Pitambar Das, Assistant Professor, Department of Mathematics

Netaji Nagar Day College

Topic for – Paper MTMA CC-11(Probability & Statistics)

Statistical Inference

Statistics is a branch of Mathematics that deals with the collection, analysis, interpretation and the presentation of the numerical data. In other words, it is defined as the collection of quantitative data. The main purpose of Statistics is to make an accurate conclusion using a limited sample about a greater population.

Types of Statistics

Statistics can be classified into two different categories. The two different types of Statistics are:

- Descriptive Statistics
- Inferential Statistics

In Statistics, **descriptive statistics** describe the data, whereas **inferential statistics** help you make predictions from the data. In inferential statistics, the data are taken from the sample and allows you to generalize the population. In general, inference means "guess", which means making inference about something. So, statistical inference means, making inference about the population. To take a conclusion about the population, it uses various statistical analysis techniques.

Statistical Inference

Statistical inference is the process of analysing the result and making conclusions from data subject to random variation. It is also called inferential statistics. Hypothesis testing and confidence intervals are the applications of the statistical inference. Statistical inference is a method of making decisions about the parameters of a population, based on random sampling. It helps to assess the relationship between the dependent and independent variables. The purpose of statistical inference to estimate the uncertainty or sample to sample variation. It allows us to provide a probable range of values for the true values of something in the population. The components used for making statistical inference are:

- Sample Size
- Variability in the sample

- Size of the observed differences

Types of Statistical Inference

There are different types of statistical inferences that are extensively used for making conclusions. They are:

- One sample hypothesis testing
- Confidence Interval
- Pearson Correlation
- Bi-variate regression
- Multi-variate regression
- Chi-square statistics and contingency table
- ANOVA or T-test

Statistical Inference Procedure

The procedure involved in inferential statistics are:

- Begin with a theory
- Create a research hypothesis
- Operationalize the variables
- Recognize the population to which the study results should apply
- Formulate a null hypothesis for this population
- Accumulate a sample from the population and continue the study
- Conduct statistical tests to see if the collected sample properties are adequately different from what would be expected under the null hypothesis to be able to reject the null hypothesis

Statistical Inference Solution

Statistical inference solutions produce efficient use of statistical data relating to groups of individuals or trials. It deals with all characters, including the collection, investigation and analysis of data and organizing the collected data. By statistical inference solution, people can acquire knowledge after starting their work in diverse fields. Some statistical inference solution facts are:

- It is a common way to assume that the observed sample is of independent observations from a population type like Poisson or normal

- Statistical inference solution is used to evaluate the parameter(s) of the expected model like normal mean or binomial proportion

Importance of Statistical Inference

Inferential Statistics is important to examine the data properly. To make an accurate conclusion, proper data analysis is important to interpret the research results. It is majorly used in the future prediction for various observations in different fields. It helps us to make inference about the data. The statistical inference has a wide range of application in different fields, such as:

- Business Analysis
- Artificial Intelligence
- Financial Analysis
- Fraud Detection
- Machine Learning
- Share Market
- Pharmaceutical Sector

Statistical Inference Examples

An example of statistical inference is given below.

Example: From the shuffled pack of cards, a card is drawn. This trial is repeated for 400 times, and the suits are given below:

Suit	Spade	Clubs	Hearts	Diamonds
No. of times drawn	90	100	120	90

While a card is tried at random, then what is the probability of getting a

1. Diamond cards
2. Black cards
3. Except for spade

Solution:

By statistical inference solution,

Total number of events = 400

i.e., 90+100+120+90=400

(1) The probability of getting diamond cards:

Number of trials in which diamond card is drawn = 90

Therefore, P(diamond card) = 90/400 = 0.225

(2) The probability of getting black cards:

Number of trials in which black card showed up = 90+100 =190

Therefore, P(black card) = 190/400 = 0.475

(3) Except for spade

Number of trials other than spade showed up = 90+100+120 =310

Therefore, P(except spade) = 310/400 = 0.775

Stay tuned with BYJU'S – The Learning App for more Maths-related concepts and download the app for more personalized videos.

Theory of Estimation:

One of the main objectives of Statistics is to draw inferences about a population from the analysis of a sample drawn from that population. Two important problems in statistical inference are (i) estimation and (ii) testing hypothesis.

The theory of estimation was founded by Prof. R.A. Fisher in series of fundamental papers round about 1930.

Definition: Any function of the random sample $x_1, x_2, \ldots, x_n$ that are being observed, say $T_n(x_1, x_2, \ldots, x_n)$ is called statistic. Clearly, a statistic is a random variable. If it is used to estimate an unknown parameter θ of the distribution, it is called an estimator. A particular value of the estimator, say, $T_n(x_1, x_2, \ldots, x_n)$ is called an estimate of θ.

Characteristics of Estimators

The following are some of the criteria that should by a good estimator.

(i) Unbiasedness, (ii) Consistency, (iii) Efficiency, and (iv) Sufficiency

Unbiasedness: An estimator t(x) is said to be an unbiased estimator for the population parameter θ if $E(t) = \theta$

Otherwise, t(x) is said to be biased estimator of θ

i.e., $E(t) \neq \theta$

i.e., $b(t) = E(t) - \theta$

If $b(t) > 0$, t is said to be positively biased.

If $b(t) < 0$, t is said to be negatively biased.

Example: $x_1, x_2, \dots, x_n$ is a random sample from this distribution $f(x,\theta) = \frac{1}{\theta} e^{\frac{-x}{\theta}} \quad , x > 0; \theta > 0$

Show that $\bar{x}$ is an unbiased estimator for θ.

Solution. We know $\bar{x} = \frac{1}{n}\sum_1^n x_i$

$$E(\bar{x}) = \frac{1}{n}\sum_{i=1}^n E(x_i)$$

$$\therefore E(x) = \int_0^{\infty} x.\frac{1}{\theta}.e^{\frac{-x}{\theta}} dx$$

$$= \frac{1}{\theta}\int_0^{\infty} e^{\frac{-1.x}{\theta}} x^{2-1} dx$$

$$= \frac{1}{\theta} \frac{\overline{|2}}{\left(\frac{1}{\theta}\right)^2} = \frac{\theta^2}{\theta} = \theta$$

Hence $\bar{x}$ is an unbiased estimator for θ.

Example: $x_1, x_2, \dots, x_n$ is a random sample from a normal population $N(\mu, \sigma^2)$. Show that $\bar{x}$ and s^2 are unbiased estimator for μ and σ^2 respectively.

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

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ (Sample mean)

$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ (Sample mean Square)

Solution. $E(\bar{x}) = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)$

$= \frac{1}{n}\sum_{i=1}^{n} E(x_i)$

$= \frac{1}{n}\sum_{i=1}^{n} n$ Since $E(x_i) = \mu$

$= \frac{n\mu}{n}$

$= \mu$

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

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} {x_i}^2 - n\bar{x}^2\right]$$

$$E(s^2) = \frac{1}{n-1}\left[\sum_{i=1}^{n} E({x_i}^2) - nE(\bar{x}^2)\right]$$

We know that $V(x_i) = E({x_i}^2) - \left(E(x_i)\right)^2$

$$E({x_i}^2) = V(x_i) + \left(E(x_i)\right)^2$$

$$= \sigma^2 + \mu^2$$

$$E(\bar{x})^2 = V(\bar{x}) + [E(\bar{x})]^2$$

$$= V\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] + [E(\bar{x})]^2$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} V(x_i) + \mu^2$$

$$= \frac{1}{n^2}. n\sigma^2 + \mu^2$$

$$= \frac{\sigma^2}{n} + \mu^2$$

$$E(s^2) = \frac{1}{n-1}\left[\sum_{i=1}^{n} E(x_i{}^2) - nE(\bar{x}^2)\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right]$$

$$= \frac{1}{n-1}\left[n(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right]$$

$$= \frac{1}{n-1}[n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2]$$

$$= \frac{1}{n-1}[(n-1)\sigma^2]$$

$$= \sigma^2$$

Hence $E(\bar{x}) = \mu$

$E(s^2) = \sigma^2$

(ii) **Consistency**: An estimator t is said to be consistent estimator of a parameter θ if $E(t) = \theta$ and $V(t) \to 0$ as $n \to \infty$.

Example. $x_1, x_2, \ldots, x_n$ is a random sample from a normal population $N(\mu, \sigma^2)$. Show that sample variance S^2 is biased and consistent estimator for σ^2.

Where $S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$

$$E(S^2) = E\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right]$$

$$= \frac{1}{n}E[\sum_{i=1}^{n} x_i{}^2 - n\bar{x}^2]$$

$$= \frac{1}{n}[\sum_{i=1}^{n} E(x_i{}^2) - nE(\bar{x}^2)]$$

$$=\frac{1}{n}\left[\sum_{i=1}^{n}(\mu^2+\sigma^2)-n\left(\mu^2+\frac{\sigma^2}{n}\right)\right]$$

$$=\frac{1}{n}\left[n\mu^2+n\sigma^2-n\mu^2-\sigma^2\right]$$

$$=\frac{1}{n}\left(n\sigma^2-\sigma^2\right)$$

$$=\sigma^2\left(1-\frac{1}{n}\right)$$

$$E(S^2)=\left(1-\frac{1}{n}\right)\sigma^2\neq\sigma^2$$

Hence S^2 is biased for σ^2

$$\sum_{i=1}^{n}\frac{(x_i-\overline{x})^2}{\sigma^2}\sim\chi_{n-1}^2$$

$$\frac{nS^2}{\sigma^2}\sim\chi_{n-1}^2$$

$$V\left(\frac{nS^2}{\sigma^2}\right)=2(n-1)$$

$$V(S^2)=\frac{2(n-1)\sigma^4}{n^2}$$

$$=2\left(\frac{1}{n}-\frac{1}{n^2}\right)\sigma^4$$

$E(S^2)=\left(1-\frac{1}{n}\right)\sigma^2\rightarrow\sigma^2$ as $n\rightarrow\propto$

$V(S^2)=2\left(\frac{1}{n}-\frac{1}{n^2}\right)\sigma^4\rightarrow 0$ as $n\rightarrow\propto$

Property of Efficiency: A estimator t_1 is said to be more efficient estimator than other estimator t_2 for estimating the unknown parameter θ.

If $V(t_1) < V(t_2)$

The efficiency of the estimator t_1 with respect to the estimator t_2 is defined as

$$e = \frac{V(t_1)}{V(t_2)}, 0 \leq e \leq 1$$

Where e = 1, $V(t_1) = V(t_2)$, both are equally efficient.

Example. $x_1, x_2, \ldots, x_n$ is a random sample from a normal population $N(\mu, \sigma^2)$. Show that the sample mean $\bar{x}$ is more efficient than sample median $\tilde{x}$ for estimating the population mean μ.

Solution. $E(x_i) = \mu$, $E(\bar{x}) = \mu$, $Var(\bar{x}) = \frac{\sigma^2}{n}$

$$E(\tilde{x}) = \mu, Var(\tilde{x}) = \frac{1}{4nf_1^{\,2}}$$

f_1 = median ordinate

$$f_1 = f(x) \vdots x = \mu$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}; -\infty < x < \propto$$

$$f_1 = f(x) \vdots x = \mu = \frac{1}{\sigma\sqrt{2\pi}}$$

$$V(\tilde{x}) = \frac{1}{4n}\sigma^2 . 2\pi$$

$$= \frac{\pi}{2} . \frac{\sigma^2}{n}$$

$$V(\tilde{x}) = \frac{\pi}{2} V(\bar{x})$$

$$V(\bar{x}) < V(\tilde{x})$$

Sufficiency: An estimator / statistics t is said to be sufficient for estimating the unknown parameter θ if the conditional probability distribution of $x_1, x_2, \ldots, x_n$ given t, is independent of θ.

Example. $x_1, x_2, \ldots, x_n$ is a random sample from a Poisson distribution with parameter λ; $p(x) = \dfrac{e^{-\lambda}\lambda^{x}}{x!}$. Find the sufficient for λ.

Solution.

$$L = \prod_{i=1}^{n} p(x_i)$$

$$= \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{\lfloor x_i}$$

$$= \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} \lfloor x_i}$$

$$= g_{\lambda}^{(t(x))} \cdot h(x)$$

Where

$$g_{\lambda}^{(t(x))} = = e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}$$

$$h(x) = \frac{1}{\prod_{i=1}^{n} \lfloor x_i}$$

$t = \sum_{i=1}^{n} x_i = n\overline{x}$ is sufficient for λ.

Or $\overline{x}$ is sufficient for λ.