UNIT IV

Sampling Distribution

By

Dr. Amrita Jha

Department of Mathematics

ISHLS, Indus University

Ahmedabad

Sampling Distribution

Topics to be covered

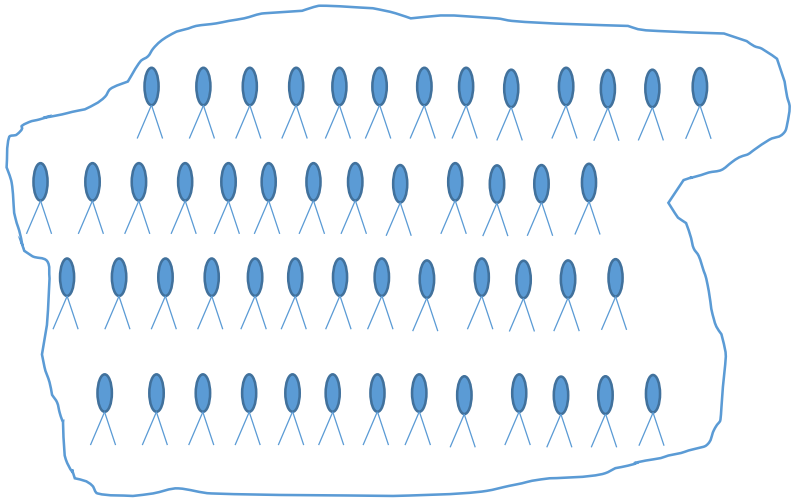
- Introduction
- Population and Sample
- Sampling distribution
- Point Estimation
- Sampling Distribution of Mean
- Sampling Distribution for the Difference between Two Means
- Sampling Distribution of Proportion
- Sampling Distribution for the Difference between Two Proportions

Sampling Distribution :Introduction

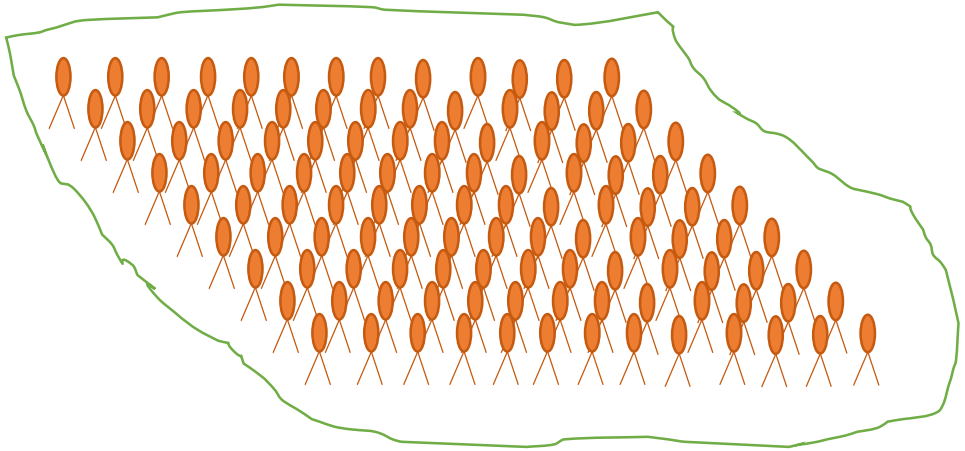
- **Sampling distribution** of a statistic is the theoretical probability distribution of the statistic which is easy to understand and is used in inferential or inductive statistics.
- A **statistic** is a random variable since its value depends on **observed sample values** which will differ from sample to sample.
- Its particular value depends on a given **set of sample** values.
- Thus determination of **sampling distribution** of a statistic is essentially a mathematical problem.

Sampling Distribution : Introduction: Example

A

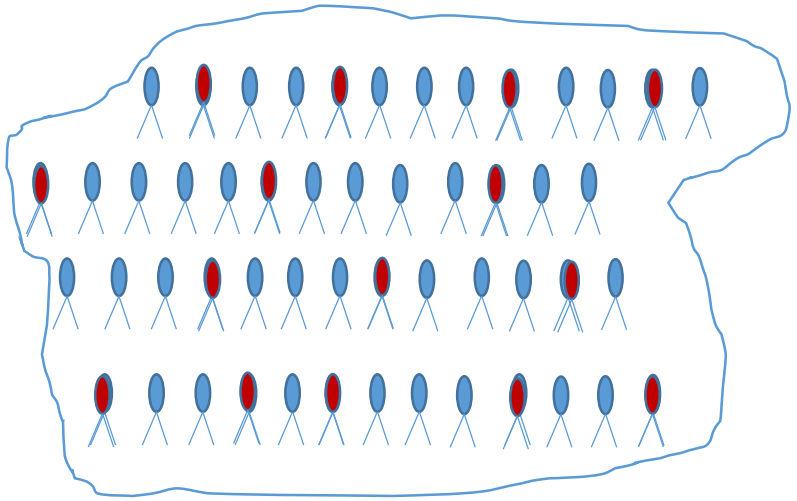


B

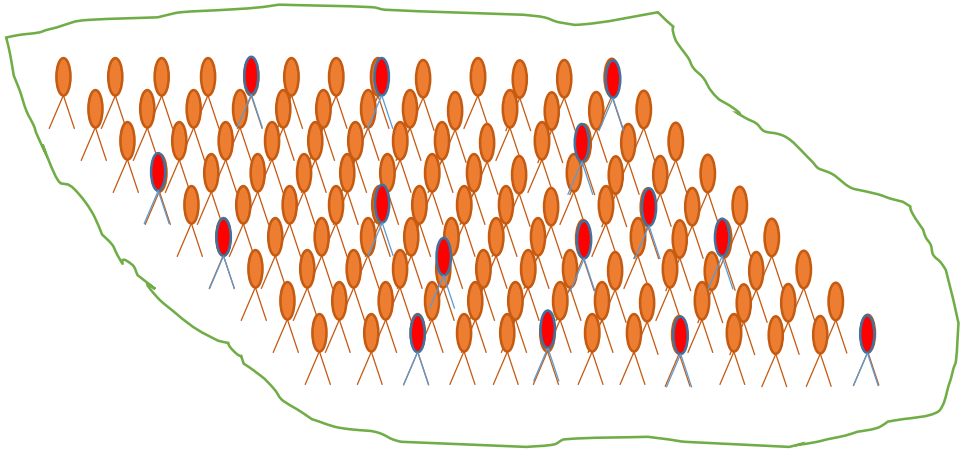


Sampling Distribution : Introduction: Example

A



B



S1

p_1

compare

p_2

S2

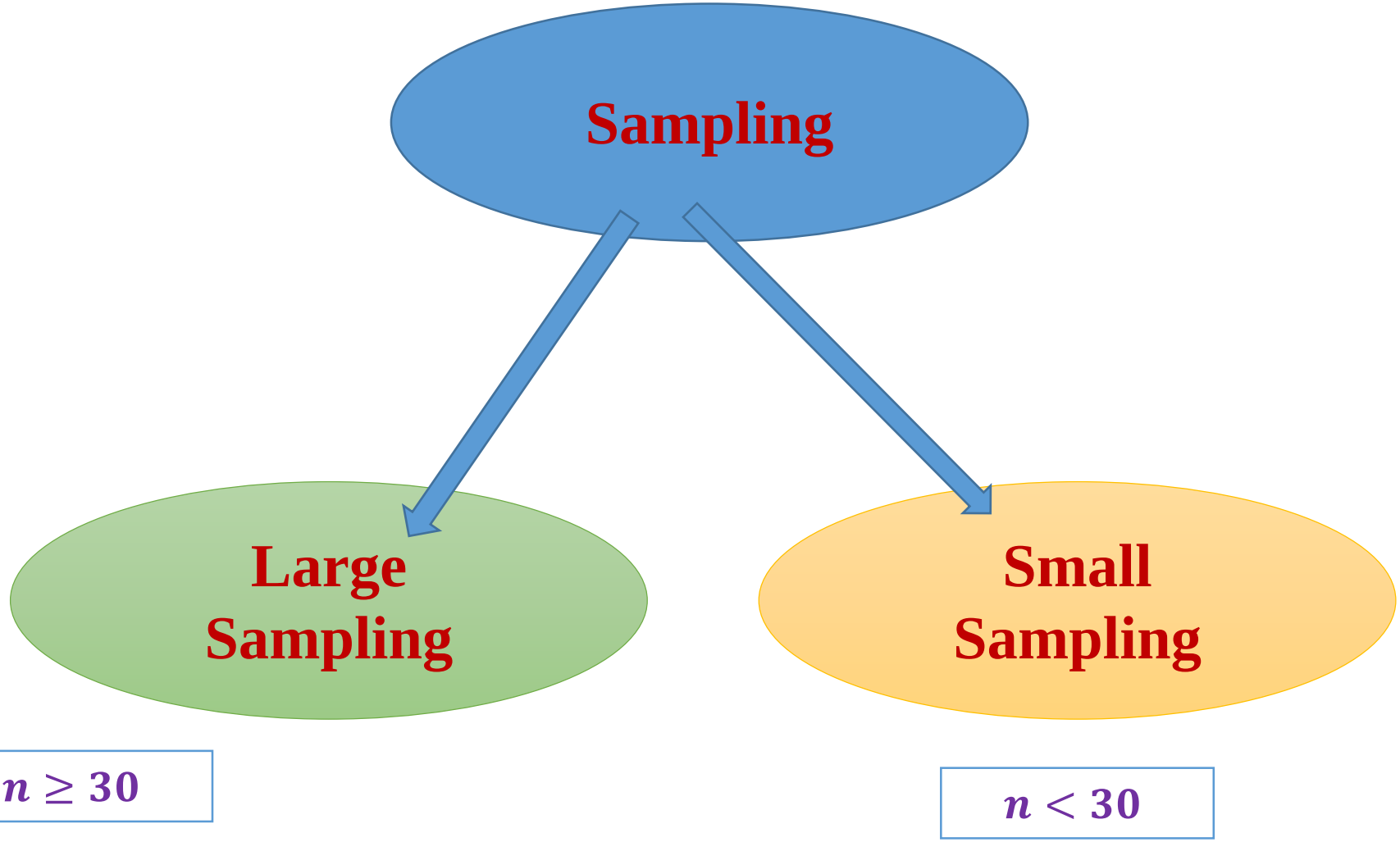
Sampling Distribution :Introduction

- Thus in **Sampling distribution**, statistician is mainly concerned with the analysis of data about the characteristics of persons or objects or observations.

Sampling Distribution :Introduction

- **Population and Sample:** Population is the set of collection or totality of object, animate or inanimate, actual or hypothetical, under study. Thus mainly population consists of sets on numbers, measurements or observations which are of interest.
- **Size:** Size of the population N is the number of objects or observations in the population.
- Population may be finite or infinite depending on the **size N** being **finite or infinite.**
- A finite subset of the population known as sample. Size of the sample is **denoted by n.**
- Sampling is the process of drawing samples from a given population.
- **Example:** Population of India, Population of any state, engineering colleges affiliated to AICTE, cars produced in India,

Sampling Distribution :Introduction



Sampling Distribution :Introduction

- **Statistical Inference:** Statistical inference or inductive statistics deals with the methods of drawing (arriving at) valid or logical generalization and predictions about the population using the information contained in the sample alone, with an indication of the accuracy of such inferences.

- **Parameters:** Statistical measures or constants obtained from the population are known as population parameters or simply parameters.

- Example: Population Mean, Population variance.

- Statistical quantities computed from sample observations are known as **sample statistics or statistics.**

Sampling Distribution :Introduction

- **Notation:** μ, σ, p represent the **population** mean, **population** standard deviation , **population** proportion.
- $\bar{X}, s, P$ represent the **sample** mean, **sample** standard deviation , **sample** proportion.
- Population f(x) is a population whose probability distribution is f(x).
- Example: if f(x) is binomial, Poisson or normal, then the corresponding population is known as binomial population, Poisson population or normal population.

Sampling Distribution :Introduction

- **Statistic:** Statistic is real valued function of the random sample. Statistic is a function of samples observation only and is itself a random variable.
- Sample mean and sample variance are two important statistics which are measures of a random sample $X_1, X_2, X_3, \ldots . X_n$ of size n.

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

$$Sample\ variance\ = S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X}^2)}{(n-1)} = \frac{n\sum X_i^2 - (\sum X_i)^2}{n(n-1)}$$

Sample standard deviation is the positive square root of the sample variance.

Degree of Freedom (dof): Degree of freedom of a statistic is a positive integer, denoted by $\boldsymbol{\nu}$, equal to $\boldsymbol{n-k}$ where **n** is the number of independent observations of the random sample

k is the number of population parameters which are calculated using the sample data.

Thus $\boldsymbol{dof\ \nu = n-k}$ is the difference between n the sample size and k the number of independent constraints imposed on the observations in the sample.

Point Estimation

- **Point estimation** is a form of statistical inference.
- In **point estimation** we use the data from the sample to compute a value of a sample statistic that serves as an estimate of a population parameter.
- We refer $\overline{X}$ as the **point estimator** of the population mean μ.
- s is the **point estimator** of the population standard deviation σ.
- $\hat{p}$ is the **point estimator** of the population proportion p.

Sampling Distribution

Relationships between the population distribution and Sampling Distribution of the Sample Mean

- The mean of the sample means is exactly equal to the population mean
- The dispersion of the sampling distribution of sample means is narrower than the population distribution
- The sampling distribution of sample means tends to become a bell- shaped and to approximate

Sampling Distribution of the Sample Mean

- The probability distribution of $\bar{X}$ is called its sampling distribution. It list the various values that $\bar{X}$ can assume and the probability of each value of $\bar{X}$. In general, the probability distribution of a sample statistic is called the sampling distribution.
- If a population is normal with mean μ and standard deviation σ, the sampling distribution of $\bar{X}$ is also normally distributed with $\bar{X}$.

$$\mu_{\hat{X}=}\ \mu \qquad and \qquad \sigma_{\hat{X}} = \frac{\sigma}{\sqrt{n}}$$

Z-value of the sampling distribution of $\bar{X}$

$$\boldsymbol{Z = \frac{(\bar{X} - \mu)}{\sigma/\sqrt{n}}}$$

Sampling Distribution

- **Central Limit Theorem:** if $\bar{X}$ is the mean of a sample of size n drawn from a population with mean μ and finite variance σ^2 then the standardized sample mean

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Is a random variable whose distribution function approaches that of the standard normal distribution $N(Z; 0,1)$as $n \to \infty$.

Sampling Distribution

Properties and shape of the sampling distribution of the sample Mean

- If $n \geq 30$, $\bar{X}$ is normally distributed, where $\bar{X}$:

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

Note: if σ^2 is unknown then it is estimated as s^2

- If $n < 30$ and variance is known, $\bar{X}$ is normally distributed

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

- If $n < 30$ and variance is unknown, t distribution with $n - 1$ degree of freedom is use

$$\mathrm{T} = \frac{\bar{X} - \mu}{\sqrt{\frac{s^2}{n}}} \sim t_{n-1}$$

Sampling Distribution

❑ **Example 1:** Determine the mean and standard deviation of the sampling distribution of means of 300 random sample each of size n=36 are drawn from a population of N=1500 which is normally distributed with mean $\mu = 22.4\ and\ S.D.\sigma = 0.048.$ If the sampling is done

a. with replacement and b. without replacement.

Solution:

a. with replacement: $\mu_{\bar{x}} = \mu = 22.4$ and

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.048}{\sqrt{36}} = 0.008$$

b. without replacement: $\mu_{\bar{x}} = \mu = 22.4$ and

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{0.048}{\sqrt{36}}\sqrt{\frac{1500-36}{1500-1}} = 0.0079\ ,\ \sigma_{\bar{x}} \approx 0.008$$

Sampling Distribution

❑**Example 2:** Suppose a population has mean $\mu = 8\ and\ s.d.\ \sigma = 3$. Suppose a random sample of size n=36 is selected What is the probability that the sample mean is between 7.8 and 8.2 ?

Solution: Even if the population is not normally distributed, the central limit theorem can be used $n > 30$. So the sampling distribution $\bar{x}$ is approximately normal.

Mean $\mu_{\bar{x}} = \mu = 8$ and

Standard Deviation $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{36}} = 0.5$

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Now Probability the sample mean lies between (7.8 and 8.2):

$$P(7.8 < Z < 8.2) = P\left(\frac{7.8-8}{3/\sqrt{36}} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \frac{8.2-8}{3/\sqrt{36}}\right) = P(-0.4 < Z < 0.4) = 0.3108$$

Sampling Distribution

❑ **Example 3:** The amount of time required to change of oil and filter of any vehicle is normally distributed with a mean of $\mu = 45$ minutes and standard deviation of $\sigma = 10$ minutes. A random sample of 16 cars is selected.

a. What is the standard error of the sample mean to be?

b. What is the probability of the sample mean between 45 and 52 minutes ?

c. What is the probability of the sample mean between 39 and 48 minutes ?

Solution: X: the amount of time required to change the oil and filter of any vehicles.

$$X = N(45, 10^2) \qquad n = 16$$

$\bar{X}$: the mean amount of time required to change the oil and filter of any vehicles.

$$\bar{X} = N\left(45, \frac{10^2}{16}\right)$$

Sampling Distribution

a. Standard error= Standard deviation,

$$\sigma = \frac{10}{\sqrt{16}} = 2.5$$

using formula $\boldsymbol{Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}$

b. $P(45 < Z < 52) = P\left(\frac{45-45}{2.5} < \boldsymbol{Z} < \frac{52-45}{2.5}\right) = P(0 < \boldsymbol{Z} < 2.8) = 0.4978$

c. $P(39 < Z < 48) = P\left(\frac{39-45}{2.5} < \boldsymbol{Z} < \frac{48-45}{2.5}\right) = P(-2.4 < \boldsymbol{Z} < 1.2)$

$= 0.4918 + 0.3849$

$= 0.8767$

Sampling Distribution for the difference between Two Mean

- Suppose we have two populations, X_1 and X_2 which are normally distributed. X_1 has mean μ_1 and variance σ_1^2 while X_2 has mean μ_2 and variance σ_2^2. These two distributions can be written as:

$$X_1 = N(\mu_1, \sigma_1^2) \qquad \text{and} \qquad X_2 = N(\mu_2, \sigma_2^2)$$

Now we are interested in finding out what is the sampling distribution of the difference between two sample means, the distribution of $\bar{X}_1 - \bar{X}_2$

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \ \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

Sampling Distribution

- **Example 4**: A taxi company purchases two brands of tiers, brans A and brand B. It is known that the mean distance travelled before the tiers wear out is 36300 km for brand A with standard deviation of 200 km while mean distance travelled before the tiers wear out is 36100 km for brand B with standard deviation of 300 km. A random sample of 36 tires of brand A and 49 tires of brand B are taken. What is the probability that the
- **a.** difference between the mean distance travelled before the tires of brand A and brand B wear out is at most 300 km.
- **b.** mean distance travelled by tiers with brand A is larger than the mean distance travelled by tires with brand B before the tiers wear out ?

Sampling Distribution

Solution:

- $\bar{X}_1$= the mean distance travelled before the tiers of brand A wear out
- $\bar{X}_2$= the mean distance travelled before the tiers of brand B wear out

$$\bar{X}_1 - \bar{X}_2 \sim N\left(36300 - 36100, \frac{200^2}{36} + \frac{300^2}{49}\right)$$

$$\bar{X}_1 - \bar{X}_2 \sim N\ (200,\ 2947.846)$$

a. $P(|\bar{X}_1 - \bar{X}_2| \leq 300) = P(-300 \leq \bar{X}_1 - \bar{X}_2 \leq 300)$

$$= P\left(\frac{-300 - 200}{\sqrt{2947.846}} \leq Z \leq \frac{300 - 200}{\sqrt{2947.846}}\right)$$

$=P(-9.21 \leq Z \leq 1.84)=$ **0.9671**

b. $P(\bar{X}_1 > \bar{X}_2) = P(\bar{X}_1 - \bar{X}_2 > 0)$

$$= P\left(Z > \frac{0-200}{\sqrt{2947.846}}\right)$$

$=P(Z > -3.68)=$ **0.9999**

Sampling Distribution

Sampling Distribution of Sample Proportion

- The population and sample proportion are denoted by p and $\hat{p}$ respectively, are calculated as:

$$p = \frac{X}{N} \qquad \text{and} \qquad \hat{p} = \frac{x}{n}$$

where

- N=total number of elements in the population ;
- X= number of elements in the population that posses the specific characteristics
- n=total number of elements in the sample; and
- x= number of elements in the sample that posses the specific characteristics
- For the large value of n (n $\geq$ 30), the sampling distribution is very closely normally distributed. $\hat{p} = N\left(p, \frac{pq}{n}\right)$
- **Mean and Standard Deviation of Sample Proportion**

$$\mu_{\hat{p}=} p \qquad \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$$

Sampling Distribution

- **Example 5**: If the true proportion of voters we support proposition A is $\hat{p}$= 0.40. what is the probability that a sample of size 200 yields a sample proportion between 0.40 and 0.45 ? If $p = 0.40$ and $n = 200$, what is the probability P(0.40≤ $\hat{p}$ ≤ 0.45) ?

  Solution:

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.4(1-0.4)}{200}} = 0.03464$$

$$\text{P}(0.40 \leq \hat{p} \leq 0.45) = P\left(\frac{0.40-0.40}{0.03464} < \boldsymbol{Z} < \frac{0.45-0.40}{0.03464}\right)$$

$$= P(0 < \boldsymbol{Z} < 1.44) = 0.4251$$

Sampling Distribution

- **Example 6:** The national survey of Engagement shows about 87% of freshmen and seniors rate their college experience as "good" or "excellent". Assume this result is true for the current population of freshmen and seniors. Let $\hat{p}$ be the proportion of freshmen and seniors in a random sample of 900 who hold this view. Find the mean and standard deviation.

Solution: Let p be the proportion of all freshmen and seniors who rate their college experience as "good" or "excellent". Then

$p = 0.87$ and $q = 1 - p = 1 - 0.87 = 0.13$

The Mean of the sample distribution $\hat{p}$ is: $\mu_{\hat{p}} = p = 0.87$

The Standard deviation of $\hat{p}$ is: $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.87(0.13)}{900}} = 0.11$

Sampling Distribution for the difference between Two Proportions

- Suppose we have two binomial populations with proportion of success p_1 and p_2 respectively. Samples of size n_1 are taking from population 1 and samples of size n_2 are taking from population 2. Then $\hat{p}$:

$$\hat{P}_1 \sim N\left(p_1, \frac{p_1(1-p_1)}{n_1}\right)$$

$$\hat{P}_2 \sim N\left(p_2, \frac{p_2(1-p_2)}{n_2}\right)$$

$$\hat{P}_1 - \hat{P}_2 \sim N\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right)$$

Sampling Distribution

- **Example 7:** A certain change in a process for manufacture of component parts was considered. It is fund that 75 out of 1500 items from existing procedure were found to be defective and 80 out of 2000 items from the new procedure were found to be defective. If one random sample of size 49 items were taken from the existing procedure and a random sample of 64 items were taken from the new procedure, what is the probability that

a. The proportion of the defective items from the new procedure exceeds the proportion of the defective items from the existing procedure

b. Proportion differs by at most 0.015 ?

c. The proportion of the defective items from the new procedure exceeds proportion of the defective items from the existing procedure by at least 0.02 ?

Sampling Distribution

- **Solution:**
- $\hat{P}_N$: The proportion of defective items from the new procedure
- $\hat{P}_E$: The proportion of defective items from the existing procedure

$$p_N = \frac{80}{2000} = 0.04 \qquad p_E = \frac{75}{1500} = 0.05$$

$$\hat{P}_N \sim N\left(0.04, \frac{0.04(0.96)}{64}\right) \qquad \hat{P}_E \sim N\left(0.05, \frac{0.05(0.95)}{49}\right)$$

$$\hat{P}_N - \hat{P}_E \sim N\left(0.04 - 0.05, \frac{0.05(0.95)}{49} + \frac{0.04(0.96)}{64}\right) \qquad \hat{P}_N - \hat{P}_E \sim N(-0.01, 0.0016)$$

Sampling Distribution

a. $P(\hat{P}_N > \hat{P}_E) = P(\hat{P}_N - \hat{P}_E > 0)$

$$= P\left(Z > \frac{0-(-0.01)}{\sqrt{0.0016}}\right)$$

$$= P(Z > 0.25) = \mathbf{0.4013}$$

b. $P\left(\left|\hat{P}_N - \hat{P}_E\right| \leq 0.015\right) = P\left(-0.015 \leq \hat{P}_N - \hat{P}_E \leq 0.015\right)$

$$= P\left(\frac{-0.015-(-0.01)}{\sqrt{0.0016}} \leq Z \leq \frac{0.015-(-0.01)}{\sqrt{0.0016}}\right)$$

$$= P(-0.125 \leq Z \leq 0.625) = \mathbf{0.2838}$$

c. $P(\hat{P}_N - \hat{P}_E \geq 0.02) = P(\hat{P}_N - \hat{P}_E \geq 0.02)$

$$= P\left(Z \geq \frac{0.02-(-0.01)}{\sqrt{0.0016}}\right)$$

$$= P(Z \geq 0.75) = \mathbf{0.2266}$$

Thank You