Econ 3790: Business and Economics Statistics

Instructor: Yogesh Uppal
Email: yuppal@ysu.edu

Chapter 7, Part A
Sampling and Sampling Distributions

- Simple Random Sampling
- Point Estimation
- Introduction to Sampling Distributions
- Sampling Distribution of $\bar{x}$

Statistical Inference

- The sample results provide only estimates of the values of the population characteristics.
- With proper sampling methods, the sample results can provide "good" estimates of the population characteristics.
- A parameter is a numerical characteristic of a population.

Statistical Inference

- The purpose of statistical inference is to obtain information about a population from information contained in a sample.
- A population is the set of all the elements of interest.
- A sample is a subset of the population.

Simple Random Sampling: Finite Population

- A simple random sample of size *n* from a finite population of size *N* is a sample selected such that each possible sample of size *n* has the same probability of being selected.

Simple Random Sampling: Finite Population

- Replacing each sampled element before selecting subsequent elements is called sampling with replacement.
- Sampling without replacement is the procedure used most often.
- In large sampling projects, computer-generated random numbers are often used to automate the sample selection process.

Point Estimation

- 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 to $\bar{x}$ as the point estimator of the population mean μ.
- s is the point estimator of the population standard deviation σ.
- $\bar{p}$ is the point estimator of the population proportion p.

Sampling Error

- When the expected value of a point estimator is equal to the population parameter, the point estimator is said to be unbiased.
- The absolute value of the difference between an unbiased point estimate and the corresponding population parameter is called the sampling error.
- Sampling error is the result of using a subset of the population (the sample), and not the entire population.
- Statistical methods can be used to make probability statements about the size of the sampling error.

Sampling Error

- The sampling errors are:
  - $|\bar{x} - \mu|$ for sample mean
  - $|s - \sigma|$ for sample standard deviation
  - $|\bar{p} - p|$ for sample proportion

Air Quality Example

- Let us suppose that the population of air quality data consists of 191 observations.

- How would you determine the following population parameters: <u>mean, standard deviation, proportion of cities with good air quality.</u>

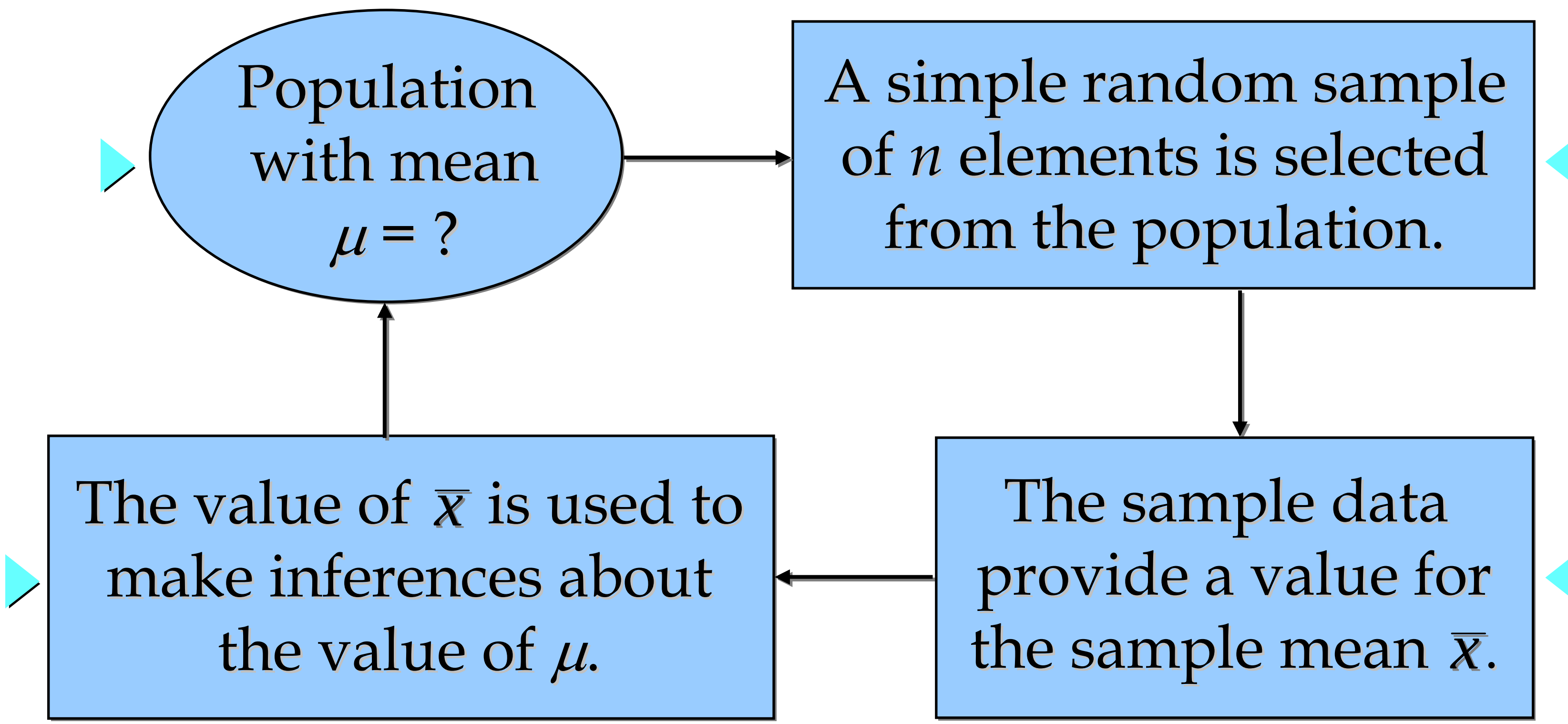
Air Quality Example

- How about picking a random sample from this population representing the air quality?
  - We shall use SPSS to do this random sampling for us.
- How would you use this sample to provide point estimates of the population parameters?

Summary of Point Estimates Obtained from a Simple Random Sample

Population Parameter	Parameter Value	Point Estimator	Point Estimate
μ = Population mean SAT score	40.9	$\bar{x}$ = Sample mean SAT score	….
σ = Population std. deviation for SAT score	20.5	s = Sample std. deviation for SAT score	…..
p = Population pro- portion	.62	$\bar{p}$ = Sample pro- portion wanting campus housing	….

- Process of Statistical Inference



Sampling Distribution of $\bar{x}$

The sampling distribution of $\bar{x}$ is the probability distribution of all possible values of the sample mean $\bar{x}$.

- **Expected Value of $\bar{x}$**

$$E(\bar{x}) = \mu$$

where:

μ = the population mean

Sampling Distribution of $\bar{x}$

Standard Deviation of $\bar{x}$

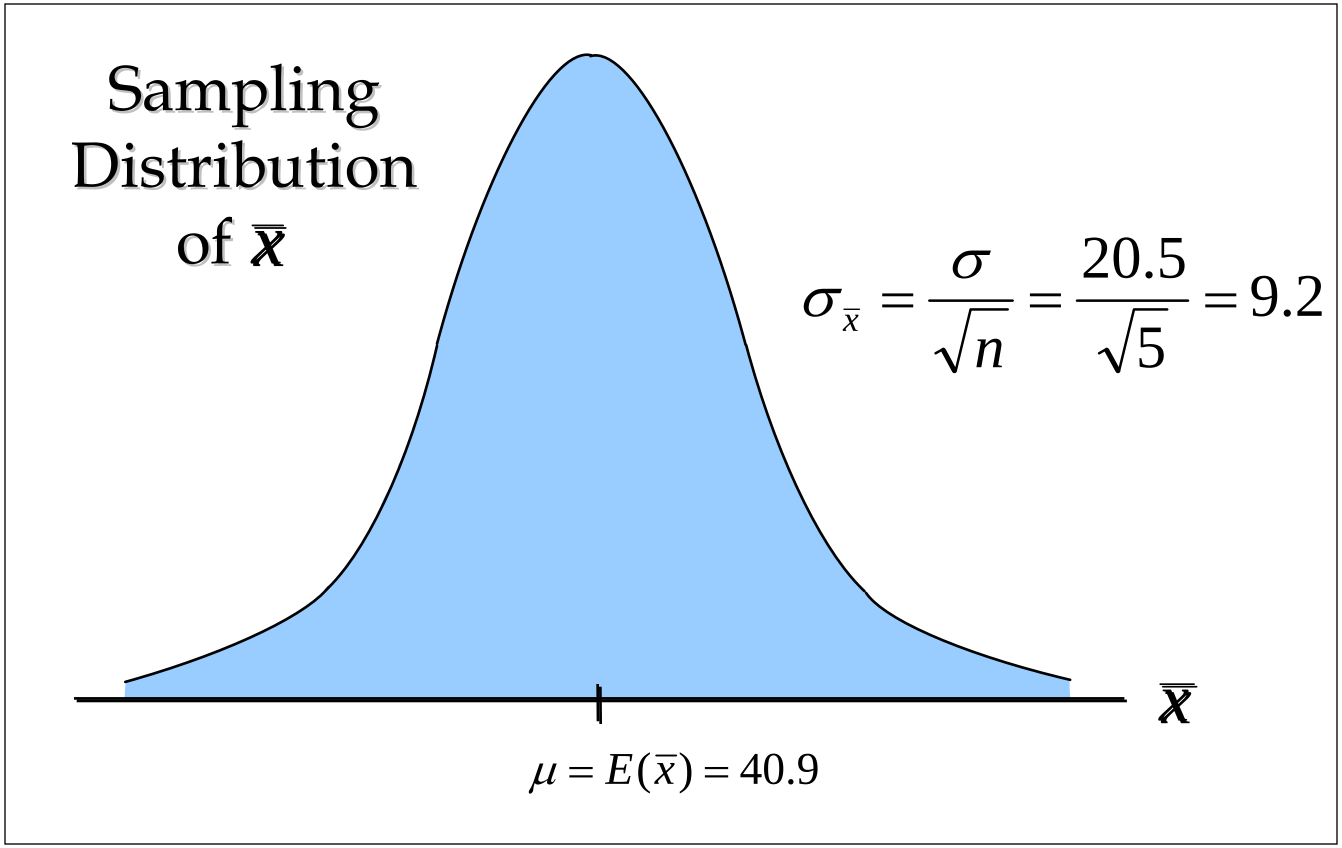
Finite Population	Infinite Population
$\sigma_{\bar{x}} = \left(\frac{\sigma}{\sqrt{n}}\right)\sqrt{\frac{N-n}{N-1}}$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

- 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 also referred to as the standard error of the mean.

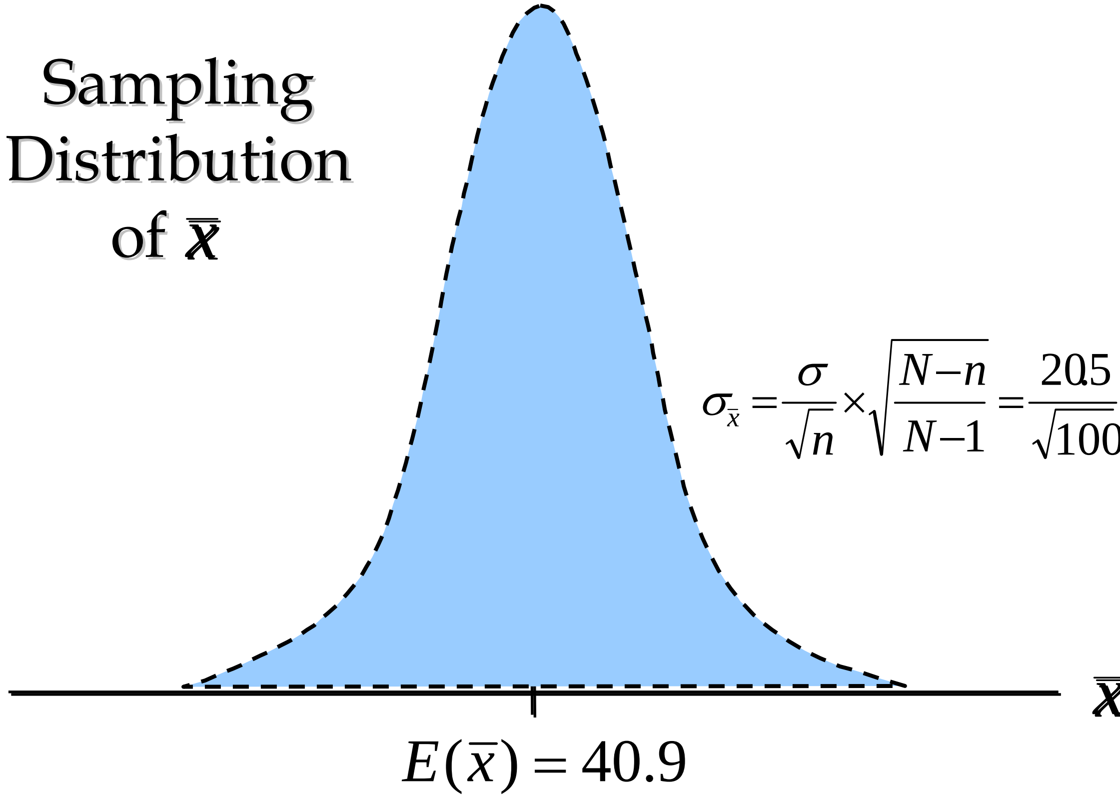
The Shape of Sampling Distribution of $\bar{x}$

- If the shape of the distribution of x in the population is *normal*, the shape of the sampling distribution of $\bar{x}$ is *normal* as well.
- If the shape of the distribution of x in the population is *approximately normal*, the shape of the sampling distribution of $\bar{x}$ is *approximately normal* as well.
- If the shape of the population is not approximately normal then
  - If n is small, the shape of the sampling distribution of $\bar{x}$ is unpredictable.
  - If n is large ($n \geq 30$), the shape of the sampling distribution of $\bar{x}$ can be assumed to be *approximately normal*.

Sampling Distribution of $\bar{x}$ for the air quality example when the population is (almost) infinite



Sampling Distribution of $\bar{x}$ for the air quality example when the population is finite

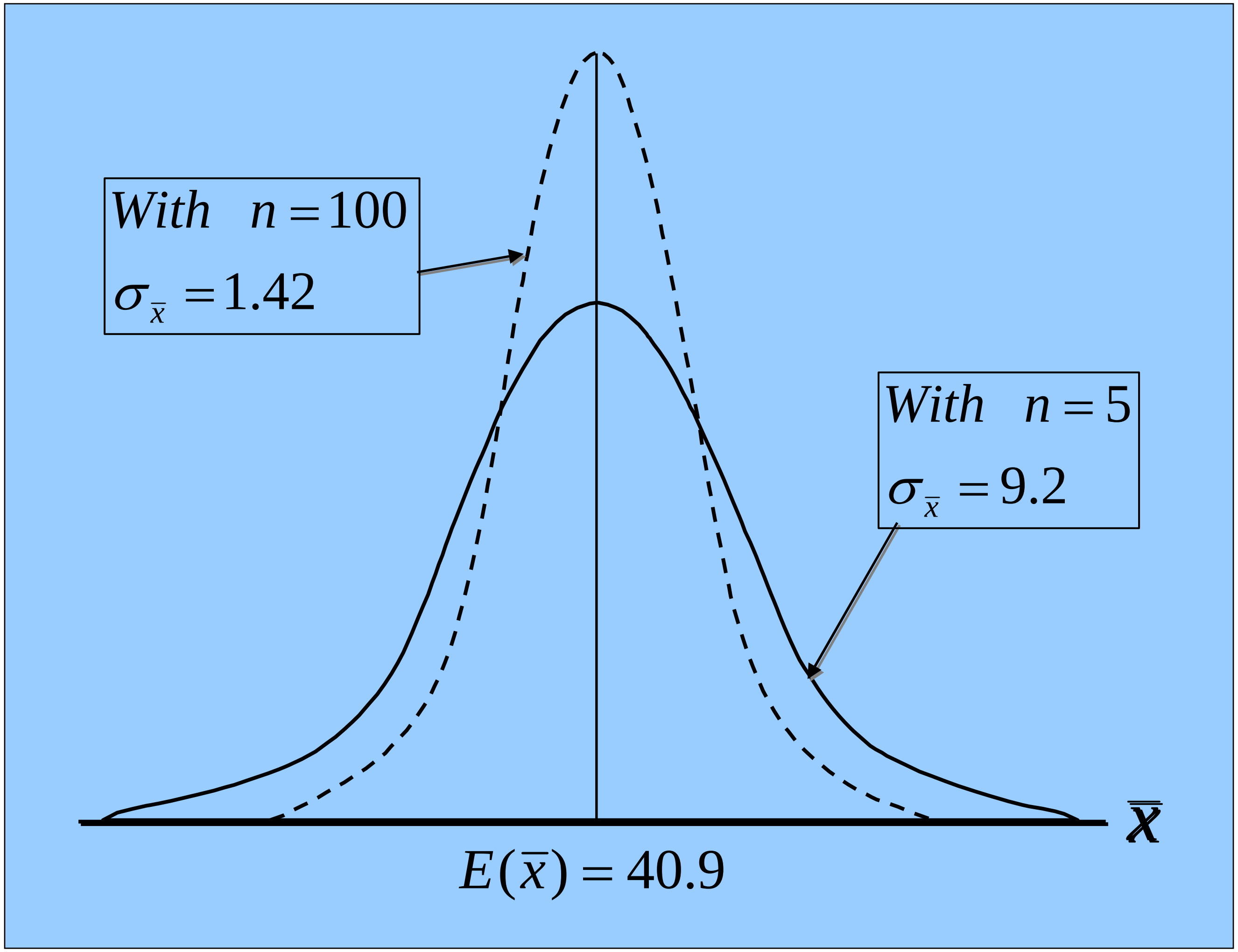


$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}} = \frac{20.5}{\sqrt{100}} \times \sqrt{\frac{191-100}{191-1}} = 1.42$$

Relationship Between the Sample Size and the Sampling Distribution of $\bar{x}$

- $E(\bar{x}) = \mu$ regardless of the sample size. In our example, $E(\bar{x})$ remains at 40.9.
- Whenever the sample size is increased, the standard error of the mean $\sigma_{\bar{x}}$ is decreased. With the increase in the sample size to n = 100, the standard error of the mean decreases.

Relationship Between the Sample Size and the Sampling Distribution of $\bar{x}$



Sampling Distribution of $\bar{x}$

- If we use a large random sample (n>30), then the sampling distribution of $\bar{x}$ can be approximated by the normal distribution.

- If the sample is small, then the sampling distribution of $\bar{x}$ can be normal only if we assume that our population has a normal distribution.

Sampling Distribution of $\bar{x}$ for the air quality Index when n = 5.

- What is the probability that a simple random sample of 5 applicants will provide an estimate of the population mean air quality index that is within +/-2 of the actual population mean, μ?

- In other words, what is the probability that $\bar{x}$ will be between 38.9 and 42.9?

Sampling Distribution of $\bar{x}$ for the air quality Index when n = 100.

- What is the probability that a simple random sample of 100 applicants will provide an estimate of the population mean air quality index that is within +/-2 of the actual population mean, μ?

Relationship Between the Sample Size and the Sampling Distribution of $\bar{x}$

- Because the sampling distribution with n = 100 has a smaller standard error, the values of $\bar{x}$ have less variability and tend to be closer to the population mean than the values of $\bar{x}$ with n = 5.

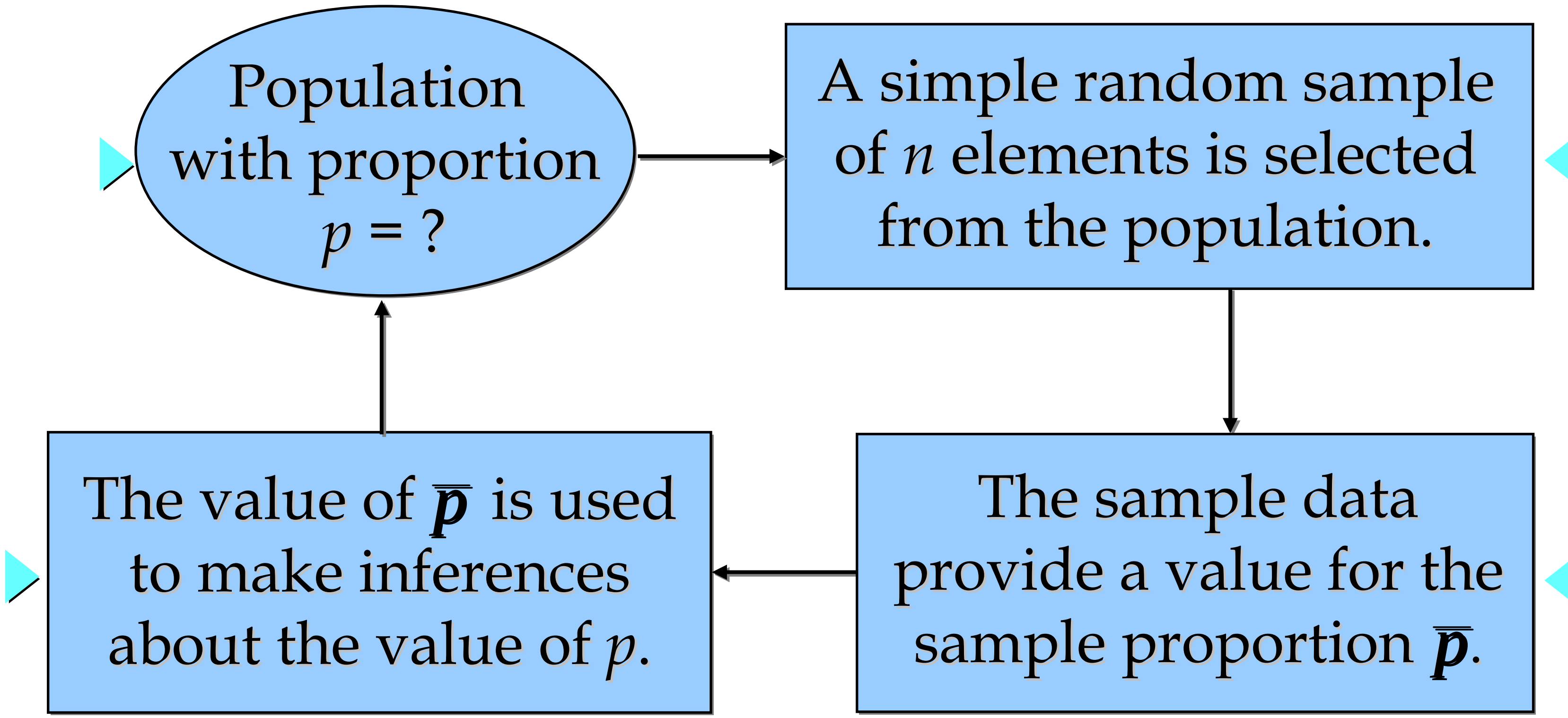
- Basically, a given interval with smaller standard error (larger n) will cover more area under the normal curve than the same interval with larger standard error (smaller n).

Chapter 7, Part B
Sampling and Sampling Distributions

- Sampling Distribution of $\bar{p}$

Sampling Distribution of $\bar{p}$

- Making Inferences about a Population Proportion



- Sampling Distribution of $\bar{p}$

  The sampling distribution of $\bar{p}$ is the probability distribution of all possible values of the sample proportion $\bar{p}$.

- Expected Value of $\bar{p}$

$$E(\bar{p}) = p$$

where:

p = the population proportion

Sampling Distribution of $\bar{p}$

Standard Deviation of $\bar{p}$

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

$\sigma_{\bar{p}}$ is referred to as the standard error of the proportion.

Form of Sampling Distribution of $\bar{p}$

The sampling distribution of $\bar{p}$ can be approximated by a normal distribution whenever the sample size is large: Central Limit Theorem (CLT).

The sample size is considered large whenever these conditions are satisfied:

$np \geq 5$ and $n(1 - p) \geq 5$