Statistics and Their Distributions :

Population : Total set of observations or collection of some measurements is known as population.

Sample : A subset of population is known as sample. and the process of selecting sample from population is known as sampling.

Random Sample : Let $X_1, X_2 \cdots X_n$ be n i.i.d (identically independent distributed) random variables each having the same probability distribution then we say that $(X_1, X_2 \cdots, X_n)$ is a random sample from the population ~~$f(x,\theta)$~~ with probability distribution function $f(x,\theta)$.

The joint dist function of $(X_1, X_2 \cdots X_n)$ is given as

$$f(x_1, x_2 \ldots x_n) = f(x_1;\theta) \cdot f(x_2;\theta) \cdots f(x_n;\theta) \quad \because X_i \text{ are independent}$$

$$= \prod_{i=1}^{n} f(x_i;\theta)$$

Statistic : Any function of random sample which does not contain any unknown parameter is called statistic.

$$T = T(X_1, X_2 \cdots X_n).$$

Sample Mean :- Let $X_1, X_2 \ldots X_n$ be the sample of size n from the population $f(x, \theta)$ then the sample mean is defined as

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Note :- Sample mean $\bar{x}$ is statistic.

Sample Variance : Sample variance is denoted as S^2 and defined as

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

Note :- Here in sample variance, we have divided it by $(n-1)$ in place of n because if we divide it by n then sample variance will not be unbiased for population variance and we will loose very important characteristics in statistics. Since it is little bit deep concept so just skip these things and remember the formula of sample variance.

V. Imp

Q: Let $X_1, X_2 \ldots X_n$ are i.i.d and follow the $N(\mu, \sigma^2)$ where μ and σ^2 are charecteristic of population then find the sampling distribution of $\overline{X}$.

Sol^n

$\because \bar{X} = \frac{1}{n}(X_1 + X_2 \cdots + X_n)$

$$\boxed{\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i}$$

Note that each X_i are Independent and have the same distribution (iid) then $\mathrm{cov}(X_i, X_j) = 0 \quad \forall\, i \neq j$

Note :- We know that if X_i ; $i = 1, 2 \cdots n$ follows the normal distribution then any linear combinat of these X_i also follows the normal distribution. (We prove it later)

Now, We have to find the distribution of $\bar{X}$.

$$\bar{X} = \frac{1}{n}(X_1 + X_2 \cdots + X_n)$$

$$E(\bar{X}) = E\left(\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right)$$

$$= \frac{1}{n} E(X_1 + X_2 + \cdots + X_n)$$

$$= \frac{1}{n}\left(E(X_1) + E(X_2) + \cdots + E(X_n)\right)$$

$$= \frac{1}{n}(\mu + \mu + \cdots + \mu) \qquad \because X_i \sim N(\mu, \sigma^2)$$

$$E(X_i) = \mu$$

$$V(X_i) = \sigma^2$$

$$= \frac{1}{\cancel{n}} \cancel{n}\mu$$

$$= \mu$$

④

Now variance of $\bar{X}$.

Note: $V(aX+bY) = a^2 V(X) + b^2 V(Y) + 2ab\, cov(X,Y)$

$$V(\bar{X}) = V\left(\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right)$$

$$= \frac{1}{n^2} V(X_1 + X_2 + \cdots + X_n) \quad (\because V(aX) = a^2 V(X))$$

$$= \frac{1}{n^2}\left(V(X_1) + V(X_2) + \cdots + V(X_n) + 0 + 0 \cdots\right)$$

(Since all X_i are indep. so covariance will be zero)

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

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

$$V(\bar{X}) = \frac{\sigma^2}{n}$$

Thus $\bar{X} \sim N\left(\mu, \sigma^2/n\right)$

$Z \sim N\left((a_1 + a_2 + \cdots a_n)\cdot\mu, (a_1^2 + a_2^2 + \cdots a_n^2)\sigma^2\right)$

Result: If $X_i \sim N(\mu, \sigma^2)$ and X_i are iid then any linear combination (l.c) also follow the same dist.

Proof: Consider $Z = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n$ (l.c)

$$E(Z) = E(a_1 X_1 + a_2 X_2 + \cdots + a_n X_n)$$

$$= a_1 E(X_1) + a_2 E(X_2) \cdots + a_n E(X_n)$$

$$= a_1 \mu + a_2 \mu + \cdots + a_n \mu = (a_1 + a_2 \cdots + a_n)\mu$$

$$V(Z) = V(a_1 X_1 + a_2 X_2 + \cdots + a_n X_n)$$

$$= a_1^2 V(X_1) + \cdots + a_n^2 V(X_n)$$

$$= (a_1^2 + a_2^2 \cdots + a_n^2)\sigma^2$$

Thus $Z \sim N\left((a_1 + a_2 \cdots + a_n)\mu, (a_1^2 + a_2^2 + \cdots + a_n^2)\sigma^2\right)$