FAIRFIELD Institute of Management & Technology
(Affiliated to GGSIP University, New Delhi)
तेजस्वि नावधीतमस्तु
ISO 9001:2008 & 14001:2004
NAAC ACCREDITED
'A' Grade Institute by DHE, Govt. of NCT Delhi, Affiliated to GGSIP University Delhi
and Approved by Bar Council of India & NCTE

**B.A. (H)-Economics**

**Sem- 2nd**

**Subject- Statistical Methods-II ( BA ECO 104)**

# Syallbus

**Unit- 1: Theory of Distribution**

Introduction to probability distribution- Normal, Bernoulli, Poiusson- Negative, Binomial, Uniform, Chi Square, Exponential Distribution, Central Limit Theorem.

**Unit- II: Sampling**

Census and sample survey, Sample Selection Methods, Simple Random Sampling- With and without replacement, Systematic sampling, Properties of estimates and their variances.

**Unit-III: Point and Interval Estimation**

Point estimation, properties of estimators, Cramer Rao inequality, Methods of estimation and their properties, Introduction to methods of moments, Least Squares, Maximum likelihood-interval estimation, Confidence interval.

**Unit –IV: Hypothesis Testing**

Null and alternative hypothesis- critical region, Type I and Type II error, level of significance, p-value, Power of test- ANOVA, Inferences based on mean and variance- One way classification.

# Unit 1: Probability Distribution

## Introduction to Probability Distributions

Probability distributions describe the probability of observing a particular event. Let X be a random variable. We would like to specify the probabilities of events such as $\{X = x\}$ and $\{a \le X \le b\}$. If we can specify all probabilities involving X, we say that we have specified the probability distribution of X. Probability distributions are generally divided into two classes.

A **discrete probability distribution** (applicable to the scenarios where the set of possible outcomes is discrete, such as a coin toss or a roll of dice) can be encoded by a discrete list of the probabilities of the outcomes, known as a probability mass function.

On the other hand, a **continuous probability distribution** (applicable to the scenarios where the set of possible outcomes can take on values in a continuous range (e.g. real numbers), such as the temperature on a given day) is typically described by probability density functions (with the probability of any individual outcome actually being 0). The normal distribution is a commonly encountered continuous probability distribution.

## The Binomial Distribution

The Binomial Distribution Now we are ready to write down an expression for the probability distribution that describes the likelihood of r events (e.g. heads) occurring in a total of m events (e.g. coin flips) where the probability of an r-event occurring is p while the probability of it not occurring is $(1 - p)$. Since the individual events occur independently, the probability of a subset of r events amongst many m is the product of individual probabilities. If r occur, then $m - r$ don't and the probability is $p^r (1 - p)^{m-r}$. For the total probability of a particular event occurring (e.g. 2 heads), we multiply the probability that the event occurs by the number of ways that event can occur. The complete formula for the probability distribution is then given by

$P^r$ = m! divided by $(m - r)!r!$ multiplied by $(1 - p)^{m-r} p^r$.

This distribution is called the binomial distribution. It describes the probability that r events occur among a total of m independent events. Note that it is a discrete distribution; it is defined only at integral values of the variable r.

Properties of Binomial Distribution

1. The expectation is EX = np.

2. The variance of X is Var(X) = np(1 − p).

3. The probability generating function of X is $G(z) = (1 - p + zp)^n$

**Bernoulli Distribution**

We say that X has a Bernoulli distribution with success probability p if X can only assume the values 0 and 1, with probabilities P(X = 1) = p = 1 − P(X = 0) .

Following are some properties:

1. The expectation is EX = 0P(X = 0)+1P(X = 1) = 0×(1−p)+1×p = p.

2. The variance is Var(X) = EX2−(EX)2 = EX−(EX)2 = p−p2 = p(1−p). (Note that X2 = X).

3. The PGF is given by G(z) = z0(1 − p) + z1p = 1 − p + zp.

**Poisson Distribution**

A random variable X for which P(X = x) = λx divided by x!multiplied by $e^{-\lambda}$, x = 0, 1, 2,...

(for fixed λ > 0) is said to have a Poisson distribution. We write X ∼ Poi(λ). The Poisson distribution is used in many probability models and may be viewed as the "limit" of the Bin(n, μ/n) for large n.

1. The PGF is derived as: $G(z)=e^{-\lambda(1-z)}$ .

2. It follows that the expectation is EX = G (1) = $\lambda$. The intuitive explanation is that the mean number of successes of the corresponding coin flip experiment is np = n($\lambda$/n) = $\lambda$.

3. The variance is n($\lambda$/n)(1 − $\lambda$/n) → $\lambda$. For the Poisson distribution the variance and expectation are the same.

**Exponential Distribution**

A random variable X with probability density function f, given by $f(x) = \lambda e^{-\lambda x}$, $x \geq 0$

is said to have an exponential distribution with parameter $\lambda$. We write X ∼ Exp($\lambda$). The exponential distribution can be viewed as a continuous version of the geometric distribution. The exponential distribution is often used to model the failure time of manufactured items in production lines, say, light bulbs. If X denotes the (random) time to failure of a light–bulb of a particular make, then the exponential distribution postulates that the probability of survival of the bulb decays exponentially fast – to be precise, P(X > x) = e −$\lambda$ x. Notice that the bigger the value of $\lambda$, the faster the decay. This indicates that for large $\lambda$ the average time of failure of the bulb is smaller.

**Chi Square Distribution**

Definitions

Chi-square distribution

A distribution obtained from the multiplying the ratio of sample variance to population variance by the degrees of freedom when random samples are selected from a normally distributed population

Contingency Table

Data arranged in table form for the chi-square independence test

Expected Frequency

The frequencies obtained by calculation.

Goodness-of-fit Test

A test to see if a sample comes from a population with the given distribution.

Independence Test

A test to see if the row and column variables are independent.

Observed Frequency

The frequencies obtained by observation. These are the sample frequencies.

A random variable X is said to have a chi-square distribution with n ($\in \{1, 2, ...\}$) degrees of freedom if X ~ Gam(n/2, 1/2). We write X ~ $\chi 2$ n.

We mention a few properties of the $\Gamma$-function. 1. $\Gamma(a + 1) = a\,\Gamma(a)$, for a $\in$ R+. 2. $\Gamma(n)=(n - 1)!$ for n = 1, 2,.... 3. $\Gamma(1/2) = \sqrt{\pi}$. Chi-square distribution is encountered when we deal with collections of values that involve adding up squares. Variances of samples require us to add a collection of squared quantities and thus have distributions that are related to chi-square distribution. If we take each one of a collection of sample variances, divide them by the known population variance and multiply these quotients by (n – 1), where n means the number of items in the sample, we shall obtain a chi-square distribution. Thus,( $\sigma^2 s/ \sigma^2 p$) ( n– 1) would have the same distribution as chi-square distribution with (n – 1) degrees of freedom. Chi-square distribution is not symmetrical and all the values are positive. One must know the degrees of freedom for using chi-square distribution. This distribution may also be used for judging the significance of difference between observed and expected frequencies and also as a test of goodness of fit. The generalised shape of $\chi 2$ distribution depends upon the d.f.

**Uniform Distribution**

We say that a random variable X has a uniform distribution on the interval [a, b], if it has density function f, given by

$f(x) = 1 \div b - a$ , $a \le x \le b$ . We write X ~ U[a, b]. X can model a randomly chosen point from the interval [a, b], where each choice is equally likely.

**Negative Binimial Distribution**

In probability theory and statistics, the negative binomial distributionis a discrete probability distribution of the number of successes in a sequence of independent and identically distributed Bernoulli trials before a specified (non-random) number of failures (denoted *r*) occurs. For example, if we define a 1 as failure, all non-1s as successes, and we throw a dice repeatedly until the third time 1 appears (*r* = three failures), then the probability distribution of the number of non-1s that had appeared will be a negative binomial.

Suppose there is a sequence of independent Bernoulli trials. Thus, each trial has two potential outcomes called "success" and "failure". In each trial the probability of success is *p* and of failure is (1 − *p*). We are observing this sequence until a predefined number *r* of failures has occurred. Then the random number of successes we have seen, *X*, will have the **negative binomial** (or **Pascal**) distribution:

**The Normal Distribution**

A random variable *X* whose distribution has the shape of a **normal curve** is called a **normal random variable**.

**Properties of a Normal Distribution**

- Bell-shaped
- Symmetric about mean
- Continuous
- Never touches the x-axis
- Total area under curve is 1.00
- Approximately 68% lies within 1 standard deviation of the mean, 95% within 2 standard deviations, and 99.7% within 3 standard deviations of the mean. This is the Empirical Rule mentioned earlier.
- Data values represented by x which has mean mu and standard deviation sigma.
- Probability Function given by $$p(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

**Normal Probabilities**

Computing Normal Probabilities

There are several different situations that can arise when asked to find normal probabilities.

| Situation | Instructions |
|---|---|
| Between zero and any number | Look up the area in the table |
| Between two positives, or Between two negatives | Look up both areas in the table and subtract the smaller from the larger. |
| Between a negative and a positive | Look up both areas in the table and add them together |
| Less than a negative, or Greater than a positive | Look up the area in the table and subtract from 0.5000 |
| Greater than a negative, or Less than a positive | Look up the area in the table and add to 0.5000 |

This can be shortened into two rules.

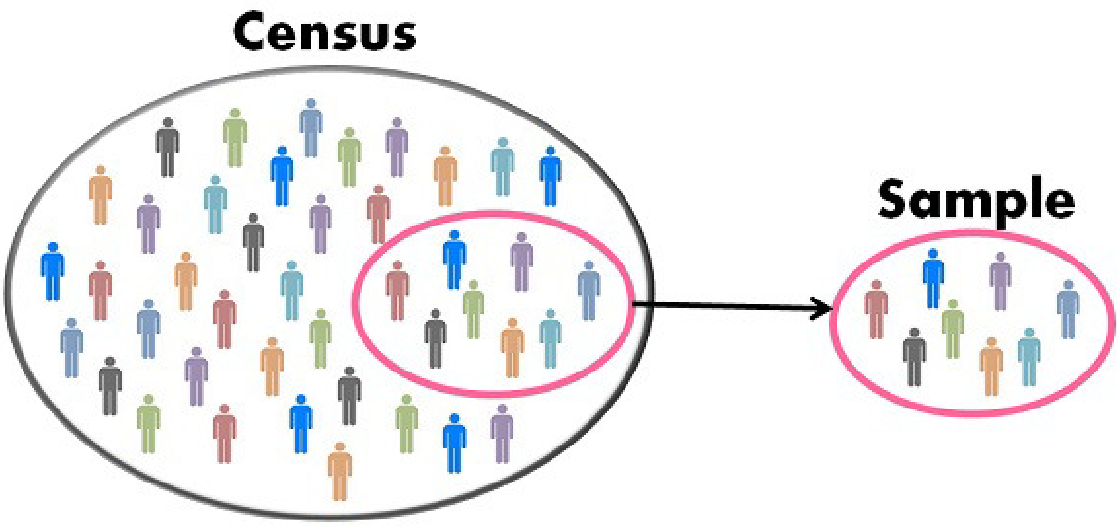
1. If there is only one z-score given, use 0.5000 for the second area, otherwise look up both z-scores in the table
2. If the two numbers are the same sign, then subtract; if they are different signs, then add. If there is only one z-score, then use the inequality to determine the second sign (< is negative, and > is positive).

**Central Limit Theorem**

As n gets large (n > 30), the shape of the sampling distribution will become more and more like a normal distribution, irrespective of the shape of the parent population. The theorem which explains this sort of relationship between the shape of the population distribution and the sampling distribution of the mean is known as the central limit theorem. This theorem is by far the most important theorem in statistical inference. It assures that the sampling distribution of the mean approaches normal distribtion as the sample size increases. In formal terms, we may say that the central limit theorem states that "the distribution of means of random samples taken from a population having mean $\mu$ and finite variance $\sigma^2$ approaches the normal distribution with mean $\mu$ and variance $\sigma^2$ /n as n goes to infinity. The significance of the central limit theorem lies in the fact that it permits us to use sample statistics to make inferences about population parameters without knowing anything about the shape of the frequency distribution of that population other than what we can get from the sample.

# Unit II: Sampling

## Difference between Census and Sampling




Census and sampling are two methods of collecting survey data about the population that are used by many countries. Census refers to the quantitative research method, in which all the members of the population are enumerated. On the other hand, the sampling is the widely used method, in statistical testing, wherein a data set is selected from the large population, which represents the entire group.

Census implies complete enumeration of the study objects, whereas Sampling connotes enumeration of the subgroup of elements chosen for participation. These two survey methods are often contrasted with each other, and so this article makes an attempt to clear the differences between census and sampling, in detail; Have a look.

Content: Census Vs Sampling

Comparison Chart

| BASIS FOR COMPARISON | CENSUS | SAMPLING |
|---|---|---|
| Meaning | A systematic method that collects and records the data about the members of the population is called Census. | Sampling refers to a portion of the population selected to represent the entire group, in all its characteristics. |
| Enumeration | Complete | Partial |
| Study of | Each and every unit of the population. | Only a handful of units of the population. |
| Time required | It is a time consuming process. | It is a fast process. |
| Cost | Expensive method | Economical method |
| Results | Reliable and accurate | Less reliable and accurate, due to the margin of error in the data collected. |

| BASIS FOR COMPARISON | CENSUS | SAMPLING |
|---|---|---|
| Error | Not present. | Depends on the size of the population |
| Appropriate for | Population of heterogeneous nature. | Population of homogeneous nature |

**Sample Selection Methods**

**Non-probability sampling**:

Non-probability sampling is that sampling procedure which does not afford any basis for estimating the probability that each item in the population has of being included in the sample. Non-probability sampling is also known by different names such as deliberate sampling, purposive sampling and judgement sampling. In this type of sampling, items for the sample are selected deliberately by the researcher; his choice concerning the items remains supreme. In other words, under non-probability sampling the organisers of the inquiry purposively choose the particular units of the universe for constituting a sample on the basis that the small mass that they so select out of a huge one will be typical or representative of the whole. For instance, if economic conditions of people living in a state are to be studied, a few towns and villages may be purposively selected for intensive study on the principle that they can be representative of the entire state. Thus, the judgement of the organisers of the study plays an important part in this sampling design. In such a design, personal element has a great chance of entering into the selection of the sample.

**Probability sampling**

Sample has a known probability of being selected.

In probability sampling it is possible to both determine which sampling units belong to which sample and the probability that each sample will be selected. The following sampling methods are examples of probability sampling:

- Simple Random Sampling (SRS)
- Stratified Sampling
- Cluster Sampling
- Systematic Sampling
- Multistage Sampling

Probability sampling is also known as 'random sampling' or 'chance sampling'. Under this sampling design, every item of the universe has an equal chance of inclusion in the sample. It is, so to say, a lottery method in which individual units are picked up from the whole group not deliberately but by some mechanical process. Here it is blind chance alone that determines whether one item or the other is selected. The results obtained from probability or random sampling can be assured in terms of probability i.e., we can measure the errors of estimation or the significance of results obtained from a random sample, and this fact brings out the superiority of random sampling design over the deliberate sampling design. Random sampling ensures the law of Statistical Regularity which states that if on an average the sample chosen is a random one, the sample will have the same composition and characteristics as the universe. This is the reason why random sampling is considered as the best technique of selecting a representative sample. Random sampling from a finite population refers to that method of sample selection which gives each possible sample combination an equal probability of being picked up and each item in the entire population to have an equal chance of being included in the sample.

**Sampling with Replacement and Sampling without Replacement**

**Sampling with replacement:**

Consider a population of potato sacks, each of which has either 12, 13, 14, 15, 16, 17, or 18 potatoes, and all the values are equally likely. Suppose that, in this population, there is exactly one sack with each number. So the whole population has seven sacks. If I sample two with replacement, then I first pick one (say 14). I had a 1/7 probability of choosing that one. Then I replace it. Then I pick another. Every one of them still has 1/7 probability of being chosen. And there are exactly 49 different possibilities here (assuming we distinguish between the first and second.) They are: (12,12), (12,13), (12, 14), (12,15), (12,16), (12,17), (12,18), (13,12), (13,13), (13,14), etc.

**Sampling without replacement:**

Consider the same population of potato sacks, each of which has either 12, 13, 14, 15, 16, 17, or 18 potatoes, and all the values are equally likely. Suppose that, in this population, there is exactly one sack with each number. So the whole population has seven sacks. If I sample two without replacement, then I first pick one (say 14). I had a 1/7 probability of choosing that one. Then I pick another. At this point, there are only six possibilities: 12, 13, 15, 16, 17, and 18. So there are only 42 different possibilities here (again assuming that we distinguish between the first and the second.) They are: (12,13), (12,14), (12,15), (12,16), (12,17), (12,18), (13,12), (13,14), (13,15), etc.

What's the Difference?

When we sample with replacement, the two sample values are independent. Practically, this means that what we get on the first one doesn't affect what we get on the second. Mathematically, this means that the covariance between the two is zero.

In sampling without replacement, the two sample values aren't independent. Practically, this means that what we got on the for the first one affects what we can get for the second one. Mathematically, this means that the covariance between the two isn't zero. That complicates the computations. In particular, if we have a SRS (simple random sample) without replacement, from a population with variance $\sigma^2$, then the covariance of two of the different sample values is $\frac{-\sigma^2}{N-1}$, where *N* is the population size.

## Systematic Sampling

In some instances, the most practical way of sampling is to select every ith item on a list. Sampling of this type is known as systematic sampling. An element of randomness is introduced into this kind of sampling by using random numbers to pick up the unit with which to start. For instance, if a 4 per cent sample is desired, the first item would be selected randomly from the first twenty-five and thereafter every 25th item would automatically be included in the sample. Thus, in systematic sampling only the first unit is selected randomly and the remaining units of the sample are selected at fixed intervals. It is an easier and less costlier method of sampling and can be conveniently used even in case of large populations. But there are certain dangers too in using this type of sampling. If there is a hidden periodicity in the population, systematic sampling will prove to be an inefficient method of sampling. For instance, every 25th item produced by a certain production process is defective. If we are to select a 4% sample of the items of this process in a systematic manner, we would either get all defective items or all good items in our sample depending upon the random starting position. If all elements of the universe are ordered in a manner representative of the total population, i.e., the population list is in random order, systematic sampling is considered equivalent to random sampling. But if this is not so, then the results of such sampling may, at times, not be very reliable. In practice, systematic sampling is used when lists of population are available and they are of considerable length.

## Properties of Estimates and Their Variances

The sample statistic is calculated from the sample data and the population parameter is inferred (or estimated) from this sample statistic. Let me say that again: Statistics are calculated, parameters are estimated. The estimation of the population parameter is done from the sample statistic. There are two types of estimates: Point Estimates and Interval Estimates. In most statistical research studies, population parameters are usually unknown and have to be estimated from a sample. As such the methods for estimating the population parameters assume an important role in statistical anlysis. The random variables (such as X s and σ2 ) used to estimate population parameters, such as μ σ and p 2 are conventionally called as 'estimators', while specific values of these (such as X = 105 or σ s 2 = 2144 . ) are referred to as 'estimates' of the population parameters. The estimate of a population parameter may be one single value or it could be a range of values. In the former case it is referred as point estimate, whereas in the latter case it is termed as interval estimate.

A good estimator must satisfy three conditions:

- Unbiased: The expected value of the estimator must be equal to the mean of the parameter
- Consistent: The value of the estimator approaches the value of the parameter as the sample size increases
- Relatively Efficient: The estimator has the smallest variance of all estimators which could be used.

## Unit III: Point and Interval Estimation

### Introduction to Estimation

One area of concern in inferential statistics is the estimation of the population parameter from the sample statistic. It is important to realize the order here. The sample statistic is calculated from the sample data and the population parameter is inferred (or estimated) from this sample statistic. Let me say that again: Statistics are calculated, parameters are estimated.

We talked about problems of obtaining the value of the parameter earlier in the course when we talked about sampling techniques.

Another area of inferential statistics is sample size determination. That is, how large of a sample should be taken to make an accurate estimation. In these cases, the statistics can't be used since the sample hasn't been taken yet.

**Definitions**

Confidence Interval

An interval estimate with a specific level of confidence

Confidence Level

The percent of the time the true mean will lie in the interval estimate given.

Consistent Estimator

An estimator which gets closer to the value of the parameter as the sample size increases.

Degrees of Freedom

The number of data values which are allowed to vary once a statistic has been determined.

Estimator

A sample statistic which is used to estimate a population parameter. It must be unbiased, consistent, and relatively efficient.

Interval Estimate

A range of values used to estimate a parameter.

Maximum Error of the Estimate

The maximum difference between the point estimate and the actual parameter. The Maximum Error of the Estimate is 0.5 the width of the confidence interval for means and proportions.

Point Estimate

A single value used to estimate a parameter.

Relatively Efficient Estimator

The estimator for a parameter with the smallest variance.

T distribution

A distribution used when the population variance is unknown.

Unbiased Estimator

An estimator whose expected value is the mean of the parameter being estimated.

## Point Estimation

**Point estimation** involves the use of sample data to calculate a single value (known as a statistic) which is to serve as a "best guess" or "best estimate" of an unknown (fixed or random) population parameter. It is the process of finding an approximate value of some parameter—such as the mean (average)—of a population from random samples of the population.

## Properties of an estimator:

It is desirable for a point estimate to be:

(1) Consistent. The larger the sample size, the more accurate the estimate.

(2) Unbiased. The expectation of the observed values of many samples ("average observation value") equals the corresponding population parameter. For example, the sample mean is an unbiased estimator for the population mean.

(3) Most efficient or best unbiased—of all consistent, unbiased estimates, the one possessing the smallest variance (a measure of the amount of dispersion away from the estimate). In other words, the estimator that varies least from sample to sample. This generally depends on the particular distribution of the population.

## Cramer Rao Inequality

In estimation theory and statistics, the Cramér–Rao bound (CRB), Cramér–Rao lower bound (CRLB), Cramér–Rao inequality, Frechet-Darmois-Cramér-Rao inequality, or information inequality expresses a lower bound on the variance of estimators of a deterministic (fixed, though unknown) parameter. This term is named in honor of Harald Cramér, Calyampudi Radhakrishna Rao, Maurice Frechet and Georges Darmois all of whom independently derived this limit to statistical precision in the 1940s.

In its simplest form, the bound states that the variance of any unbiased estimator is at least as high as the inverse of the Fisher information.

The Cramer-Rao Inequality provides a lower bound for the variance of an unbiased estimator of a parameter. It allows us to conclude that an unbiased estimator is a minimum variance unbiased estimator for a parameter. Cramer Rao inequality provides lower bound for the estimation error variance. Minimum attainable variance is often larger than CRLB. We need to know the probability density function to evaluate CRLB. Often we don't know this information and cannot evaluate this bound. If the data is multivariate Gaussian or with known distribution, we can evaluate it. If the estimator reaches the CRLB, it is called efficient. MVUE (minimum variance *bound* unbiased estimator ) may or may not be efficient. If it is not, we have to use other tools than CRLB to find it.

**Methods of Estimation and their Properties:**

- Maximum likelihood estimators.
- Bayes estimators.
- Method of moments estimators.
- Cramér–Rao bound.
- Minimum mean squared error (MMSE), also known as Bayes least squared error (BLSE)
- Maximum a posteriori (MAP)
- Minimum variance unbiased estimator (MVUE)

**Methods of Moment Estimator:**

• Advantage: simplest approach for constructing an estimator.

• Disadvantage: usually are not the "best" estimators possible.

• Principle: Equate the kth population moment E[Xk] with the kth sample moment and solve for the unknown parameter ! 1 n Xi k n.

**Maximum likelihood estimators**

• Before an experiment is performed the outcome is unknown. Probability allows us to predict unknown outcomes based on known parameters: ! P(Data I") • For example: ! P(x I n, p) = ( ) p x (1" p) n n"x

In statistics, maximum likelihood estimation (MLE) is a method of estimating the parameters of a statistical model given observations, by finding the parameter values that maximize the likelihood of making the observations given the parameters. MLE can be seen as a special case of the maximum a posteriori estimation (MAP) that assumes a uniform prior distribution of the parameters, or as a variant of the MAP that ignores the prior and which therefore is unregularized.

The method of maximum likelihood corresponds to many well-known estimation methods in statistics. For example, one may be interested in the heights of adult female penguins, but is unable to measure the height of every single penguin in a population due to cost or time constraints. Assuming that the heights are normally distributed with some unknown mean and variance, the mean and variance can be estimated with MLE while only knowing the heights of some sample of the overall population. MLE would accomplish this by taking the mean and variance as parameters and finding particular parametric values that make the observed results the most probable given the model.

In general, for a fixed set of data and underlying statistical model, the method of maximum likelihood selects the set of values of the model parameters that maximizes the likelihood function. Intuitively, this maximizes the "agreement" of the selected model with the observed data, and for discrete random variables it indeed maximizes the probability of the observed data under the resulting distribution. Maximum likelihood estimation gives a unified approach to estimation, which is well-defined in the case of the normal distribution and many other problems.

In statistics, **maximum likelihood estimation** (**MLE**) is a method of estimating the parameters of a statistical model given observations, by finding the parameter values that maximize the likelihood of making the observations given the parameters. MLE can be seen as a special case of the maximum a posteriori estimation (MAP) that assumes a uniform prior distribution of the parameters, or as a variant of the MAP that ignores the prior and which therefore is unregularized.

**Least Squares Method:**

The method of least squares is a standard approach in regression analysis to approximate the solution of overdetermined systems, i.e., sets of equations in which there are more equations than unknowns. "Least squares" means that the overall solution minimizes the sum of the squares of the residuals made in the results of every single equation.

The most important application is in data fitting. The best fit in the least-squares sense minimizes *the sum of squared residuals* (a residual being: the difference between an observed value, and the fitted value provided by a model). When the problem has substantial uncertainties in the independent variable (the *x* variable), then simple regression and least-squares methods have problems; in such cases, the methodology required for fitting errors-in-variables models may be considered instead of that for least squares.

Least-squares problems fall into two categories: linear or ordinary least squaresand nonlinear least squares, depending on whether or not the residuals are linear in all unknowns. The linear least-squares problem occurs in statistical regression analysis; it has a closed-form solution. The nonlinear problem is usually solved by iterative refinement; at each iteration the system is approximated by a linear one, and thus the core calculation is similar in both cases.

**Interval Estimation**

Interval estimates provide a range of values for a parameter value, within which we have a stated degree of confidence that the parameter lies.

**Confidence interval**:

An interval of plausible values for the parameter being estimated, where degree of plausibility specified by a "confidence level". The **confidence level** is the frequency (i.e., the proportion) of possible confidence intervals that contain the true value of their corresponding parameter. In other words, if confidence intervals are constructed using a given confidence level in an infinite number of independent experiments, the proportion of those intervals that contain the true value of the parameter will match the confidence level.

Confidence intervals consist of a range of values (interval) that act as good estimates of the unknown population parameter. However, the interval computed from a particular sample does not necessarily include the true value of the parameter. Since the observed data are random samples from the true population, the confidence interval obtained from the data is also random. If a corresponding hypothesis test is performed, the confidence level is the complement of the level of significance; for example, a 95% confidence interval reflects a significance level of 0.05. If it is hypothesized that a true parameter value is 0 but the 95% confidence interval does not contain 0, then the estimate is significantly different from zero at the 5% significance level.
The desired level of confidence is set by the researcher (not determined by data). Most commonly, the 95% confidence level is used. However, other confidence levels can be used, for example, 90% and 99%.
Factors affecting the width of the confidence interval include the size of the sample, the confidence level, and the variability in the sample. A larger sample size normally will lead to a better estimate of the population parameter.

# Unit IV: Hypothesis Testing

## Hypothesis

A statistical hypothesis or simply hypothesis is an assumption or claim either about the parameters of populations (single parameter value or several values of parameter) or about the population (from of an entire probability distribution). Any statement concerning a parameter is a hypothesis about a parameter.

Example 1: The claim $\mu$ = Rs. 10,000, where $\mu$ is the average salary of a newly graduated student is an example of hypothesis.

## Hypothesis Testing

Whenever we do hypothesis testing, we formulate two opposing or contradictory hypothesis. In example 1, $\mu = 10{,}000$ and $\mu \neq 10{,}000$ are two opposing statements based on the sample findings; we decide which of two hypothesis is correct. The hypothesis that we test based on the initial assumption that it is true is called **null hypothesis**. It is denoted by Ho. It is also called the testable proposition. The opposing or the contradictory hypothesis is the **alternative hypothesis**. It is also the counter proposition to null hypothesis and is denoted by $H_1$.

## Type I & Type II Error

| | State of Nature | |
|---|---|---|
| Decision | $H_0$ True | $H_0$ False |
| Reject $H_0$ | Type I Error<br>alpha | Correct Assessment |
| Fail to reject $H_0$ | Correct Assessment | Type II Error |

| | | |
|---|---|---|
| | | beta |

Type I Error: Rejecting Null hypothesis when it is true.

Type II Error: Accepting Null hypothesis when it is false.

Which of the two errors is more serious? Type I or Type II ?

Since *Type I is the more serious error* (usually), that is the one we concentrate on. We usually pick alpha to be very small (0.05, 0.01). Note: alpha is not a Type I error. Alpha is the *probability of committing* a Type I error. Likewise beta is the *probability of committing* a Type II error.

**Conclusions**

Conclusions are sentence answers which include whether there is enough evidence or not (based on the decision), the level of significance, and whether the original claim is supported or rejected.

Conclusions are based on the original claim, which may be the null or alternative hypotheses. The decisions are always based on the null hypothesis

| | | |
|---|---|---|
| | Original Claim | |
| Decision | $H_0$<br>"REJECT" | $H_1$<br>"SUPPORT" |
| Reject $H_0$<br>"SUFFICIENT" | There is sufficient evidence at the alpha level of | There is sufficient evidence at the alpha level of significance to support |

| | | |
|---|---|---|
| | significance to reject the claim | the claim |
| Fail to reject $H_0$ "INSUFFICIENT" | There is insufficient evidence at the alpha level of significance to reject the claim | There is insufficient evidence at the alpha level of significance to support the claim |

**Definitions**

**Null Hypothesis ( $H_0$ )**

Statement of zero or no change. If the original claim includes equality (<=, =, or >=), it is the null hypothesis. If the original claim does not include equality (<, not equal, >) then the null hypothesis is the complement of the original claim. The null hypothesis *always* includes the equal sign. The decision is based on the null hypothesis.

**Alternative Hypothesis ($H_1$ or $H_a$ )**

Statement which is true if the null hypothesis is false. The type of test (left, right, or two-tail) is based on the alternative hypothesis.

**Type I error**

Rejecting the null hypothesis when it is true (saying false when true). Usually the more serious error.

**Type II error**

Failing to reject the null hypothesis when it is false (saying true when false).

**alpha**

Probability of committing a Type I error.

**beta**

Probability of committing a Type II error.

**Test statistic**

Sample statistic used to decide whether to reject or fail to reject the null hypothesis.

**Critical region**

Set of all values which would cause us to reject $H_0$

**Critical value(s)**

The value(s) which separate the critical region from the non-critical region. The critical values are determined independently of the sample statistics.

**Significance level ( alpha )**

The probability of rejecting the null hypothesis when it is true. alpha = 0.05 and alpha = 0.01 are common. If no level of significance is given, use alpha = 0.05. The level of significance is the complement of the level of confidence in estimation.

**Decision**

A statement based upon the null hypothesis. It is either "reject the null hypothesis" or "fail to reject the null hypothesis". We will never accept the null hypothesis.

**Conclusion**

A statement which indicates the level of evidence (sufficient or insufficient), at what level of significance, and whether the original claim is rejected (null) or supported (alternative).

We outline the tests of hypothesis by the following steps :

1. Formulate Ho and Ha; and specify $\alpha$.
2. Give the formula for computing the value of test statistics.
3. Determine the rejection region for the specified $\alpha$.
4. Compute the value of test statistic using the sample data.
5. Check whether the value of test statistic falls in the rejection region. Accordingly, decide whether Ho should be rejected and conclude.

**The Level Of Significance:**

This is a very important concept in the context of hypothesis testing. It is always some percentage (usually 5%) which should be chosen with great care, thought and reason. In case we take the significance level at 5 per cent, then this implies that H0 will be rejected when the sampling result (i.e., observed evidence) has a less than 0.05 probability of occurring if H0 is true. In other words, the 5 per cent level of significance means that researcher is willing to take as much as a 5 per cent risk of rejecting the null hypothesis when it (H0 ) happens to be true. Thus the significance level is the maximum value of the probability of rejecting H0 when it is true and is usually determined in advance before testing the hypothesis.

**Measuring the Power of a Hypothesis Test**

As stated above we may commit Type I and Type II errors while testing a hypothesis. The probability of Type I error is denoted as $\alpha$ (the significance level of the test) and the probability of Type II error is referred to as $\beta$ . Usually the significance level of a test is assigned in advance and once we decide it, there is nothing else we can do about $\alpha$. But what can we say about $\beta$? We all know that hypothesis test cannot be foolproof; sometimes the test does not reject H0 when it happens to be a false one and this way a Type II error is made. But we would certainly like that $\beta$ (the probability of accepting H0 when H0 is not true) to be as small as possible. Alternatively, we would like that $1 - \beta$ (the probability of rejecting H0 when H0 is not true) to be as large as possible. If $1 - \beta$ is very much nearer to unity (i.e., nearer to 1.0), we can infer that the test is working quite well, meaning thereby that the test is rejecting H0 when it is not true and if $1 - \beta$ is very much nearer to 0.0, then we infer that the test is poorly working, meaning thereby that it is not rejecting H0 when H0 is not true. Accordingly $1 - \beta$ value is the measure of how well the test is working or what is technically described as the power of the test. In case we plot the values of $1 - \beta$ for each possible value of the population parameter (say $\mu$ , the true population mean) for which the H0 is not true (alternatively the Ha is true), the resulting curve is known as the power curve associated with the given test. Thus power curve of a hypothesis test is the curve that shows the conditional probability of rejecting H0 as a function of the population parameter and size of the sample. The function defining this curve is known as the power function. In other words, the power function of a test is that function defined for all values of the parameter(s) which yields the probability that H0 is rejected and the value of the power function at a specific parameter point is called the power of the test at that point. As the population parameter gets closer and closer to hypothesised value of the population parameter, the power of the test (i.e., 1 – β) must get closer and closer to the probability of rejecting H0 when the population parameter is exactly equal to hypothesised value of the parameter. We know that this probability is simply the significance level of the test, and as such the power curve of a test terminates at a point that lies at a height of α (the significance level) directly over the population parameter. Closely related to the power function, there is another function which is known as the operating characteristic function which shows the conditional probability of accepting H0 for all values of population parameter(s) for a given sample size, whether or not the decision happens to be a correct one. If power function is represented as H and operating characteristic function as L, then we have L = 1 – H. However, one needs only one of these two functions for any decision rule in the context of testing hypotheses.

**P-Value Approach**

The P-Value Approach, short for Probability Value, approaches hypothesis testing from a different manner. Instead of comparing z-scores or t-scores as in the classical approach, you're comparing probabilities, or areas.

The level of significance (alpha) is the area in the critical region. That is, the area in the tails to the right or left of the critical values.

The p-value is the area to the right or left of the test statistic. If it is a two tail test, then look up the probability in one tail and double it.

If the test statistic is in the critical region, then the p-value will be less than the level of significance. It does not matter whether it is a left tail, right tail, or two tail test. This rule always holds.

Reject the null hypothesis if the p-value is less than the level of significance.

You will fail to reject the null hypothesis if the p-value is greater than or equal to the level of significance.

The p-value approach is best suited for the normal distribution when doing calculations by hand. However, many statistical packages will give the p-value but not the critical value. This is because it is easier for a computer or calculator to find the probability than it is to find the critical value.

Another benefit of the p-value is that the statistician immediately knows at what level the testing becomes significant. That is, a p-value of 0.06 would be rejected at an 0.10 level of significance, but it would fail to reject at an 0.05 level of significance. Warning: Do not decide on the level of significance after calculating the test statistic and finding the p-value.

Here is a proportion to help you keep the order straight. Any proportion equivalent to the following statement is correct.

| |
|---|
| The test statistic is to the p-value as the critical value is to the level of significance. |

**Hypothesis Testing Of Means**

Mean of the population can be tested presuming different situations such as the population may be normal or other than normal, it may be finite or infinite, sample size may be large or small, variance of the population may be known or unknown and the alternative hypothesis may be two-sided or one sided. Our testing technique will differ in different situations. We may consider some of the important situations.

1. Population normal, population infinite, sample size may be large or small but variance of the population is known, Ha may be one-sided or two-sided: In such a situation z-test is used for testing hypothesis of mean.

2. Population normal, population finite, sample size may be large or small but variance of the population is known, Ha may be one-sided or two-sided: In such a situation z-test is used.
3. Population normal, population infinite, sample size small and variance of the population unknown, Ha may be one-sided or two-sided: In such a situation t-test is used.
4. Population normal, population finite, sample size small and variance of the population unknown, and Ha may be one-sided or two-sided: In such a situation t-test is used
5. Population may not be normal but sample size is large, variance of the population may be known or unknown, and Ha may be one-sided or two-sided: In such a situation we use z-test.

**Hypothesis Testing For Comparing a Variance to Some Hypothesised Population Variance**

The test we use for comparing a sample variance to some theoretical or hypothesised variance of population is different than z-test or the t-test. The test we use for this purpose is known as chi square test and the test statistic symbolised as $\chi 2$ , known as the chi-square value. Then by comparing the calculated value of $\chi 2$ with its table value for (n – 1) degrees of freedom at a given level of significance, we may either accept H0 or reject it. If the calculated value of $\chi 2$ is equal to or less than the table value, the null hypothesis is accepted; otherwise the null hypothesis is rejected. This test is based on chi-square distribution which is not symmetrical and all the values happen to be positive; one must simply know the degrees of freedom for using such a distribution.

**Testing the Equality of Variances of Two Normal Populations**

When we want to test the equality of variances of two normal populations, we make use of F-test based on F-distribution.

When we use the F-test, we presume that

(i) the populations are normal;

(ii) samples have been drawn randomly;

(iii) observations are independent; and

(iv) there is no measurement error.

The object of F-test is to test the hypothesis whether the two samples are from the same normal population with equal variance or from two normal populations with equal variances. F-test was initially used to verify the hypothesis of equality between two variances, but is now mostly used in the context of analysis of variance.

**The Basic Principle Of ANOVA**

The basic principle of ANOVA is to test for differences among the means of the populations by examining the amount of variation within each of these samples, relative to the amount of variation between the samples. In terms of variation within the given population, it is assumed that the values of (Xij) differ from the mean of this population only because of random effects i.e., there are influences on (Xij) which are unexplainable, whereas in examining differences between populations we assume that the difference between the mean of the jth population and the grand mean is attributable to what is called a 'specific factor' or what is technically described as treatment effect. Thus while using ANOVA, we assume that each of the samples is drawn from a normal population and that each of these populations has the same variance. We also assume that all factors other than the one or more being tested are effectively controlled. This, in other words, means that we assume the absence of many factors that might affect our conclusions concerning the factor(s) to be studied. In short, we have to make two estimates of population variance viz., one based on between samples variance and the other based on within samples variance. Then the said two estimates of population variance are compared with F-test, wherein we work out.

F = Estimate of population variance based on between samples variance divided by Estimate of population variance based on within samples variance

This value of F is to be compared to the F-limit for given degrees of freedom. If the F value we work out is equal or exceeds the F-limit value, we may say that there are significant differences between the sample means.