PL 07 - SAMPLING & SAMPLING DISTRIBUTION

Video 1 : Inferential Statistics : POINT ESTIMATION

- From the samples, we must make Inferences about the entire population.

- The inferences made using samples are by definition incomplete, therefore the sample characteristics will always have some error built-in.

- E.g. Highway Paving Samples

Specimen	Viscosity
1	3193
2	3124
3	3153
4	3154
5	3093
6	2466
7	3355
8	2979
9	3182
10	3227
11	3256
12	3332
13	3204
14	3282
15	3170

Sample # 1

- Sample Mean: $\bar{x} = 3210.73$
- Sample S.D : $s = 117.61$

- Population Mean $= \mu$
- Population SD $= \sigma$
- Population Proportion $= p$

- Sample Mean $= \bar{x}$
- Sample SD $= s$
- Sample proportion $= \bar{p}$

- Point Estimates are NEVER Perfect.
- There is always an ERROR Component.
- Error component is expressed as a CONFIDENCE INTERVAL.

To conclude :

POINT ESTIMATION is taking samples of larger population and using those as an ESTIMATION of the overall pop^n.

In doing this, there is always an ERROR involved.

Video 2 : Inferential Statistics : SAMPLING DISTRIBUTIONS

- Highway Paving Inc needs Asphalt at a viscosity of 3200.
- The QC specialists takes 15 samples/specimens of the material and tests viscosity to ensure the batch has uniform viscosity.

Note : (1) No way to test every ounce of asphalt (population)
(2) ∴ company must take samples.
(3) From those sample, HWP must make inferences about entire batch.
(4) The Inferences made using samples, will always have some ERROR.

- Sample #1 : $\bar{x}_1 = 3210.73$, $s_1 = 117.61$ (previous page)

This one Sample, may not correctly reflect the actual mean. So, we will take few more samples & repeat the process.

- sample #2 : $\bar{x}_2 = 3150.13$, $s_2 =$

⋮

- Sample #9 : $\bar{x}_9 = 3023.59$, $s_9 =$

- Now, create & Analyze the distribution of the Sample mean.

Sample #	Sample Mean ($\bar{x}$)
1	3210.73
2	3150.13
3	3345.5
4	⋮
5	⋮
6	⋮
7	⋮
8	3413.01
9	3023.59

$\longrightarrow E(\bar{x}) = 3217.08$

Range	Frequency
2950 - 3049	I
3050 - 3149	I
3150 - 3249	IIII
3250 - 3349	II
3350 - 3449	I

↓
here, we have created a Histogram of sample mean.

This is called SAMPLING DISTRIBUTION :

↳ It is a distribution of Sample Mean themselves.

Now, EXPECTED Value of Sampling Distribution = $E(\bar{x})$ should be equal to Population mean = μ

$\boxed{\text{To conclude:}}$

If we take many Random Samples from the population each with its own sample mean and then create a distribution based on all those samples means, the mean of that sampling distribution is equal to the mean of the Population.

•• -- The expected value of the sampling distribution of $\bar{x}$ is at best going to be an estimate of μ.

•• -- We would have to take every conceivable sample from the population to match the population mean perfectly.

•• -- The best we are going to be able to do is FIND AN INTERVAL estimate for the population mean μ

•• -- Our interval estimate will be influenced by sample size and degree of "confidence" we are satisfied with.

Video3 : Inferential Statistics : Standard Error of the Mean

- HWP Inc, expects the asphalt viscosity as 3200.

- Expected value of Sampling Distribution, = 3217.50 (previously derived).

- $\boxed{E(\bar{x}_n) = \mu \quad \text{as } n \to \text{"large"}}$ | Here, $E(\bar{x})$ = the expected value of $\bar{x}$, μ = the population mean

- STANDARD DEVIATION of $\bar{x}$ SAMPLING DISTRIBUTION

Standard error of the mean is the standard deviation of the sampling distribution.

Standard error of the Mean ; $\boxed{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$

$\sigma_{\bar{x}}$ = standard deviation of $\bar{x}$
σ = standard deviation of population
n = sample size

Standard Deviation of Samp Dist"

- As 'n' Becomes larger, Standard Error of mean **Decreases**.

Sample Size (n) ↑, Standard Error ↓

∴ Larger sample size, up to a point, generates a better approximation

∴ $\sigma_{\bar{x}}$ is same for all the samples (even with different mean) of same sample size.

Video 4 : Inferential Statistics : Sample Mean Proximity to Population mean (μ)

Q.) How close is the sample mean $\bar{x}$ to the population mean μ?

→ Standard error of the mean - is a Standard Deviation (to be precise, the Standard deviation of the Sampling Dis)

→ If standard deviation of population is known, we use Z-distribution else t-distribution.

→ The standard error of the mean is inversely related to Sample Size.

eg Standard Error & Sample Size

n	$\sigma_{\bar{x}}$	Z-Score -3	-2	-1	0	1	2	3
15	38.7	- -	- -	3161.3	3200	3238.7	- -	- -
135	12.9	- -	- -	3187.1	3200	3212.9	- -	- -
500	6.71	- -	- -	3193.3	3200	3206.7	- -	- -

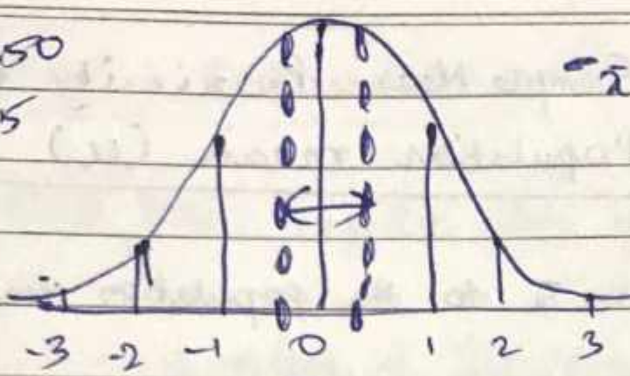
Q) For each of our sample size, what is the probability that the sample mean is within 15 of the population mean $\mu = 3200$?

↓

$P(3185 \leq \bar{x} \leq 3215)$

$P(x) = 3185 \le \bar{x} \le 3215$

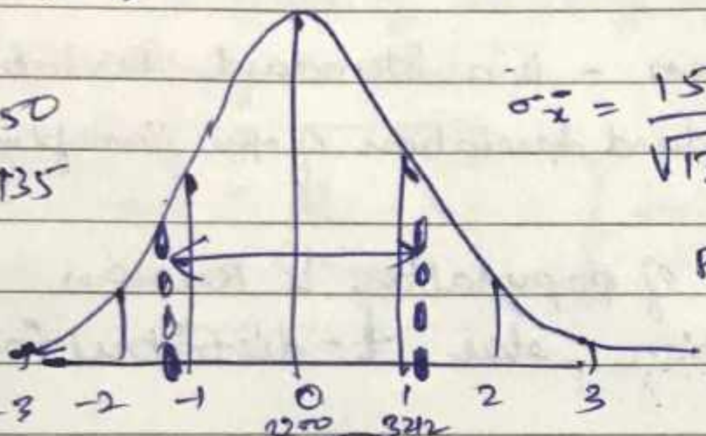
$\sigma = 150$
$n = 15$



$$\sigma_{\bar{x}} = \frac{150}{\sqrt{15}} = 38.7$$

$P(x) = 0.302$

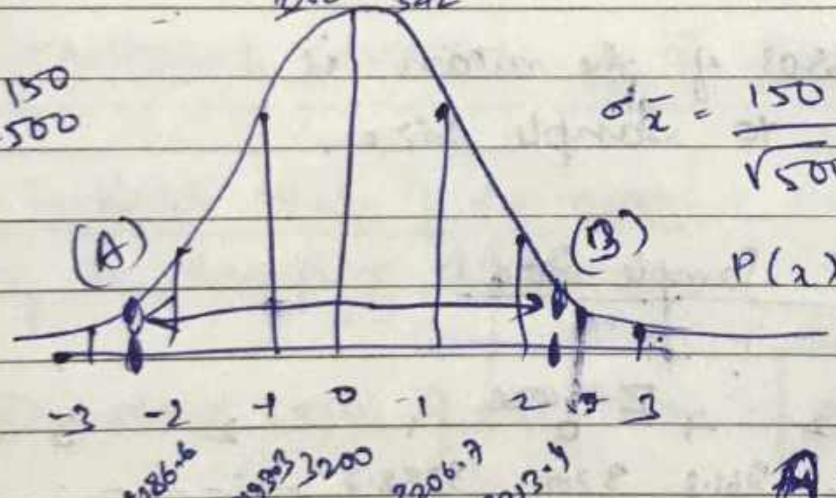
$\sigma = 150$
$n = 135$



$$\sigma_{\bar{x}} = \frac{150}{\sqrt{135}} = 12.9$$

$P(x) = .755$

$\sigma = 150$
$n = 500$



$$\sigma_{\bar{x}} = \frac{150}{\sqrt{500}} = 6.71$$

$P(x) = 0.975$

$$A = \frac{3185 - 3200}{6.7} = -2.239$$

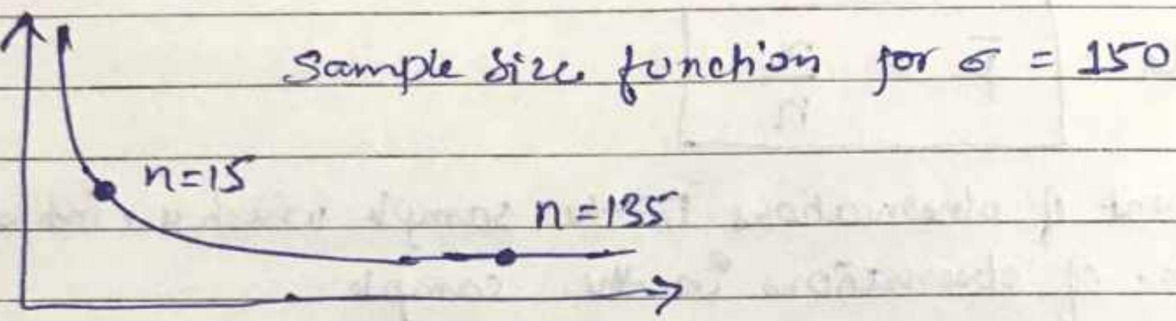
$$B = \frac{3215 - 3200}{6.7} = +2.239$$

P_1 = NORM.DIST(-2.239, 0, 1, True)
P_2 = NORM.DIST(+2.239, 0, 1, True)

Req Prop $= P_2 - P_1$
$= 0.987422 - 0.012578$
$= 0.974844$
$= .975$

- SAMPLE SIZE DIMINISHING RETURNS

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \Rightarrow \boxed{f(x) = \frac{150}{\sqrt{x}}}$$



This function represents Standard Error of Mean.

Video 5: Inferential Statistics: Sample Proportions, $\bar{p}$

- Sample mean = $\bar{x}$
  Sample proportion = $\bar{p}$

- Formula : $\boxed{\bar{p} = \frac{x}{n}}$

$\begin{cases} x = \text{count of observations in the sample which u r interested in} \\ n = \text{no. of observations in the sample} \end{cases}$

- Expected value of $\bar{p}$

$E(\bar{p}) = p$ (proportion of population)

The expected value of the sample proportion, $\bar{p}$, the mean of all potential values of $\bar{p}$, is equal to the population porportion, p.

- Standard error of $\bar{p}$

-- Standard deviation of the sample proportion is called standard error of the sample proportion.

-- Formula: FINITE Population | Infinite Population

FINITE Population	Infinite Population
$\sigma_{\bar{p}} = \sqrt{\frac{N-n}{N-1}}\sqrt{\frac{p(1-p)}{n}}$	$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}$

↓ population correction factor

-- If $\frac{n}{N} \leq 0.05$, we use Infinite population,

If $\frac{n}{N} > 0.05$, we use the correction.

-- FORMULA

$$\boxed{Z = \frac{\bar{p} - E(\bar{p})}{\sigma_{\bar{p}}} = \frac{\bar{p} - p}{\sqrt{\frac{p(1-p)}{n}}}}$$

Similar like, sample mean

$$\boxed{Z = \frac{\bar{x} - E(\bar{x})}{\sigma_{\bar{x}}}}$$

- (i) $np \geq 5$ (ii) $n(1-p) \geq 5$

If both the conditions true, we can use the normal distribution.

-- Similarly sample mean, we can find Probability of sample Proportions.

• To conclude

(i) Find sample proportion $\bar{p} = \frac{x}{n}$

(ii) Find standard error of proportion using sample proportion & the sample size

↳ The standard error of proportion is standard deviation of the sampling distribution.

(iii) Next, make sure that we could use the normal approximation for sampling distribut.

↳ using the two tests mentioned.

(iv) If yes, use the standard error to find z - scores.

$$z = \frac{p - \bar{p}}{\sigma_{\bar{p}}}$$

(v) Use z - score cumulative probabilities to find probability intervals.

P(z) = NORM.DIST (z, 0, 1, TRUE)

Video 6 : Inferential Statistics : Sampling Distrobution for proportions.

- Survey: 25 students are tasked with contacting 30 students to ask each of them demographic information and survey question about cafe.

  In the end, Researchess will end up with 25 samples each containing the yes/no responses for 30 students. The no. of 'yes' response out of 30 is the quantity of interest.

- Sample #1 : ~~Mean of~~

  Mean of sample #1 to sample #25 = $\bar{x} = 0.612$
  standard deviation (error) = $\bar{s} = 0.0922$

- Using empirical rule, $\pm 1\sigma = 68.28\%$
  $\pm 2\sigma = 95.44\%$
  $+ 3\sigma = 99.74\%$

- Effect of sample size on S.E (standard error)

$$\sigma_{\bar{p}} = \sqrt{\frac{P(1-P)}{n}}$$

[n - Sample size should not be increaced infinitely to minimize $\sigma_{\bar{p}}$.]