence Interval} = \bar{x} \pm t_{\alpha/2,\nu} \cdot SE = 60,139.7 \pm 2.131 \cdot 911.485$$

$$\text{Confidence Interval} = 60,139.7 \pm 1949.08$$

$$\text{Confidence Interval} = (58,190.62, 62,088.78)$$

6. An Izod impact test was performed on 20 specimens of PVC pipe. The sample mean is ẍ = 1.25 and the sample standard deviation is s = 0.25. Find a 99% lower confidence bound on Izod impact strength.

$$\text{Lower confidence bound} = \bar{x} - t_{\alpha,n-1} \times \frac{s}{\sqrt{n}}$$

Given :

Sample mean (ẍ) = 1.25

Sample standard deviation (s) = 0.25

Sample size (n) = 20

Confidence level = 99%

For a **99% lower confidence bound**, the critical value corresponds to the one-tailed 1% significance level (α=0.01) with n−1=19 degrees of freedom. We need $t_{0.01,19}$.

From a **t-distribution table**:

- $t_{0.01,19} \approx 2.539$

$$\text{Standard error} = \frac{s}{\sqrt{n}} = \frac{0.25}{\sqrt{20}} \approx \frac{0.25}{4.472} \approx 0.0559$$

Calculate the Lower Confidence Bound

$$\text{Lower confidence bound} = 1.25 - 2.539 \times 0.0559 \approx 1.25 - 0.142$$

$$\text{Lower confidence bound} \approx 1.108$$

The 99% lower confidence bound on the Izod impact strength is approximately **1.108**.

7. A postmix beverage machine is adjusted to release a certain amount of syrup into a chamber where it is mixed with carbonated water. A random sample of 25 beverages was found to have a mean syrup content of ẍ = 1.10 fluid ounce and a standard deviation of s = 0.015 fluid ounce. Find a 95% CI on the mean volume of syrup dispensed.

$$\text{CI} = \bar{x} \pm t_{\alpha/2,n-1} \times \frac{s}{\sqrt{n}}$$

Given

- Sample mean (ẍ) = 1.10 fluid ounces
- Sample standard deviation (s) = 0.015 fluid ounces
- Sample size (n) = 25
- Confidence level = 95%

For a **95% confidence bound**, the critical value corresponds to the one-tailed 1% significance level (α=0.05, α /2=0.025) with n−1=24 degrees of freedom. We need $t_{0.025,24}$.

From a **t-distribution table**:

- $t_{0.025,24} \approx 2.064$

Calculate the Standard Error of the Mean

$$\text{Standard error} = \frac{s}{\sqrt{n}} = \frac{0.015}{\sqrt{25}} = \frac{0.015}{5} = 0.003$$

Calculate the Confidence Interval

$$\text{Margin of error} = t_{\alpha/2,n-1} \times \text{Standard error} = 2.064 \times 0.003 \approx 0.0062$$

$$\text{CI} = 1.10 \pm 0.0062$$

Confidence Interval

$$\text{Lower bound} = 1.10 - 0.0062 = 1.0938$$

$$\text{Upper bound} = 1.10 + 0.0062 = 1.1062$$

The 95% confidence interval for the mean volume of syrup dispensed is approximately **(1.0938, 1.1062)** fluid ounces.

3. Confidence Interval on the Variance and Standard Deviation of a Normal Distribution

Formula :

Confidence Interval on the Variance

If s^2 is the sample variance from a random sample of n observations from a normal distribution with unknown variance σ^2, then a **100(1 – α)% confidence interval on σ^2** is

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}} \qquad (8\text{-}19)$$

One-Sided Confidence Bounds on the Variance

The 100(1 – α)% lower and upper confidence bounds on σ^2 are

$$\frac{(n-1)s^2}{\chi^2_{\alpha,n-1}} \le \sigma^2 \quad \text{and} \quad \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\alpha,n-1}}$$

respectively.

1. Determine the values of the following percentiles: $\chi^2_{0.05\ 10}$, $\chi^2_{0.05\ 15}$, $\chi^2_{0.01\ 12}$, $\chi^2_{0.95\ 20}$ $\chi^2_{0.99\ 18}$

   $\chi^2_{0.05\ 10}$. This represents the critical value at the 95th percentile (5% significance level) with 10 degrees of freedom. Therefore, $\chi^2_{0.05\ 10}$= 18.307

   $\chi^2_{0.05\ 15}$, : This is the critical value at the 95th percentile with 15 degrees of freedom. Therefore $\chi^2_{0.05\ 15}$= 24.99

$\chi^2_{0.01\,12}$, This is the critical value at the 99th percentile (1% significance level) with 12 degrees of freedom. Therefore, $\chi^2_{0.01\,12}=26.217$

$\chi^2_{0.95\,20}$ This is the critical value at the 5th percentile with 20 degrees of freedom. Therefore, $\chi^2_{0.95\,20}= 10.85$

$\chi^2_{0.99\,18}$ This is the critical value at the 99th percentile with 18 degrees of freedom. Therefore, $\chi^2_{0.99\,18}=34.805$

2. Determine the χ^2 percentile that is required to construct each of the following CIs:
   (a) Confidence level = 95%, degrees of freedom = 24, one sided (upper)
   (b) Confidence level = 99%, degrees of freedom = 9, one-sided (lower)
   (c) Confidence level = 90%, degrees of freedom = 19, two-sided.

   (a)For a **one-sided upper bound** at a 95% confidence level, we need the critical value at the **5% significance level** (α=0.05, 1- α =0.95) for the upper tail.
   $\chi^2_{0.95,24}$ =13.85.

(b) For a **one-sided lower bound** at a 99% confidence level, we need the critical value at the **1% significance level** (α=0.01) for the lower tail.
   For lower bounds, we look at the χ^2 value such that the left area under the curve is 1% (or the 1st percentile).
   $\chi^2_{0.01,9}$=21.67

(c) For a **two-sided** 90% confidence interval, the critical values are found for $\alpha/2$=0.05 for both tails (i.e., 5% in each tail).
   $\chi^2_{0.05,19}$=30.14 (lower bound)
   $\chi^2_{0.95,19}$=10.12 (upper bound)

3. A rivet is to be inserted into a hole. A random sample of n = 15 parts is selected, and the hole diameter is measured. The sample standard deviation of the hole diameter measurements is s = 0.008 millimeters. Construct a 99% lower confidence bound for σ^2.

   Lower confidence : $\frac{(n-1)s^2}{\chi^2_{\alpha,n-1}} \leq \sigma$

   Given :

- Sample size (n) = 15
- Sample standard deviation (s) = 0.008 millimeters
- Significance level for a 99% confidence bound (α) = 0.01

   Calculate the Sample Variance
   $s2=(0.008)_2=0.000064$
   Determine the Critical Value

For a one-sided lower confidence bound at a 99% confidence level with n−1=14 degrees of freedom:
We need the critical value $\chi^2_{0.01,14}$ =29.14
Calculate the Lower Confidence Bound
Lower confidence bound for σ^2=(15−1)·0.000064 / (29.14)= 0.000896/29.14
=0.00003mm^2

The 99% lower confidence bound for the population variance σ^2 is approximately =0.00003

4. The sugar content of the syrup in canned peaches is normally distributed. A random sample of n = 10 cans yields a sample standard deviation of s = 4.8 milligrams. Calculate a 95% two-sided confidence interval for σ.

To calculate a 95% two-sided confidence interval for the population standard deviation (σ) based on the sample standard deviation (s), we can use the **Chi-square distribution**.

Formula for confidence interval on variance σ^2

$$\frac{(n-1)s^2}{\chi^2_{1-\alpha/2,\,n-1}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{\alpha/2,\,n-1}}$$

Then take square roots for **σ**: The formula for the confidence interval for the population

$$\left(\sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}}, \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}}\right)$$

standard deviation σ is:

Where:

- n is the sample size (10 cans),
- s is the sample standard deviation (4.8 milligrams),
- $\chi_{\alpha/2,n-1}$ $\chi_{\alpha/2,n-1}$ are the critical values of the Chi-square distribution with n−1 degrees of freedom (9 degrees of freedom in this case) for the given significance level α=0.05 (for a 95% confidence interval).

Step 1: Find the Chi-square critical values

For $\alpha = 0.05$ and $n - 1 = 9$ degrees of freedom:

- $\chi^2_{\alpha/2,9} = \chi^2_{0.025,9}$
- $\chi^2_{1-\alpha/2,9} = \chi^2_{0.975,9}$

Using a Chi-square table or calculator, the critical values are approximately:

- $\chi^2_{0.025,9} \approx 2.700$
- $\chi^2_{0.975,9} \approx 16.919$

Step 2: Apply the formula

Now, we can calculate the confidence interval for σ:

$$\left(\sqrt{\frac{(10-1)\times(4.8)^2}{16.919}}, \sqrt{\frac{(10-1)\times(4.8)^2}{2.700}}\right)$$

$$\left(\sqrt{12.26}, \sqrt{76.8}\right)$$

$$(3.5, 8.77)$$

The 95% confidence interval for the population standard deviation σ is approximately **(3.5, 8.77)** milligrams.

5. The percentage of titanium in an alloy used in aerospace castings is measured in 51 randomly selected parts. The sample standard deviation is s = 0.37. Construct a 95% two sided confidence interval for σ.
   Solve as above example

Problems on Prediction and tolerance intervals:

1. A research engineer for a tire manufacturer is investigating tire life for a new rubber compound and has built 16 tires and tested them to end-of-life in a road test. The sample mean and standard deviation are 60,139.7 and 3645.94 kilometers. Find a 95% confidence interval on mean tire life. Compute a 95% prediction interval on the life of the next tire of this type tested under conditions that are similar to those employed in the original test. Compare the length of the prediction interval with the length of the 95% CI on the population mean.

To find the 95% confidence interval for the mean tire life, we will use the formula for the confidence interval when the sample size is small (i.e., $n < 30$) and the population standard deviation is unknown:

$$\text{Confidence Interval} = \bar{x} \pm t_{\alpha/2,\nu} \cdot \frac{s}{\sqrt{n}}$$

Where:

- $\bar{x} = 60,139.7$ (sample mean),
- $s = 3645.94$ (sample standard deviation),
- $n = 16$ (sample size),
- $t_{\alpha/2,\nu}$ is the critical value from the t-distribution with $\nu = n - 1 = 16 - 1 = 15$ degrees of freedom and $\alpha = 0.05$ (for a 95% confidence level. $\alpha/2 = 0.025$).

Confidence Interval :

Step 1: Find the t-critical value

For a 95% confidence interval with 15 degrees of freedom, the t-critical value $t_{0.025,15}$ is approximately **2.131** (you can find this value in a t-distribution table or using statistical software).

Step 2: Calculate the standard error

The standard error of the mean is given by:

$$SE = \frac{s}{\sqrt{n}} = \frac{3645.94}{\sqrt{16}} = \frac{3645.94}{4} = 911.485$$

Step 3: Calculate the confidence interval

Now we can calculate the confidence interval:

$$\text{Confidence Interval} = \bar{x} \pm t_{\alpha/2,\nu} \cdot SE = 60,139.7 \pm 2.131 \cdot 911.485$$

$$\text{Confidence Interval} = 60,139.7 \pm 1949.08$$

$$\text{Confidence Interval} = (58,190.62, 62,088.78)$$

Prediction Interval :

Formula,

$$\bar{x} - t_{\alpha/2,n-1} s \sqrt{1 + \frac{1}{n}} \leq X_{n+1} \leq \bar{x} + t_{\alpha/2,n-1} s \sqrt{1 + \frac{1}{n}}$$

95% prediction interval for the next tire

$$\bar{x} \pm t_{0.975,15}\, s\sqrt{1+\tfrac{1}{n}} = 60{,}139.7 \pm 2.131\,(3645.94)\sqrt{1+\tfrac{1}{16}}$$

$$\sqrt{1+\tfrac{1}{16}} = \sqrt{1.0625} = 1.030776 \Rightarrow 2.131 \times 3645.94 \times 1.030776 \approx 8{,}009$$

$$\boxed{(52{,}131,\ 68{,}148)\ \text{km}}$$

CI length for the mean: 2×1,943≈3,885 km

Prediction interval length: 2×8,009≈16,0182km

2. A research engineer for a tire manufacturer is investigating tire life for a new rubber compound and has built 16 tires and tested them to end-of-life in a road test. The sample mean and standard deviation are 60,139.7 and 3645.94 kilometers.

a. Find a 95% confidence interval on mean tire life.
b. Compute a 95% tolerance interval on the life of the tires that has confidence level 95%.
c. Compare the length of the tolerance interval with the length of the 95% CI on the population mean.
d. Which interval is shorter?

a. Find a 95% confidence interval on mean tire life : **Same as above problem**

b. Compute a 95% tolerance interval on the life of the tires that has confidence level 95%.

60,139.7+2.903 (3645.94) and 60,139.7-2.903 (3645.94)

That is. 60,139.7±2.903* 3645.94

Tolerance interval is from 49,555.681 to 70,723.86

c. **Length of the tolerance interval is 21,168.32**
95% CI for the mean: length ≈ 62,088.78-58,190.62=3898.16
d. Which interval is shorter?
CI is shorter