Second Part:

Statistical Inference: The process of reaching conclusions concerning population parameters is a statistical investigation to make an inference about one or more population parameters.

Estimation: is the process of reaching conclusions concerning population parameters using sampling.

Estimator: An **estimator** is a rule (**a formula or a function**) for calculating an **estimate** of a given quantity (**Parameter**) based on observed data **(sample data)** or measurements contained in a sample.

In short: an **estimator** is a formula or a function that maps samples into the parameter space; it is used to make an estimate.
Estimate is the value of that function taken on a particular set of data (a sample).

***For example*:**

An **estimator** is a function that maps a random sample to the parameter estimate:

$$\hat{\Theta} = t(X_1, X_2, \ldots, X_n)$$

Note that an estimator of *n* random variables $X_1, X_2, \ldots, X_n$ is a random variable $\hat{\Theta}$. For instance, an estimator is the sample mean:

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

An **estimate** $\hat{\theta}$ is the result of applying the estimator function to a lowercase observed sample $x_1, x_2, \ldots, x_n$:

$$\hat{\theta} = t(x_1, x_2, \ldots, x_n)$$

For instance, an estimate of the observed sample $x_1, x_2, \ldots, x_n$ is the sample mean:

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

Using descriptive and inferential statistics, two types of estimation and estimates about the population parameter can be defined: point and interval estimation.

- **Point Estimation**: is where estimation procedure uses the sample information to arrive at a single number that is intended to be close to the true value of the unknown parameter θ.

Point estimate: A point estimate of a population parameter is a single value of a statistic. For example, the sample mean $\bar{x}$ is a point estimate of the population mean μ. Similarly, the sample proportion $\hat{p}$ is a point estimate of the population proportion P.

Interval Estimation: is the estimation of two values within which the estimated value of the unknown parameter θ lies.

Interval estimate: An interval estimate is defined by two numbers, between which a population parameter is said to lie. For example, a < x < b is an interval estimate of the population mean μ. It indicates that the population mean is greater than a but less than b.

Note: Both types of estimates are important for gathering a clear idea of where a parameter is likely to lie.

Methods of Point Estimation: Two methods of point estimations are described: The Method of Moments (MM) and Maximum Likelihood Estimation (MLE).

1. ***The Method of Moments or Moment Method (MM)***:

What is a sample moment? : Suppose that a R.S. of observations is taken from a population, let it be denoted by $(x_1, x_2, .., x_n)$ each with probability (1/n). the rth sample moment about the origin is denoted by m'_r where $m'_r = \sum \frac{x^k_i}{n}$

(The rth moment for a population about the origin is denoted by μ'_r, $\mu'_r = E(X^r)$) .

Def: let $X_1, X_2, .., X_n$ be a R.S. from a Pop. Whose density function involves n unkown parameters, $\theta_1, \theta_2, \dots., \theta_n$, the moment estimators $\acute{\theta}_1, \acute{\theta}_2, \dots., \acute{\theta}_n$ of $\theta_1, \theta_2, \dots., \theta_n$ respectively, are the solutions of the simultaneous equations ;

$m'_r(x_1, x_2, .., x_n) = \mu'_r(\theta_1, \theta_2, \dots., \theta_n), r = 1,2, .., k.$

This method of obtaining estimators of unknown parameters from the sample moments is called the Method of Moments, MM.

***Note*:** This method is an easy technique for estimating (up to 3) parameters but in many cases the obtained estimator does not satisfy the properties of good estimate.

The method of Maximum Likelihood Estimator:

Let X be a R.V. having the P.d.f $f(x, \theta)$ where θ is unknown parameter, let $X_1, X_2, .., X_n$ be a R.S. of size n from a Pop corresponding to the R.V. X. The joint p.d.f. for this R.S. is defined as

$$L(\theta) = f(x_1, \theta) f(x_2, \theta) \dots f(x_n, \theta) \quad \dots. \qquad (1)$$

$L(\theta)$ is called the likelihood function of the R.S.

***Def*:**
Let L(θ) be the likelihood function of the R.S obtained from a population with unknown parameter θ . if Ω denotes the set of all possible values of θ, and if $\acute{\theta}$ is a value in Ω, such that $L(\acute{\theta}) \geq L(\theta)$ for all θ in Ω, then $\acute{\theta}$ is called a Maximum Likelihood Estimator (MLE) for θ.
This method is called the method of **Maximum Likelihood**.

How to determine the ML estimation: there are two cases the regular case and non regular case. Each one will be studied separately.

Examples; H.W and Assignments.

Estimators and Properties

Let θ be a parameter of a distribution, whose true (and unknown) value is θ_0.

An estimator is a function of the observations in the sample, and whose value will be used as a guess of the true but unknown value θ_0 of the parameter θ. The value taken by an estimator on a given sample is called an estimate (a "statistic") of (θ_0). We will denote:

Estimates are expected to be close to the true value θ_0. But the sample being random, an estimator is a random variable. So it can never been said with certainty that an estimate is close to the true value of a parameter of the distribution (or of the model).

So it appears that Estimation theory will focus not on estimates, but also on the properties of estimators considered as random variables, that is, their probability distributions, or some restricted aspects (mostly mean, variance and other properties).

Note: parameter of a distribution may have several different estimators to choose from. So the two central questions of Estimation theory are :

1. What are the properties that make a statistic a "good" estimator of a parameter ?
2. How can we design a good estimator for a given parameter ?

The Goodness of a Point Estimators

One way to measure the goodness of any point estimation procedure is in terms of the distances between the estimates it generates and the target parameter. This quality which varies randomly in repeated sampling is called the Error of Estimation. Therefore, the goodness of the point estimation

can be evaluated from a frequency dis. Of the values of estimates obtained from a sampling and then note how closely this distribution clusters about target parameter.

Def: The error of estimation ε is the distance between an estimator and its target parameter.

$\varepsilon = |\theta^* - \theta|$. θ^* is a random variable (error of estimation) is also a random quantity we cannot say how large or small it will be for a parameter estimate, it is possible to make probability statement about it.

Properties of Good Estimator:

Using different methods of estimation can lead to different estimators. Criteria for deciding which are good estimators are required. Before listing the qualities of a good estimator, it is important to understand that they are random variables. For example, suppose that we take a sample of size 5 from a uniform distribution and calculate $\bar{X}$. Each time we repeat the experiment we will probably get a different sample of 5 and therefore a different $\bar{X}$. The behavior of an estimator for different random samples will be described by a probability distribution.

As a first condition it seems reasonable to ask that the distribution of the estimator be centered around the parameter it is estimating. If not it will tend to overestimate or underestimate θ.

A second property an estimator should possess is precision. An estimator is precise if the dispersion of its distribution is small. These two concepts are integrated in the definitions of unbiasedness and efficiency:

Unbiasedness: An estimator θ^* is unbiased estimator for a parameter θ, if $E[\theta^*] = \theta$, otherwise θ^* is said to be biased estimator.

Examples: show that $\overline{X}$ is an unbiased estimator but S^2 is a biased estimator.

Biased estimators can be converted to unbiased estimators using the following theorem.

Theorem: Let θ^* be a biased estimator for the parameter θ, if $E[\theta^*] = a\theta + b, \text{where } a, b$ are constants, then $\hat{\theta} = \frac{\theta^* - b}{a}$ ia an unbiased estimator for θ.

Proof;

Example; find an unbiased estimator for σ^2 .

***Def*:** the bias of a point estimator θ^* is given by $B(\theta^*) = E(\theta^*) - \theta^*$

Note: given two unbiased estimators for a parameter θ, the one with smallest variance is selected as it is the preferred one.

To characterized the goodness of a point estimator, rather than using the bias and variance , the expected value $E[\theta^* - \theta]^2$, the average of the square of the distance between the estimator and its target parameter is used.

Def: The mean square error of a point estimation θ^* is the expected value of ($\theta^* - \theta$) and denoted by MSE. $MSE(\theta^*) = E[\theta^* - \theta]^2$.

The mean square error of an estimator is a function of both its variance and its bias.

If $B(\theta^*)$ denotes the bias of the estimator θ^* , it can be shown that :

$MSE(\theta^*) = V(\theta^*) + (B(\theta^*))^2$.

Efficiency:

If there are two unbiased estimators, say $\hat{\theta}$, $\acute{\theta}$, for a parameter θ, which one do we use?

The one with the smallest mean square error is selected. i.e., if

$E[(\hat{\theta} - \theta)^2] < E[(\acute{\theta} - \theta)^2]$, then $\hat{\theta}$ is preferred.

Therefore, if $E[(\hat{\theta})] = E[(\acute{\theta})] = \theta$, then the one with smallest variance is selected.

Def: An estimator T (could be $\widehat{\theta}$, θ^*,) with the following properties: $E(T) = \theta$, where

$T = \sum_1^n a_i x_i$, i.e. T is a linear function of the samples.

$Var(T) \leq Var(\acute{T})$ where $\acute{T}$ is any other estimate for θ, is called best linear unbiased estimate of θ.

Note: unbisedness and minimum variance are the most desirable properties that a statistic should have if that statistic is to be called a "best" statistic or estimator.

Rao-Cramer Inequality helps decide among several estimators, it provides a lower bound for the variance of every unbiased estimator.

Def: let $\hat{\theta}$ be unbiased estimator for a parameter θ, the statistic $\hat{\theta}$ is called an efficient statistic for θ iff the variance of $\hat{\theta}$ attains the Rao-Cramer lower bound for the variance of $\hat{\theta}$.

i.e.: $\sigma^2_{\hat{\theta}} \geq \frac{1}{nE[(\frac{\partial \ln f(x,\theta)}{\partial \theta})^2]}$. $\frac{1}{nE[(\frac{\partial \ln f(x,\theta)}{\partial \theta})^2]}$ is called the Rao-Cramer lower bound for θ.

Note: for a point estimator, $E[(\frac{\partial \ln f(x,\theta)}{\partial \theta})^2] = -E[\left(\left(\partial^2 \frac{\ln f(x,\theta)}{\partial \theta}\right)\right] / \partial \theta^2$

Theorem: let $X_1, X_2, .., X_n$ be a R.S with p.d.f. f(x, θ), let $\hat{\theta}$ be unbiased estimator for a parameter θ. Then, if $Var(\widehat{\theta}) \geq \frac{1}{nE[(\frac{\partial \ln f(x,\theta)}{\partial \theta})^2]}$, then $\hat{\theta}$ is called a MVUE for θ.

The Inequality is called the Rao-Cramer Inequality.

Example: A sample of 1000 students are randomly selected from a certain college; it shows that X=560 have passed the exam. estimate the proportion of success in the population. what is the probability that the error of estimation is less than 2σ. Show that $\widehat{P} = \frac{X}{n}$ is a consistent estimator for P.

Relative Efficiency: Given two unbiased estimators $\hat{\theta}_1, \hat{\theta}_2$ for a parameter θ, with, variances Var $(\hat{\theta}_1)$, Var $(\hat{\theta}_2)$ respectively, then the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$ denoted by $eff(\hat{\theta}_1, \hat{\theta}_2) = \frac{Var(\hat{\theta}_2)}{Var(\hat{\theta}_1)}$

Therefore, efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$ > 1 iff Var $(\hat{\theta}_2) >$ Var $(\hat{\theta}_1)$.

***Interval Estimation*:**

In statistics, interval estimation is the use of sample data to calculate an interval of possible values of an unknown population parameter, in contrast to point estimation, which is a single number. Neyman (1937) identified interval estimation ("estimation by interval") to be different than point `estimation ("estimation by unique estimate").

Interval estimation is the estimation of unknown parameter through a random interval known as **Confidence Interval (C.I)** and the end points of the interval is known as **Confidence Limits**. i.e., the value of θ will be covered by two values say a,b, (b>a) depending on the sample $X_1, X_2, .., X_n$ with a define probability (1 – α) known as **Confidence Coefficient**; it is the probability that a C.I will enclose θ.

If we choose different samples randomly from a population, we may construct different C.I, then the parameter θ is expected to lie within those set of intervals with $100(1-\alpha)\%$ confidence, α being the level of significance.

Note: for a point estimate, it is impossible to determine how far the estimate deviates from the true value of the parameter. If we need to obtain an estimate a parameter with some measure of possible error to be attached to the estimator.

Hence, interval estimation is a rule specifying a method for using the sample measurements to calculate two numbers that form the ends of the interval; this interval should satisfies:

1. It should contain the specific parameter θ .
2. It should be relatively narrow.

Therefore, the main objective is to find an interval estimator capable of generating narrow intervals that have high probability of enclosing θ.

Suppose that $\hat{\theta}_L, \hat{\theta}_U$ or $(\hat{\theta}_a, \hat{\theta}_b)$ are the random lower and upper condifence limits respectively for a parameter θ, if $p\{\hat{\theta}_L < \theta < \hat{\theta}_U\} = 1 - \alpha$, then the probability $(1 - \alpha)$ is called the confidence coefficient and $[\hat{\theta}_a, \hat{\theta}_b]$ is called a two-sided C.I.

It is possible to form a one-sided interval such that $p\{\hat{\theta}_L \leq \theta\} = 1 - \alpha$, C.I = $[\hat{\theta}_L, \infty)$

Or $p\{\theta \leq \hat{\theta}_U\} = 1 - \alpha$, hence C.I = $[-\infty, \hat{\theta}_U]$. We can construct C.Is for all parameters, using random samples.

How to construct C.I for parameters:

Confidence Interval for Means

i. (if σ^2 is unknown) :

Suppose that the outcome of a random experiment is a R.V. that has a normal distribution with known variance σ^2, but unknown mean μ , (μ is the target parameter).

In order to obtain some information about μ, the experiment is repeated n-times independently and under identical conditions.

Let $X_1, X_2, .., X_n$, be a random sample from the normal population, $\{X \sim N(\mu, \sigma^2)\}$, σ^2 is known, then the MLE for the parameter μ is $\overline{X}$. Then C.I for μ will be

$[\overline{X} - Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}, \overline{X} + Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}]$ is a $100(1-\alpha)\%$ confidence interval for μ where:

$\hat{\theta}_L = \overline{X} - Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}$, $\hat{\theta}_U = \overline{X} + Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}$

Note: C.I for μ is depending on $\overline{X} \mp$ constant(n), as n increases $Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}$ decreases resulting in a shorter C.I with the same confidence coefficient $(1 - \alpha)$. A shorter C.I indicates that we have more reliance in $\overline{X}$ as an estimate of μ. It is possible to increase the C.I as short as possible by increasing the sample size n.

For a fixed sample size n, the length of the C.I can also be shortened by increasing the confidence coefficient $(1-\alpha)$.

(If σ^2 is unknown n is large: S^2, the variance of a R.S of size n>2, (converges statistically to σ^2) is used as an estimated value for σ^2, then the C.I for μ is $[\bar{X} \mp Z_{\frac{\alpha}{2}} * \frac{S}{\sqrt{n}}]$, where $[\bar{X} \mp Z_{\frac{\alpha}{2}} * \sigma/\sqrt{n}] \cong [\bar{X} \mp Z_{\frac{\alpha}{2}} * \frac{S}{\sqrt{n}}]$. (If σ^2 is known and n is small) : Z is replaced by T- distribution.

Hence, $[\bar{X} \mp Z_{\frac{\alpha}{2}} * \sigma/\sqrt{n}] \cong [\bar{X} \mp T_{\frac{\alpha}{2},} * \frac{S}{\sqrt{n}}]$, with (n-1) degrees of freedom.

Similarly, the C.I for all other parameters could be constructed.

Examples; H.Ws and Assignments

Definition:(Meaning of significant)

In English significant means important but in statistics it means probability true (not due to chance). Highly significant means very probable.

If we choose different samples randomly from a population, we may construct different C.I, then the parameter θ is expected to lie within those set of intervals with $100(1-\alpha)\%$ confidence, or α being the level of significance.

Note: for a point estimate, it is impossible to determine how far the estimate deviates from the true value of the parameter. If we need to obtain an estimate a parameter with some measure of possible error to be attached to the estimator.

Confidence Interval for Means

ii. (if σ^2 is unknown) :

Suppose that the outcome of a random experiment is a R.V. that has a normal distribution with known variance σ^2, but unknown mean μ , (μ is the target parameter).

In order to obtain some information about μ, the experiment is repeated n-times independently and under identical conditions.

Let $X_1, X_2, .., X_n$, be a random sample from the normal population, $\{ X \sim N(\mu, \sigma^2)\}$, σ^2 is known, then the MLE for the parameter μ is $\bar{X}$. Then C.I for μ will be

$[\overline{X} - Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}, \overline{X} + Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}]$ is a $100(1-\alpha)\%$ confidence interval for μ where:

$\hat{\theta}_L = \overline{X} - Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}$, $\hat{\theta}_U = \overline{X} + Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}$

Note: C.I for μ is depending on $\overline{X} \mp$ constant(n), as n increases $Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}$ decreases resulting in a shorter C.I with the same confidence coefficient $(1-\alpha)$. A shorter C.I indicates that we have more reliance in $\overline{X}$ as an estimate of μ. It is possible to increase the C.I as short as possible by increasing the sample size n.

For a fixed sample size n, the length of the C.I can also be shortened by increasing the confidence coefficient $(1-\alpha)$.

(If σ^2 is unknown n is large: S^2, the variance of a R.S of size n>2, (converges statistically to σ^2) is used as an estimated value for σ^2, then the C.I for μ is $[\overline{X} \mp Z_{\frac{\alpha}{2}} * \frac{S}{\sqrt{n}}]$, where $[\overline{X} \mp Z_{\frac{\alpha}{2}} * \sigma/\sqrt{n}] \cong [\overline{X} \mp Z_{\frac{\alpha}{2}} * \frac{S}{\sqrt{n}}]$.

(If σ^2 is known and n is small) : Z is replaced by T- distribution.

Hence, $[\overline{X} \mp Z_{\frac{\alpha}{2}} * \sigma/\sqrt{n}] \cong [\overline{X} \mp T_{\frac{\alpha}{2}, *} \frac{S}{\sqrt{n}}]$, with (n-1) degrees of freedom.

Similarly, the C.I for all other parameters could be constructed.

Examples; H.Ws and Assignments.