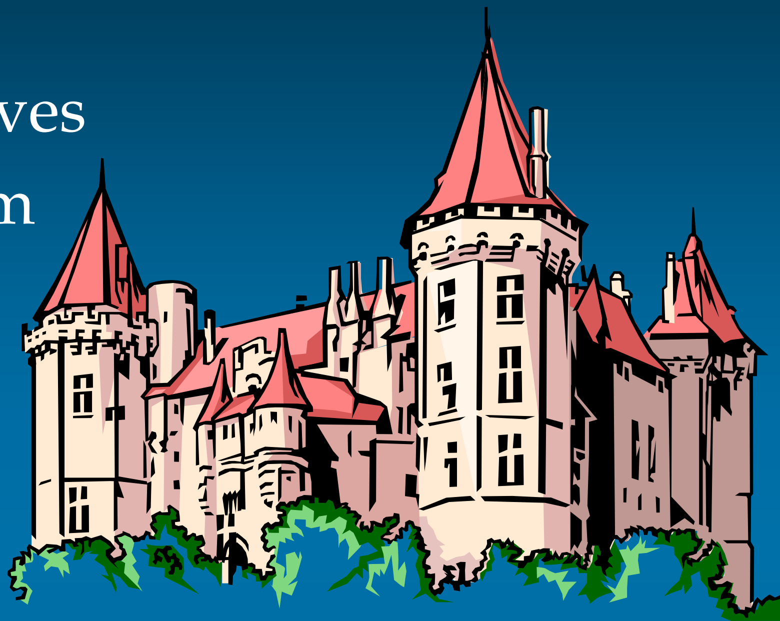
Chapter 7
Sampling and Sampling Distributions

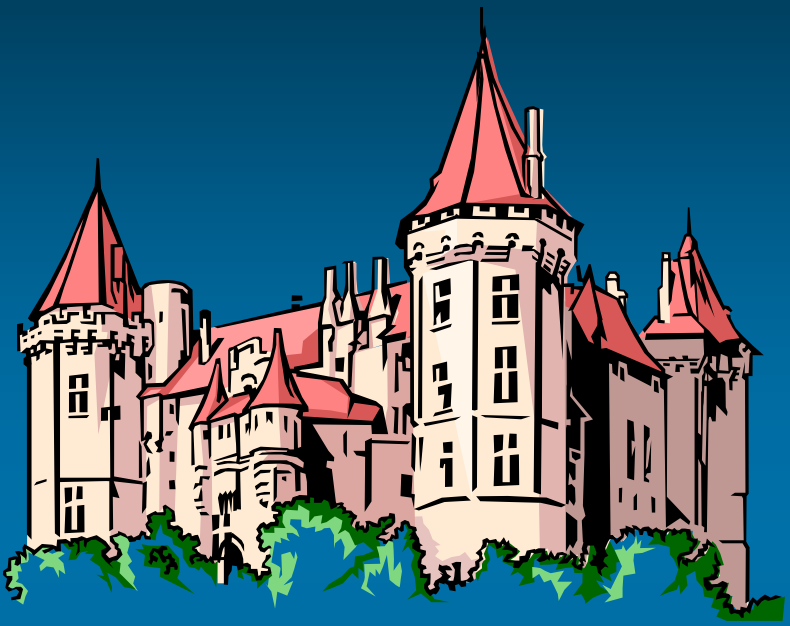
- Simple Random Sampling
- Point Estimation
- Introduction to Sampling Distributions
- Sampling Distribution of $\bar{x}$

Example: St. Andrew's

- St. Andrew's College receives 900 applications annually from prospective students. The application form contains a variety of information including the individual's scholastic aptitude test (SAT) score and whether or not the individual desires on-campus housing.



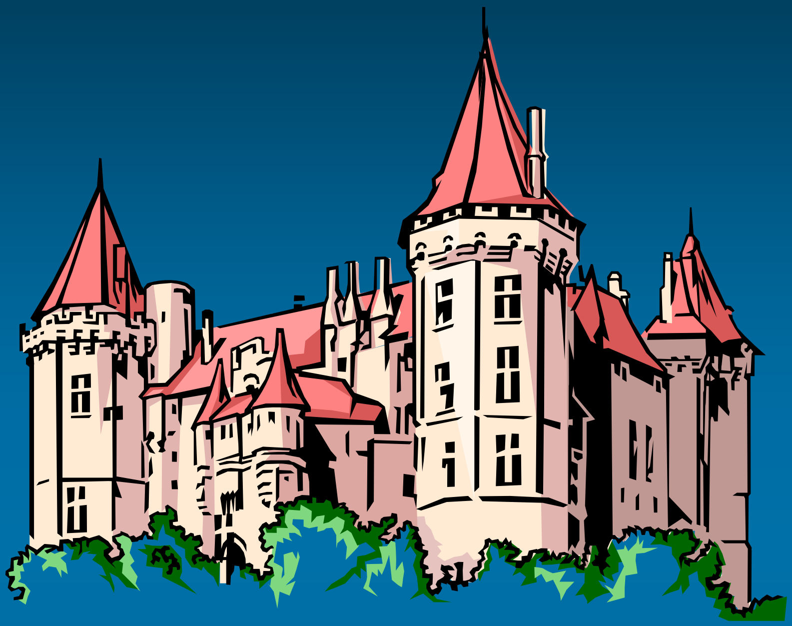
Example: St. Andrew's



- The director of admissions would like to know the following information:
  - the average SAT score for the 900 applicants, and
  - the proportion of applicants that want to live on campus.

Example: St. Andrew's

- We will now look at three alternatives for obtaining the desired information.
  - Conducting a census of the entire 900 applicants
  - Selecting a sample of 30 applicants, using a random number table
  - Selecting a sample of 30 applicants, using Excel



Conducting a Census

- If the relevant data for the entire 900 applicants were in the college's database, the population parameters of interest could be calculated using the formulas presented in Chapter 3.
- We will assume for the moment that conducting a census is practical in this example.

Conducting a Census

- Population Mean SAT Score

$$\mu = \frac{\sum x_i}{900} = 990$$

- Population Standard Deviation for SAT Score

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{900}} = 80$$

- Population Proportion Wanting On-Campus Housing

$$p = \frac{648}{900} = .72$$

Point Estimation

- q $\bar{x}$ as Point Estimator of μ

$$\bar{x} = \frac{\sum x_i}{30} = \frac{29,910}{30} = 997$$

- q s as Point Estimator of σ

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{29}} = \sqrt{\frac{163,996}{29}} = 75.2$$

- q $\bar{p}$ as Point Estimator of p

$$\bar{p} = 20/30 = .68$$

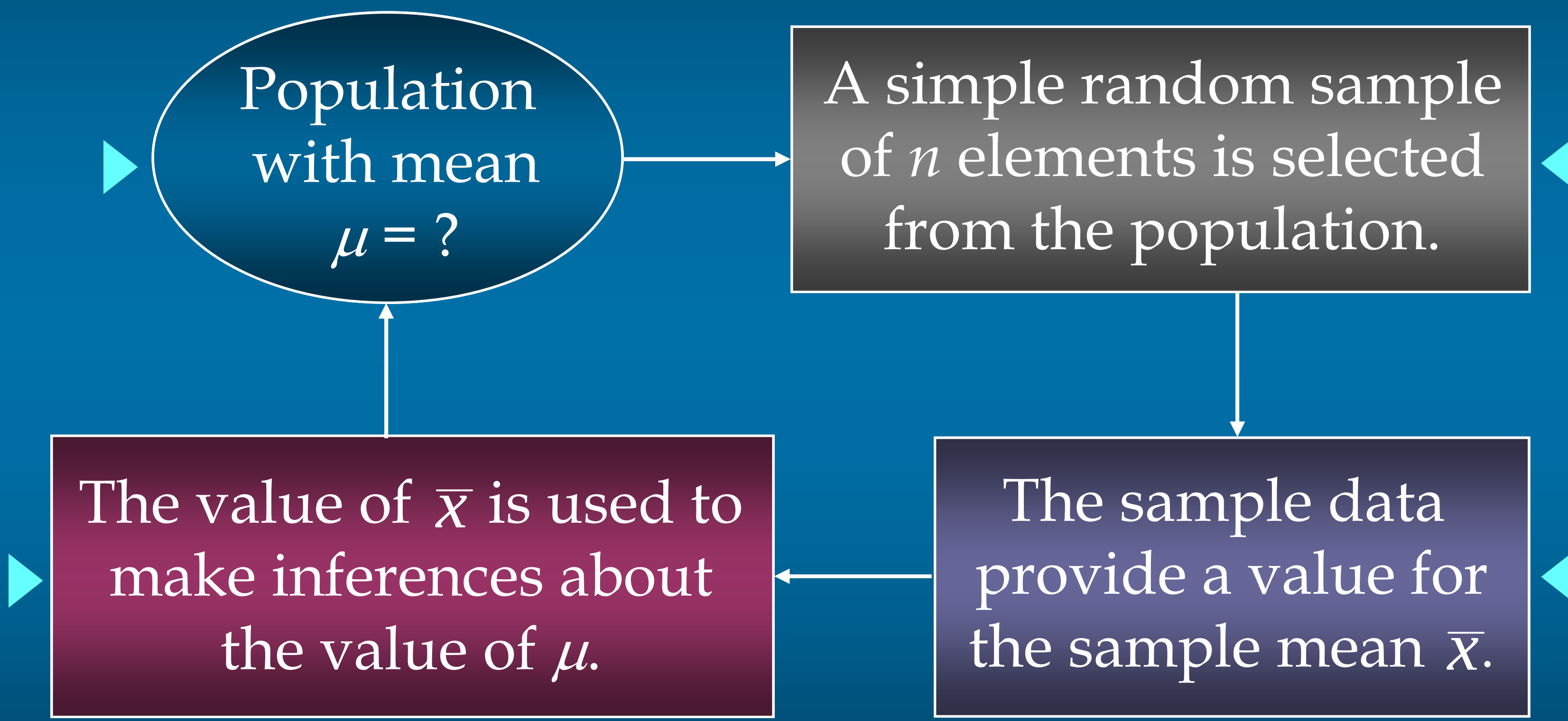
Note: Different random numbers would have identified a different sample which would have resulted in different point estimates.

Summary of Point Estimates Obtained from a Simple Random Sample

Population Parameter	Parameter Value	Point Estimator	Point Estimate
μ = Population mean SAT score	990	$\bar{x}$ = Sample mean SAT score	997
σ = Population std. deviation for SAT score	80	s = Sample std. deviation for SAT score	75.2
p = Population proportion wanting campus housing	.72	$\bar{p}$ = Sample proportion wanting campus housing	.68

Sampling Distribution of $\bar{x}$

- Process of Statistical Inference



Sampling Distribution of $\bar{x}$

The <u>sampling distribution of $\bar{x}$</u> 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
$\sigma_{\bar{x}} = \left(\frac{\sigma}{\sqrt{n}}\right)\sqrt{\frac{N-n}{N-1}}$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

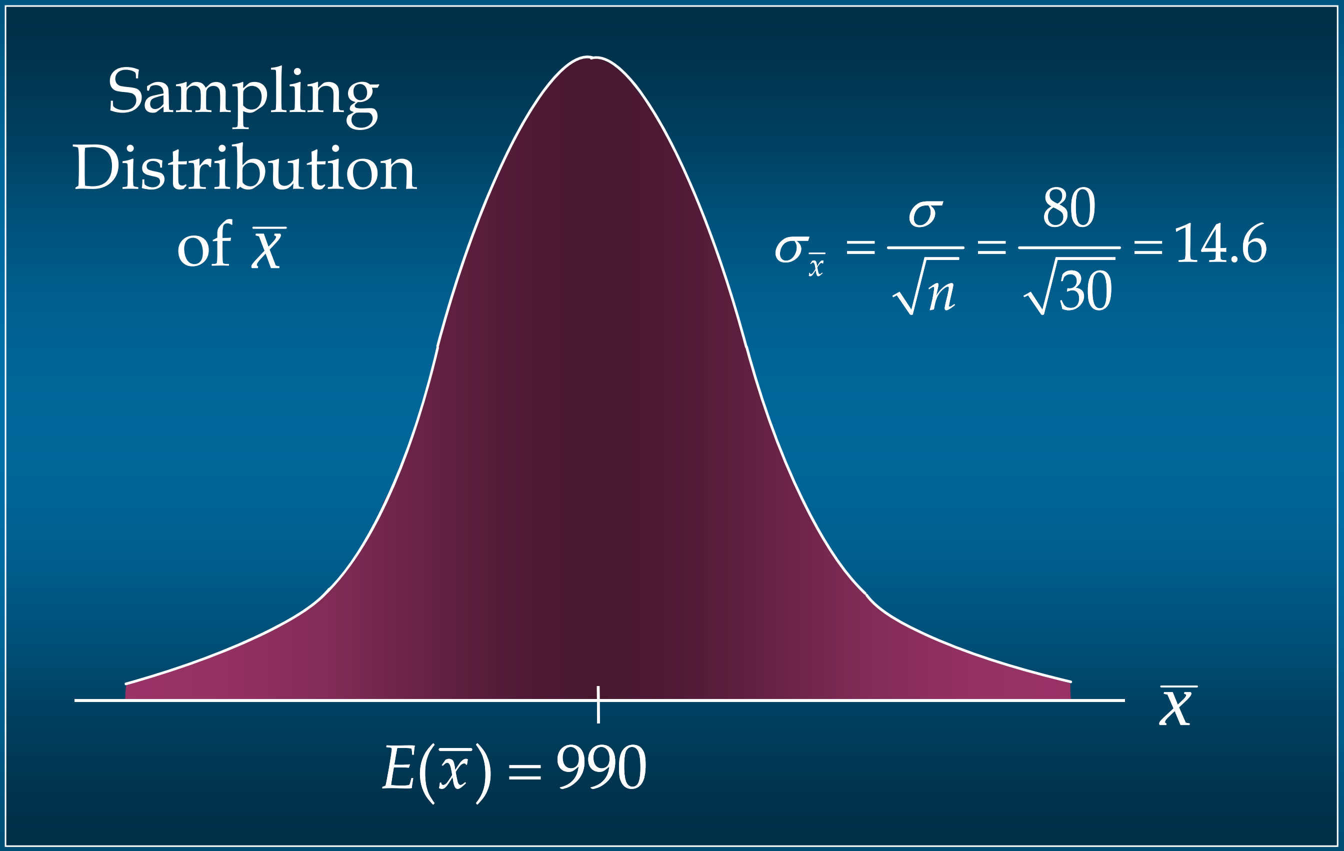
- A finite population is treated as being infinite if $n/N \leq .05$.
- $\sqrt{(N-n)/(N-1)}$ is the finite correction factor.
- $\sigma_{\bar{x}}$ is referred to as the standard error of the mean.

Form of the Sampling Distribution of $\bar{x}$

- If we use a large ($n \geq 30$) simple random sample, the central limit theorem enables us to conclude that the sampling distribution of $\bar{x}$ can be approximated by a normal distribution.

- When the simple random sample is small ($n < 30$), the sampling distribution of $\bar{x}$ can be considered normal only if we assume the population has a normal distribution.

Sampling Distribution of $\bar{x}$ for SAT Scores



Sampling Distribution of $\bar{x}$ for SAT Scores

With a mean SAT score of 990 and a standard deviation of 80, what is the probability that a simple random sample of 30 applicants will provide an estimate of the population mean SAT score that is within +/–10 of the actual population mean μ?

In other words, what is the probability that $\bar{x}$ will be between 980 and 1000?

Sampling Distribution of $\bar{x}$ for SAT Scores

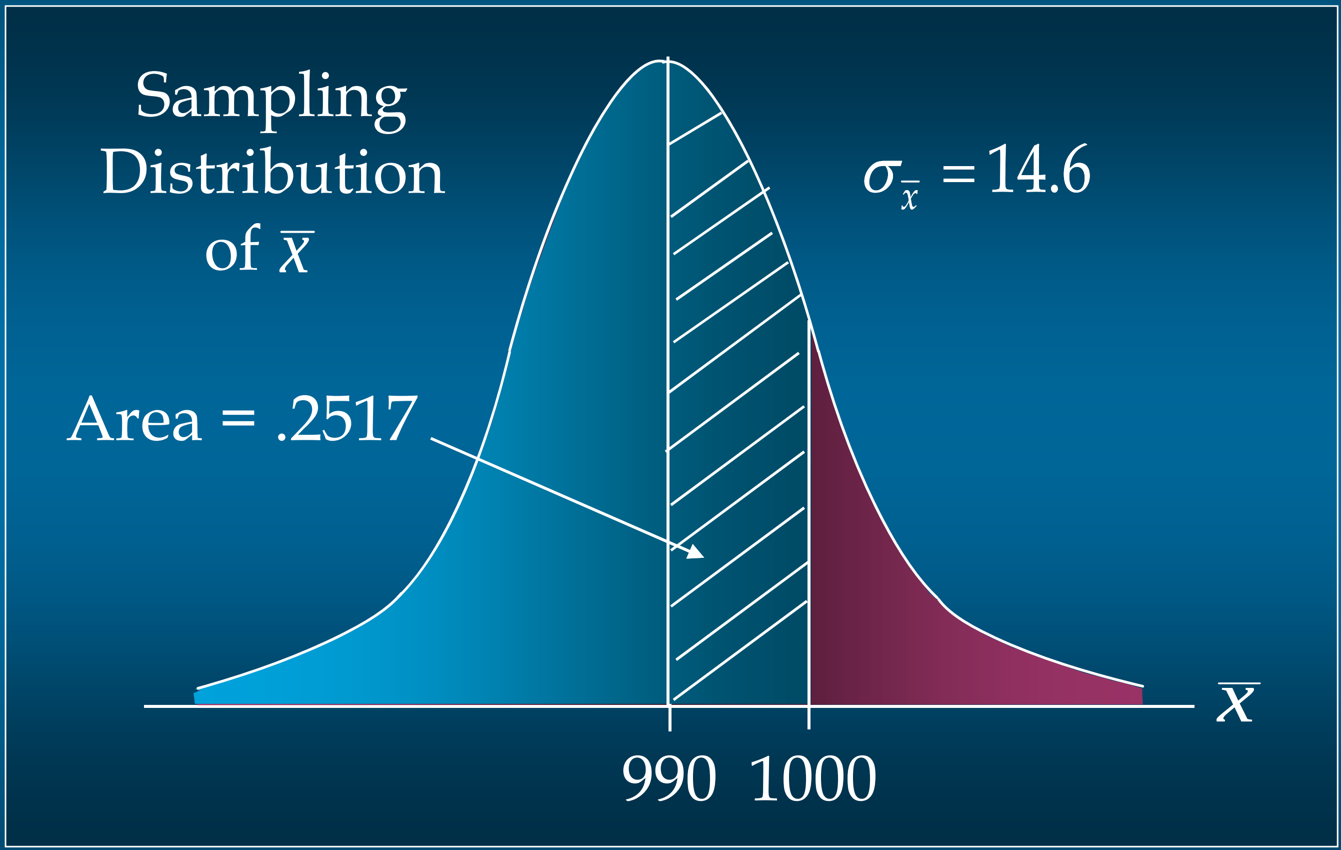
- Step 1: Calculate the *z*-value at the upper endpoint of the interval.
  - $z = (1000 - 990)/14.6 = .68$
- Step 2: Find the area under the curve between the mean and the upper endpoint.
  - .2517

Sampling Distribution of $\bar{x}$ for SAT Scores

Probabilities for the Standard Normal Distribution

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.	.	.	.	.	.	.	.	.	.	.
.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
.	.	.	.	.	.	.	.	.	.	.

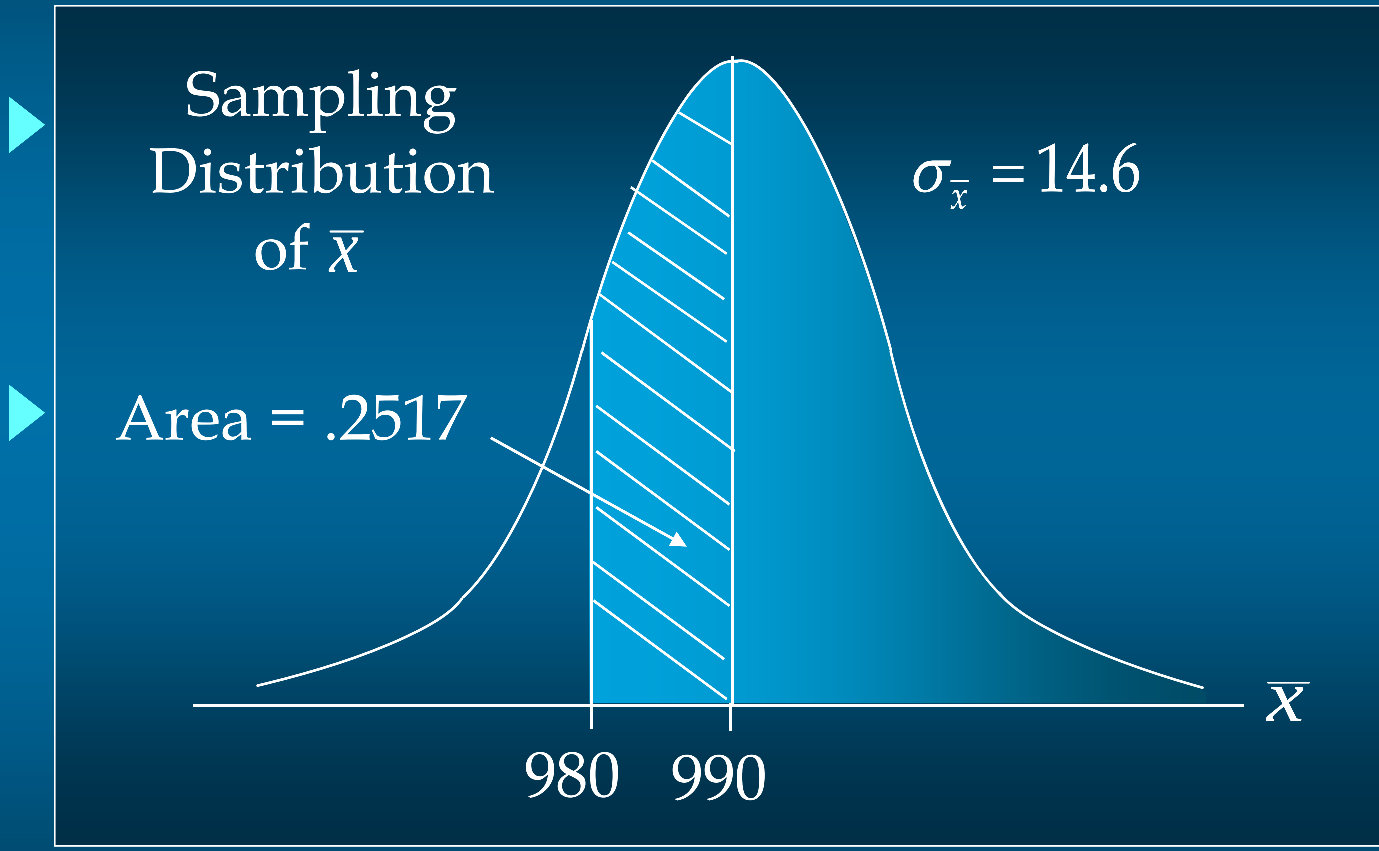
Sampling Distribution of $\overline{x}$ for SAT Scores



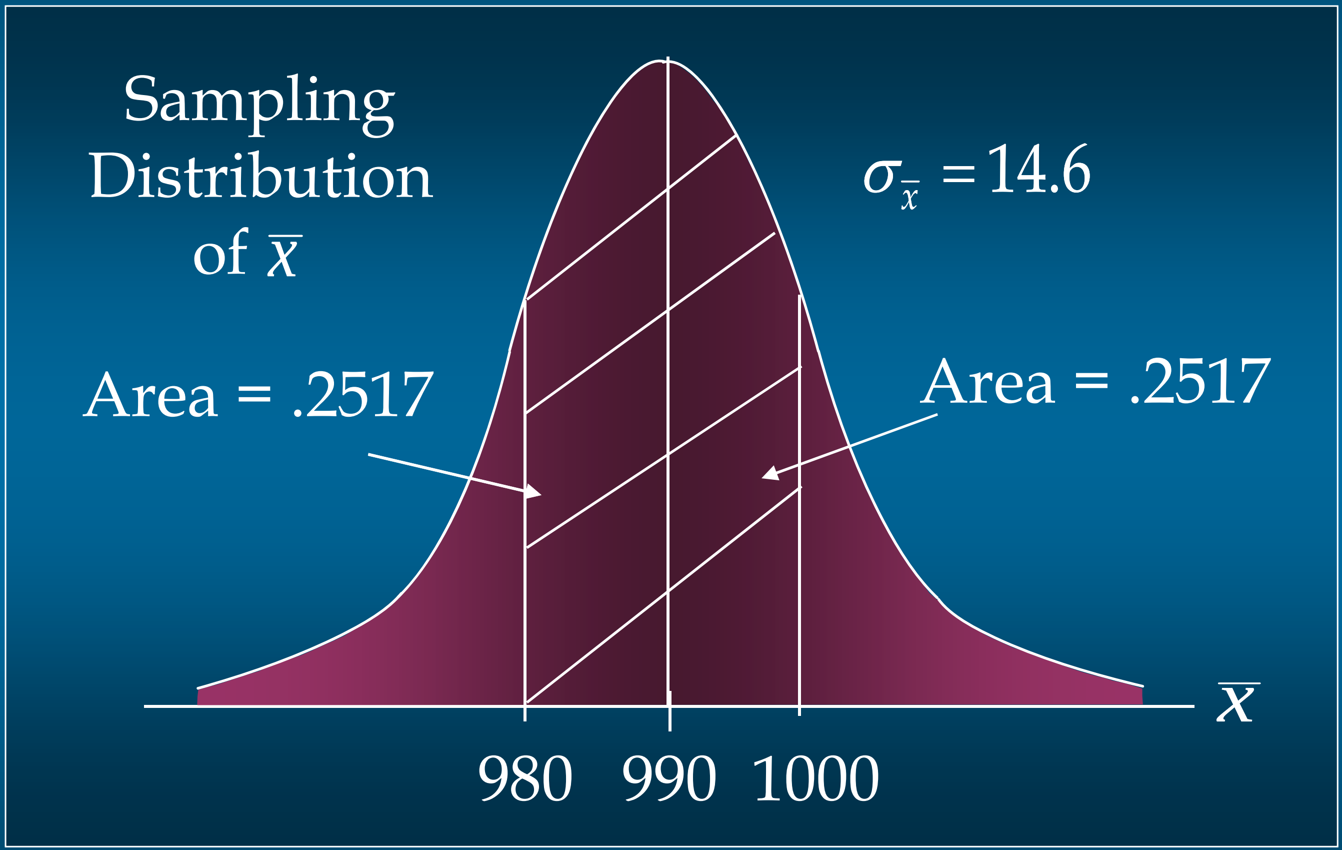
Sampling Distribution of $\overline{x}$ for SAT Scores

- Step 3: Calculate the *z*-value at the lower endpoint of the interval.
  - z = (980 0 990)/14.6= - .68
- Step 4: Find the area under the curve ehwz hhq wkh p hdq and the lower endpoint.
  - = .2517

Sampling Distribution of $\bar{x}$ for SAT Scores



Sampling Distribution of $\bar{x}$ for SAT Scores



Sampling Distribution of $\bar{x}$ for SAT Scores

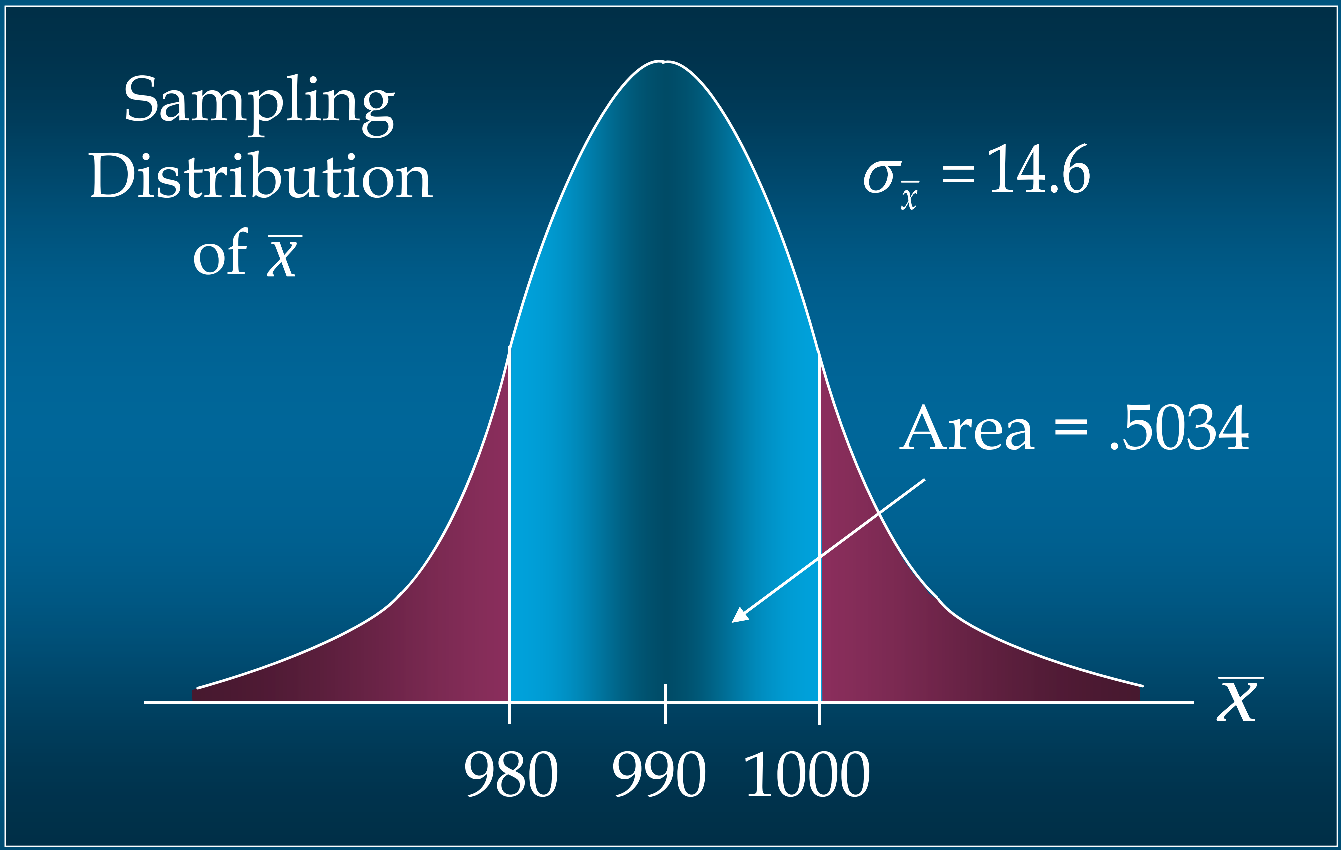
- Step 5: Calculate the area under the curve between the lower and upper endpoints of the interval.
  - $P(-.68 \leq z \leq .68) =$

$$= .2517 \;.\; .2517$$
$$= .5034$$

The probability that the sample mean SAT score will be between 980 and 1000 is:

$$P(980 \leq \bar{x} \leq 1000) = .5034$$

Sampling Distribution of $\bar{x}$ for SAT Scores

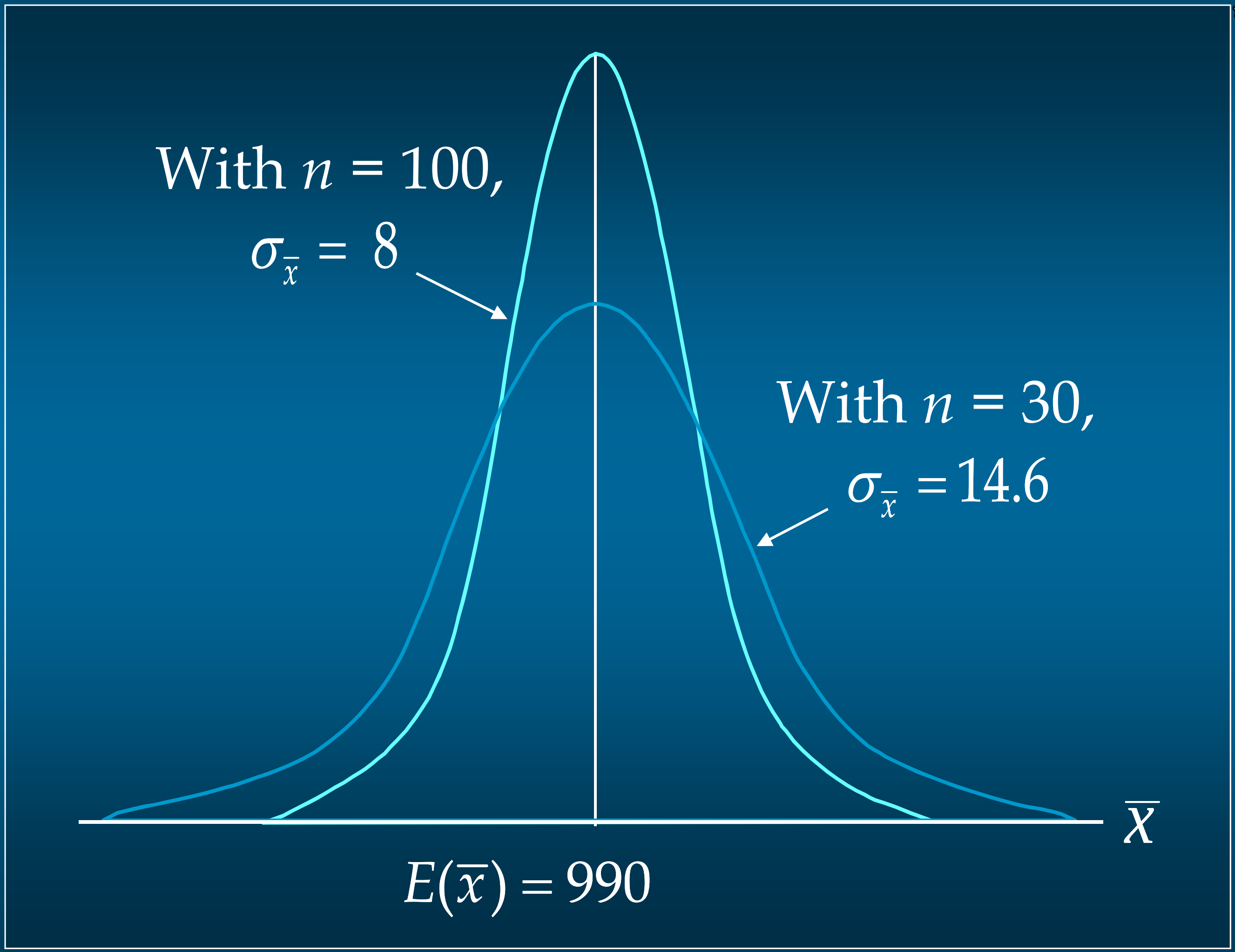


Relationship Between the Sample Size and the Sampling Distribution of $\bar{x}$

- Suppose we select a simple random sample of 100 applicants instead of the 30 originally considered.
- $E(\bar{x}) = \mu$ regardless of the sample size. In our example, $E(\bar{x})$ remains at 990.
- 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 is decreased to:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{80}{\sqrt{100}} = 8.0$$

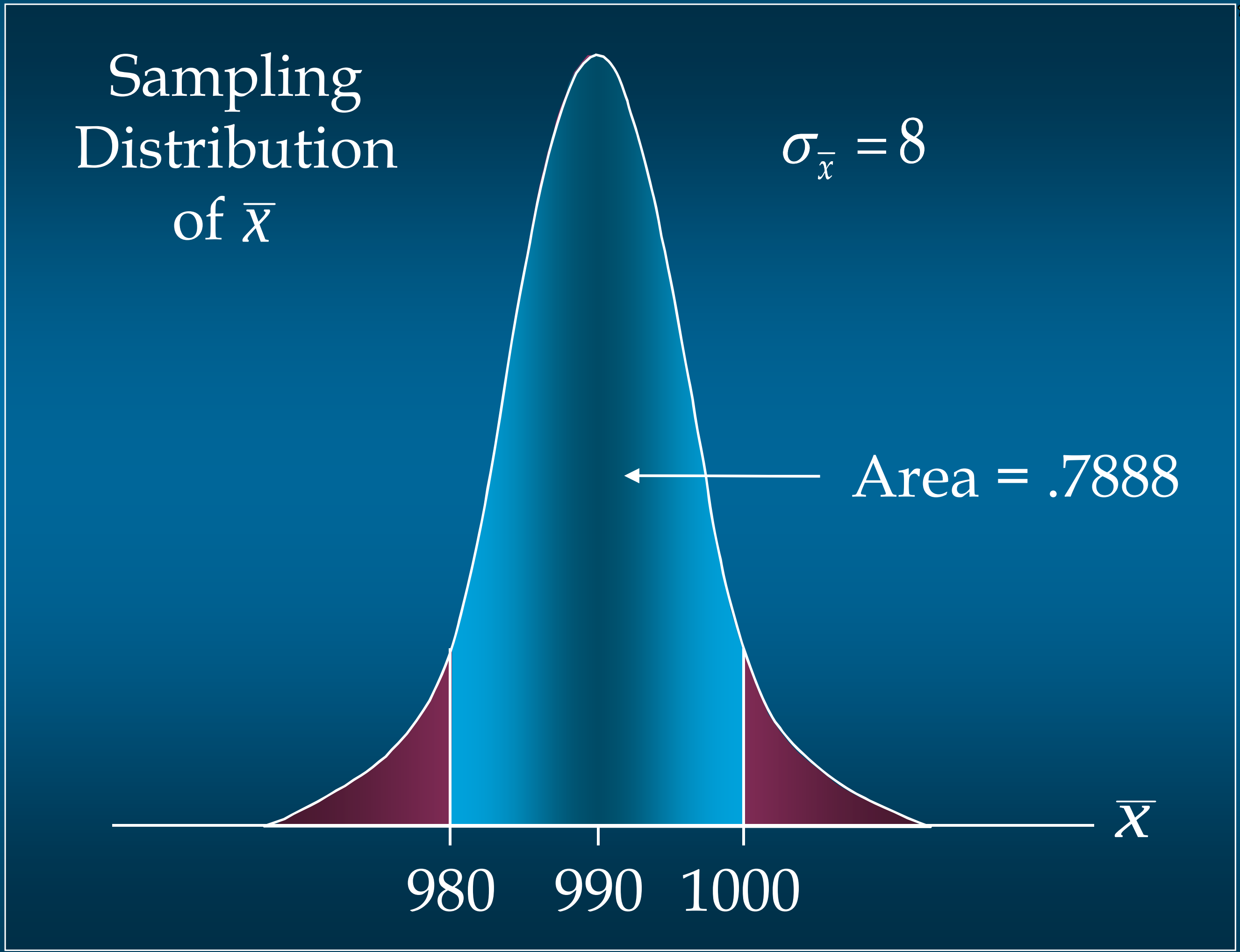
Relationship Between the Sample Size and the Sampling Distribution of $\bar{x}$



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

- Recall that when $n = 30$, $P(980 \le \bar{x} \le 1000) = .5034$.
- We follow the same steps to solve for $P(980 \le \bar{x} \le 1000)$ when $n = 100$ as we showed earlier when $n = 30$.
- Now, with $n = 100$, $P(980 \le \bar{x} \le 1000) = .7888$.
- 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 $n = 30$.

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

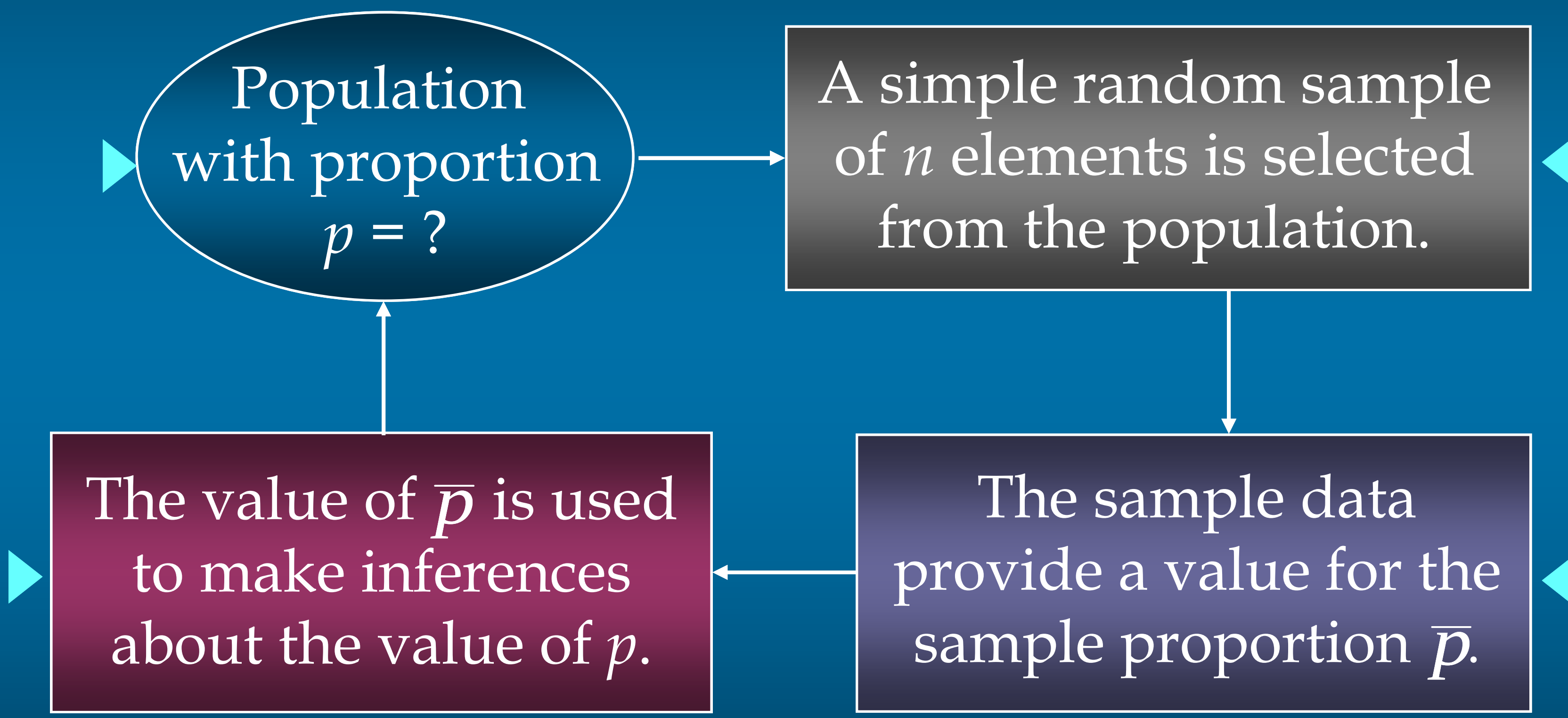


Chapter 7
Sampling and Sampling Distributions

- Sampling Distribution of $\bar{p}$
- Other Sampling Methods

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

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

Infinite Population

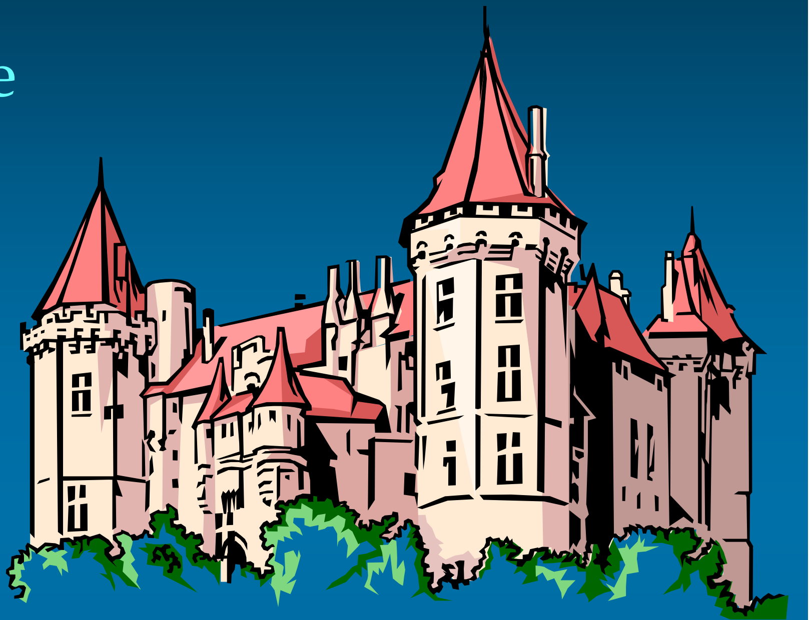
$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}$$

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

- A finite population is treated as being infinite if $n/N \leq .05$.

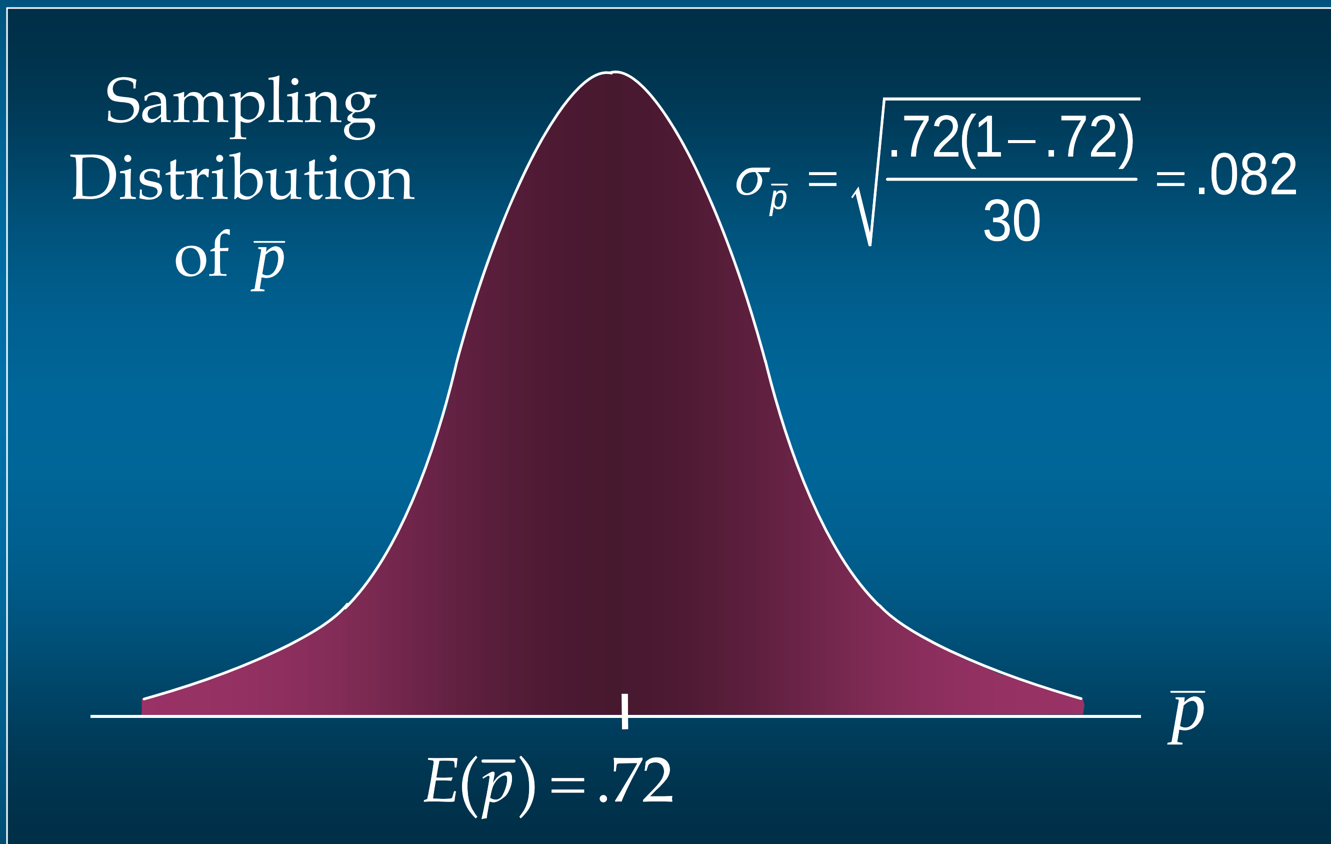
Sampling Distribution of $\bar{p}$

- Example: St. Andrew's College
  - Recall that 72% of the prospective students applying to St. Andrew's College desire on-campus housing.



  - What is the probability that a simple random sample of 30 applicants will provide an estimate of the population proportion of applicant desiring on-campus housing that is within plus or minus .05 of the actual population proportion?

Sampling Distribution of $\bar{p}$



Sampling Distribution of $\bar{p}$

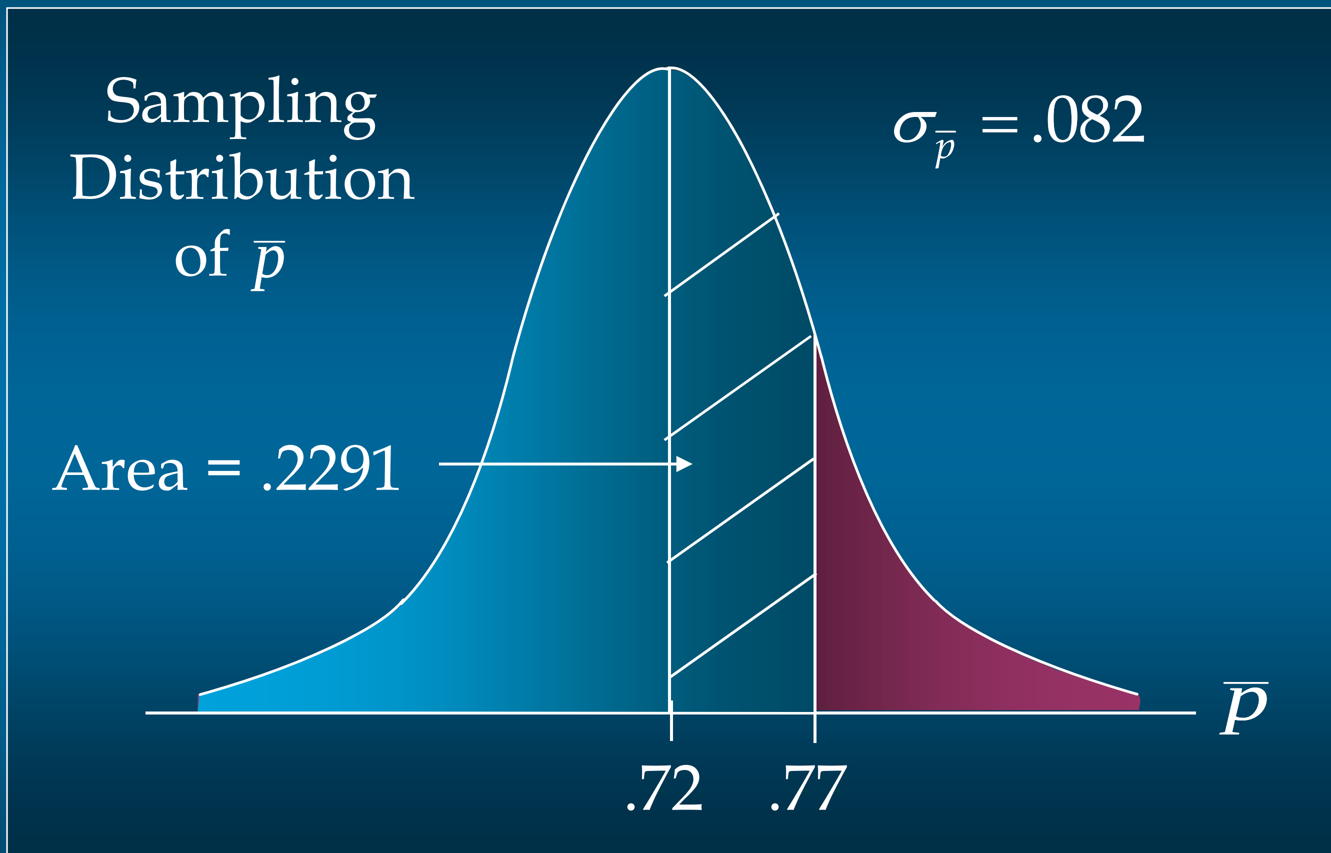
- Step 1: Calculate the z-value at the upper endpoint of the interval.
  - z = (.77 0 .72)/.082 = .61
- Step 2: Find the area under the curve ehwz hhq wkh p hdq and upper endpoint.
  - .2291

Sampling Distribution of $\bar{p}$

Probabilities for
the Standard Normal Distribution

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.	.	.	.	.	.	.	.	.	.	.
.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
.	.	.	.	.	.	.	.	.	.	.

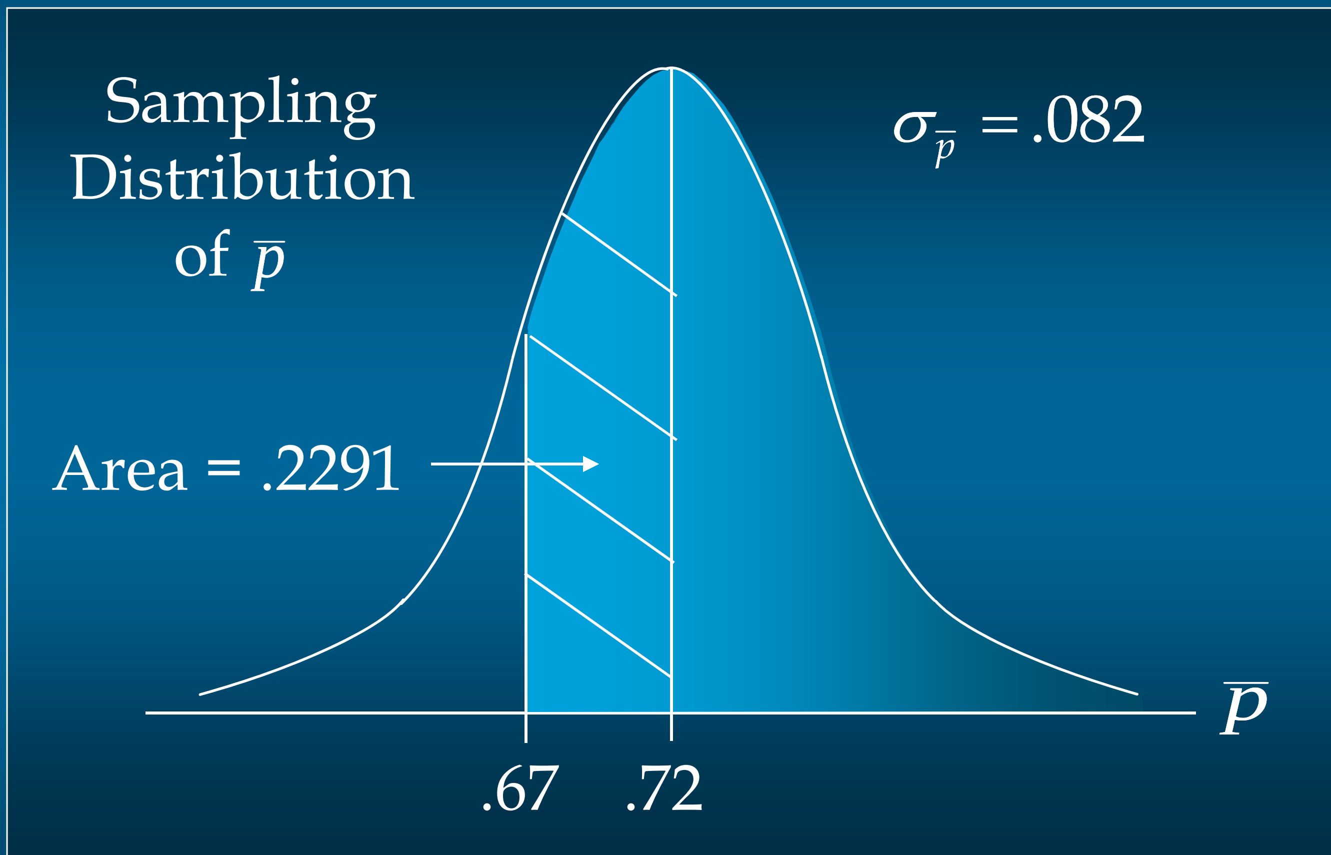
Sampling Distribution of $\bar{p}$



Sampling Distribution of $\bar{p}$

- Step 3: Calculate the *z*-value at the <u>lower</u> endpoint of the interval.
  - *z* = (.67 0 .72)/.082 = - .61
- Step 4: Find the area under the curve ehwz hhq wkh p hdq and the <u>lower</u> endpoint.
  - .2291

Sampling Distribution of $\bar{p}$



Sampling Distribution of $\bar{p}$

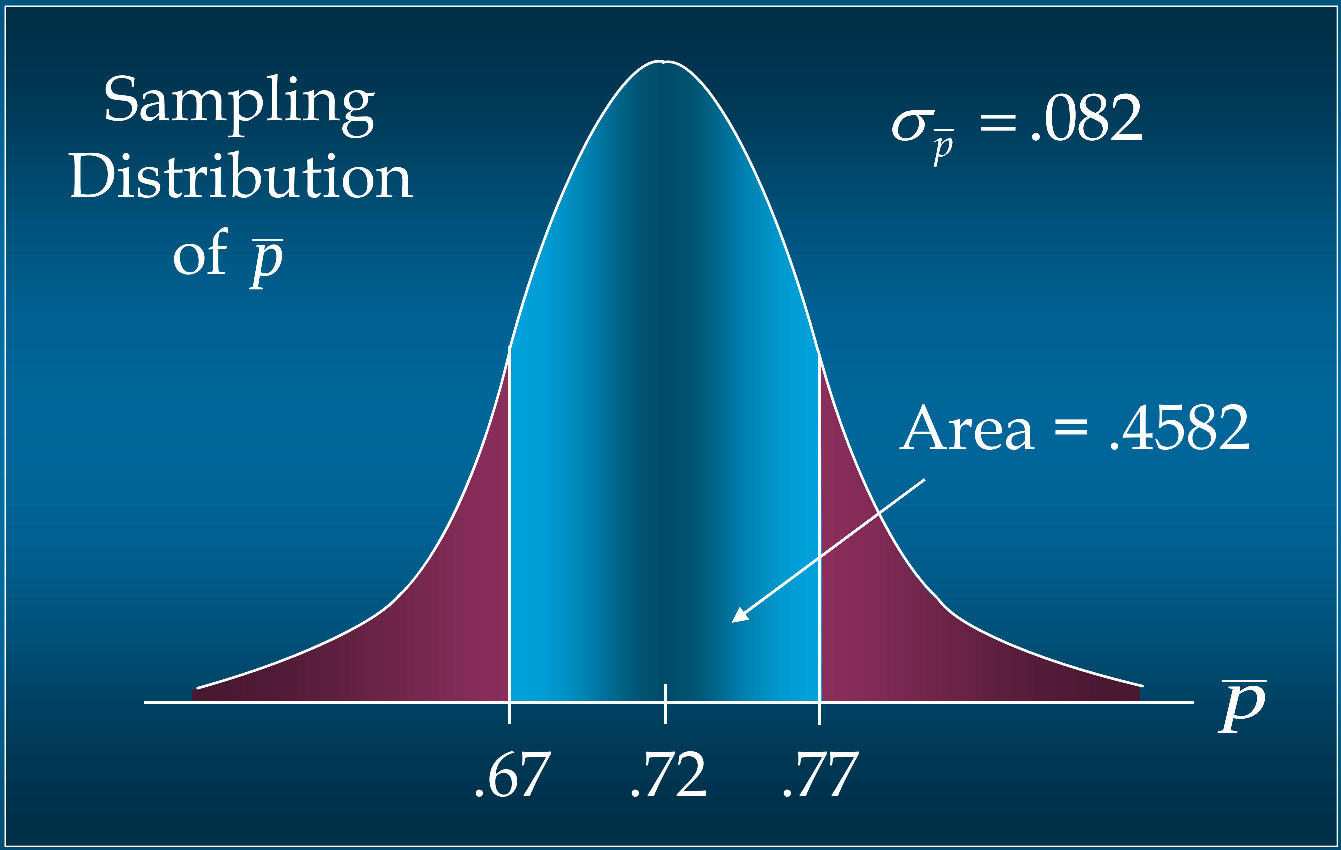
- Step 5: Calculate the area under the curve between the lower and upper endpoints of the interval.
  - $P(-.61 \leq z \leq .61) =$

$$= .2291 \;.\; .2291$$
$$= .4582$$

The probability that the sample proportion of applicants wanting on-campus housing will be within +/-.05 of the actual population proportion :

$$P(.67 \leq \bar{p} \leq .77) = .4582$$

Sampling Distribution of $\bar{p}$



Other Sampling Methods

- Stratified Random Sampling
- Cluster Sampling
- Systematic Sampling
- Convenience Sampling
- Judgment Sampling

Stratified Random Sampling

- The population is first divided into groups of elements called strata.
- Each element in the population belongs to one and only one stratum.
- Best results are obtained when the elements within each stratum are as much alike as possible (i.e. a homogeneous group).

Stratified Random Sampling

- A simple random sample is taken from each stratum.
- Formulas are available for combining the stratum sample results into one population parameter estimate.
- <u>Advantage</u>: If strata are homogeneous, this method is as "precise" as simple random sampling but with a smaller total sample size.
- <u>Example</u>: The basis for forming the strata might be department, location, age, industry type, and so on.

Cluster Sampling

- The population is first divided into separate groups of elements called clusters.
- Ideally, each cluster is a representative small-scale version of the population (i.e. heterogeneous group).
- A simple random sample of the clusters is then taken.
- All elements within each sampled (chosen) cluster form the sample.

Cluster Sampling

- Example: A primary application is area sampling, where clusters are city blocks or other well-defined areas.
- Advantage: The close proximity of elements can be cost effective (i.e. many sample observations can be obtained in a short time).
- Disadvantage: This method generally requires a larger total sample size than simple or stratified random sampling.

Systematic Sampling

- If a sample size of n is desired from a population containing N elements, we might sample one element for every n/N elements in the population.
- We randomly select one of the first n/N elements from the population list.
- We then select every n/Nth element that follows in the population list.

Systematic Sampling

- This method has the properties of a simple random sample, especially if the list of the population elements is a random ordering.
- Advantage: The sample usually will be easier to identify than it would be if simple random sampling were used.
- Example: Selecting every 100^{th} listing in a telephone book after the first randomly selected listing

Convenience Sampling

- It is a nonprobability sampling technique. Items are included in the sample without known probabilities of being selected.
- The sample is identified primarily by convenience.
- Example: A professor conducting research might use student volunteers to constitute a sample.

Convenience Sampling

- Advantage: Sample selection and data collection are relatively easy.
- Disadvantage: It is impossible to determine how representative of the population the sample is.

Judgment Sampling

- The person most knowledgeable on the subject of the study selects elements of the population that he or she feels are most representative of the population.
- It is a nonprobability sampling technique.
- Example: A reporter might sample three or four senators, judging them as reflecting the general opinion of the senate.

Judgment Sampling

- Advantage: It is a relatively easy way of selecting a sample.
- Disadvantage: The quality of the sample results depends on the judgment of the person selecting the sample.