Chapter-1

Sampling Distribution

1a What is a sampling distribution, and why is it important in statistics.

Ans: A Sampling distribution is a probability distribution of a Statistics obtained through a large number of samples drawn from a specific Population.

Sample distributions are important in statistics because they provide a major simplification enroute to statistical inference.

More specifically, they allow analytical considerations to be based on the probability distribution of a statistic, rather than on the joint probability distribution of all the individual sample values.

For example:- χ^2, F and t distribution are sampling distribution.

* The distribution of sample statistics called sampling distribution.

1.b: Difference between Sampling distribution and Parent distribution.

Ans:

Parent Distribution (Population)	Sampling Distribution
1. The distribution of the entire Population.	1. The distribution of a sample Statistics. (Sample mean).
2. The actual data of the Population.	2. Repeated samples taken from the Population.
3. Parameters are mean(μ), Standard deviation(σ).	3. Mean of Sample means (μ) Standard error($\sigma_{\bar{x}}$)
4. Can be any distribution. Shape.	4 Approaches normality as n increase.
5. Size fixed (entire Population)	5. Changes based on Sample size n.
6. Example: Heights of all adults in a city.	6. Example: The distribution of sample mean from repeated height Samples (size 30)

1.c: Define the Chi-square distribution and explain how it is derived from standard normal variables.

Ans: The sum of squares of n independent standard normal variables is called chi-squares (χ^2) ~~variable~~ distribution with n degrees of freedom.

Derived from standard normal variables:-

Let $z_1, z_2 \cdots z_n$ be n independent standard normal variables then chi-square denoted by χ^2, is defined as

$$\chi_n^2 = \sum_{i=1}^{n} z_i^2$$

$$= z_1^2 + z_2^2 + z_3^2 + \cdots z_n^2$$

However, if $x_1, x_2, \cdots x_n$ are n independently and identically distributed random variables each of which is normally distributed with mean μ and variance σ^2, then

$$\chi_n^2 = \sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma}\right)^2$$

is a chi-square (χ^2) distribution with n degree of freedom.

1.d: What are the Properties of the Chi-square distribution? List at least five.

Ans: Properties of χ^2 distribution:-

(I) χ^2 is a continuous type of distribution and its range is 0 to ∞. i.e $0 < \chi^2 < \infty$.

(II) This distribution contains only one parameter which is the degree of freedom of the distribution.

(III) The mean and variance of χ^2 distribution for n d.f. is n and 2n respectively.

(IV) The mode of χ^2 distribution for n d.f is (n-2).

(V) The moment generating function of χ^2 distribution for n d.f. is $\frac{1}{(1-2t)^{n/2}}$ or $(1-2t)^{-n/2}$.

(VI) χ^2 distribution tends to normal distribution for large degree of freedom.

(VII) It is positively skwed distribution for smaller values of n.

(VIII) The distribution becomes symmetrical as n tends to infinity ($n \to \infty$)

* Application of Chi-square distribution

(i) To test if the hypothetical value of the population variance is $\sigma^2 = \sigma_0^2$.

(ii) To test the goodness of fit.

(iii) To test the independence of attributes.

(iv) To test the homogenity of independent estimates of the population variance.

(v) To test the homogenity of independent estimates of the population correlation cofficent.

1.e: How does the chi-square distribution behave as the degree od freedom increase?

Ans: As the degrees of freedom(k) increase, the chi-square distribution:

1. Becomes more Symmetric → Initially right skewed, but approches a normal distribution for large k.

2. Mean Increases → The mean is k, So it Shifts right as k grows.

3. Variance Increases → The variance is 2k, making the distribution spread out more.

4. Peak shift Right → The mode moves from near zero to k-2.

5. Approximates normal Distribution → For k>30, it behaves like N(k,2k).

1.f: How is the chi-square distribution used to test the goodness of fit?

Ans: The chi-square goodness of-fit-test checks if observed data matches an expected distribution.

Steps:-

1. Set hypothesis:-

H_0: observed data follows the expected distribution.

H_1: Observed data does not follow the expected distribution.

2. Calculate Expected Frequencies:-

$E_i = N \times P_i$ (where, N = total observations, P_i = expected proportion)

3. Compute Chi-square Statistics:-

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

4. Compare with critical value/P-value:-

Use chi-square table with $k-1$ degree of freedom.

if P-value $< \alpha$, Reject H_0

1.9 Derive the moment generating function (MGF) of the Chi-square distribution.

Ans:

Derivation:- Let $x_1, x_2 \cdots x_n$ be n independent random variable from $N(\mu, \sigma^2)$ i.e $x_i \sim N(\mu, \sigma^2)$; $i = 1, 2, 3 \cdots n$, x_i are independent.

Now, we want to find the distribution of

$$\chi_n^2 = \Sigma y_i = \Sigma \left(\frac{x_i - \mu}{\sigma}\right)^2$$ by mgf technique.

Hence, the mgf of χ^2 is given by

$$M_{\chi^2}(t) = M_{\Sigma y_i}(t) = \prod_{i=1}^{n} M_{y_i}(t) \quad [y_i \text{ are independent}]$$

$$M_{\chi^2}(t) = \prod_{i=1}^{n} \left[M_{\left(\frac{x_i - \mu}{\sigma}\right)^2}(t) \right]$$

$$= \prod_{i=1}^{n} E\left[e^{t\left(\frac{x_i - \mu}{\sigma}\right)^2} \right]$$

$$= \prod_{i=1}^{n} E[e^{tu^2}] \quad \left(u = \frac{x_i - \mu}{\sigma}\right)$$

$$= \prod_{i=1}^{n} \left\{ \int_{-\infty}^{\infty} e^{tu^2} f(u) \cdot du \right\}$$

$$= \prod_{i=1}^{n} \left\{ \int_{-\infty}^{\infty} e^{tu^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du \right\}$$

$$= \prod_{i=1}^{n} \left\{ \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} e^{tu^2 - \frac{1}{2}u^2} du \right\}$$

[Since the integrand is an even function of u]

$$= \prod_{i=1}^{n} \left\{ \frac{\sqrt{2}}{\sqrt{\pi}} \int_0^\infty e^{-(1/2-t)u^2} du \right\}$$

Let $z = (1/2 - t)u^2$ $\quad dz = (\frac{1}{2}-t)u^2 = (\frac{1}{2}-t).2u.du$

$$\Rightarrow u^2 = \frac{z}{\frac{1-2t}{2}} \qquad \Rightarrow 2u.du = \frac{dz}{\left(\frac{1-2t}{2}\right)}$$

$$\Rightarrow u = \sqrt{\frac{2z}{1-2t}} \qquad \Rightarrow du = \frac{dz}{\cancel{2}u\frac{(1-2t)}{\cancel{2}}}$$

$$\therefore du = \frac{dz}{(1-2t)u}$$

$$\Rightarrow du = \frac{dz}{(1-2t)\sqrt{\frac{2z}{(1-2t)}}} = \frac{dz}{\sqrt{1-2t}\sqrt{2z}}$$

$$\therefore M_{x^2}(t) = \prod_{i=1}^{n} \left\{ \frac{\sqrt{2}}{\sqrt{\pi}} \int_0^\infty e^{-z} \cdot \frac{dz}{\sqrt{1-2t}\sqrt{2z}} \right\}$$

$$= \prod_{i=1}^{n} \left\{ \frac{\cancel{\sqrt{2}}}{\sqrt{\pi}\sqrt{1-2t}.\cancel{\sqrt{2}}} \int_0^\infty e^{-z} . z^{-1/2} dz \right\}$$

$$= \prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{\pi}\sqrt{1-2t}} \int_0^\infty e^{-z} z^{1/2-1} dz \right.$$

$$= \prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{\pi}\sqrt{1-2t}} \sqrt{1/2} \right\} \quad \left[\because \sqrt{n} = \int_0^\infty e^{-x} x^{n-1} dx \right]$$

$$= \prod_{i=1}^{n} \left\{ \frac{1}{\cancel{\sqrt{\pi}}\sqrt{1-2t}} . \cancel{\sqrt{\pi}} \right\}$$

$$= \prod_{i=1}^{n} \left\{(1-2t)^{-1/2}\right\}$$

$$= (1-2t)^{-n/2}$$

$$\therefore M_{x^2}(t) = (1-2t)^{-n/2}$$

which is the mgf of gamma distribution with shape parameter $\alpha = n/2$ and Scale parameter $\beta = 2$

Therefore, the pdf x^2 distribution is as

$$f(x^2) = \frac{1}{2^{n/2}\,\Gamma(n/2)} (x^2)^{\frac{n}{2}-1} \cdot e^{-x^2/2}, \quad x^2 > 0, \ (0 < x^2 < \infty)$$

This is the pdf of x^2-variabe with 'n' degree of freedom.

1.h: Using the MGF, Find the mean and variance of the chi-square distribution.

Ans: The moment Generating function (MGF) of chi-square distribution random variable $X \sim \chi^2_n$

$$\therefore M_X(t) = (1-2t)^{-n/2} \quad \text{for } (t < \tfrac{1}{2})$$

Compute mean

The mean (first moment) is given by

$$E(X) = M'_X(0) = \frac{d}{dt} M_X(t) \Big|_{t=0}$$

Differentiate $M_x(t)$

$$M_x(t) = (1-2t)^{-n/2}$$

$$\Rightarrow M'_{(x)} t = -\frac{n}{2}(1-2t)^{-n/2-1} \cdot (-2)$$

$$= \frac{n}{(1-2t)^{n/2+1}}$$

$\boxed{\cancel{\frac{n}{2}} \text{ neglect} \quad \cancel{\frac{n}{2}=0}}$

∴ Now simplify $M'_{(x)} t = \dfrac{n}{(1-2t)^{n/2+1}}$

Evaluating $t=0$.

$$E(x] = M'_x(0) = \frac{n}{(1-0)^{n/2+1}} = n$$

$$\therefore E[x] = n$$

* Compute the variance $Var(x)$

The variance given by:-

$$Var[x] = E[x^2] - (E[x])^2$$

• To find $E[x^2]$, we compute the second derivative of $M_x(t)$.

Differentiate $M_x'(t)$ again

we have $M_x'(t) = \frac{k}{1-2t}$

Differentiate again using the quotient rule

$$M_x''(t) = \frac{k \cdot 2}{(1-2t)^2} = \frac{2k}{(1-2t)^2}$$

Evaluating at $t=0$

$$E[X^2] = M_x''(0) = \frac{2k}{1^2} = 2k$$

∴ Now, we compute the variance:

$$Var(x) = E[x^2] - (E[X])^2$$

$$Var(x) = 2k - k^2$$

∴ Thus the variance of chi-square distribution is $var(X) = 2k$

Now differentiate again

$$M'_X(t) = n(1-2t)^{-n/2-1}$$

Applying chain rule again,

$$M''_X(t) = \cancel{k}.n.(-n/2-1)(1-2t)^{-n/2-2}.(-2)$$

$$= n(n/2+1)(1-2t)^{-n/2-2}.2$$

$$= n(n/2+1).2.(1-2t)^{-n/2-2}$$

∴ Evaluating $t=0$

$$E[x^2] = M''_X(0) = n(n/2+1).2(1-0)^{-n/2-2}$$

$$= n(n/2+1).2$$

$$= \frac{n^2+2n}{2}.2$$

$$= n^2+2n$$

$$\therefore Var[X] = E(x^2) - (E[X])^2$$

$$= n^2+2n-n^2$$

$$= 2n$$

$$\therefore Var[X] = 2n$$