Statistics for Business and Economics

Chapter 6

Sampling and Sampling Distributions

Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

6.1 Tools of Business Statistics

- **Descriptive statistics**
  - Collecting, presenting, and describing data
- **Inferential statistics**
  - Drawing conclusions and/or making decisions concerning a population based only on sample data

Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Populations and Samples

- A **Population** is the set of all items or individuals of interest

▪	Examples:	All likely voters in the next election All parts produced today All sales receipts for November

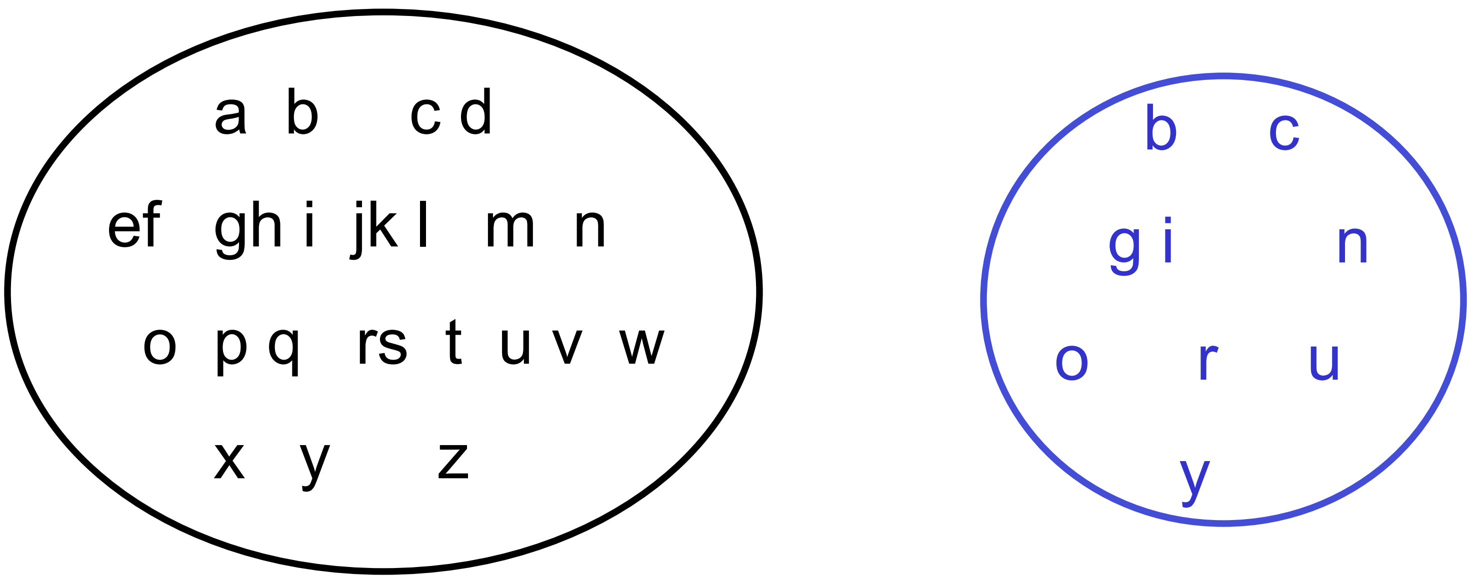
- A **Sample** is a subset of the population

▪	Examples:	1000 voters selected at random for interview A few parts selected for destructive testing Random receipts selected for audit

Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Population vs. Sample

Population **Sample**



Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

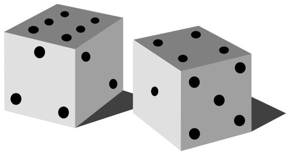
Why Sample?

- Less time consuming than a census
- Less costly to administer than a census
- It is possible to obtain statistical results of a sufficiently high precision based on samples.

Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Simple Random Samples

- Every object in the population has an equal chance of being selected
- Objects are selected independently
- Samples can be obtained from a table of random numbers or computer random number generators



- A simple random sample is the ideal against which other sample methods are compared

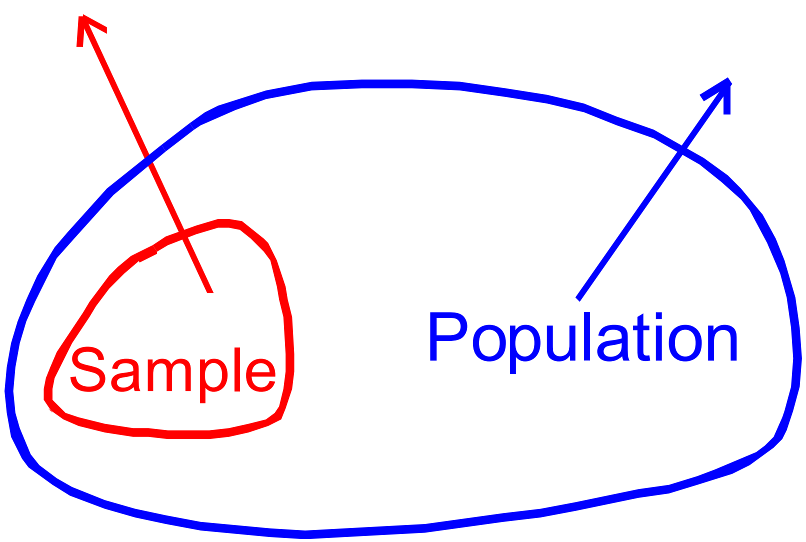
Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Inferential Statistics

- Making statements about a population by examining sample results

Sample statistics → Population parameters

(known) **Inference** (unknown, but can be estimated from sample evidence)



Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Inferential Statistics

Drawing conclusions and/or making decisions concerning a population based on sample results.

- **Estimation**
  - e.g., Estimate the population mean weight using the sample mean weight
- **Hypothesis Testing**
  - e.g., Use sample evidence to test the claim that the population mean weight is 120 pounds

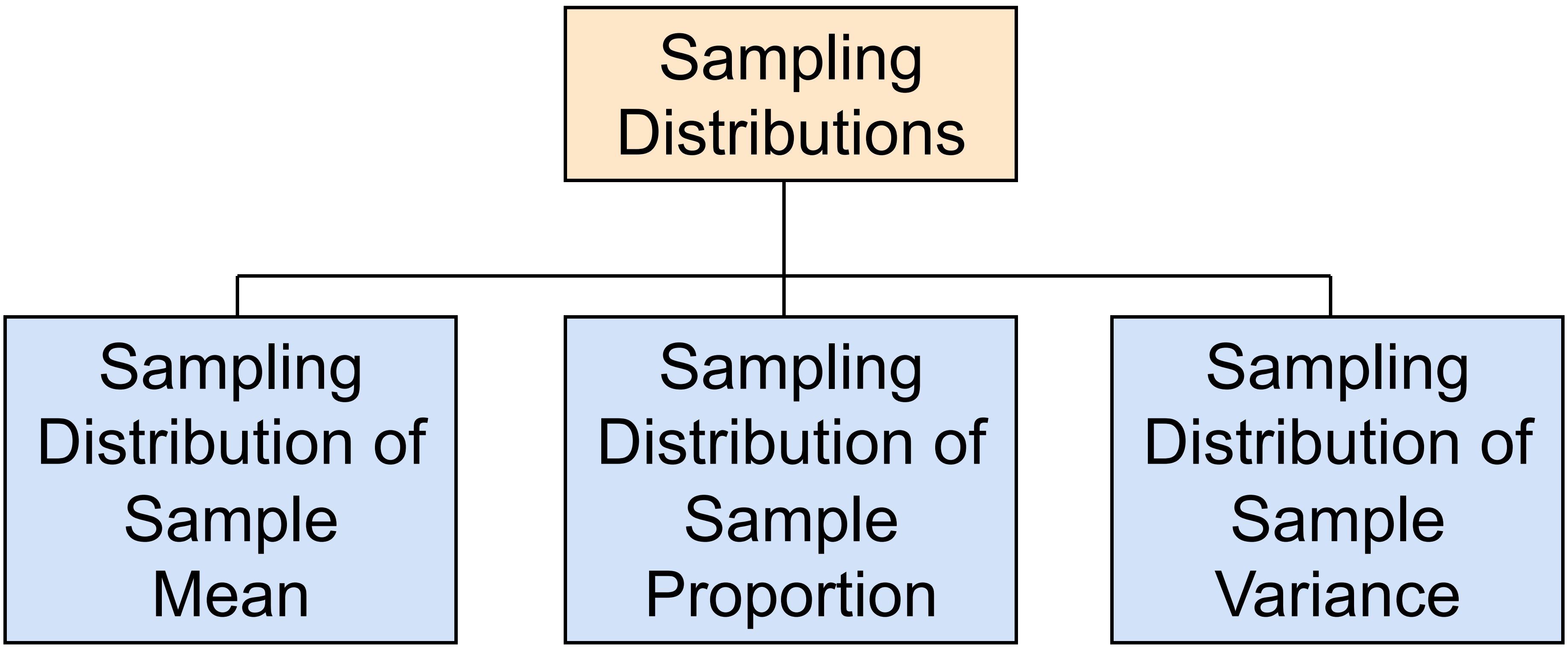
Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Sampling Distributions

- A sampling distribution is a distribution of all of the possible values of a statistic for a given size sample selected from a population

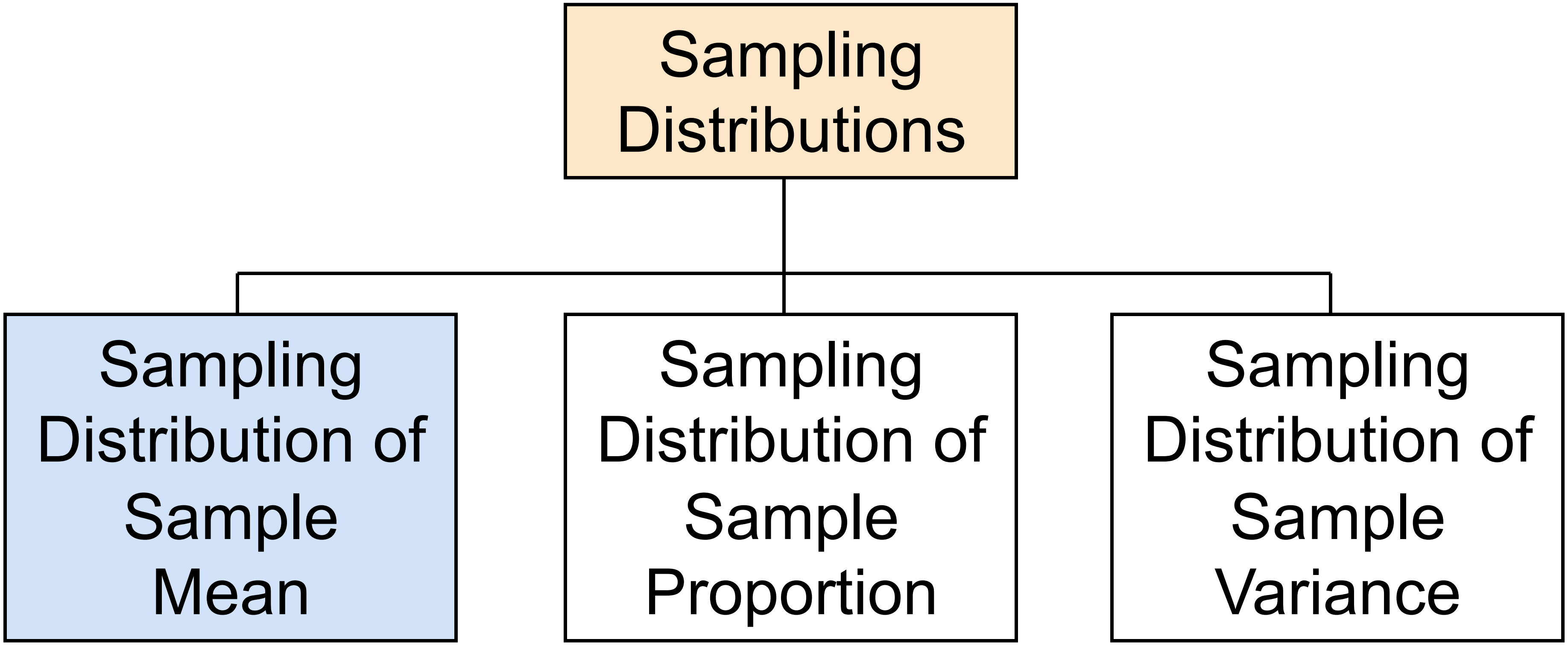
Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Chapter Outline



Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Sampling Distributions of Sample Means



Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Developing a Sampling Distribution

- **Assume there is a population …**
- Four types of people
- Random variable, X, is age of individuals
- Possible Values of X: 18, 20, 22, 24 (years)

Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

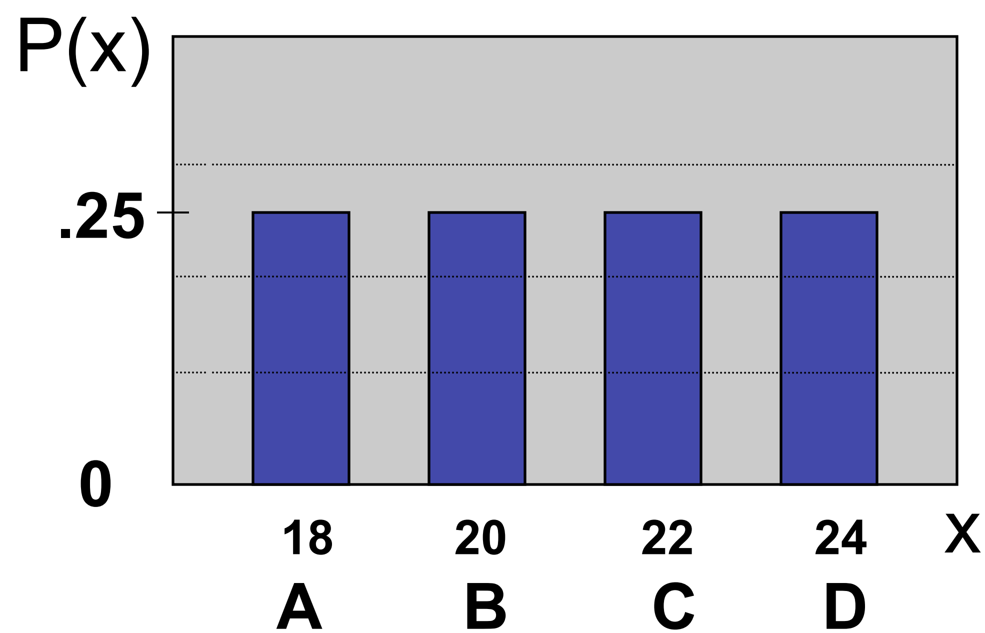
Developing a Sampling Distribution

(continued)

Summary Measures for the Population Distribution:

$$\mu = \frac{18+20+22+24}{4} = 21$$

$$\sigma^2 = \sqrt{\frac{\sum (X_i - \mu)^2}{4}} = 2.236$$



Uniform Distribution

Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Developing a Sampling Distribution

(continued)

Now consider all possible samples of size n = 2

1st Obs	2nd Observation			
	18	20	22	24
18	18,18	18,20	18,22	18,24
20	20,18	20,20	20,22	20,24
22	22,18	22,20	22,22	22,24
24	24,18	24,20	24,22	24,24

16 possible samples (sampling with replacement)

16 Sample Means

1st Obs	2nd Observation			
	18	20	22	24
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Developing a Sampling Distribution

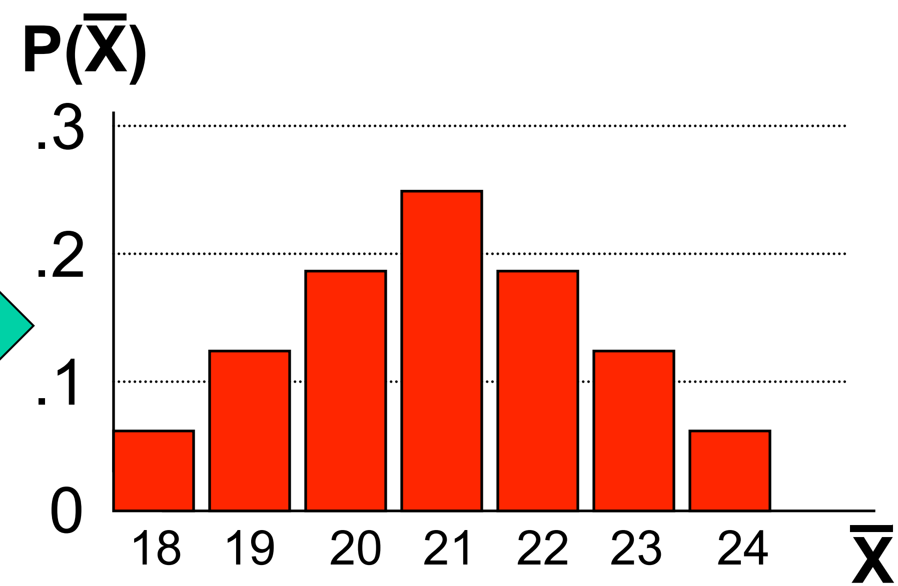
(continued)

Sampling Distribution of All Sample Means

16 Sample Means

1st Obs	2nd Observation			
	18	**20**	**22**	**24**
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

Sample Means Distribution



(no longer uniform)

Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Developing a Sampling Distribution

(continued)

Summary Measures of this Sampling Distribution:

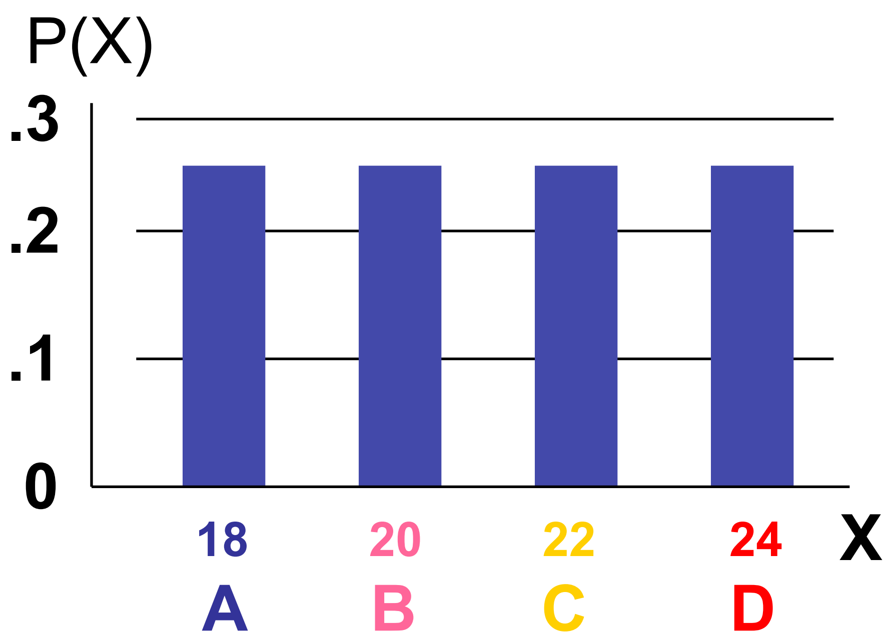
$$\mu_{\overline{X}} = \frac{18 + 2 \times 19 + 3 \times 20 + \cdots + 2 \times 23 + 24}{16} = 21$$

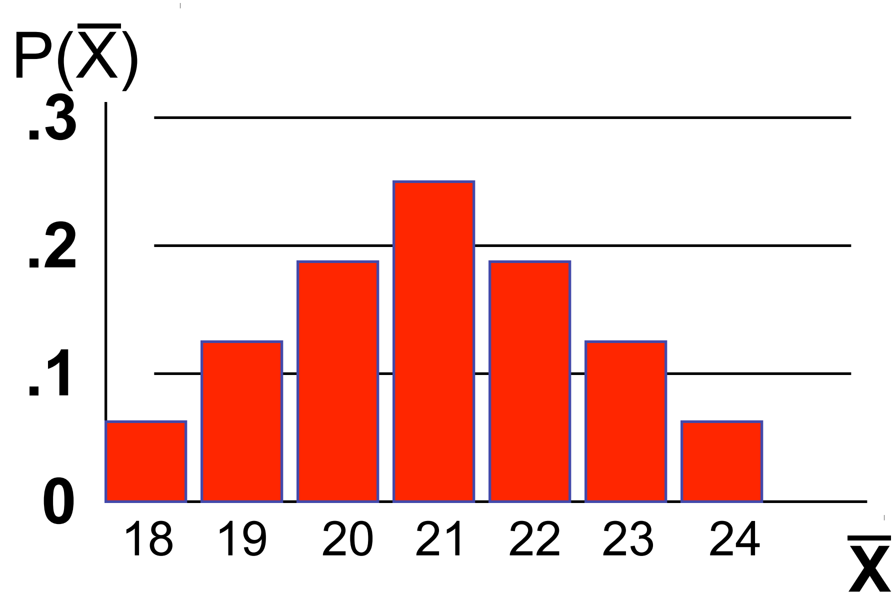
$$\sigma_{\overline{X}} = \sqrt{\frac{(18\text{-}21)^2 + 2 \times (19\text{-}21)^2 + \cdots + (24\text{-}21)^2}{16}} = 1.58$$

Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Comparing the Population with its Sampling Distribution

Population	Sample Means Distribution n = 2
$\mu = 21 \quad \sigma = 2.236$	$\mu_{\bar{X}} = 21 \quad \sigma_{\bar{X}} = 1.58$





Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Expected Value of Sample Mean

- Let X_1, X_2, . . . X_n represent a random sample from a population

- The sample mean value of these observations is defined as

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

Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Standard Error of the Mean

- Different samples of the same size from the same population will yield different sample means
- A measure of the variability in the mean from sample to sample is given by the Standard Error of the Mean:

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

- Note that the standard error of the mean decreases as the sample size increases
- If n=2, then $\sigma / \sigma_{\bar{X}} = \sqrt{2} = 1.4142$

Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

If the Population is Normal

- If a population is normal with mean μ and standard deviation σ, the sampling distribution of $\bar{X}$ is also normally distributed with

$$\mu_{\bar{X}} = \mu$$

and

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

Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Z-value for Sampling Distribution of the Mean

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

$$Z = \frac{(\bar{X} - \mu)}{\sigma_{\bar{X}}}$$

where:

$\bar{X}$ = sample mean

μ = population mean

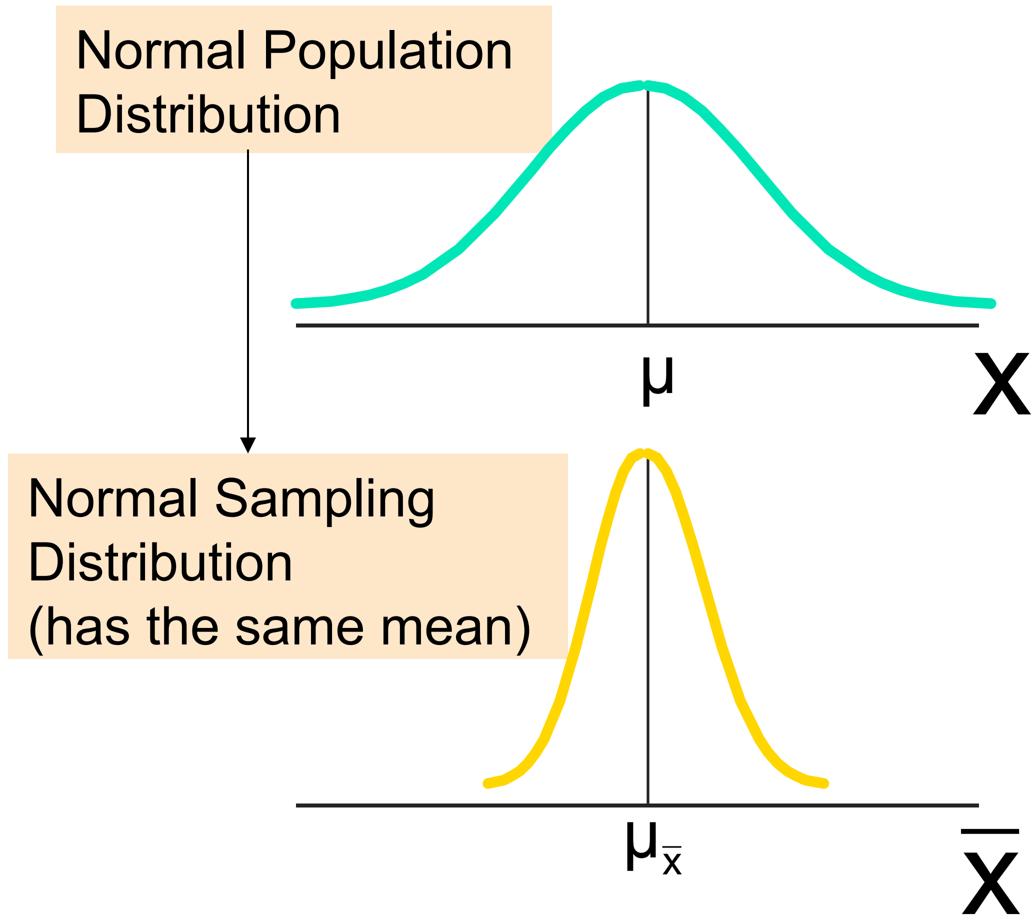
$\sigma_{\bar{x}}$ = standard error of the mean

Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Sampling Distribution Properties

$$\mu_{\bar{X}} = \mu$$

(i.e. $\bar{X}$ is unbiased)



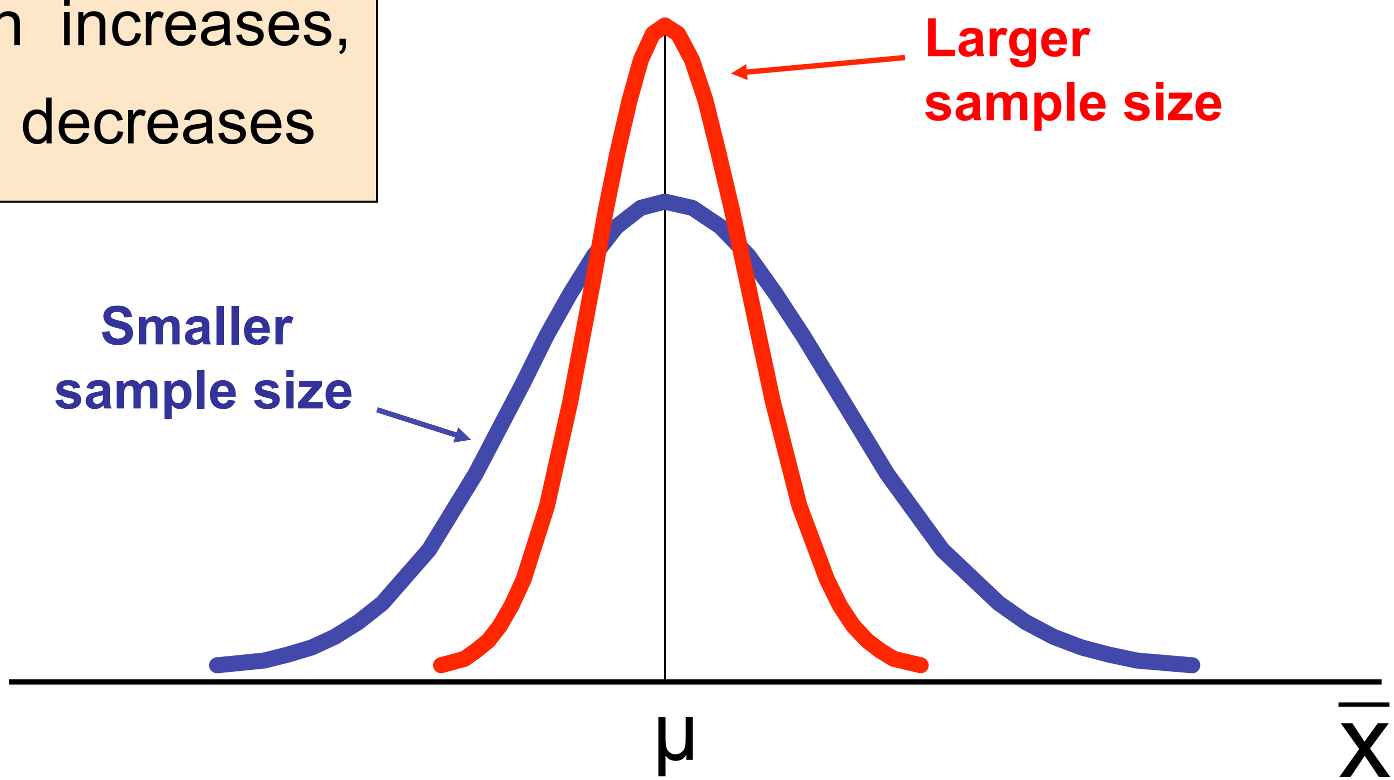
Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Sampling Distribution Properties

(continued)

- For sampling with replacement:

As n increases, $\sigma_{\bar{x}}$ decreases



Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

If the Population is **not** Normal

- We can apply the Central Limit Theorem:
  - Even if the population is not normal,
  - …sample means from the population will be approximately normal as long as the sample size is large enough.

Properties of the sampling distribution:

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

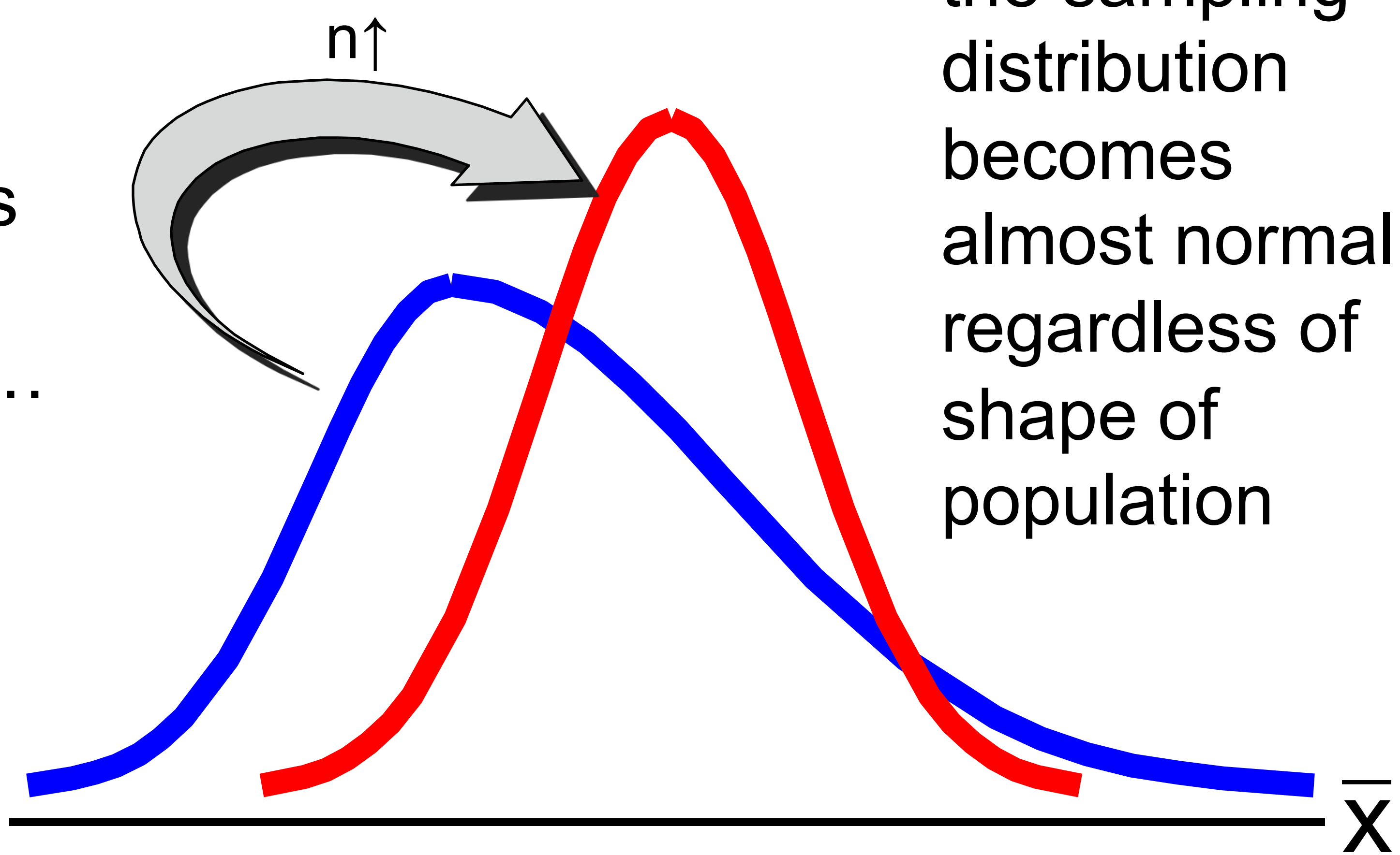
and

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

Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Central Limit Theorem

As the sample size gets large enough…



the sampling distribution becomes almost normal regardless of shape of population

Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

If the Population is **not** Normal

(continued)

Sampling distribution properties:

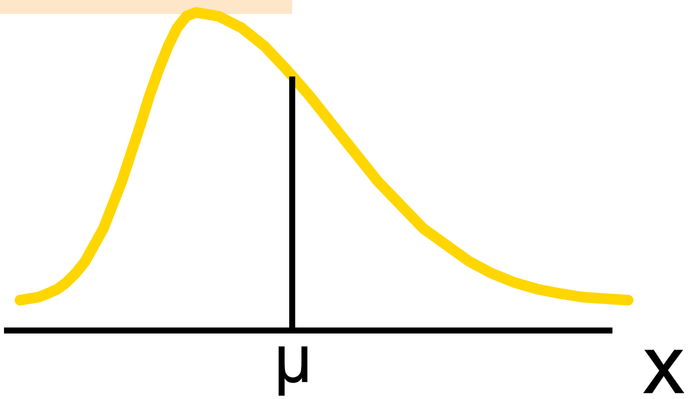
Central Tendency

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

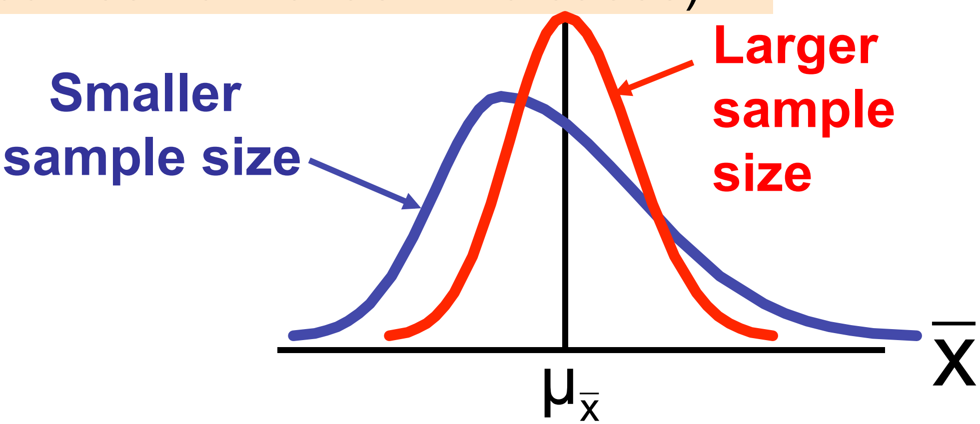
Variation

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

Population Distribution



Sampling Distribution
(becomes normal as n increases)



Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Example

- Suppose a large population has mean $\mu = 8$ and standard deviation $\sigma = 3$. 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?

Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Example

(continued)

Solution:

- Even though the population is not normally distributed, we use the central limit theorem to get an approximated solution
- … the sampling distribution of $\bar{X}$ is approximately normal
- … with mean $\mu_{\bar{x}} = 8$
- …and standard deviation $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{36}} = 0.5$

Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

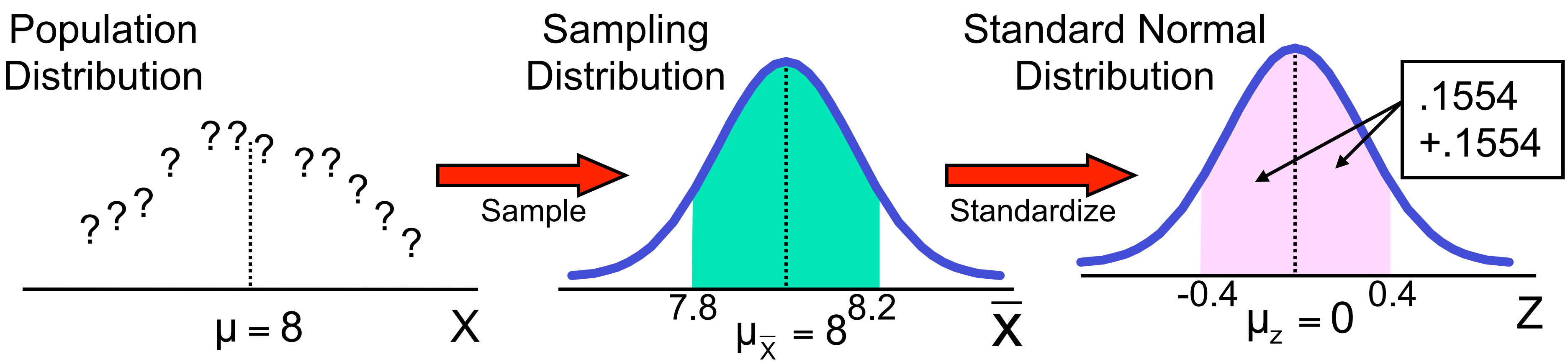
Example

(continued)

Solution (continued):

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

$$= P(-0.4 < Z < 0.4) = 0.3108$$



Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Acceptance Intervals

- Goal: determine a range within which sample means are likely to occur, given a population mean and variance
  - By the Central Limit Theorem, we know that the distribution of $\bar{X}$ is approximately normal if n is large enough, with mean μ and standard deviation $\sigma_{\bar{X}}$
  - Let $z_{\alpha/2}$ be the z-value that leaves area α/2 in the upper tail of the normal distribution (i.e., the interval $-z_{\alpha/2}$ to $z_{\alpha/2}$ encloses probability 1 – α)
  - Then

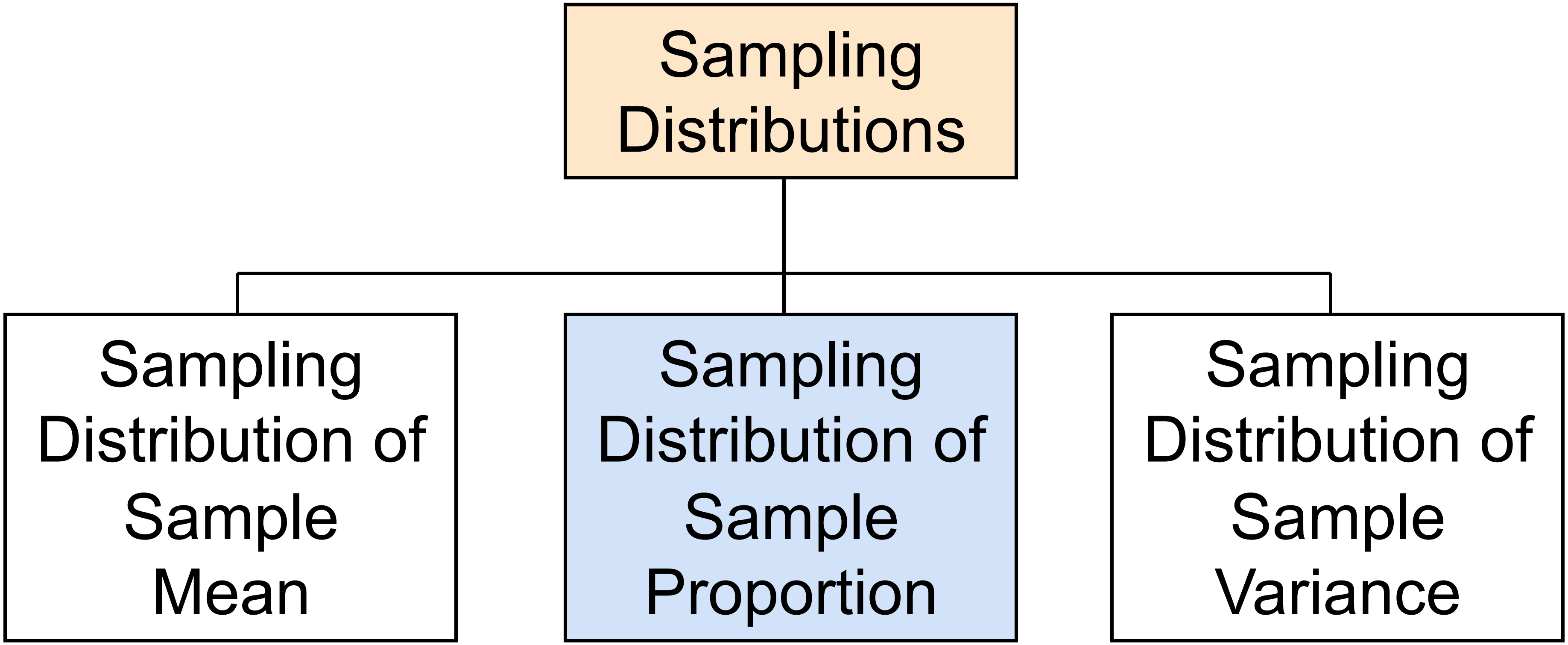
$$\mu \pm z_{\alpha/2}\sigma_{\bar{X}}$$

is the interval that includes $\bar{X}$ with probability 1 – α

Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Sampling Distributions of Sample Proportions

6.3



Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Sampling Distributions of Sample Proportions

p = the proportion of the population having some characteristic

- Sample proportion ($\hat{p}$) provides an estimate of p:

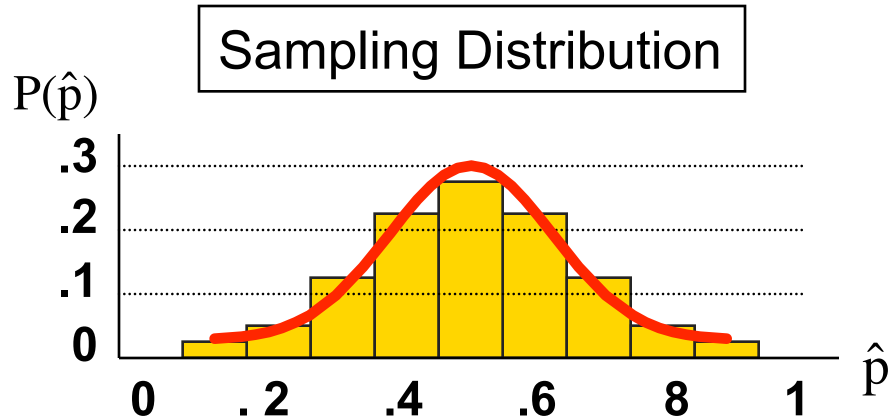
$$\hat{p} = \frac{\text{number of items in the sample having the characteristic of interest}}{\text{sample size}}$$

- $0 \leq \hat{p} \leq 1$
- $\hat{p}$ has a binomial distribution, but can be approximated by a normal distribution when n is large

Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Sampling Distribution of $\hat{p}$

- Normal approximation:



Properties:

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

and

$$\sigma^2_{\hat{p}} = \mathrm{Var}(\hat{p}) = \frac{p(1-p)}{n}$$

(where p = population proportion)

Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Z-Value for Proportions

Standardize $\hat{p}$ to a Z value with the formula:

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

Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Example

- If the true proportion of voters who support Proposition A is $p = .4$, what is the probability that a sample of size 200 yields a sample proportion between .40 and .45?

- i.e.: **if $p = .4$ and $n = 200$, what is $P(.40 \le \hat{p} \le .45)$?**

Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Example

(continued)

- **if p = .4 and n = 200, what is $P(.40 \le \hat{p} \le .45)$?**

Find $\sigma_{\hat{p}}$:

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.4(1-.4)}{200}} = .03464$$

Convert to standard normal:

$$P(.40 \le \hat{p} \le .45) = P\left(\frac{.40-.40}{.03464} \le Z \le \frac{.45-.40}{.03464}\right)$$

$$= P(0 \le Z \le 1.44)$$

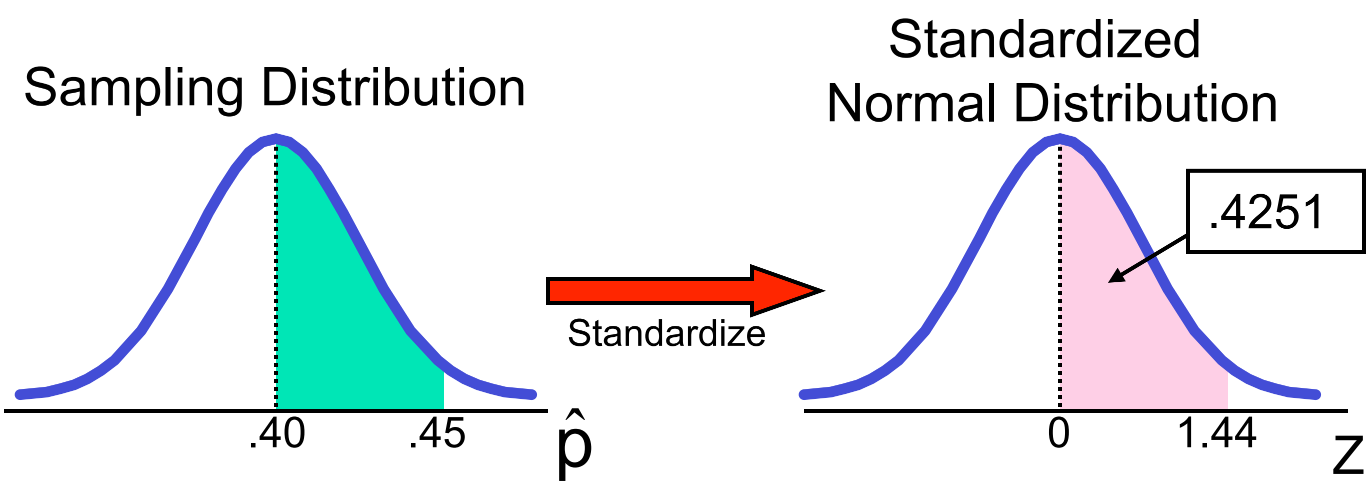
Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Example

(continued)

- **if P = .4 and n = 200, what is $P(.40 \le \hat{p} \le .45)$?**

Use standard normal table: $P(0 \le Z \le 1.44) = .4251$



Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Sampling Distributions of Sample Variance

6.4

Sampling Distributions

- Sampling Distribution of Sample Mean
- Sampling Distribution of Sample Proportion
- Sampling Distribution of Sample Variance

Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Sample Variance

- Let $x_1, x_2, \ldots, x_n$ be a random sample from a population. The sample variance is

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

- the square root of the sample variance is called the sample standard deviation

- the sample variance is different for different random samples from the same population

Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Sampling Distribution of Sample Variances

- The sampling distribution of s^2 has mean σ^2

$$E(s^2) = \sigma^2$$

- If the population distribution is normal, then

$$\mathrm{Var}(s^2) = \frac{2\sigma^4}{n-1}$$

- If the population distribution is normal then

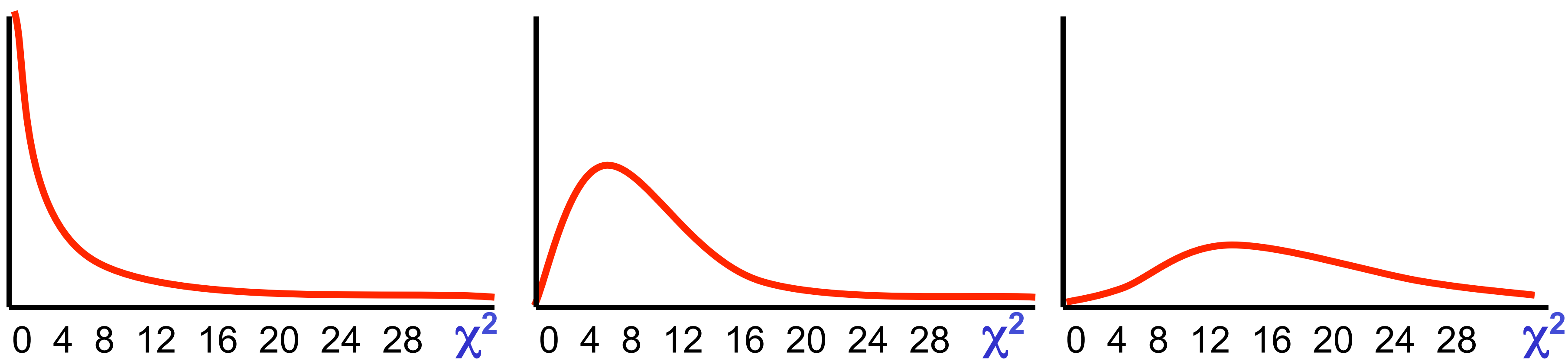
$$\frac{(n-1)s^2}{\sigma^2}$$

has a χ^2 distribution with n – 1 degrees of freedom

Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

The Chi-square Distribution

- The chi-square distribution is a family of distributions, depending on degrees of freedom:
- d.f. = n – 1



d.f. = 1 d.f. = 5 d.f. = 15

- Text Table 7 contains chi-square probabilities

Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Degrees of Freedom (df)

Idea: Number of observations that are free to vary after sample mean has been calculated

Example: Suppose the mean of 3 numbers is 8.0

Let $X_1 = 7$
Let $X_2 = 8$
What is X_3?

→

If the mean of these three values is 8.0,
then X_3 must be 9
(i.e., X_3 is not free to vary)

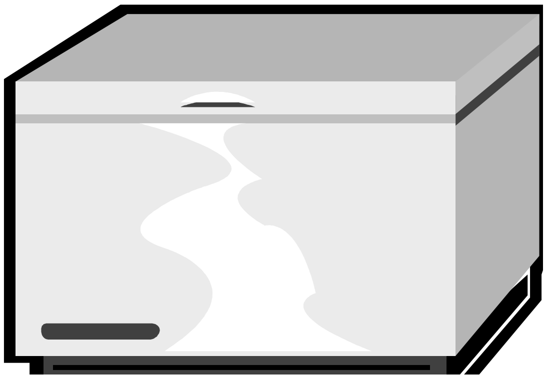
Here, n = 3, so degrees of freedom $= n - 1 = 3 - 1 = 2$

(2 values can be any numbers, but the third is not free to vary for a given mean)

Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Chi-square Example

- A commercial freezer must hold a selected temperature with little variation. Specifications call for a standard deviation of no more than 4 degrees (a variance of 16 degrees^2).
- A sample of 14 freezers is to be tested
- What is the upper limit (K) for the sample variance such that the probability of exceeding this limit, given that the population standard deviation is 4, is less than 0.05?



Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

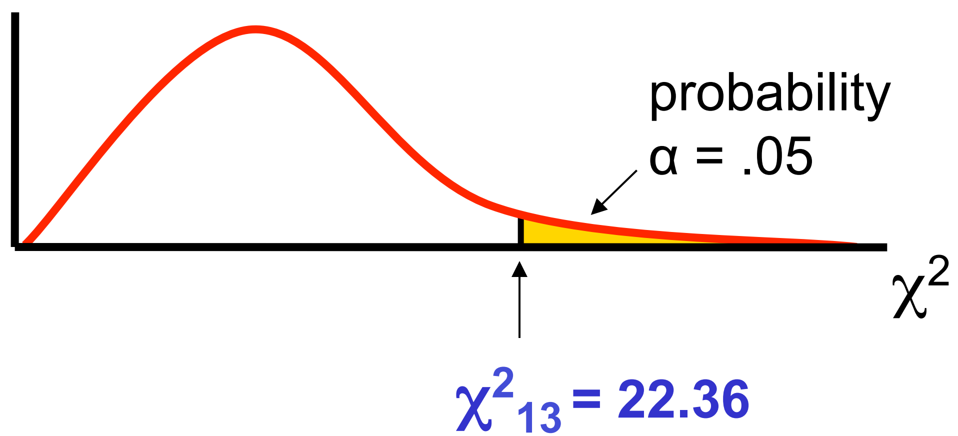
Finding the Chi-square Value

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

Is chi-square distributed with (n – 1) = 13 degrees of freedom

- Use the the chi-square distribution with area 0.05 in the upper tail:

χ^2_{13} = 22.36 (α = .05 and 14 – 1 = 13 d.f.)



Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall

Chi-square Example

(continued)

$\chi^2_{13} = 22.36$ (α = .05 and 14 – 1 = 13 d.f.)

So:

$$P(s^2 > K) = P\left(\frac{(n-1)s^2}{16} > \chi^2_{13}\right) = 0.05$$

or $$\frac{(n-1)K}{16} = 22.36$$ (where n = 14)

so $$K = \frac{(22.36)(16)}{(14-1)} = 27.52$$

If s^2 from the sample of size n = 14 is greater than 27.52, there is strong evidence to suggest the population variance exceeds 16.

Copyright © 2010 Pearson Education, Inc. Publishing as Prentice Hall