**Question: What is sample?**

Answer:

In statistics, a sample is a subset of a larger population. The purpose of taking a sample is to make inferences about the population, based on the characteristics of the sample. Sample is the representative part of a population.

**Question: What is sampling? Explain with example.**

Answer:

Sampling is a scientific process of selecting a representative part from a population and also embrace the derivation of estimates and any inferences derived from them for population.

Example: The procedure of choosing the portion of students from the entire student body is technically known as sampling.

**Question: Explain different types of sampling.**

Answer:

Sampling is a fundamental concept in statistics and research methodology, wherein a subset of individuals or items is selected from a larger population. Different sampling techniques are employed depending on the research objectives, population characteristics, and available resources. Here are some common types of sampling methods:

**Purposive sampling:** Purposive sampling is one in which the sample units are selected with definite purpose in view. This sampling suffers from. the drawback of favoritism and nepotism and does not give a representative sample of the population. For example, suppose a researcher wants to collect feedback from students on the pedagogical methods in their school. The researchers will select the brightest students who can provide relevant information for systematic investigation.

**Random sampling:** the sample units are selected at random and the drawback of purposive sampling, viz... favoritism or subjective element, is completely overcome. A random sample is one in which each unit of population has an equal chance of being included in it. Examples include drawing names out of a hat, using random number generators, or assigning each member a unique number and selecting randomly.

**Simple sampling:** Simple sampling is random sampling in which each unit of the population has an equal chance, say p, of being included in the sample and that this probability is independent of the previous drawings. Thus, a simple sample of size n from a population may. be identified with a series of n independent trials with constant probability 'p' of success for each trial.

**Stratified sampling:** The population is divided into homogeneous subgroups or strata based on certain characteristics (e.g., age, gender, income). Random samples are then drawn from each stratum proportionally or disproportionately based on their representation in the population.

**Question: What are parameters and statistic?**

Answer:

Parameters and statistics are fundamental concepts in statistics that are used to describe and summarize data.

**Parameters:**

Parameters are numerical characteristics of a population. A population in statistics refers to the entire group of individuals or items that researchers are interested in studying. Parameters are typically fixed and unknown values that describe various aspects of the population distribution.

Examples of parameters include the population mean, population variance, population standard deviation, and population proportion.

Parameters are often denoted using as μ (mu) for the population mean, σ² (sigma squared) for the population variance, and p for the population proportion.

**Statistics:**

Statistics, on the other hand, are numerical characteristics of a sample. A sample is a subset of the population that is selected for observation or study. Statistics are calculated from sample data and are used to estimate or infer the corresponding parameters.

Unlike parameters, statistics are known values since they are calculated from observed data.

Examples of statistics include the sample mean, sample variance, sample standard deviation, and sample proportion.

Statistics are often denoted as x̄ (x-bar) for the sample mean, s² (s squared) for the sample variance, and p̂ (p-hat) for the sample proportion.

**Jacobian Transformation:**

Let continuous random variables and functions as follows,

Where . Now we can find the joint distribution of by an extension of the foregoing method. For the special case , we solve the set of equations for to obtain a set offunctions,

The determinant

Will be called the **Jacobian of the transformation** and will be denoted by **J .**

For example: *Let* Find the Jacobian *J**(u, v).*

**Solution:**

Given:

We know that**,**

Therefore,

**Cumulant Generating function:** The logarithm of the moment generating function or characteristics function of a random variable is called the cumulant generating function. It is generally denoted by

CGF of a random variable is defined by:

**Properties of cumulate generating function:**

1. The cumulate generating function is infinitely differentiable and passes through the origin. Its first derivative is monotonic function from the least to the greatest upper bounds of the probability distribution. Its second derivative is positive everywhere where it is defined.
2. Cumulates accumulate: The *kth* cumulant of a sum of independent random variables is just the sum of the *kth* cumulants of the summands.
3. Cumulate also have a scaling property: The *nth* cumulant of *n* *X* is *cn* times the *nth* cumulant of *X* .

**Statistics and parameter**

**Parameter:**

In statistics, a parameter is a numerical value that describes a characteristic of an entire population. It represents a fixed, unknown attribute of the population.

Example- Population mean, Standard deviation, Variance.

**Statistic:**

In statistics, a statistic is a numerical value calculated from sample data that provides information about a specific characteristic or features of the population from which the sample was drawn. It serves as an estimate or approximation of the corresponding parameter, which represents the same characteristic for the entire population.

Example- Sample mean, Sample standard deviation, Sample variance.

**Difference between Parameter and Statistic-**

|  |  |  |
| --- | --- | --- |
| Aspect | Statistic | Parameter |
| Definition | A numerical value calculated from sample data. | A numerical value describing an entire population. |
| Based on | Derived from sample data. | Describes the entire population. |
| Variability | Varies from sample to sample. | Fixed, unknown value. |
| Purpose | Used to estimate or infer population parameters. | Provides summary information about the population. |
| Examples | Sample mean, sample standard deviation, Sample variance. | Population mean, population variance, population standard deviation. |
| Denoted by | Typically denoted using Roman letters (e.g., x̄, s). | Typically denoted using Greek letters (e.g., μ, σ). |
| Usage | Central to inferential statistics. | Essential for describing populations and making predictions. |

**Random Variable**

A random variable is a key concept in probability, closely linked to uncertainty. It represents the possible outcomes of a random experiment, which cannot be predicted beforehand. The idea behind random variables is to express experiment results in numerical terms for clearer analysis and mathematical modeling of uncertainty.

Such as – The experiment of tossing two fair coins,

Here, we can consider two Random variable in this sample space- .

Note that, a random variable denoted by or for a given probability Space is a function with Domain and counter domain the Real line.

As a definition we can define definition as a function as well as a variable, like-

**Random Variable is a real valued function which assigns a real number to each sample point in the sample space.**

**A variable which values are determined by the outcomes of the experiments is known as random variable.**

There are two kinds of random variable-

1. Desecrate Random Variable.
2. Continuous Random Variable

**Continuous Random Variables**

Suppose now that we are given the joint probability density function of the n-dimensional continuous random variable *.* Let

Again assume that the joint density of the random variables is desired, where is some integer satisfying*. If* , we will introduce additional, new random variablesfor judiciously selected functions *,….,* then we will find the joint distribution of *,…,* and finally we will find the desired marginal distribution of *,…,*  from the joint distribution of *,…,* . This use of possibly introducing additional random variables makes the transformation a transformation from an n-dimensional space to an n-dimensional space. Henceforth we will assume that we are seeking the joint distribution of (rather than the joint distribution o*f ,…,* ) when we have given the joint probability density of *.*

We will state our results first for *n = 2* and later generalize to *n > 2*. Let be given. Set We want to find the joint distribution of and for known functions and *.* Now suppose that and defines a one-to-one transformation which maps onto . *x*₁ and can be expressed in terms of *y₁* and *y₂*; so we can write, say, *x₁ =*  and *x₁ =* . Note that *X* is a subset of the plane and is a subset of  *the* plane. The determinant

will be called the Jacobian of the transformation and will be denoted by *J*. The above discussion permits us to state Theorem 13.

Theorem 13 Let *X₁* and *X₂* be jointly continuous random variables with density function . Set

Assume that:

(i) and defines a one-to-one transformation of onto

(ii) The first partial derivatives of *x₁ =*  and *x₁ =*  are continuous over .

(iii) The Jacobian of the transformation is nonzero for . Then the joint density of and is given by

**Proof:** We omit the proof; it is essentially the same as the derivation of the formulas for transforming variables in double integrals, which may be found in many advanced calculus textbooks. is that subset of the plane consisting of points for which there exists *a*  such that

**Moment Generating Function Technique**

There is another method of determining the distribution of functions of random variables which we shall find to be particularly useful in certain instances. This method is built around the concept of the moment generating function and will be called the moment-generating-function technique.

The statement of the problem remains the same. For given random variables with given density and given functions find the joint distribution of Now the joint moment generating function of , if it exists, is

If after the integration of (i) is performed, the resulting function ofcan be recognized as the joint moment generating function of some known joint distribution, it will follow that has that joint distribution by virtue of the fact that a moment generating function, when it exists, is unique and uniquely determines its distribution function.

For , this method will be of limited use to us because we can recognize only a few joint moment generating functions. the moment generating function is a function of a single argument, and we should have a better chance of recognizing the resulting moment generating function.

This method is quite powerful in connection with certain techniques of advanced mathematics which, in many instances, enable one to determine the distribution associated with the derived moment generating function.

The most useful application of the moment-generating-function technique is the Distribution of sums. There it will be used to find the distribution of sums of independent random variables.

**Example 6**

Suppose has a normal distribution with mean and variance *. Let* , and find the distribution of .

which we recognize as the moment generating function of a gamma with parametersand

**Parent Distribution Vs Sample Distribution**

Parent distribution and sample distribution are concepts used in statistics to describe different aspects of a dataset:

|  |  |
| --- | --- |
| Parent Distribution | Sample Distribution |
| The parent distribution, also known as the population distribution, is the theoretical distribution that represents all possible outcomes or observations within a population. This distribution is often characterized by certain parameters such as mean (μ), variance (), skewness, and kurtosis. | The sample distribution represents the distribution of observations or measurements within a specific sample drawn from the parent distribution. Unlike the parent distribution, the sample distribution is empirical and is based on actual observed data. |
| Characteristics:   * Parameters: The parent distribution is described by population parameters, which are typically unknown and are estimated using sample statistics. * Theoretical: It represents the idealized distribution of the entire population, which may or may not be directly observable or accessible. * Infinite or Finite: The parent distribution may represent an infinite population (such as all possible heights of adults in a country) or a finite population (such as all possible ages of students in a classroom). * Fixed: The parent distribution is fixed and does not change, although our knowledge or understanding of it may evolve over time with more data or better statistical techniques. | **Characteristics**:   * **Statistics**: The sample distribution is described by sample statistics, which are calculated from the observed data in the sample. These statistics include sample mean (x̄), sample variance (s^2), sample skewness, and sample kurtosis. * **Empirical**: It is based on real data collected from a subset (sample) of the population, rather than being a theoretical construct. * **Variability**: Sample distributions can vary from one sample to another, especially for small sample sizes, due to sampling variability. |
| Examples:   * Normal Distribution: If we're studying the IQ scores of all adults in a country, the parent distribution might be a normal distribution with a certain mean and standard deviation. * Exponential Distribution: If we're analyzing the lifetimes of a particular type of battery produced by a company, the parent distribution might be an exponential distribution with a certain rate parameter. | **Examples:**   * If we take multiple random samples of 100 adults from a population and record their heights, each sample will have its own sample distribution of heights. * Similarly, if we collect multiple samples of test scores from students in a school, each sample will have its own sample distribution of scores. |
| Inference: Statistical inference involves making conclusions or predictions about the parent distribution based on sample data. Techniques such as hypothesis testing, confidence interval estimation, and parameter estimation are used for this purpose. | **Inference**: Sample distributions are used to draw conclusions about the parent distribution or to estimate population parameters. Statistical techniques involve making inferences about the parent distribution based on characteristics observed in the sample distribution. |

**Tests of Significance**

A very important aspect of the sampling theory is the study of the tests of significance, which enable us to decide on the basis of the sample results, if

1. the deviation between the observed sample statistic and the hypothetical parameter value, or
2. the deviation between two independent sample statistics; is significant or might be attributed to chance or the fluctuations of sampling.

Since, for large n, almost all the distributions, Binomial, Poisson, Negative binomial, Hyper-geometric, chi-square can be approximated very closely by a normal probability curve, we use the normal test of significance for large samples. Some of the well-known tests of significance for studying such differences for small samples are t-test, F-test and Fisher’s z-transformation.

**Example 2**

Let there be only one given random variable, say X, which has a standard normal distribution. Suppose the distribution of is desired.

*, for y > 0,*

which can be recognized as the cumulative distribution function of a gamma distribution with parameters and

Other applications of the cumulative-distribution-function technique expounded above are given in the following three subsections

**Example 6**

Suppose *X* has a normal distribution with mean 0 and variance 1.

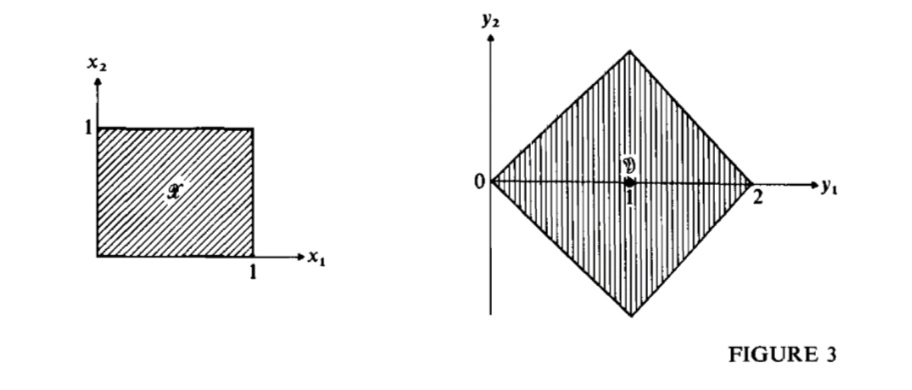
Let *,* and find the distribution of *Y*.

which we recognize as the moment generating function of a gamma with parameters and. (It is also called a chi-square distribution with one degree of freedom.)

**Example 22**

Suppose that *X1*and *X2*are independent random variables, each uniformly distributed over the interval (0, 1). Then *(x1, x2) = I (0, 1) (x1) I (0, 1) (x2). Ӿ = {(x1, x2): 0 < x1< 1* and *0 < x2 < 1}*. Let *y1 = g1(x1, x2) = x1 + x2* and *y2 = g2(x1, x2) = x2 – x1*; then *x1 = (y1 – y2) = g1-1 (y1, y2)*, and *x2 = (y1 + y2) = g2-1(y1, y2).*

*J = = = .*



and are sketched in Fig. 3. Note that the boundary of goes into the boundary, the boundary *x2 = 0* of Ӿ goes into the boundary *(y1 + y2) = 0* of *Y*, the boundary *x1 = 1* of Ӿ goes into the boundary *(y1 - y2) = 1* of *Y*, and the boundary *x2 = 1* of Ӿ goes into the boundary *(y1 + y2) = 1* of *Y*. Now the transformation is one-to-one, the first partial derivatives of *g1-1* and *g2-1*are continuous, and the Jacobian is nonzero; so

*(y1, y2) = ((y1, y2), g2-1(y1, y2))*

*= I (o, 1) () I (0, 1) ()*

*=*