DS-221. LECTURE NO. 30 SAMPLE MEAN & POP MEAN.

Population.

This is the entire possible dataset $\{x\}$

It has a countable size say N_p

It also has std. deviation and mean.

Sample:-

- It is a random subset of the population, where sampling is done with replacement.
- Its size is smaller than population size N_p.
- The sample mean of a population is X^n, and considered a random variable.

Sample Mean:-

Sample mean of population is just like the normal sample mean of random variables if.

(i) Samples are independant & identically dist.

(ii) Randomly and independanty drawn, with replacement. samples. —— (b).

$$\therefore \quad X^n = \frac{1}{N}(x_1 + x_2 + x_3 \cdots)$$

due to (b) we can also deduce expected value of sample mean as.

$$E(X)^n = \frac{1}{N}\left(E[X^{(1)}] + E[X^{(2)}] \cdots\cdots\right)$$

$E(X^n)$

(b) if our population follow some specific dist. like normal or uniform dist then every sample taken from this population will also have the same dist.

Consider the following example.

Population = $X = \{1,2,3,4,5,6,7,8,9,10,11,12,13\}$.

Draw 1st Sample

$\{x_1\} = \{4,7,10\}$

mean $\{x_1\} = 7$

Draw 2nd Sample

$\{x_2\} = \{2,6,9\}$.

mean$\{x_2\} = 5.\cancel{5}67$.

Then the Expected value of the samples we have drawn is.

$$E(x^{(n)}) = \frac{1}{N}\left[E(X^1) + E(X^2) + \cdots\right]$$

$$\frac{1}{2}(7+5.67) \boxed{\Rightarrow 6.33} \text{ Ans!}$$

Since each sample was drawn uniformly we can say that:

$$E(X^N) = \text{pop mean}\{x\}.$$

we can also say that

$(X)^N$ is an unbiased estimator of population mean.

Sample mean that is calculated after making a sample from the population, also have a distribution, it is known as the "sampling dist of the mean" which describes the behaviour of sample mean across many repeated samples.

Now the Central limit theorem states that if you make many samples of size n the mean of those samples $(\bar{X})$ will be:

∵ for smaller n it might resemble the pop. dist.

Date

Therefore summarizing:-

(i) Sample mean is a random variable because d/f samples from the population will have d/f values of the sample mean.

(ii) Expected value of the sample mean random variable => population mean.

(i) approx normally dist (regardless of the original population dist) if n is large enough.

(ii) have a mean equal to the population mean (μ).

(iii) have a variance equal to σ^2/n where σ^2 is the variance of the population.

Now further analyzing other statistical properties of our sample mean.

(i) Variance of the sample mean.

$$var\{(X^n)\} = \frac{pop\ var\{X\}}{N}$$

→ variance of population.
→ (no. of datapoints).

(ii) Std deviation of sample mean

$$std\{X^{(N)}\} = \frac{pop\ std\ dv^e\{X\}}{\sqrt{n}}$$

Important Note:

(1) the more datapoints we draw, the less the std(X^n) becomes as our estimate of pop mean get better.

this improvement is slow e.g. to half std deviation of X^n we need four times as many samples.

Date

Now previously we saw that std.dv $\{X^{(n)}\}$ require the popolation variance i.e popvar$\{x\}$ or popstd$\{x\}$.

The unbiased estimator of popstd$\{x\}$ is. apply on individual sample you generate not together.

$$\text{stdunbiased}(\{x\}) = \sqrt{\frac{1}{N-1} \sum_{x_i \in \text{sample}} (x_i - \text{mean}\{x_i\})^2}$$

to use 1 less degree of freedom because 1 is already used in estimating sample mean.

∴ we can say that:

$$\text{std}[X^{(n)}] = \frac{\text{popstd}[X]}{\sqrt{N}} \Rightarrow \frac{\text{stdunbiased}(\{x\})}{\sqrt{N}}$$

still the estimate of sample mean's std deviation. ⇒ std eer also known as (standard error).

smaller 'std error' means a good/ more accurate value.

this std error will give insight about the accuracy of sample as an estimate of the population mean.

Previously we discussed how increasing sample size would decrease std. deviation same concept apply to the std. error being reduced as the sample size increase.

$$\text{std error} \propto \frac{1}{N}$$