Central Limit Theorem, Sampling Distributions

Large and Small samples

When the sample size is more than 30, the sample is known as large sample, otherwise small sample.

Statistics and Parameters

A measurable single valued function of the observations in a sample is called a statistic. If $x_1, x_2, \ldots\ldots x_n$ is a sample drawn from a population, then a function of $x_1, x_2, \ldots\ldots x_n$ is a statistic. For example: sample mean $= \frac{x_1 + x_2 + \ldots\ldots x_n}{n}$, is a function of the sample values. So it is a statistic.

So a measure obtained from a sample is a sample statistic.

Any function of the population values is called a parameter. For example, Mean of the population is a parameter. So a parameter is a measure obtained from the population.

Central limit theorem

Let $x_1, x_2, \ldots\ldots x_n$ be n independent variables. Let all have same distribution, same mean say μ and same standard deviation say σ. Then the mean of all these variables ie $\frac{x_1 + x_2 + \ldots\ldots x_n}{n}$ follows a normal distribution with mean $= \mu$ and S.D $= \frac{\sigma}{\sqrt{n}}$ when n is large.

Central limit theorem is considered to be one of the most remarkable theorems in the entire theory of statistics. The theorem is called central because of its central position in probability distribution and statistical inference.

So condition for central limit theorem are

a) variables must be independent.
b) all variables should have common mean and common S.D.
c) all variables should have same distribution.
d) n is very large.

Sampling Distributions

Sample statistic is a random variable. As every random variable has a probability distribution, sample statistic also has a probability distribution.

The probability distribution of a sample statistic is called the sampling distribution of that statistic. For example: Sample mean is a statistic and the distribution of sample mean is a sampling distribution

Sampling distribution plays a very important role in the study of statistical inference.

Standard error:

Standard deviation of a sampling distribution of a statistic is called the standard error of that statistic.

For example, sample mean ($\bar{x}$) has a sampling distribution. The S.D of that distribution is called standard error of $\bar{x}$.

Standard error of sample mean is $\frac{\sigma}{\sqrt{n}}$ where σ is the population S.D and n is sample size.

Uses of Standard Error

Standard error plays a very important role in the large sample theory and forms the basis of the testing of hypothesis.

(1) It is used for testing a given hypothesis

(2) S.E gives an idea about the reliability of a sample. The reciprocal of S.E is a measure of reliability of the sample.

(3) S.E can be used to determine the confidence limits of population measures like mean, proportion and standard deviation.

<u>Commonly used sampling distributions</u>

1) Normal distribution Large sample
2) χ^2 - distribution
3. t - distribution small sample
4) F - distribution

<u>Normal Distribution (As a samling Distribution)</u>

When the sample is large or when population SD is known, the following sample statistics have standard Normal distribution. Suppose we denote those statistic by Z.

1) $Z = \frac{\bar{x} - \mu}{\text{S. D of } \bar{x}}$ where $\bar{x}$ is the sample mean and μ is the population mean.

2) $Z = \frac{\bar{x}_1 - \bar{x}_2}{\text{S.D of } (\bar{x}_1 - \bar{x}_2)}$ where $\bar{x}_1$ and $\bar{x}_2$ are the standard deviations of two samples.

3) $Z = \frac{p - P}{\text{S.D of } p}$ where ‘p’ is the sample proportion and ‘P’ is the population proportion.

4) $Z = \frac{p_1 - p_2}{S.D \text{ of } (P_1 - P_2)}$ where p_1 and p_2 are sample proportions.

5) $Z = \frac{s - \sigma}{S.D \text{ of } S}$ where 'S' is sample S.D and σ is population S. D.

6) $Z = \frac{S_1 - S_2}{S.D \text{ of } (S_1 - S_2)}$ where S_1 and S_2 are the standard deviations of two samples.

In all these cases Z follows a standard normal distribution. ie Z follows N (0, 1). Probability function of z is $f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ for $-\infty < z < \infty$.

Properties of 'Z' distribution

1. z-distribution is a normal distribution. So it has all the properties of a normal distribution. Further

1) its mean = 0 and SD = 1

2) $P(-t < z < 0) = P(0 < z < t) = \int_0^t \phi(z)\,dt$ where $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$

3) $\int_0^\infty \phi(z)\,dz = \int_{-\infty}^0 \phi(z)\,dz = 0.5$

Uses of the sampling distribution of z

1. To test the given population mean.
2. To test the significance of difference between two population means.
3. To test the given population proportion.
4. To test the difference between two population proportions.
5. To test the given population S. D
6. To test the difference between two population Standard deviations.

χ^2- distribution [Chi - square Distribution]

1. If Z follows a standard normal distribution, then Z^2 will follow χ^2 distribution with one degree of freedom.
2. Let 's' and 'σ' be the standard deviations of sample and population respectively. Let 'n' be the sample size. Then $\frac{ns^2}{\sigma^2}$ follows a χ^2 distribution with n – 1 degrees of freedom.
3. If $z_1, z_2, \ldots\ldots\ldots z_n$ are n standard normal variates then $z_1^2, z_2^2, \ldots\ldots z_n^2$ follows χ^2 distribution with n degrees of freedom.

Probability density function of χ^2 - distribution:

A continuous random variable χ^2 is said to follow χ^2 distribution if

$$f(\chi^2) = \frac{(\frac{1}{2})^{\frac{n}{2}}}{\left|\frac{n}{2}\right.} e^{-\frac{\chi^2}{2}} (\chi^2)^{\frac{n}{2}-1} \text{ for } 0 \le \chi^2 \le \infty$$

This distribution is known as χ^2 distribution with 'n' degrees of freedom. The parameter of the distribution is n.

Properties of χ^2 distribution

1. χ^2 distribution is a sampling distribution. It is a continuous probability distribution.
2. Parameter of χ^2 distribution is n.
3. As the degree of freedom increases, χ^2 distribution approaches to Normal Distribution.
4. Mean of χ^2 distribution is n, variance of χ^2 distribution is 2n and mode of χ^2 distribution is n – 2 where 'n' is the degree of freedom.
5. For large values of n, χ^2 distribution is symmetric.
6. Sum of two independent χ^2 variates is also a χ^2 variate.

Uses of χ^2 distribution

χ^2 is a test statistic in tests of hypotheses. Following are the uses of χ^2

1. To test the given population variance when sample is small.
2. To test the goodness of fit between observed and expected frequencies.
3. To test the independence of two attributes.
4. To test the hogoneneity of data.

<u>Students t - distribution</u>

1. Let $\overline{x}$ and s be the mean and S.D of a sample drawn from a normal population and let sample size (n) be small, then $\frac{\overline{x}-\mu}{s/\sqrt{n-1}} = t$ follows a t - distribution with n – 1 degrees of freedom.
2. Let $\overline{x}_1$, s_1 and n_1 be the mean, SD and the size of a sample. Let $\overline{x}_2$ s_2 and n_2 be the mean, SD and the size of another sample. Both the samples be small. Suppose the samples are drawn from normal populations with same mean and same variance. Then

$$\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = t$$ Follows t - distribution with $n_1 + n_2 - 2$ degrees of freedom

Probability function of t - distribution:

A random variable 't' is said to follow t - distribution if its probability density function is $f(t) = \frac{\left|\frac{n+1}{2}\right.}{\sqrt{n\pi}\left|\frac{n}{2}\right.} (1+\frac{t^2}{n})^{-\frac{n+1}{2}}$ for $-\infty < t < \infty$

O′.5 St.

Here 'n' is the degree of freedom.

Properties of t-distribution

1. t-distribution is a sampling distribution.
2. For large samples, t-distribution approaches to normal distribution.
3. All odd moments of the distribution are 0.
4. Mean = 0 and variance = $\frac{n}{n-2}$ for $n > 2$ and n is degree of freedom.
5. t-curve is maximum at $t = 0$.
6. t-curve has long tails towards the left and the right.

Uses of t distribution

The variable 't' is a statistic and it is used in many tests of hypotheses. Those tests are known as t tests and are

a) to test the given population mean when sample is small.

b) to test whether the two samples have same mean when the samples are small

c) to test whether there is difference in the observations of the two dependent samples.

d) to test the significance of population correlation coefficient.

F distribution

$$F = \frac{\dfrac{n_1 s_1^2}{n_1 - 1}}{\dfrac{n_2 s_2^2}{n_2 - 1}}$$

follows F distribution with $(n_1 - 1, n_2 - 1)$ degrees of freedom

where n_1 = size of first sample, n_2 = size of second sample.

S_1 = S.D of first Sample
S_2 = S.D of Second Sample.

Properties of f distribution

1) It is a Sampling distribution.
2) Mean = $\frac{n_2}{n_2 - 2}$
3) If F follows F distribution with (n_1, n_2) degrees of freedom then $1/F$ follows F distribution with (n_2, n_1) degrees of freedom.
4) F ~~cu~~ curve is j shaped when $n_2 \leq 2$ and bell Shaped when $n_1 > 2$.

Uses of f distribution

1) To test equality of variance of 2 populations.
2) To test equality of means of 3 or more ~~dist~~ Population.

Imp: Relations b/w Normal distribution, Standard normal distribution, χ^2, t and F Distributions

when X follows normal distribution with mean = μ and S.D = σ then $Z = \frac{x-\mu}{\sigma}$ follows Standard normal distribution (Relationship b/w Normal distribution & Standard Normal distribution) when $z_1, z_2 \ldots z_k$ are k variables which follows standard normal distribution then $\sum z_i^2$ follows χ^2 distribution with k degrees of freedom. (Relationship b/w SND and χ^2)

If Z follows standard Normal distribution and y follows χ^2 distribution with k degrees of freedom then $\dfrac{z}{\sqrt{y/k}}$ follows t distribution with k degrees of freedom (relationship b/w χ^2 and t).

If y_1, y_2 are 2 χ^2 variables with n_1 and n_2 degrees of freedom then $\dfrac{y_1/n_1}{y_2/n_2}$ follows f distribution with n_1-1 and n_2-1 degrees of freedom (relationship b/w χ^2 and F).

Sampling Distribution of Sample Mean & Sample Variance

Let $x_1, x_2, \ldots, x_n$ be an independent sample where each of $x_1, x_2, \ldots, x_n$ follows Normal distribution with mean μ & SD $= \sigma$. Then, sample mean

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$ follows Normal distribution.

To find mean of the distribution of $\bar{x}$

$$\text{Mean of } \bar{x} = E(\bar{x}) = E\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right)$$

$$= \frac{1}{n} \times E(x_1 + x_2 + \cdots + x_n) \quad \left[\because E(ax) = aE(x)\right]$$

$$= \frac{1}{n} \times \left[E(x_1) + E(x_2) + \cdots + E(x_n)\right]$$

$$\left[\because E(x+y) = E(x) + E(y)\right]$$

$$= \frac{1}{n} \times \left[\underbrace{\mu + \mu + \cdots + \mu}_{(n \text{ times})}\right]$$

$$= \frac{1}{n} \times n\mu = \underline{\underline{\mu}}$$

To find variance of distribution of $\bar{x}$

$$\text{Variance of } \bar{x} = Var(\bar{x})$$

$$= Var\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right)$$

$$= \frac{1}{n^2} \times Var(x_1 + x_2 + \cdots + x_n) \quad [\because Var(ax) = a^2 Var(x)]$$

$$= \frac{1}{n^2} \times [Var(x_1) + Var(x_2) + \cdots + Var(x_n)]$$

$$[\because Var(x+y) = Var(x) + Var(y)]$$

$$= \frac{1}{n^2} \times [\underbrace{\sigma^2 + \sigma^2 + \cdots + \sigma^2}_{(n \text{ times})}] \quad \left[\begin{array}{l}\because SD \text{ of } x_i = \sigma \\ \Rightarrow Var \text{ of } x_i = \sigma^2, \\ \text{where } i = 1, 2, \ldots, n\end{array}\right]$$

$$= \frac{1}{n^2} \times n\sigma^2 = \frac{n\sigma^2}{n^2}$$

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

$$\therefore SD \text{ of } \bar{x} = \sqrt{\frac{\sigma^2}{n}}$$

$$= \frac{\sigma}{\sqrt{n}}$$

$\therefore$ $\bar{x}$ follows Normal distribution with Mean $= \mu$ &

$$SD = \frac{\sigma}{\sqrt{n}}$$

& probability density function is,

$$f(x) = \frac{1}{\sqrt{2\pi} \times \sigma} \times e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}$$

$$= \frac{1}{\sqrt{2\pi} \times \frac{\sigma}{\sqrt{n}}} \times e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma/\sqrt{n}}\right]^2}$$

$$= \frac{\sqrt{n}}{\sqrt{2\pi} \times \sigma} \times e^{-\frac{1}{2}\left[\frac{(x-\mu)^2}{(\sigma/\sqrt{n})^2}\right]}$$

$$= \frac{\sqrt{n}}{\sqrt{2\pi} \times \sigma} \times e^{-\frac{1}{2}\left[\frac{(x-\mu)^2}{\sigma^2/n}\right]}$$

$$= \frac{\sqrt{n}}{\sqrt{2\pi} \times \sigma} \times e^{-\frac{1}{2}\left[\frac{n(x-\mu)^2}{\sigma^2}\right]}$$