Sampling Distribution of the Mean

1) Mean of the sampling Distribution of the Mean

The Mean of the sampling distribution of the mean equals the population mean (μ).

Formula:

$$\mu_M = \mu$$

Example: Suppose a population has a mean height of 165 cm. If we take repeated random samples from this population the mean of all sample means will also be 165 cm

2) Variance of the Sampling Distribution of the mean

- The variance of the sampling distribution of the mean is the population variance (σ^2) divided by the sample size (N)

Formula $$\sigma^2_M = \frac{\sigma^2}{N}$$

Implication

- larger sample size → Smaller variance of the sample mean distribution

Example

- If the population variance $\sigma^2 = 64$ and we take a sample size of $N = 4$ then $\sigma^2_M = 64/4 = 16$

3) Standard Error of the Mean (SEM).

The standard Error of the Mean is the standard deviation of the sampling distribution of the mean

Formula $$\sigma_M = \frac{\sigma}{\sqrt{N}}$$

Example: if $\sigma = 8$ and $N = 4$ then;

$$\sigma_M = 8/\sqrt{4} = 4$$

Interpretation

Smaller SEM → Sample means are more tightly clustered around the population mean

4) Central limit Theorem (CLT)

Given any population with mean μ and finite variance σ^2, the sampling distribution of the mean approaches a normal distribution as N increases, regardless the shape of the original population.

Date / /
Page No.

Key takeaways
For small N, the sampling distribution may not be Normal
for large N (typically $N \geq 30$), the distribution becomes approximately normal.

5) Practical Importance of CLT
- Why is CLT Important?
  - Allows statistical inference (eg confidence intervals, hypothesis testing) even when the population is not normal
  - Justifies using z-scores and t-test when sample sizes are sufficiently large

Real World application
- Suppose we want to estimate the average income of residents in a city.
- The income distribution may be right-skewed (some people earn much more than most)
- With large enough samples, the sample mean distribution will be normal, making it easier to analyze.

Sampling Distribution of the Difference Between means

1) Mean of the Sampling Distribution of the Difference between means
The mean of the sampling distribution of the difference between ~~mean~~ two sample means is equal to the difference between the two population means
formula

$$\mu_{M_1 - M_2} = \mu_1 - \mu_2$$

Example : If the mean test score of 12 year old is 34 and that of 10 year old is 25 then

$$\mu_{M_1 - M_2} = 34 - 25 = 9$$

2) Variance of the sampling distribution of the difference between means

The variance of the sampling distribution is the sum of the variances of the sampling distribution of the individual means

Formula

$$\sigma^2_{M_1 - M_2} = \frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2}$$

3) Standard Error of the Difference between means

The standard error is the square root of the variance

$$\sigma_{M_1 - M_2} = \sqrt{\frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2}}$$

Example : Two Mattain species

- species 1 Mean Height = 32, variance = 60, Sample size: 10
- species 2 : Mean height : 22, variance = 70, Sample size: 14

- Mean of the Distribution :

$$\mu_{M_1 - M_2} = 32 - 22 = 10$$

- standard error

$$\sigma_{M_1 - M_2} = \sqrt{\frac{60}{10} + \frac{70}{14}} = 3.317$$

4) Computing probabilities using the sampling distribution

To compute probabilities, we find the z-score for a given difference between means $z = \frac{(M_1 - M_2) - \mu_{M_1 - M_2}}{\sigma_{M_1 - M_2}}$

Date / /
Page No.

Example (Martian Species)
Probability that Species 1's mean Height exceeds species 2's by at leas 5

$$z = \frac{5 - 10}{3.317} = -1.51$$

Using a z-table, the probability is 0.934

5) Simplified formula for equal sample sizes and variances
If $n_1 = n_2$ and $\sigma_1^2 = \sigma_2^2$ then:

$$\sigma_{M_1 - M_2} = \sqrt{\frac{2\sigma^2}{n}}$$

Example (Boys vs Girls height study):
Boys: Mean: 175, variance: 64, sample size = 8
Girls: Mean 165 variance: 64, sample size 8
Mean of the difference

$$\mu_{M_1 - M_2} = 165 - 175 = -10$$

Standard Error $\sigma_{M_1 - M_2} = \sqrt{\frac{2(64)}{8}} = 4$

Probability that girls mean > boys mean

$$z = \frac{0 - (-10)}{4} = 2.5$$

from a z table the probability is 0.0062 (very unlikely)