Department of Engg. Mathematics

Course : Engg. Mathematics-IV 15MAT41. Sem.: 4th (2017-18)

Course Coordinator:

Prof. S. A. Patil

Session Objectives

Sampling Distribution Introduction.

Sampling Distribution of the Mean

Student's *t* Distribution

Sampling Distribution of the Sample Mean

The Sampling Distribution of the Sample Proportion

Normal approximation to the Binomial

Sampling Distribution Introduction

- In real life calculating parameters of populations is prohibitive because populations are very large.
- Rather than investigating the whole population, we take a sample, calculate a **statistic** related to the **parameter** of interest, and make an inference.
- The **sampling distribution** of the **statistic** is the tool that tells us how close is the statistic to the parameter.

Sample Statistics as Estimators of Population Parameters

- A **sample statistic** is a numerical measure of a summary characteristic of a sample.

A **population parameter** is a numerical measure of a summary characteristic of a population.

- An **estimator** of a population parameter is a sample statistic used to estimate or predict the population parameter.
- An **estimate** of a parameter is a *particular* numerical value of a sample statistic obtained through sampling.
- A **point estimate** is a single value used as an estimate of a population parameter.

Estimators

- The sample mean, is the most common estimator of the population mean, μ.
- The sample variance, s^2, is the most common estimator of the population variance, σ^2.
- The sample standard deviation, s, is the most common estimator of the population standard deviation, σ.
- The sample proportion, , is the most common estimator of the population proportion, p.

Sampling Distribution of X

The sampling distribution of X is the probability distribution of all possible values the random variable $\bar{X}$ may assume when a sample of size n is taken from a specified population.

Sampling Distribution of the Mean

- An example
  - A die is thrown infinitely many times. Let X represent the number of spots showing on any throw.
  - The probability distribution of X is

x	1	2	3	4	5	6
p(x)	1/6	1/6	1/6	1/6	1/6	1/6

E(X) = 1(1/6) + 2(1/6) + 3(1/6)+ ……………………= 3.5

V(X) = $(1-3.5)^2(1/6)$ + $(2-3.5)^2(1/6)$ + …………. …= 2.92

Throwing a dice twice – sampling distribution of sample mean

- Suppose we want to estimate m from the mean of a sample of size n = 2.
- What is the distribution of ?

Throwing a die twice – sample mean

Sample		Mean	Sample		Mean	Sample		Mean
1	1,1	1	13	3,1	2	25	5,1	3
2	1,2	1.5	14	3,2	2.5	26	5,2	3.5
3	1,3	2	15	3,3	3	27	5,3	4
4	1,4	2.5	16	3,4	3.5	28	5,4	4.5
5	1,5	3	17	3,5	4	29	5,5	5
6	1,6	3.5	18	3,6	4.5	30	5,6	5.5
7	2,1	1.5	19	4,1	2.5	31	6,1	3.5
8	2,2	2	20	4,2	3	32	6,2	4
9	2,3	2.5	21	4,3	3.5	33	6,3	4.5
10	2,4	3	22	4,4	4	34	6,4	5
11	2,5	3.5	23	4,5	4.5	35	6,5	5.5
12	2,6	4	24	4,6	5	36	6,6	6

Sampling Distribution of the Mean

$n = 5$

$\mu_{\bar{x}} = 3.5$

$\sigma^2_{\bar{x}} = .5833 \; (= \frac{\sigma^2_x}{5})$

$n = 10$

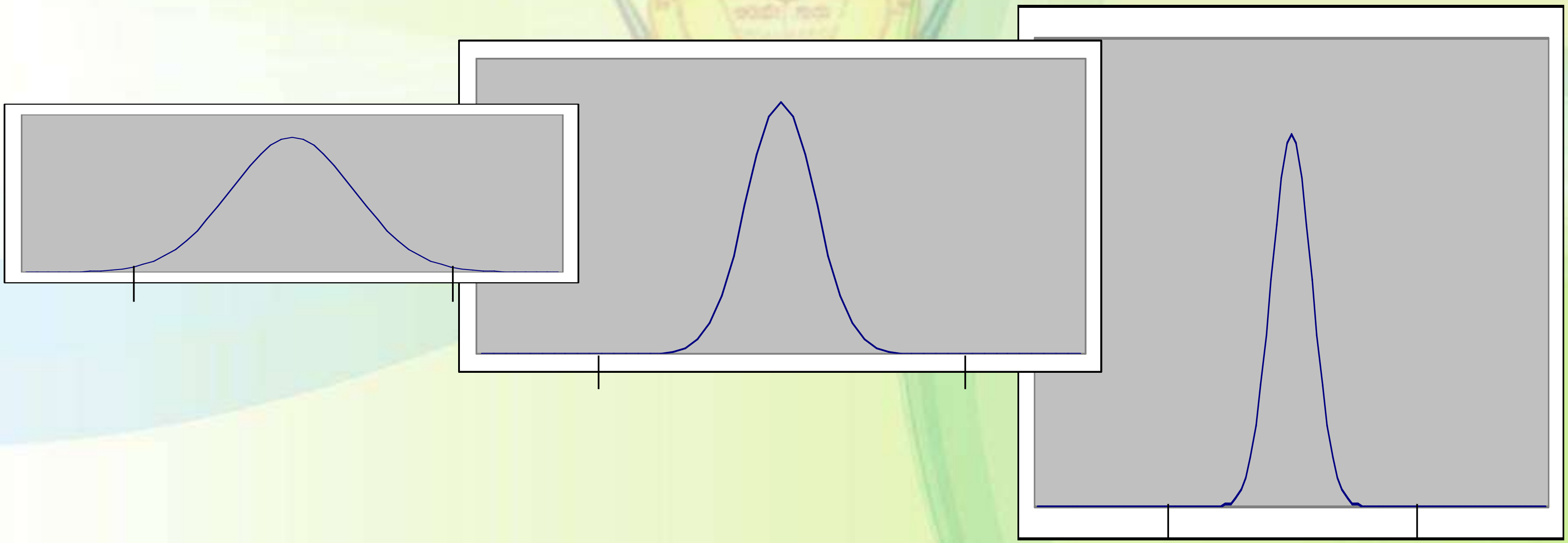
$\mu_{\bar{x}} = 3.5$

$\sigma^2_{\bar{x}} = .2917 \; (= \frac{\sigma^2_x}{10})$

$n = 25$

$\mu_{\bar{x}} = 3.5$

$\sigma^2_{\bar{x}} = .1167 \; (= \frac{\sigma^2_x}{25})$



Sampling Distribution of the Mean

$$n = 5$$
$$\mu_{\bar{x}} = 3.5$$
$$\sigma^2_{\bar{x}} = .5833 \; (= \frac{\sigma^2_x}{5})$$

$$n = 10$$
$$\mu_{\bar{x}} = 3.5$$
$$\sigma^2_{\bar{x}} = .2917 \; (= \frac{\sigma^2_x}{10})$$

$$n = 25$$
$$\mu_{\bar{x}} = 3.5$$
$$\sigma^2_{\bar{x}} = .1167 \; (= \frac{\sigma^2_x}{25})$$

Notice that $\sigma^2_{\bar{x}}$ is smaller than σ_x. The larger the sample size the smaller $\sigma^2_{\bar{x}}$. Therefore, $\bar{X}$ tends to fall closer to μ, as the sample size increases.

Relationships between Population Parameters and the Sampling Distribution of the Sample Mean

The **expected value of the sample mean** is equal to the population mean:

$$E(\overline{X}) = \mu_{\overline{X}} = \mu_X$$

The **variance of the sample mean** is equal to the population variance divided by the sample size:

$$V(\overline{X}) = \sigma^2_{\overline{X}} = \frac{\sigma^2_X}{n}$$

The **standard deviation of the sample mean, known as the standard error of the mean,** is equal to the population standard deviation divided by the square root of the sample size:

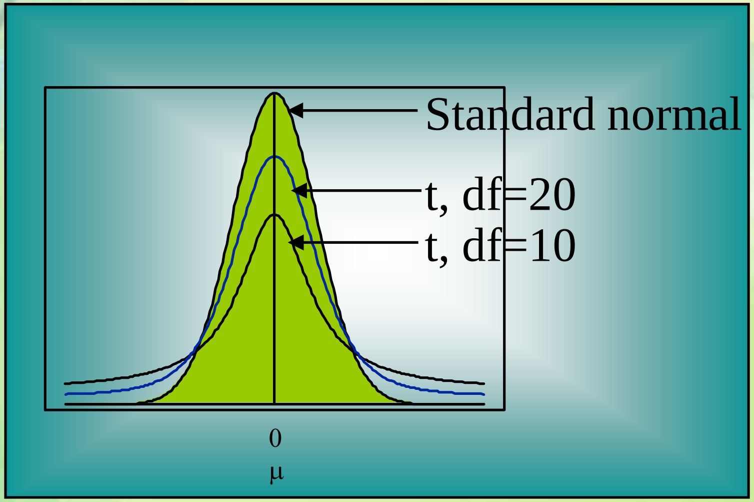
$$\text{s.e.} = SD(\overline{X}) = \sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}}$$

Student's t Distribution

If the population standard deviation, σ, is ***unknown***, replace σ with the sample standard deviation, *s*. If the population is normal, the resulting statistic:

has a ***t distribution* with *(n - 1) degrees of freedom***.

- **The *t* is a family of bell-shaped and symmetric distributions, one for each number of degree of freedom.**
- **The expected value of *t* is 0.**
- **The variance of *t* is greater than 1, but approaches 1 as the number of degrees of freedom increases.**
- **The *t* distribution approaches a standard normal as the number of degrees of freedom increases.**
- **When the sample size is small (<30) we use t distribution.**



Sampling Distributions

Finite Population Correction Factor

If the sample size is more than 5% of the population size and the sampling is done without replacement, then a correction needs to be made to the standard error of the means.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \bullet \sqrt{\frac{N-n}{N-1}}$$

Sampling Distribution of

Standard Deviation of $\bar{x}$

Finite Population

$$\sigma_{\bar{x}} = \left(\frac{\sigma}{\sqrt{n}}\right)\sqrt{\frac{N-n}{N-1}}$$

Infinite Population

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

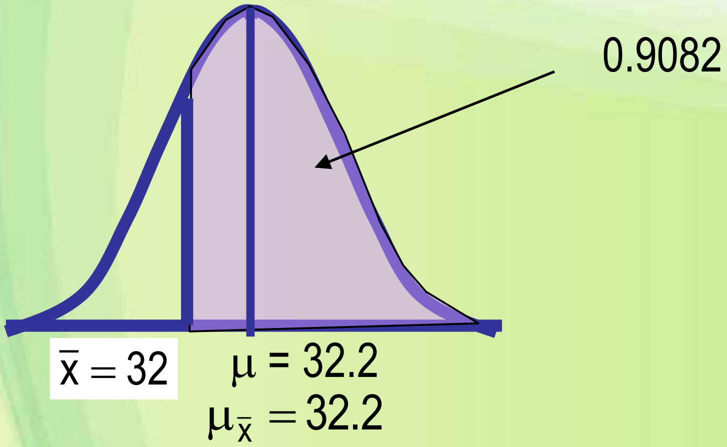
- A finite population is treated as being infinite if $n/N \leq .05$.
- $\sqrt{(N-n)/(N-1)}$ is the finite correction factor.
- $\sigma_{\bar{x}}$ is referred to as the standard error of the mean.

Sampling Distribution of the Sample Mean

- The amount of soda pop in each bottle is normally distributed with a mean of 32.2 ounces and a standard deviation of 0.3 ounces.
- Find the probability that a carton of four bottles will have a mean of more than 32 ounces of soda per bottle.
- **Solution**
  - Define the random variable as the mean amount of soda per bottle.

$$P(\bar{x} > 32) = P\left(\frac{\bar{x} - \mu}{\sigma_{\bar{x}}} > \frac{32 - 32.2}{.3/\sqrt{4}}\right)$$

$$= P(z > -1.33) = 0.9082$$



Sampling Distribution of the Sample Mean

- Example
  - Dean's claim: The average weekly income of M.B.A graduates one year after graduation is \$600.
  - Suppose the distribution of weekly income has a standard deviation of \$100. What is the probability that 25 randomly selected graduates have an average weekly income of less than \$550?
  - Solution

$$P(\bar{x} < 550) = P\left(\frac{\bar{x} - \mu}{\sigma_{\bar{x}}} < \frac{550 - 600}{100/\sqrt{25}}\right)$$

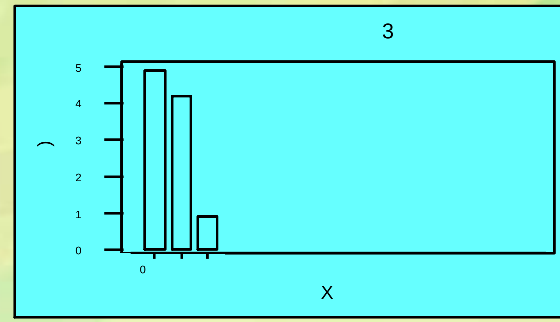
$$= P(z < -2.5) = 0.0062$$

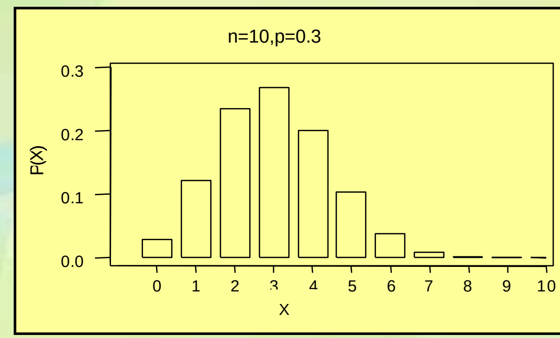
The Sampling Distribution of the Sample Proportion,

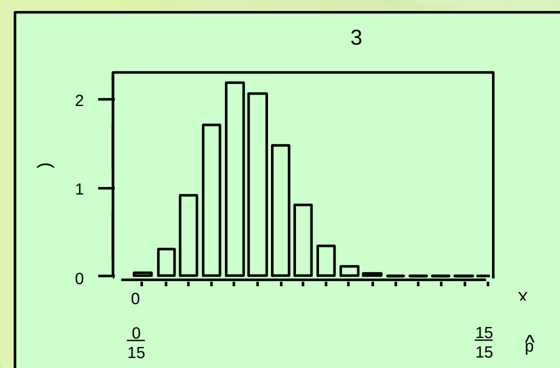
The **sample proportion** is the percentage of successes in *n* binomial trials. It is the number of successes, *X*, divided by the number of trials, *n*.

Sample proportion: $\hat{p} = \frac{X}{n}$

As the sample size, *n*, increases, the sampling distribution of approaches a **normal distribution** with mean *p* and standard deviation







Normal approximation to the Binomial

- Normal approximation to the binomial works best when
  - the number of experiments (sample size) is large, and
  - the probability of success, p, is close to 0.5.

- For the approximation to provide good results two conditions should be met:

$$np \quad 5; \quad n(1 - p) \quad 5$$

- Example
  - A state representative received 52% of the votes in the last election.
  - One year later the representative wanted to study his popularity.
  - If his popularity has not changed, what is the probability that more than half of a sample of 300 voters would vote for him?

- Example
  - Solution
  - The number of respondents who prefer the representative is binomial with n = 300 and p = .52. Thus, np = 300(.52) = 156 and n(1-p) = 300(1-.52) = 144 (both greater than 5)

Queries ….?