1. Estimation of parameter
2. Unbiased estimates
3. Efficient estimates
4. Point Estimates and intervals
5. Confidence intervals estimation of population parameters
6. Probable errors

---

a) Statistical hypothesis
b) Test of hypothesis and significance/decision rules
c) Type-I errors, Type-II errors
d) Level of Significance
e) Test involving normal distribution
f) One tailed and two tailed test
g) Special test

---

i. Operating Characteristic curve
ii. The power of a test
iii. P-Value of hypothesis test
iv. Control chart
v. Test involving sample differences
vi. Test involving binomial distributions

---

1. **Estimation of parameter**

   In most statistical studies, the population parameters are unknown and must be estimated. Therefore, developing methods for estimating as accurately as possible the values of population parameters is an important part of statistical analysis.

   (Estimators = random variables used to estimate population parameters

   Estimates = specific values of the population parameters

   Estimates need not have a single value; instead, the estimate could be a range of values

   Point estimate = estimate that specifies a single value of the population

   Interval estimate = estimate that specifies a range of values)

2. **Unbiased estimates**

   Expected value = the true value of the parameter,

   For example, $E(\bar{x}) = \mu$, $E(s^2) = \sigma^2$.

3. **Efficient estimates:**

   An estimator with smaller variance is called efficient.

4. **Point Estimates and intervals**

   Estimators are the generalized mathematical expressions for the calculation of sample statistic, and an estimate is a specific outcome of one estimation.

   Estimates take on a single numerical value and are, therefore, referred to as **point estimates.**

   It is a fixed number specific to that sample.

   It has no sampling distribution.

   We would then say that for example, the **point estimate for the population mean** $\hat{\mu} = \bar{x}$, where the ^ on top of the population parameter μ indicates that , that is the parameter we are estimating; i.e., the ^ signifies that that value is an estimate (also called estimator). In words, the expression $\hat{\mu} = \bar{x}$

   - An estimation of population parameter given by two numbers between which population parameter is expected to lie along with an associated level of confidence.
   - An interval estimate is arrange of values used to estimate population parameters.
   - An interval estimation is better than point estimation because it indicates the range.
   - **Point estimate ± Reliability factor × Standard error**

5. Confidence intervals estimation of population parameters:

In contrast, a **confidence interval** (CI) specifies a range that contains the parameter in which we are interested (1 –) % at the time.

The $(1 - \alpha)$ % is known as the degree of confidence.

Confidence intervals are generally expressed as Lower confidence limit or upper confidence limit.

To estimate an actual population mean μ, we can only obtain x¯, the mean of a sample randomly selected from the population of interest. We can use x¯ to find a range of values: Lower value < population mean (μ) < Upper value

The range of values is called a "**confidence interval**."

The purpose of taking a random sample from a lot or population and computing a statistic, such as the mean from the data, is to approximate the mean of the population. How well the sample statistic estimates the underlying population value is always an issue. A confidence interval addresses this issue because it provides a range of values which is likely to contain the population parameter of interest.

Confidence intervals are constructed at a *confidence level*, such as 95 %, selected by the user. What does this mean? It means that if the same population is sampled on numerous occasions and interval estimates are made on each occasion, the resulting intervals would bracket the true population parameter in approximately 95 % of the cases. A confidence stated at a 1−α level can be thought of as the inverse of a significance level, α.

6. Probable errors

S.E enables us to determine the **probable limits** within which the population parameter may be expected to lie.

For example: the probable limits for population proportion P are given by $p \pm 3\sqrt{\frac{pq}{n}}$

(If Population P is not known to us then taking sample proportion p us an estimate of P)

The limit for P at level of significance α are given by $p \pm z_\alpha\sqrt{\frac{pq}{n}}$

i. 95% confidence limits for P are given by $p \pm 1.96\sqrt{\frac{pq}{n}}$

ii. 99% confidence limits for P are given by $p \pm 2.58\sqrt{\frac{pq}{n}}$

Note:

- Properties of good estimate:

  Unbiasedness: An estimate is called unbiased if it is expected to draw, such a value from sample which is equal to population parameter being estimated.

  Efficiency: An estimator with smaller variance is called efficient

  Consistency: An estimator is said to be consistence if its expected value get closer and closer to the parameter being estimated as the sample size increase.

  Sufficiency: An estimator is sufficient if it conveys al the informations the sample can furnish for the estimation of the parameter being estimated.

a) Statistical hypothesis:

Hypothesis Testing:

Hypothesis as a word means a guess or assumed idea.

It is difficult to assert the population parameter like mean of the population

So a sample from the population is selected and difference between the sample statistic and population parameter is analyzed.

Such a process of analyzing the difference between the statistic and parameter is called **TEST OF HYPOTHESIS.**

Hypothesis testing allows us to use a sample to decide between two statements made about a Population characteristic.

Population Characteristics are things like "The mean of a population" or "the proportion of the population who have a particular property".

These two statements are called the Null Hypothesis and the Alternative Hypothesis.

Definitions

H_0: ***The Null Hypothesis*** This is the hypothesis or claims that is initially assumed to be true.

H_A: ***The Alternative Hypothesis*** This is the hypothesis or claim which we initially assume to be false but which we may decide to accept if there is sufficient evidence.

H_0 and H_A take the following form:

- **Null Hypothesis** H_0 : Population Characteristic = Hypothesised Value
- **Alternative Hypothesis** ~ three possibilities

1. **Upper tailed test:** H_A: Population Characteristic > Hypothesised

2. **Lower tailed test:** H_A : Population Characteristic < Hypothesised Value

3. **Two tailed test:** H_A : Population Characteristic ≠ Hypothesised Value NOTE: The same Hypothesized value must be used in the Alternative Hypothesis as in the Null Hypothesis

b) Test of hypothesis and significance/decision rules

The Five Step Model

1. Make assumptions and meet test requirements.
2. State the H_0 and H_1.
3. Select the Sampling Distribution and Determine the Critical Region.
4. Calculate the test statistic.
5. Make a Decision and Interpret Results.

If test statistic > tabulated value or z_α at given L.O.S. then we reject Null hypothesis Otherwise we accept it.

i.e. if test statistic < z_α at given L.O.S. then we accept Null hypothesis

	Mean	Proportion
Single	$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$	$z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$ Q= 1-P
Double	$z = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}}$	$z = \frac{p_1 - p_2}{\sqrt{PQ(\frac{1}{{n_1}^2} + \frac{1}{{n_2}^2})}}$ P= $\frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$, Q= 1 - P

c) Type-I errors, Type-II errors:

Type I errors: We reject the Null Hypothesis even though the Null Hypothesis is true.

P (Reject H_0/ H_0 is true)

Type II errors: We do not reject the Null Hypothesis when in fact the Null Hypothesis is false and the Alternative is true.

P (Accept H_0 / H_0 is False) = P (Accept H_0 / H_A is True)

Test Result	Ho True	Ho false
Ho True	Correct decision	Type I error
Ho False	Type II error	Correct decision

d) Level of Significance:

The probability of a Type I error is called the **Level Of Significance** of the Hypothesis Test and is denoted by α -alpha.

The probability of a Type II error is denoted by β.

$$\alpha = P(Type\ I\ \ Error) \quad \beta = P(Type\ II\ \ Error)$$

We choose α to be small (.01, .05 or .1) but we cannot completely eliminate the probability of a Type I error, as we mentioned already α and β are related *the only way to reduce* β *without increasing* α *is to increase the sample size.*

Type I errors are generally considered the more serious so in our testing procedure we control the probability of these errors (α) and usually are unaware of the probability of Type II errors (β).

Rejection Regions (R.R)/ Critical Region (C.R):

The region where null hypothesis is rejected.

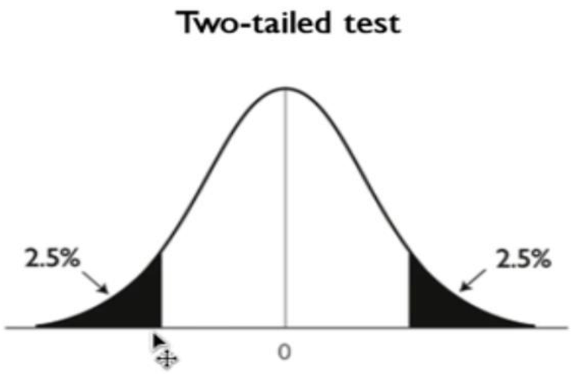
- One way of performing a Hypothesis test is to compute a rejection region. If we find that our test statistic (which we measure from our sample) is in this region then we reject our NULL Hypothesis.
- The computation of the rejection region is mathematical and involves us using statistical tables like the Normal tables.
- What is important is the idea that the rejection region is a region far away from our Null hypothesis.
- And that it is unlikely that we would observe a sample with a value of the test statistic (for example the sample mean or sample proportion) this far away from the Null Hypothesized value if that Null Hypothesis was true.

e) One tailed and two tailed test:

H_0 and H_A take the following form:

- **Null Hypothesis** H_0 : Population Characteristic = Hypothesised Value
- **Alternative Hypothesis** ~ three possibilities

- **Upper tailed test:** **H_A**: Population Characteristic > Hypothesised
- **Lower tailed test:** **H_A** : Population Characteristic < Hypothesised Value
- **Two tailed test:** **H_A** : Population Characteristic ≠ Hypothesised Value
- NOTE: The same Hypothesized value must be used in the Alternative Hypothesis as in the Null Hypothesis

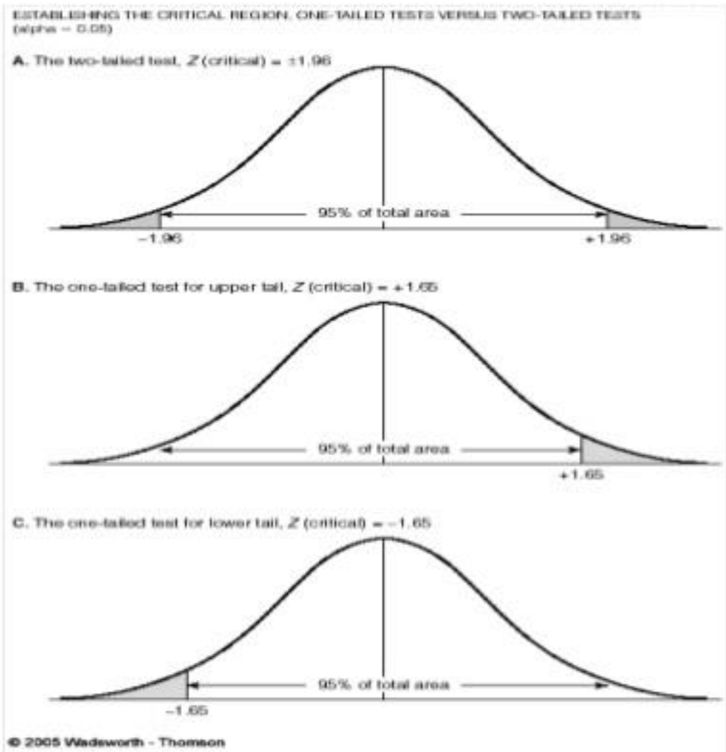


The Curve for Two- vs. One-tailed Tests at α = .05:

Two-tailed test:
"is there a **significant difference**?"

One-tailed tests:
"is the sample mean **greater** than μ or P_u?"

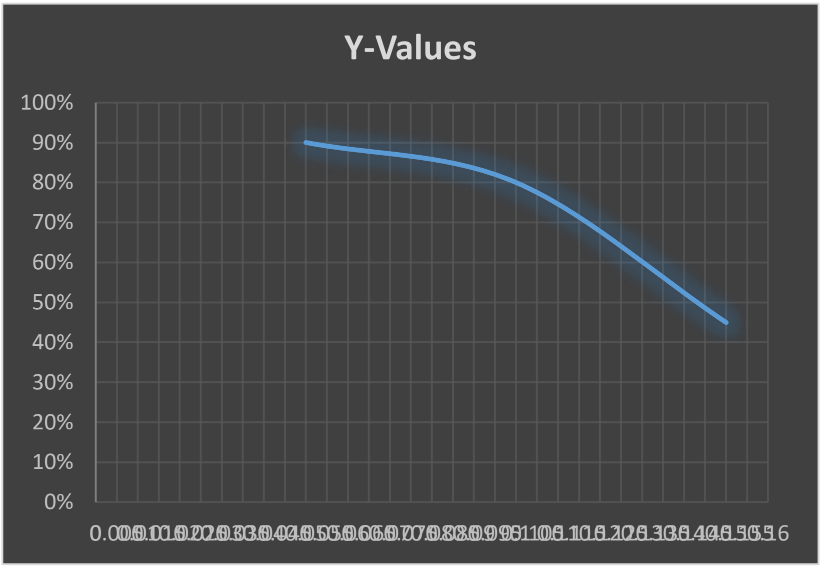
"is the sample mean **less** than μ or P_u?"



i. Operating Characteristic curve:

An operating characteristic (OC) curve is a chart that's display probability of acceptance versus percentage of defective items or loss. With no defects, we'll surely have 100% acceptance! But take a look at 5%. At this point, there is still 90% acceptance. Then the curve start dropping.

As a rule, if the curve is steeper(rising or falling sharply), it indicates a better sampling plan.



ii. The power of a test:

The strength of a test is known as power function or power of the test it is define as

Power of the test = 1 – p(type –II error)

OR Power of the test = $1 - \beta$

iii. P-Value of hypothesis test:

- The P-value approach involves determining "likely" or "unlikely" by determining the probability-assuming the nul hypothesis were true of observing a more extreme test statistic in the direction of the alternative hypothesis than the one observed.
- If p-Value is small (say less than or equal to) then it is unlikely, and P-value is more than α then it is likely.

- If P-value is less than or equal to α then Null hypothesis is rejected in favour of alterative hypothesis.
- If P-value > α then Null hypothesis is not rejected.
- Using P-value approach to conducting any hypothesis test are;

a) Specify Null and Alternative hypothesis
b) Calculate the value of test statistic by using the sample data and assuming null hypothesis is true.
c) Using the known distribution of test statistic calculate the P-value.
d) Set the significance level α, The probability of making type-I error to be small-1%, 5%, 10% L.O.S. compose P-value to α
e) If P-value > α do not reject null hypothesis.

iv. Control chart;

The **control chart** is a graph used to study how a process changes over time. Data are plotted in time order. A **control chart** always has a central line for the average, an upper line for the upper **control** limit and a lower line for the lower **control** limit. These lines are determined from historical data.

Upper Control chart

Central Line

Lower Control chart

Time or order of production

Control limits also known as Natural Process Limits, are horizontal lines drawn on a statistical process control chart. Usually at distance of $\pm$ 3 standard deviation of the plotted statistic from the statistic's mean.

Control Chart also known as Shewhart Chart or process behaviour chars, are a statistical process control tool used to determine if a manufacturing or business process is in a state of control.

A control chart always has a central line for the averages, an upper line for the upper control limit and lower line for the lower control limit. These lines are determined from the historical data.

Control chars re used routinely monitor quality.
In general the chart contains centre line that represent the mean value for the in-control chart process. Two other horizontal lines, called UPPER CONTROL LINE (UCL) and LOWER CONTROL LINE (LCL)

Note: The Primary Statistical process Control (SPC) tool for six sigma initiative is the control chart a graphical tracking of a process input or an output over time.
In a control chart these track measurements are visually compared to decision limits calculated from probabilities of the actual process performance.
SPC is an industry-standard methodology fro measuring and quality during the manufacturing process. Quality data in the form of process or product measurements are obtained in real time during manufacturing.