UNIT-V

The inductive inference may be termed as the logic of drawing statistically valid conclusions about the population characteristics on the basis of a sample drawn from it in a scientific manner. We know the utility of sample method over complete enumeration (census) method and described the various sampling techniques of obtaining representative samples from the population. In this, we shall develop the technique which enables us to generalize the results of the sample to the population; to find how far these generalizations are valid, and also to estimate the population parameters along with the degree of confidence. The answers to these and many other related problems are provided by a very important branch of statistics, known as ***Statistical Inference.***

Parameter The statistical constant of the population like mean (μ), variance(σ^2), skewness (β_1), kurtosis (β_2), moments (μ_r), correlation coefficient ($\dot{\rho}$) etc. are known as parameters. Obviously, parameters are function of the population values. Generally, the population parameters are unknown and their estimates provided by the appropriate sample statistics are used.

Parameter Space: Let us consider a random variable X with probability density function *P.d.f* $f(x,\theta)$. In most common applications, though not always, the functional form of the population distribution is assumed to be known except for the value of some unknown parameters θ which may take any value on a set Θ. This is expressed by writing the *p.d.f* in the form $f(x,\theta), \theta \in \Theta$.The set Θ, which is the set of all possible values of θ is called the ***parameter space.***

Sampling distribution of a statistic:

sampling distribution describes the manner in which a statistic or a function of statistic which is/are a functions of the random sample variate values $x_1, x_2, x_3. . . , x_n$ will vary from one sample to another of the sample size

e.g t, Z and F distributions

If we draw a sample of size n from a given finite population of size N , then total number of possible sample is

$Nc_n = N!/n!(N-n)! = k(\text{say})$

For each of these k samples we can compute some statistic $t=t(x_1, x_2, x_3. . . , x_n)$ in the particular the mean(bar) the variance(s^2) etc as given below

Sample No.	**Statistics**		
	t	**x(bar)**	$\mathbf{s^2}$
1	t_1	x_1(bar)	s_1^2
2	t_2	x_2(bar)	s_2^2
3	t_3	x_3(bar)	s_3^2
.	.	.	.
.	.	.	.
.	.	.	.
K	t_k	x_k(bar)	s_k^2

The set of the values of the statistic so obtained, one for each sample, constitutes what is called sampling distribution of the statistics. Statistics t may be regarded random variable which can take the values $t_1,t_2,t_3,\ldots t_k$ and we can compute the various statistical constants like mean, variance, skewness, kurtosis for its distribution.

For e.g the mean and variance of the sampling distribution of the statistics t is given by

$$t\ (bar) = 1/k(t_1+t_2+t_3\ldots t_k) = 1/k\sum t_i$$

$$\text{and } var(t) = 1/k[(t_1-t)^2 + k(t_2-t)^2 + k(t_3-t)^2 + \ldots + k[(t_k-t)] = 1/k\sum(ti-t(bar)] = 0$$

Types of Estimation

Point estimation and Interval estimation*:** A particular value of a statistic which is used to estimate a given parameter is known as a ***point estimate or estimator of the parameter. Also instead of estimating a single value of a parameter from sample values, a range t_1, t_2 of numbers, which constitute an interval, determined with the help of sample values and supposed to include the parameter θ with certain confidence level $\gamma=1-\alpha$ is known as confidence interval. t_1 and t_2 ($t_1<t_2$)are called the lower and upper limits of the ***interval estimate***.

Estimator and Estimate: A known function $T = t(x_1,x_2,x_3.\ .\ .\ ,x_n)$ of the observable variates of a random sample $x_1,x_2,x_3.\ .\ .\ ,x_n$ whose values are used to obtain the estimate of a parameter θ or a function of θ, is called an **estimator**. An estimator is itself a random variable. If $x_1,x_2,x_3.\ .\ .\ ,x_n$ are the values of the random sample X_1, X_2, $X_3,\ldots$, X_n, the value $t\ (x_1,x_2,x_3.\ .\ .\ ,x_n)$ of the estimator T_n is known as an estimate of the parameter θ. For example $1/n\sum X_i$ $(i=1,2,3\ldots,n)$is an estimator whereas x(Bar) = $1/n\sum x_i$ is an **estimate.**

Requirements of a good estimator.

A good estimator is one which is close to the true value of the parameter as possible. The following are some of the criteria which should be satisfied by a good ***estimator.***

(i) Unbiased (ii) Consistency (iii) Efficiency (iv) Sufficiency

Unbiased: An estimator $T_n = T(x_1, x_2, x_3. . . , x_n)$ is said to be an unbiased estimator of $\gamma(\theta)$ if $E(T_n) = \gamma(\theta)$, for all $\theta \varepsilon \Theta$

We have seen in sampling from a population with mean μ and variance σ^2, E(x) $= \mu$ and $E(s^2) \neq \sigma^2$but $E(s^2)= \sigma^2$. Hence there is a reason to prefer

$S^2 = 1/n-1^{\sum (x-x)2}$,to sample variance $s^2 = 1/n^{\sum (x-x)2}$

Example: x_1, x_2, x_3. . . ,x_n is a random sample from a normal population N(μ, 1).Show that $t=1/n\sum xi^2$ is an unbiased estimator of $\mu^2 +1$.

Sol. We have $E(x_i) =\mu$, $V(x_i) =1$;i=1,2,3,…, n

Now $E(x_i^2)= V(x_i)+\{ E(x_i)\}^2 = 1+\mu^2$

$E(t) = E(1/n\sum x_i^2) =1/n\sum E(x_i^2) =1/n\sum(1+\mu^2) =1+\mu^2$

Hence t is an unbiased estimator of $1+\mu^2$.

(ii) ***Consistency***. A statistic $t=t_n=t(x_1,x_2,x_3. . . ,x_n)$ based on a sample of size n is said to be a consistent estimator of the parameter θ if it converges in probability to θ, if $t_n \rightarrow \theta$, as $n\rightarrow\infty$.

Symbolically,

$$\lim_{n\rightarrow \theta} P(t_n\rightarrow \theta)=1$$

For any distribution, sample mean x(bar) is a consistent estimator of the population mean, sample proportion 'p' is a consistent estimator of population proportion and sample variance s^2 is a consistent estimator of population variance σ^2.

Remarks.1. A consistent estimator need not be biased. For example, the sample variance s^2 is a consistent estimator of the population variance but it is not unbiased.

2. Obviously, consistency is a property concerning the behaviour of an estimator for indefinitely large values of sample size n, as $n\rightarrow\infty$.It does not take into consideration the behaviour of the statistic for finite values of n.

Efficiency. An unbiased estimator T_n is said to be efficient than any other estimator T_n^* of $\gamma(\theta)$ if and only if

$$\mathbf{V(T_n) < V}\ (T_n^*)$$

Also relative efficiency of T_n as compared to T_n^* is given as

$$R.E = V\ (T_n^*)/\ V(T_n)$$

Crammer gave the term efficient estimator to mean a minimum variance unbiased estimator (MVUE) or best unbiased estimator. Hence, MVU estimator is unbiased and also among the class of unbiased estimators it possesses minimum variance.

A MVU estimator is unique in the sense that $V(T_n^*) = V(T_n)$ $T_n = T_n^*$

Example. A random sample (X_1, X_2, X_3, X_4, X_5) of size 5 is drawn from a Normal population with unknown mean μ. Consider the following estimators.

(i) $t_1 = X_1 + x_2 + x_3 + x_4 + x_5/5$, (ii) $t_2 = X_1 + X_2/2 + X_3$, (iii) $t_3 = 2X_1 + X_2/3$. State give reasons, the estimator which is best among t_1, t_2 and t_3.

Sol: we are given

$E(X_i) = \mu$, $var(X_i) = \sigma^2$,(say); $Cov(x_i,x_j) = 0$, $(i \neq j = 1,2,3,\ldots,n)$…(1)

Using eq. no.(1), we get

$Var(t_1) = 1/25\{var(X_1) + var(X_2) + var(X_3) + var(X_4) + var(X_5)\} = 1/5\sigma^2$

$Var(t_2) = 1/4\{var(X_1) + var(X_2)\} + var(X_3) = 1/2\,\sigma^2 + \sigma^2 = 3/2\,\sigma^2$

$var(t_3) = 1/9\{4var(X_1) + var(X_2)\} = 1/9(4\sigma^2 + \sigma^2) = 5/9\,\sigma^2$

Since $V(t_1)$ is least, t_1 is the best estimator (in case of least variance.

Sufficiency. This is the last property that a good estimator should posses. A statistic $t = t(x_1, x_2, x_3, \ldots, x_n)$ is said to be a sufficient estimator of parameter θ if it contains all the information in the sample regarding the parameter. In other words, a sufficient statistic utilizes all the information that a given sample can furnish about the parameter. If $t = t(x_1, x_2, x_3, \ldots, x_n)$ is a statistic based on a random sample of size from a population with probability function or *pdf* (x, θ) then it is a sufficient estimator of θ if the conditional distribution $P[x_1 \cap x_2 \cap x_3 \cap, \ldots, \cap x_n / t = k]$ does not depend on θ.

The sample mean x(bar) is sufficient estimator of population mean μ and sample proportion ‘p’ is a sufficient estimator of population proportion P.

Properties of sufficient estimators:

1. If a sufficient estimator exists for some parameter then, it is also the most efficient estimator.

2. It is always consistent.

3. It may or may not be biased.

4. A minimum variance unbiased estimator (M.V.U.E) for a parameter exists if and only if there exists a sufficient estimator for it.

Factorization Theorem (Neymann). The necessary and sufficient condition for a distribution to admit sufficient statistic is provided by the 'factorization theorem' due to Neymann.

Statement. T =t(x) is sufficient for θ if and only if the joint density function L (say), of the sample values can be expressed in the form:

$$L = g_\theta [t(x)].h(x)$$

Where as indicated $g_\theta [t(x)]$ depends on θ and x only through the value of t(x) and h(x) is independent of θ.

Example. Let $x_1, x_2, x_3. . . ,x_n$ be a random sample from a uniform population on [0, θ]. Find a sufficient estimator of θ.

Sol. We have

$$L = \prod f(x_i, \theta) = 1/ \theta^n; 0<xi< \theta \ldots.(1)$$

If t=max $(x_1, x_2, x_3. . . ,x_n) = x_n$, then p.d.f of t is given by

$g(t, \theta) = n\{F(x_n)\}^{n-1}.f(x(n))$

We have $F(x) = P(X \le x) = \int^x{}_0 f(x, \theta)dx = \int^x{}_0 1/\theta \, dx = x/\theta$

$g(t, \theta) = n\{x_{(n)}/\theta\}^{n-1}(1/\theta) = n/\theta^n [x_{(n)}]^{n-1}$

Rewriting equation 1, we have

$L = n[x_{(n)}]^{n-1}/\theta^n \, .1/n[x_{(n)}]^{n-1} = g(t,\theta).\, h(x_1,x_2,x_3 . . . ,x_n)$

Hence by Factorization Neymann criterion, the statistic t=x(n), is sufficient estimator for θ.

Method of maximum likelihood estimation.

Maximum likelihood principle is due a R.A Fisher in 1921.

Let $X_1, X_2, X_3,\ldots, X_n$ be n independent observations from $f(x;\theta)$ where θ is a single unknown parameter. The joint probability density function of the sample is called likelihood function (LF) and is written as

$L(x/\theta) = f(x_1, \theta)\, f(x_2, \theta)\, f(x_3, \theta)\ldots f(x_n, \theta)$

According to maximum likelihood principle, one should take the value of estimator θ within the admissible range of θ which makes $L(x/\theta)$ maximum. For this, the method of maxima-minima is used. If $L(x/\theta)$ is differentiable twice i.e if the first and second derivative of $L(x/\theta)$ exists, put $L^{\backslash}(x/\theta) = 0$ and solve for θ. Also, for maxima check that $L^{\backslash\backslash}(x/\theta)$ is negative for a value of θ obtained by $L^{\backslash}(x/\theta)$. If so, solution of $L(x/\theta)$ provides the maximum likelihood estimate of θ. In practice it is better to take logarithm of $L(x/\theta)$ and then differentiate and solve it. This makes estimation process easier.

If θ is a K dimensional parametric vector, i.e $\theta = (\theta_1,\theta_2,\theta_3,\ldots,\theta_k)$, then the estimator $(\theta^{\wedge}{}_1,\theta^{\wedge}{}_2,\theta^{\wedge}{}_3,\ldots,\theta^{\wedge}{}_k)$, which maximize $L(x/\theta_1,\theta_2,\theta_3,\ldots,\theta_k)$, can be obtained by differentiating partially the Log{ $L(x/\theta_1,\theta_2,\theta_3,\ldots,\theta_k)$, with respect to $\theta_1,\theta_2,\theta_3,\ldots,\theta_k$ respectively and equating them to zero. The solution of k equations provides the estimates of $\theta_1,\theta_2,\theta_3,\ldots,\theta_k$.

Notationally,

$\partial/\partial\theta_1$ {Log $L(x/\theta_1,\theta_2,\theta_3,\ldots,\theta_k)$}=0

$\partial/\partial\theta_2$ {Log $L(x/\theta_1,\theta_2,\theta_3,\ldots,\theta_k)$}=0

.

.

. $\partial/\partial\theta_k$ {Log $L(x/\theta_1,\theta_2,\theta_3,\ldots,\theta_k)$}=0

In this we get k equations in k unknowns. These equations are often called the likelihood equations. Solving these equations one gets the maximum likelihood estimates of θ_1, θ_2, $\theta_3,\ldots$, θ_k. To show that $\hat{\theta}$, the k-dimensional vector of estimates provides the supremum of $L(x/\theta)$, it is enough to show that the matrix

$(\partial^2 \text{Log } L/\partial\theta_i\, \partial\theta_j)\theta=\hat{\theta}$ is negative definite.

Example: If X is a poisson variate with parameter μ, find the maximum likelihood estimate of μ.

Ans. $P(x;\mu)= e^{-\mu}\mu^{x}/x!$

for x=0,1,2,…n.

The likelihood function

$L(x!/\mu)=\prod e^{-\mu}\mu^{x_i}/x_i!$

for i=1,2,3,…n

$\text{Logl} =\sum \text{Log}_e e^{-\mu}+\sum x_i \log\mu-\sum \log_e(x_i!)$

$=-n\mu+\sum x_i \log\mu-\sum \log_e(x_i!)$

$\partial \text{Log } L/\partial\mu=-n+\sum x_i/\hat{\mu}-0=0$

$\sum x_i / \hat{\mu} = n$

or $\hat{\mu} = \sum x_i / n =$ ***x(bar)***

Sample mean is the maximum likelihood estimate of μ.

Methods of moments:

let ***$x_1, x_2, x_3, \ldots, x_n$ be*** n random samples for distribution function $f_\theta(x))/p_\theta(x)$where θ is population parameters or it can be written as

$f(\theta_1,\theta_2,\theta_3,\ldots,\theta_k)$ where $\theta_1,\theta_2,\theta_3,\ldots,\theta_k \in \Theta$

then by the method of moments

$m'r=\mu'r$

where

$m'r$ are the sample moments and μ/r are the population moments..

Example: if $x \sim U(0, \theta)$. Obtain estimate of θ by moments method.

Sol. $x \sim U(0, \theta)$

$f(x,\theta)=1/\theta;\ 0<x<\theta$

$\mu'r = m'r$

$\mu'1 = m'1 \ldots (1)$

$\mu'1 = E(x)= = \int^{\theta}_{0} x\ 1/\theta\ dx=1/\theta[x^2/2]^{\theta}_{0} = \theta^2/2\theta= \theta/2$

from eq.(1)

$\theta/2 =1/n\sum xi$

$\theta/2 = x(bar)$

or $\hat{\theta} = 2(x\ bar)$