Multivariate Data Analysis – BIA 652
Class 2 – Review of Univariate II

Amir H Gandomi; PhD
Assistant Professor, Stevens Institute of Technology
a.h.gandomi@stevens.edu

Course Information

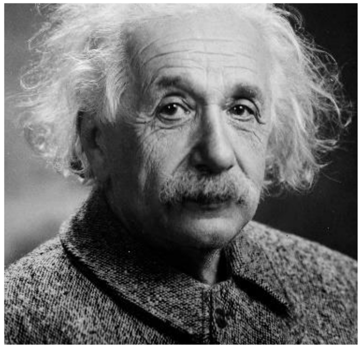
- Text:
  - Practical Multivariate Analysis, 5^{th} Ed., by A.A. Afifi, S.J. May, and V.A. Clark, Chapman Hall, New York: 2012
  - http://www.crcpress.com/product/isbn/9781439816806
- SAS, R, Python Bootcamps: Expected in first weeks
- SAS tutorial should be available on CANVAS
- SAS Loading on Laptops: at Hanlon Lab
- I have invited all of you to DataCamp and you should have free access to entire contents for about six months!

Course Information

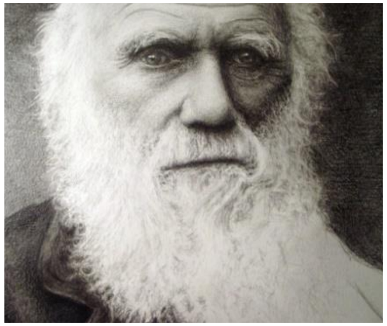
- Grading Assistant:
  - Mr. Haochen Liu
  - hliu56@stevens.edu
- Office: Babbio 616
- Office Hours:
  - Tuesdays from 2 - 6 pm
  - Or By Appointment

Team Pictures
Teams No More Than 3
(A picture of you, and your name)

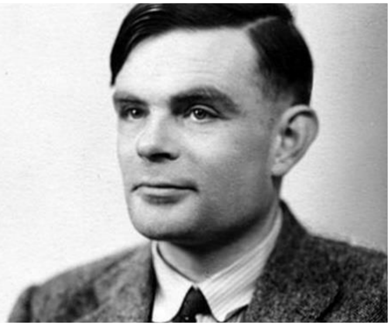
Albert Einstein



Charles Darwin



Alan Turing



Find a dataset for your term project

If you do not have your own datasets, you can look at:

- https://www.nature.com/sdata/
- https://www.journals.elsevier.com/data-in-brief
- http://www.economics-ejournal.org/special-areas/data-sets
- http://rmets.onlinelibrary.wiley.com/hub/journal/10.1002/(ISSN)2049-6060/
- Some companies may make their data available!

Class 2
Review + Intro

- Continue Review of Univariate Statistics and Matrix Algebra
- Introduction to Multivariate Analysis Ch 1-5.

More Univariate Review

Sampling and Sample Distributions

Statistics for Business and Economics

8th Edition

Chapter 6

Sampling and Sampling Distributions

Chapter Goals

After completing this chapter, you should be able to:

- Describe a simple random sample and why sampling is important
- Explain the difference between descriptive and inferential statistics
- Define the concept of a sampling distribution
- Determine the mean and standard deviation for the sampling distribution of the sample mean, $\overline{X}$
- Describe the Central Limit Theorem and its importance
- Determine the mean and standard deviation for the sampling distribution of the sample proportion, $\hat{p}$
- Describe sampling distributions of sample variances

Introduction

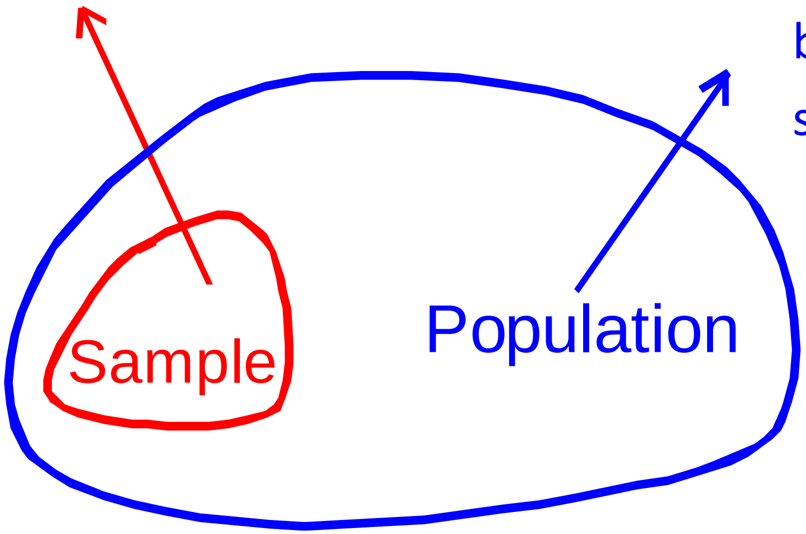
- **Descriptive statistics**
  - Collecting, presenting, and describing data
- **Inferential statistics**
  - Drawing conclusions and/or making decisions concerning a population based only on sample data

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)



6.1

Sampling from a Population

- 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

Inferential Statistics

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

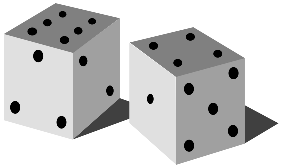
- **Estimation**
  - e.g., Estimate the population mean weight using the sample mean weight
- **Hypothesis Testing**
  - e.g., Use sample evidence to test the claim that the population mean weight is 120 pounds

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.

Simple Random Sample

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



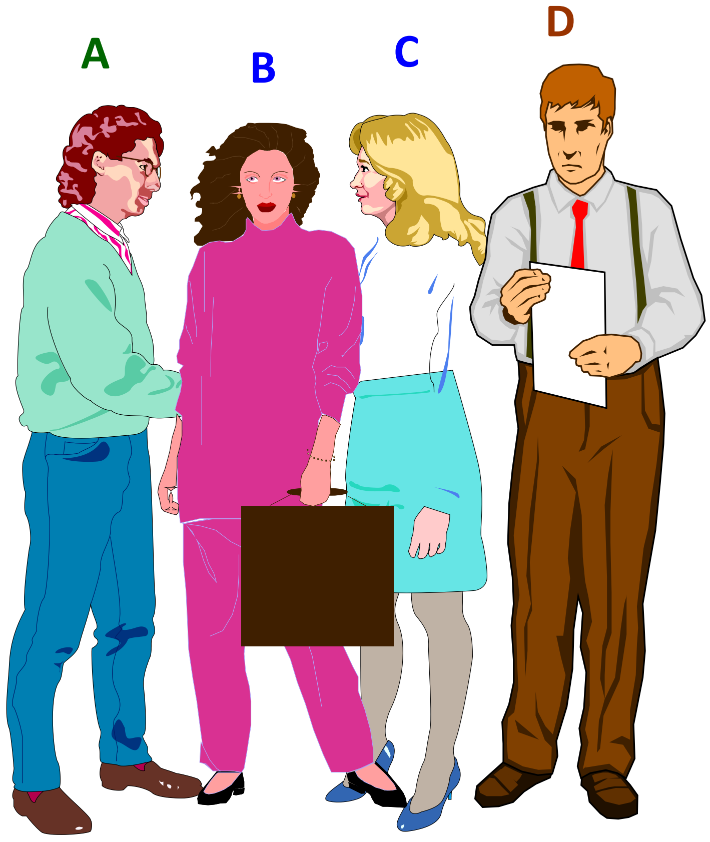
- A simple random sample is the ideal against which other sampling methods are compared

Sampling Distributions

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

Developing a Sampling Distribution

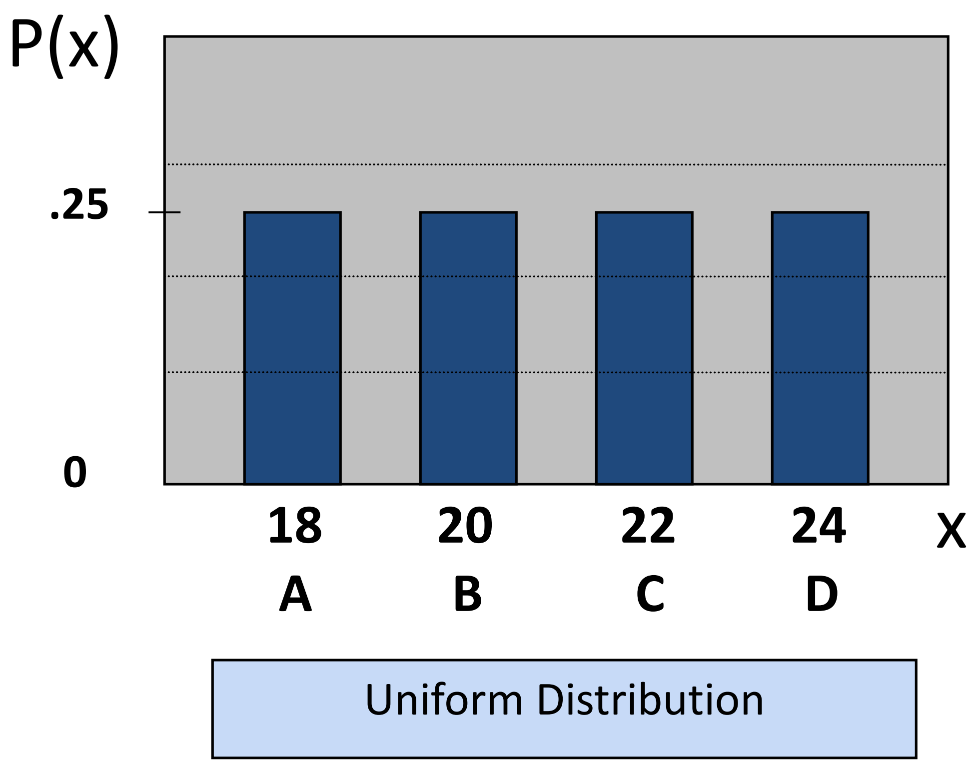
- **Assume there is a population …**
- Population size N=4
- Random variable, X, is age of individuals
- Values of X: 18, 20, 22, 24 (years)



Developing a Sampling Distribution

(continued)

In this example the Population Distribution is uniform:



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

Developing a Sampling Distribution

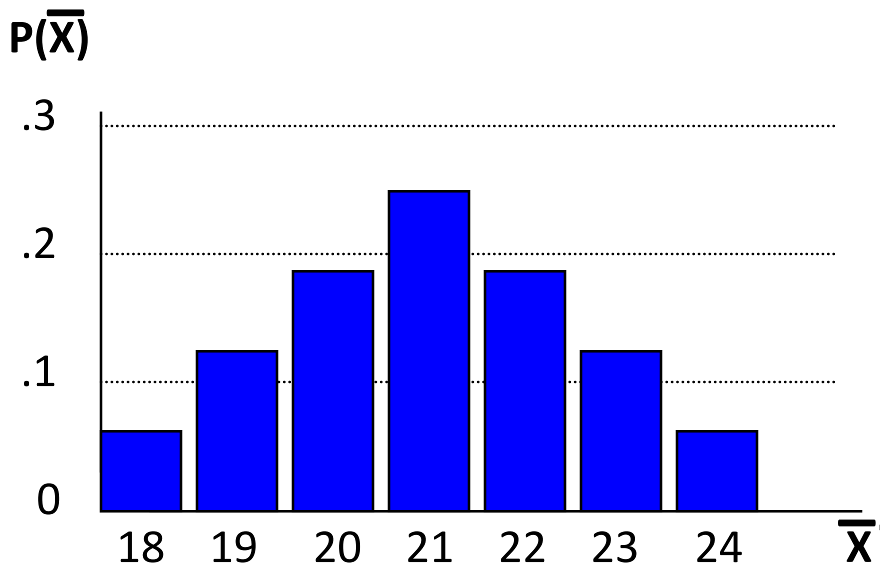
(continued)

Sampling Distribution of All Sample Means

16 Sample Means

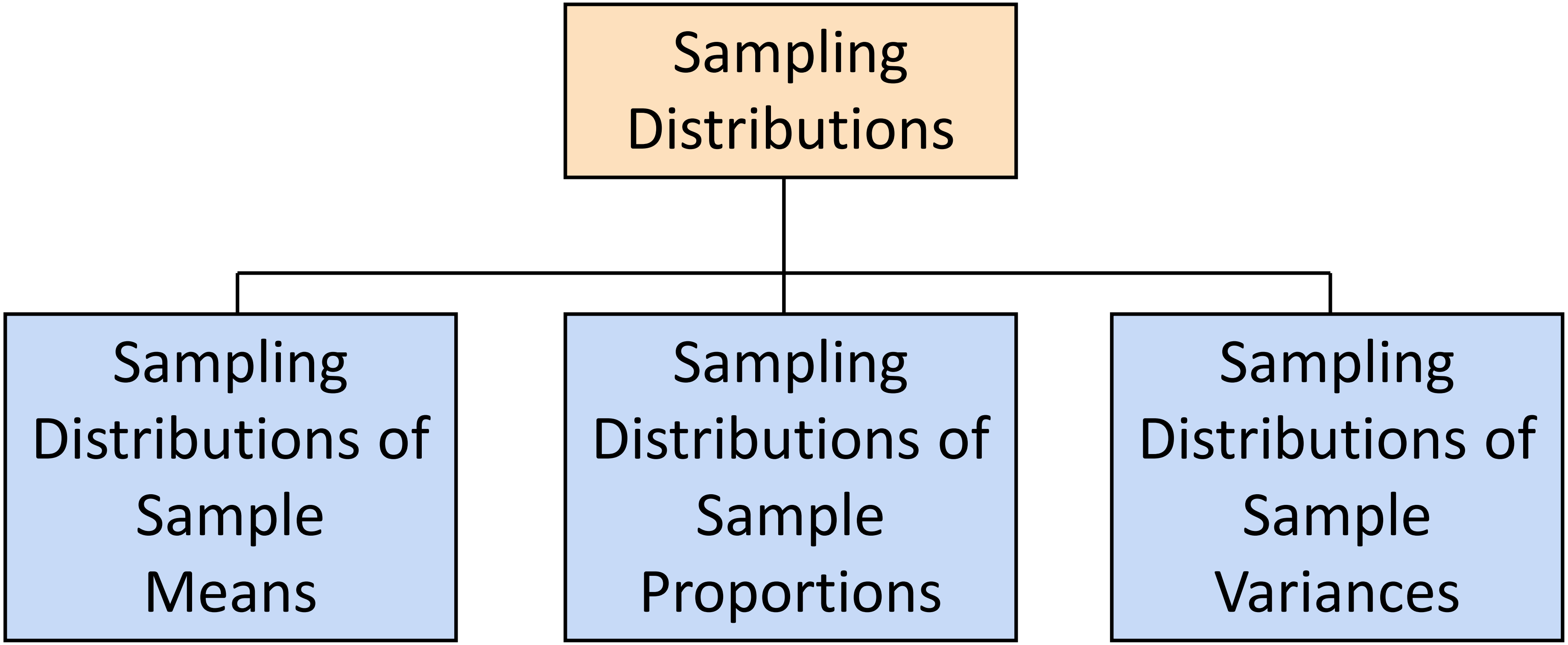
1st Obs	2nd Observation			
	18	20	22	24
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

Distribution of Sample Means



(no longer uniform)

Chapter Outline



Sample Mean

- Let $X_1, X_2, \ldots, X_n$ represent a random sample from a population

- The sample mean value of these observations is defined as

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

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:

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

- Note that the standard error of the mean decreases as the sample size increases

Comparing the Population with its Sampling Distribution

Population

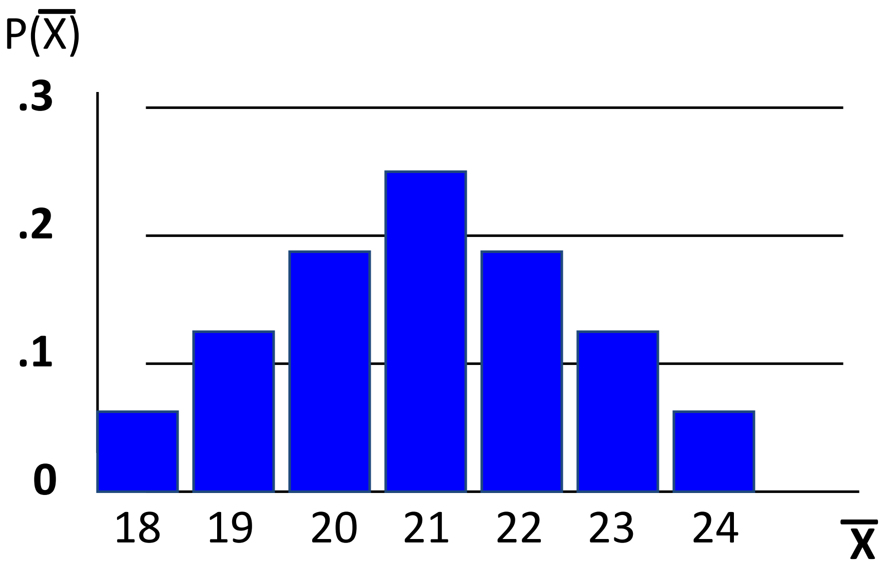
N = 4

$$\mu = 21 \qquad \sigma = 2.236$$

Sample Means Distribution

n = 2

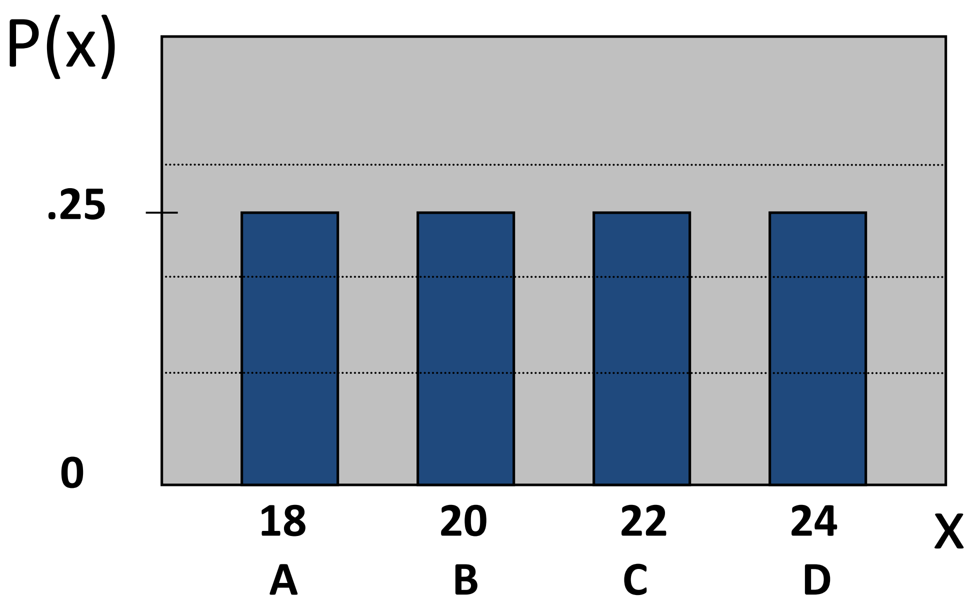
$$\mu_{\bar{X}} = 21 \qquad \sigma_{\bar{X}} = 1.58$$



Developing a Sampling Distribution

(continued)

Summary Measures for the Population Distribution:



Uniform Distribution

$$\mu = \frac{\sum X_i}{N}$$

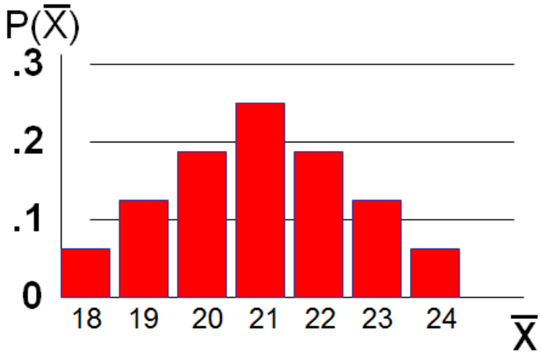
$$= \frac{18 + 20 + 22 + 24}{4} = 21$$

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

Developing a Sampling Distribution

(continued)

Summary Measures of the Sampling Distribution:



$$E(\bar{X}) = \frac{\sum \bar{X}_i}{N} = \frac{18 + 19 + 21 + \cdots + 24}{16} = 21 = \mu$$

$$\sigma_{\bar{X}} = \sqrt{\frac{\sum (\bar{X}_i - \mu)^2}{N}}$$

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

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 $\boxed{\mu_{\bar{X}} = \mu}$ and $\boxed{\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}}$

- If the sample size n is not large relative to the population size N, then $\boxed{\mu_{\bar{X}} = \mu}$ and $\boxed{\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}} \sqrt{\dfrac{N-n}{N-1}}}$

Standard Normal Distribution for the Sample Means

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

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

where: $\bar{X}$ = sample mean

μ = population mean

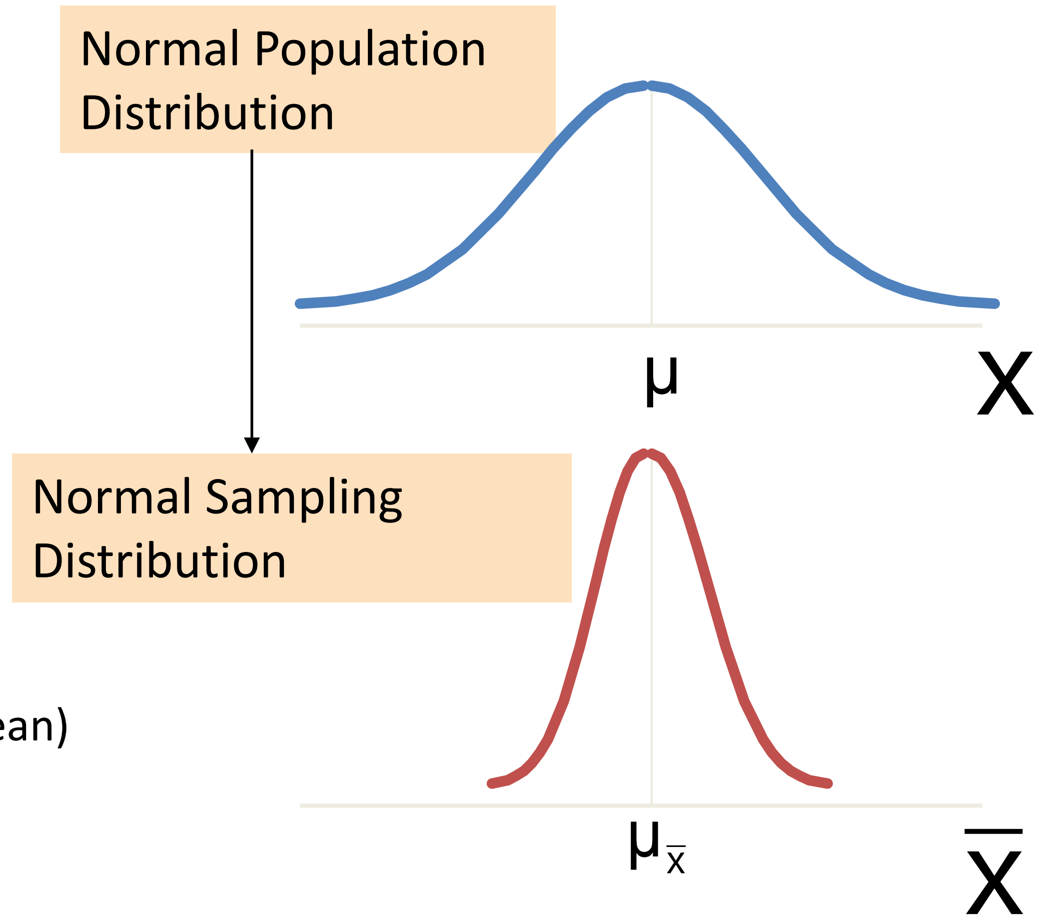
$\sigma_{\bar{X}}$ = standard error of the mean

Z is a standardized normal random variable with mean of 0 and a variance of 1

Sampling Distribution Properties

$$\boxed{E[\overline{X}] = \mu}$$

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



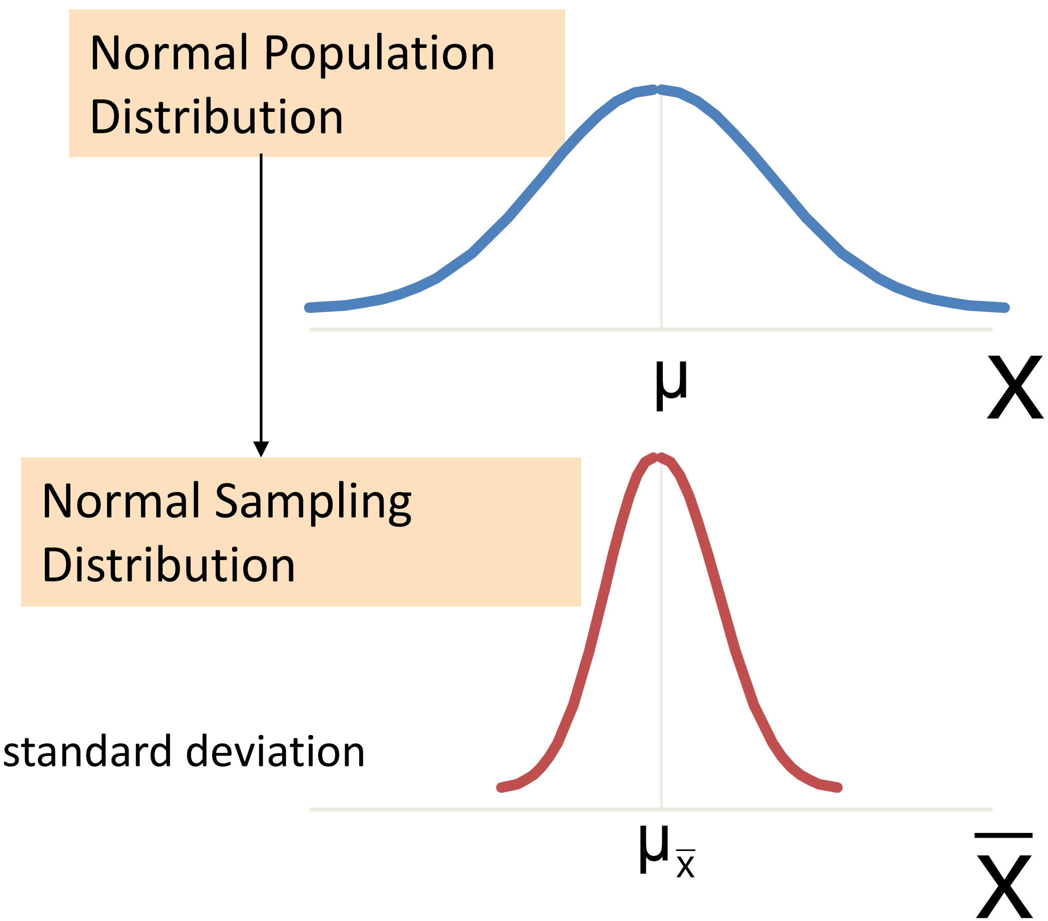
(both distributions have the same mean)

Sampling Distribution Properties

(continued)

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

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



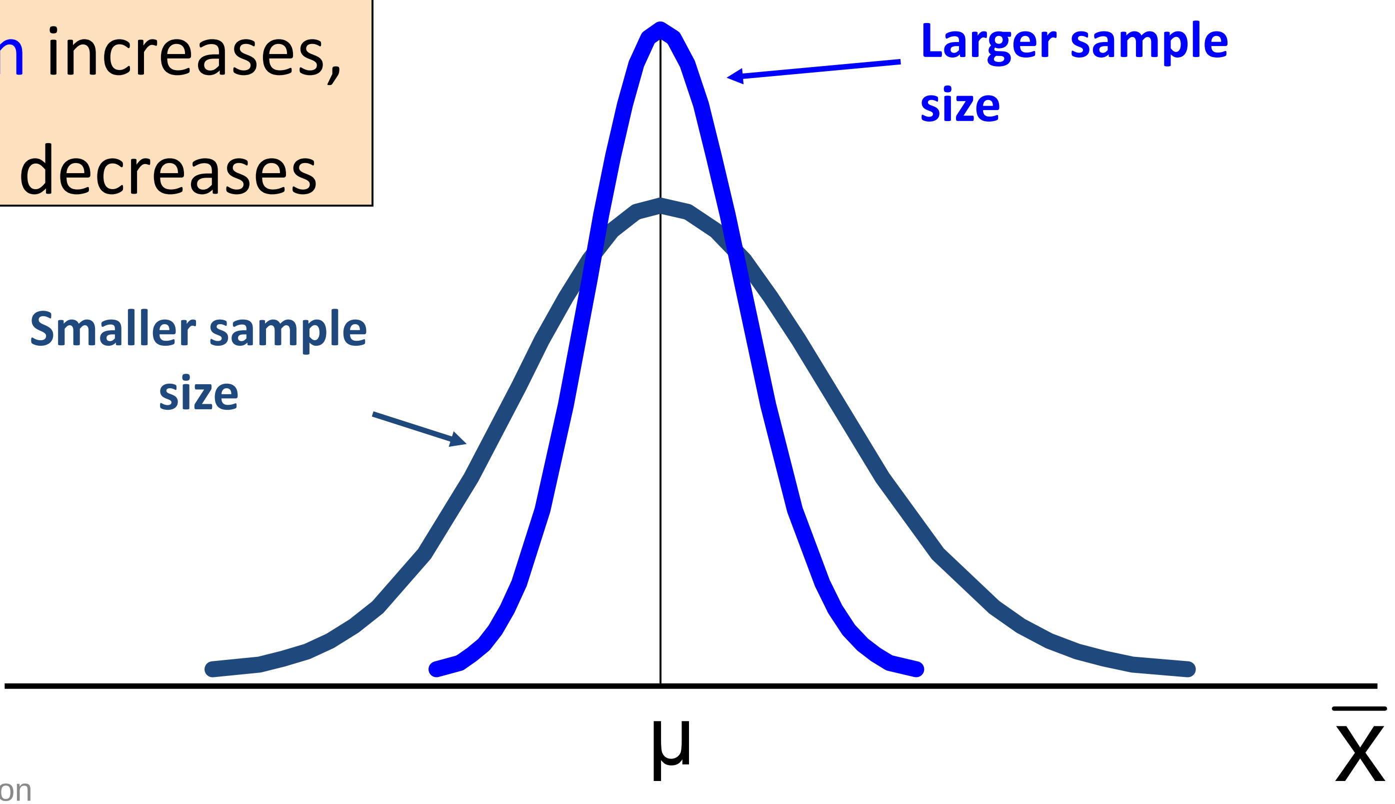
(the distribution of $\bar{X}$ has a reduced standard deviation

Sampling Distribution Properties

(continued)

As n increases,

$\sigma_{\bar{x}}$ decreases



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.

Central Limit Theorem

(continued)

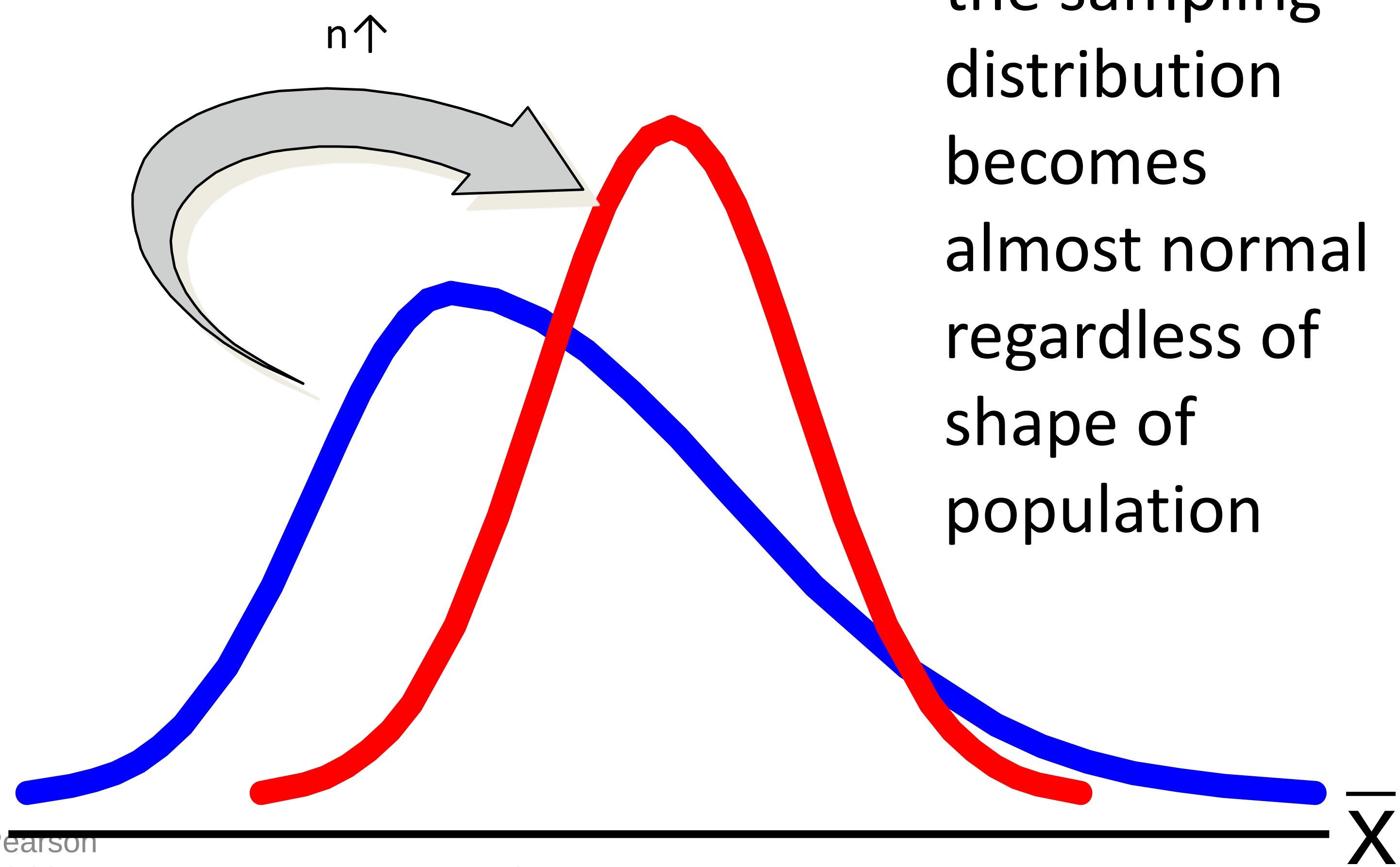
- Let $X_1, X_2, \ldots, X_n$ be a set of n independent random variables having identical distributions with mean μ, variance σ^2, and $\bar{X}$ as the mean of these random variables.
- As n becomes large, the central limit theorem states that the distribution of

$$\boxed{Z = \frac{\bar{X} - \mu_x}{\sigma_{\bar{X}}}}$$

approaches the standard normal distribution

Central Limit Theorem

As the sample size gets large enough…



the sampling distribution becomes almost normal regardless of shape of population

If the Population is **not** Normal

(continued)

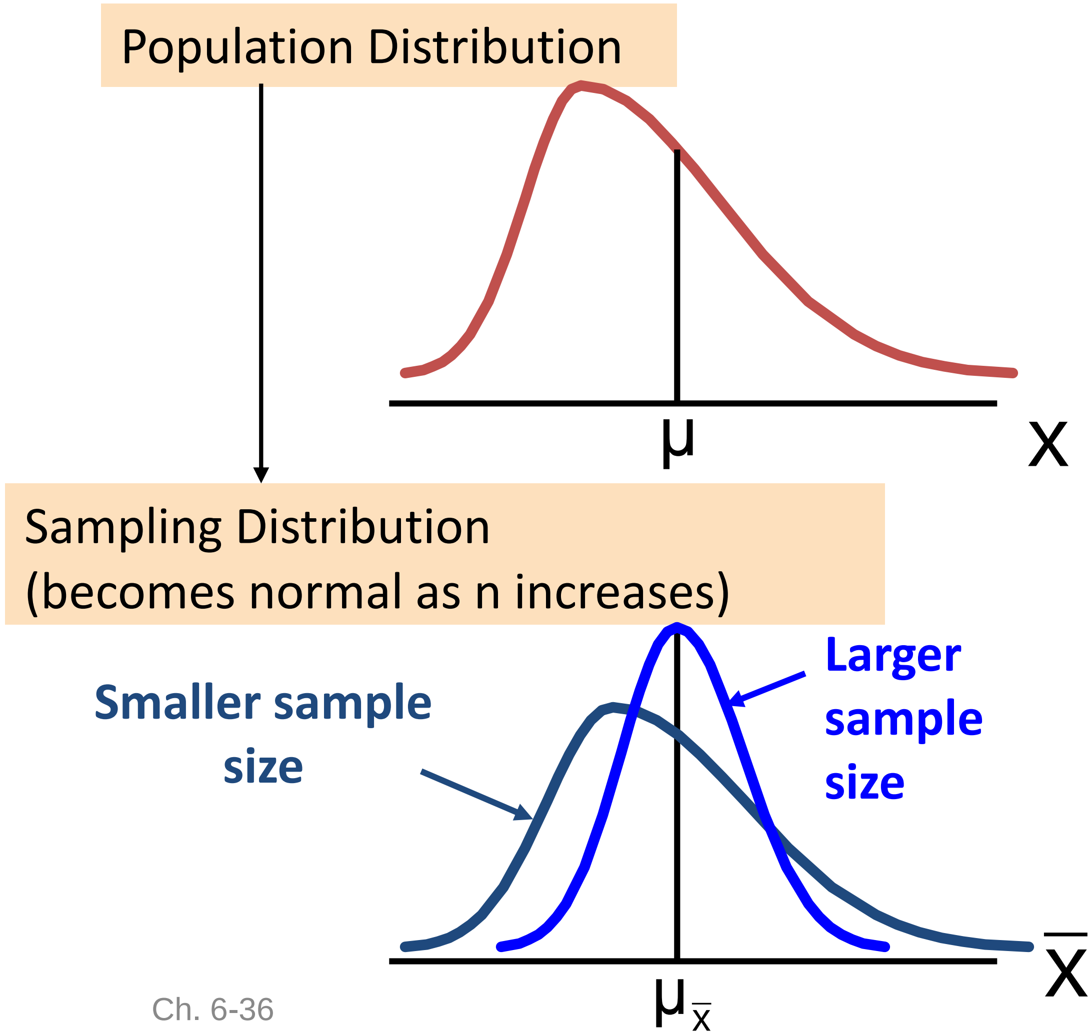
Sampling distribution properties:

Central Tendency

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

Variation

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



How Large is Large Enough?

- For most distributions, $n > 25$ will give a sampling distribution that is nearly normal
- For normal population distributions, the sampling distribution of the mean is always normally distributed

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 mean($\bar{X}$) is approximately normal if n is large enough, with mean μ and standard deviation
  - 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

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

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

Sampling Distributions of Sample Proportions

P = the proportion of the population having some characteristic

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

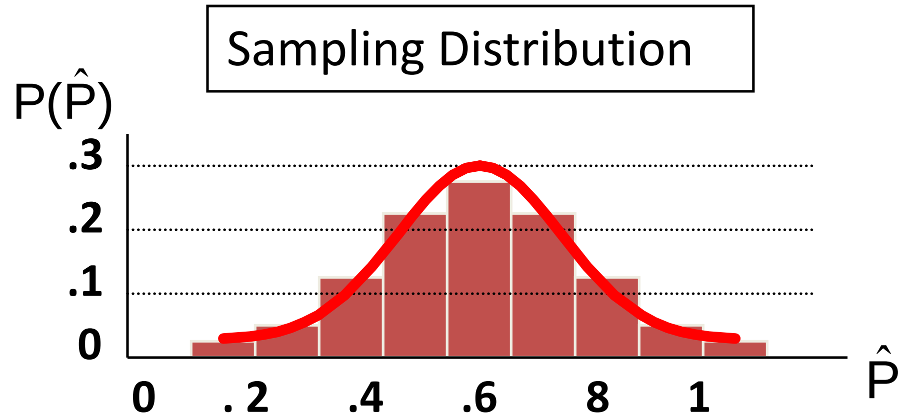
$$\hat{p} = \frac{X}{n} = \frac{\text{number of items in the sample having the characteristic of interest}}{\text{sample size}}$$

- $0 \le \hat{p} \le 1$
- $\hat{p}$ has a binomial distribution, but can be approximated by a normal distribution when $\underbrace{nP(1-P)}_{\text{Variance of binomial random variable X}} > 5$

Variance of binomial random variable X

Sampling Distribution of $\hat{p}$

- Normal approximation:



Properties: $\boxed{E(\hat{p}) = P}$ and $\boxed{\sigma_{\hat{p}} = \sqrt{\frac{P(1-P)}{n}}}$

(where P = population proportion)

Sampling Distribution of $\hat{p}$

- Mean of sampling distribution of $\hat{p}$

$$E(\hat{p}) = E\left(\frac{X}{n}\right) = \frac{1}{n} \times E(X) = \frac{1}{n} \times nP = P$$

- Variance of sampling distribution of $\hat{p}$

$$\sigma_{\hat{p}} = \sqrt{Var\left(\frac{X}{n}\right)} = \sqrt{\frac{1}{n^2} \times Var(X)} =$$

$$\sqrt{\frac{1}{n^2} \times nP(1-P)} = \sqrt{\frac{P(1-P)}{n}}$$

Z-Value for Proportions

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

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

Where the distribution of Z is a good approximation to the standard normal distribution if $nP(1-P) > 5$

Example

- If the true proportion of voters who support Proposition A is P = 0.4, what is the probability that a sample of size 200 yields a sample proportion between 0.40 and 0.45?

- i.e.: **if P = 0.4 and n = 200, what is**

  $$P(0.40 \le \hat{p} \le 0.45)\ ?$$

Example

(continued)

- **if P = 0.4 and n = 200, what is**

  P(0.40 ≤ $\hat{p}$ ≤ 0.45) ?

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

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

Convert to standard normal:

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

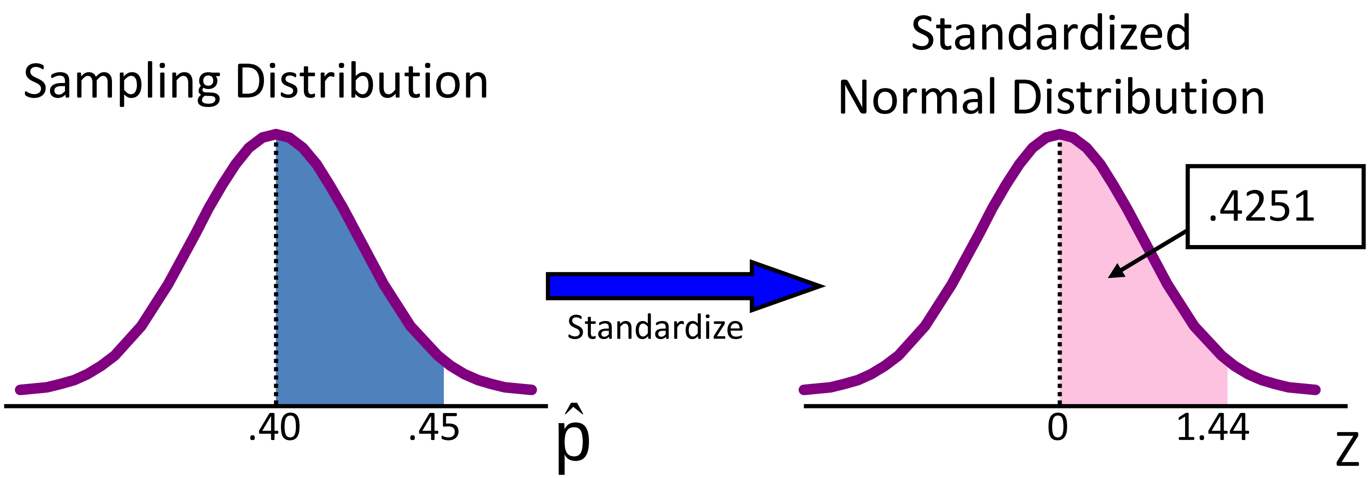
$$= P(0 \le Z \le 1.44)$$

Example

(continued)

- **if P = 0.4 and n = 200, what is $P(0.40 \le \hat{p} \le 0.45)$?**

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



Sample Variance

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

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

- 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

Sampling Distribution of Sample Variances

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

$$\boxed{E[s^2] = \sigma^2}$$

- If the population distribution is normal, then

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

Chi-Square Distribution of Sample and Population Variances

- If the population distribution is normal then

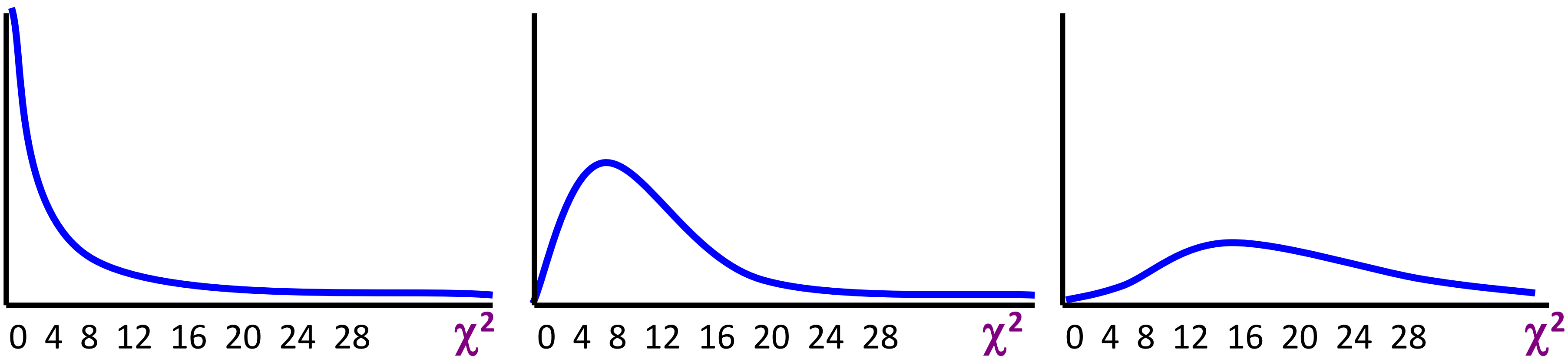
$$\chi^2_{n-1} = \frac{(n-1)s^2}{\sigma^2}$$

has a chi-square (χ^2) distribution

with $n - 1$ degrees of freedom

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

- See web for chi-square probabilities

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)

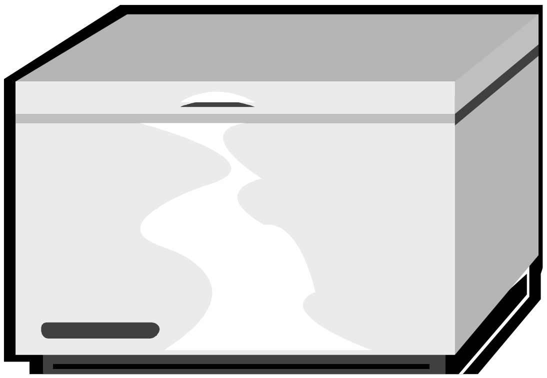
Here, n = 3, so degrees of freedom $= n - 1 = 3 - 1 = 2$

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

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?



Finding the Chi-square Value

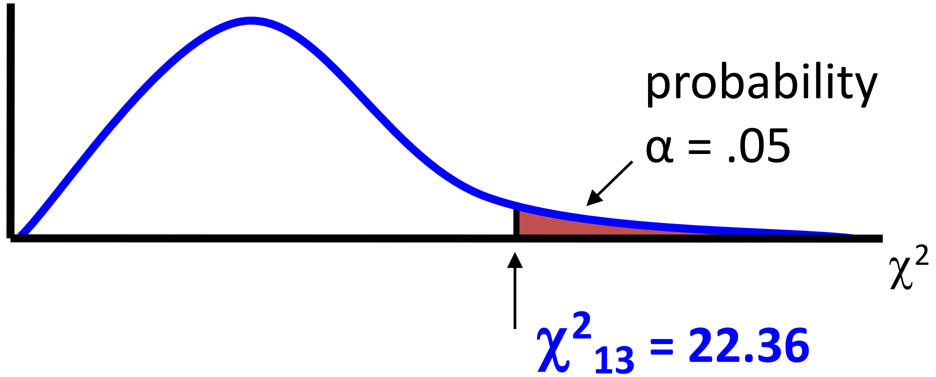
$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

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.)



Chi-square Example

(continued)

χ^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.

Chapter Summary

- Introduced sampling distributions
- Described the sampling distribution of sample means
  - For normal populations
  - Using the Central Limit Theorem
- Described the sampling distribution of sample proportions
- Introduced the chi-square distribution
- Examined sampling distributions for sample variances
- Calculated probabilities using sampling distributions

Estimation

Statistics for Business and Economics

8th Edition

Chapter 7

Estimation: Single Population

Chapter Goals

After completing this chapter, you should be able to:

- Distinguish between a point estimate and a confidence interval estimate
- Construct and interpret a confidence interval estimate for a single population mean using the Z distribution
- Form and interpret a confidence interval estimate for a single population proportion
- Create confidence interval estimates for the variance of a normal population
- Determine the required sample size to estimate a mean or proportion within a specified margin of error

Confidence Intervals

Contents of this chapter:

- Confidence Intervals for the Population Mean, μ
  - when Population Variance σ^2 is Known
- Confidence Intervals for the Population Proportion, P (large samples)
- Confidence interval estimates for the variance of a normal population
- Finite population corrections
- Sample-size determination

7.1

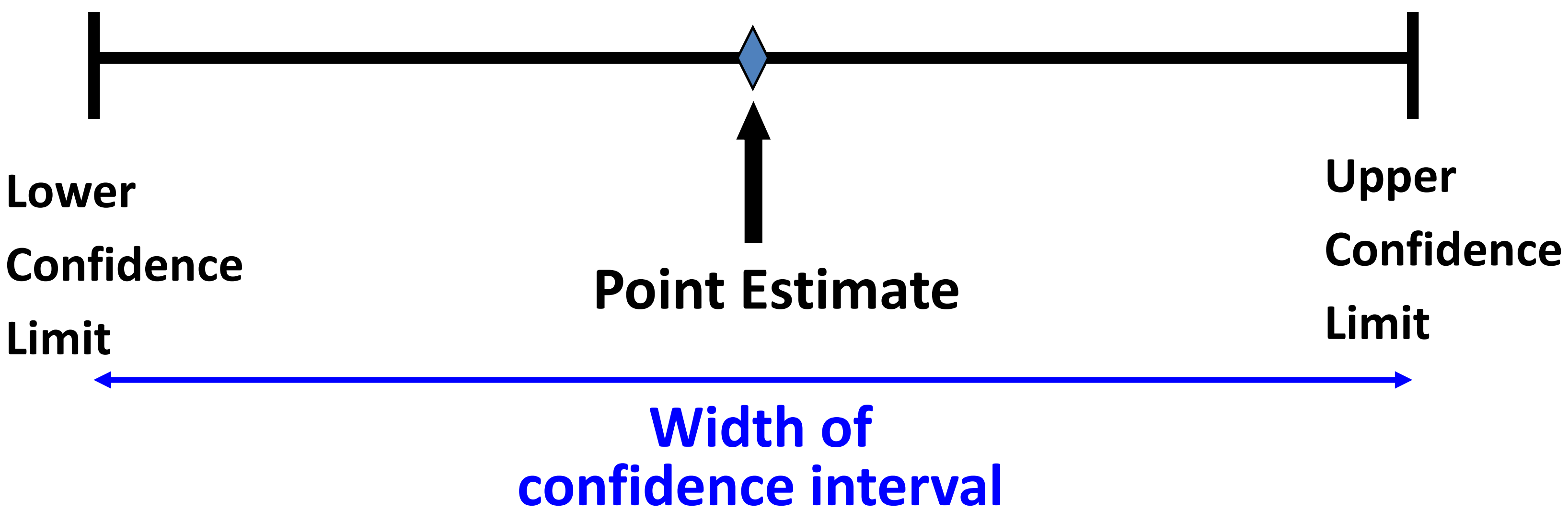
Properties of Point Estimators

- An estimator of a population parameter is
  - a random variable that depends on sample information . . .
  - whose value provides an approximation to this unknown parameter

- A specific value of that random variable is called an estimate

Point and Interval Estimates

- A point estimate is a single number,
- a confidence interval provides additional information about variability



Point Estimates

We can estimate a Population Parameter …		with a Sample Statistic (a Point Estimate)
Mean	μ	$\bar{x}$
Proportion	P	$\hat{p}$

Unbiasedness

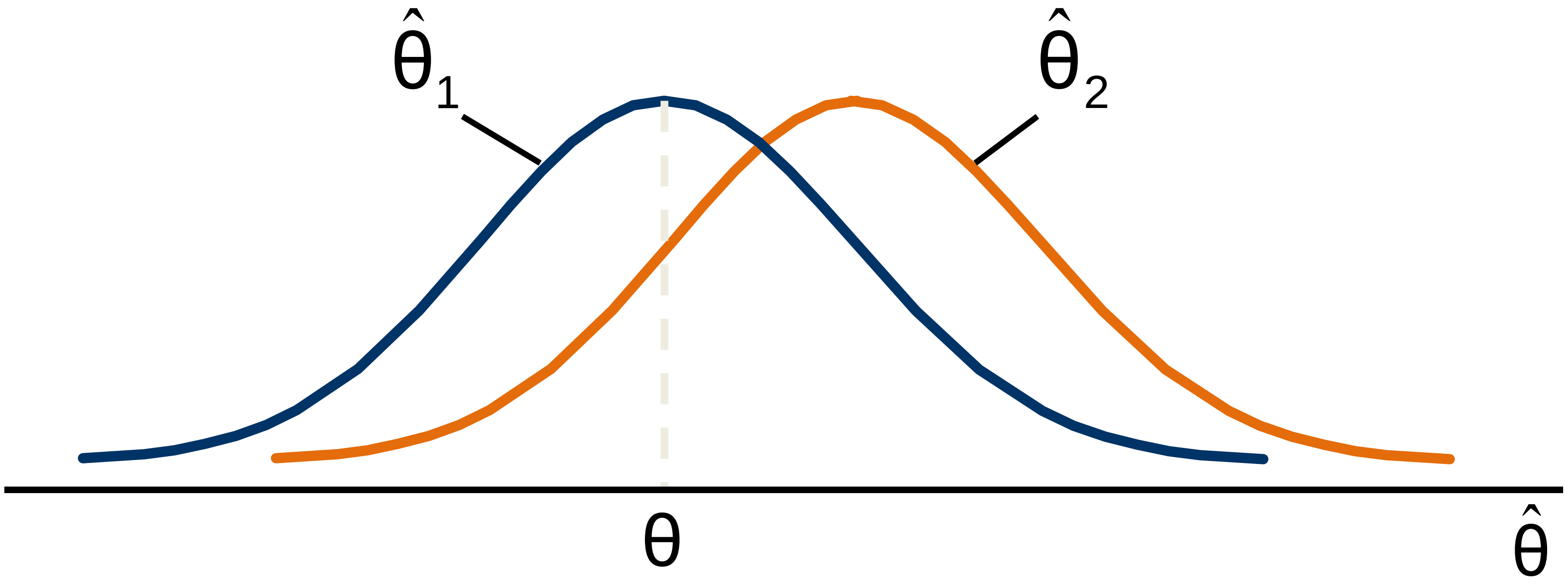
- A point estimator $\hat{\theta}$ is said to be an unbiased estimator of the parameter θ if its expected value is equal to that parameter: $\boxed{E(\hat{\theta}) = \theta}$

- Examples:
  - The sample mean $\bar{x}$ is an unbiased estimator of μ
  - The sample variance s^2 is an unbiased estimator of σ^2
  - The sample proportion $\hat{p}$ is an unbiased estimator of P

Unbiasedness

(continued)

- $\hat{\theta}_1$ is an unbiased estimator, $\hat{\theta}_2$ is biased:



Bias

- Let $\hat{\theta}$ be an estimator of θ

- The bias in $\hat{\theta}$ is defined as the difference between its mean and θ

$$\boxed{\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta}$$

- The bias of an unbiased estimator is 0

Confidence Interval Estimation

- How much uncertainty is associated with a point estimate of a population parameter?
- An interval estimate provides more information about a population characteristic than does a point estimate
- Such interval estimates are called confidence interval estimates

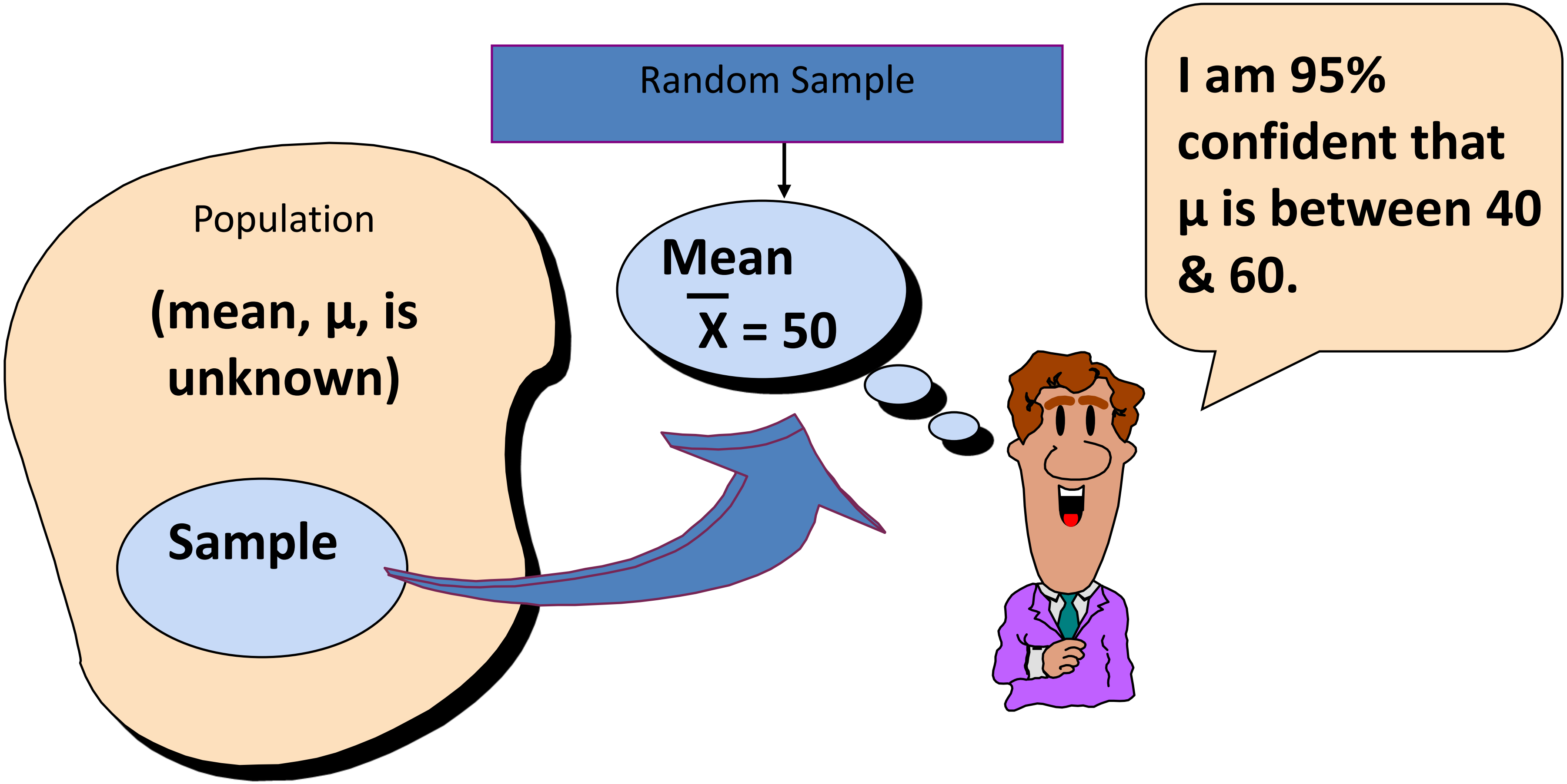
Confidence Interval Estimate

- An interval gives a range of values:
  - Takes into consideration variation in sample statistics from sample to sample
  - Based on observation from 1 sample
  - Gives information about closeness to unknown population parameters
  - Stated in terms of level of confidence
    - Can never be 100% confident

Confidence Interval and Confidence Level

- If $P(a < \theta < b) = 1 - \alpha$ then the interval from a to b is called a $100(1 - \alpha)\%$ confidence interval of θ.
- The quantity $100(1 - \alpha)\%$ is called the confidence level of the interval
  - α is between 0 and 1
  - In repeated samples of the population, the true value of the parameter θ would be contained in $100(1 - \alpha)\%$ of intervals calculated this way.
  - The confidence interval calculated in this manner is written as $a < \theta < b$ with $100(1 - \alpha)\%$ confidence

Estimation Process



Confidence Level, $(1-\alpha)$

(continued)

- Suppose confidence level = 95%
- Also written $(1 - \alpha) = 0.95$
- A relative frequency interpretation:
  - From repeated samples, 95% of all the confidence intervals that can be constructed of size n will contain the unknown true parameter
- A specific interval either will contain or will not contain the true parameter
  - No probability involved in a specific interval

General Formula

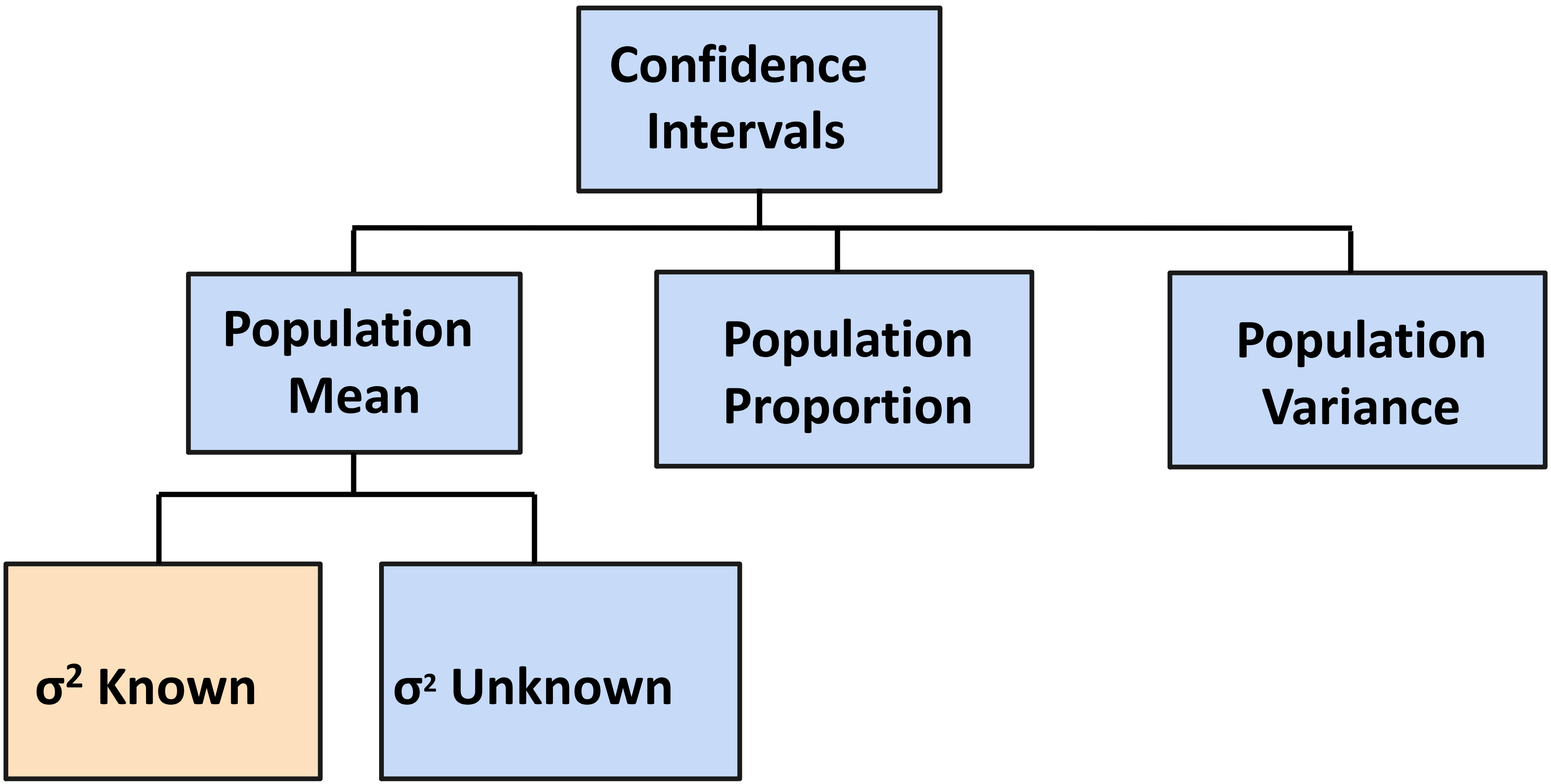
- The general form for all confidence intervals is:

$$\boxed{\hat{\theta} \pm ME}$$

$$\boxed{\textbf{Point Estimate} \pm \textbf{Margin of Error}}$$

- The value of the margin of error depends on the desired level of confidence

Confidence Intervals



(From normally distributed populations)

7.2

Confidence Interval Estimation for the Mean (σ^2 Known)

- Assumptions
  - Population variance σ^2 is known
  - Population is normally distributed
  - If population is not normal, use large sample
- Confidence interval estimate:

$$\boxed{\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}}$$

(where $z_{\alpha/2}$ is the normal distribution value for a probability of $\alpha/2$ in each tail)

Confidence Limits

- The confidence interval is

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

- The endpoints of the interval are

$$\text{UCL} = \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Upper confidence limit

$$\text{LCL} = \bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Lower confidence limit

Margin of Error

- The confidence interval,

$$\boxed{\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}}$$

- Can also be written as $\boxed{\bar{x} \pm ME}$

  where ME is called the margin of error

$$\boxed{ME = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}}$$

- The interval width, w, is equal to twice the margin of error

Reducing the Margin of Error

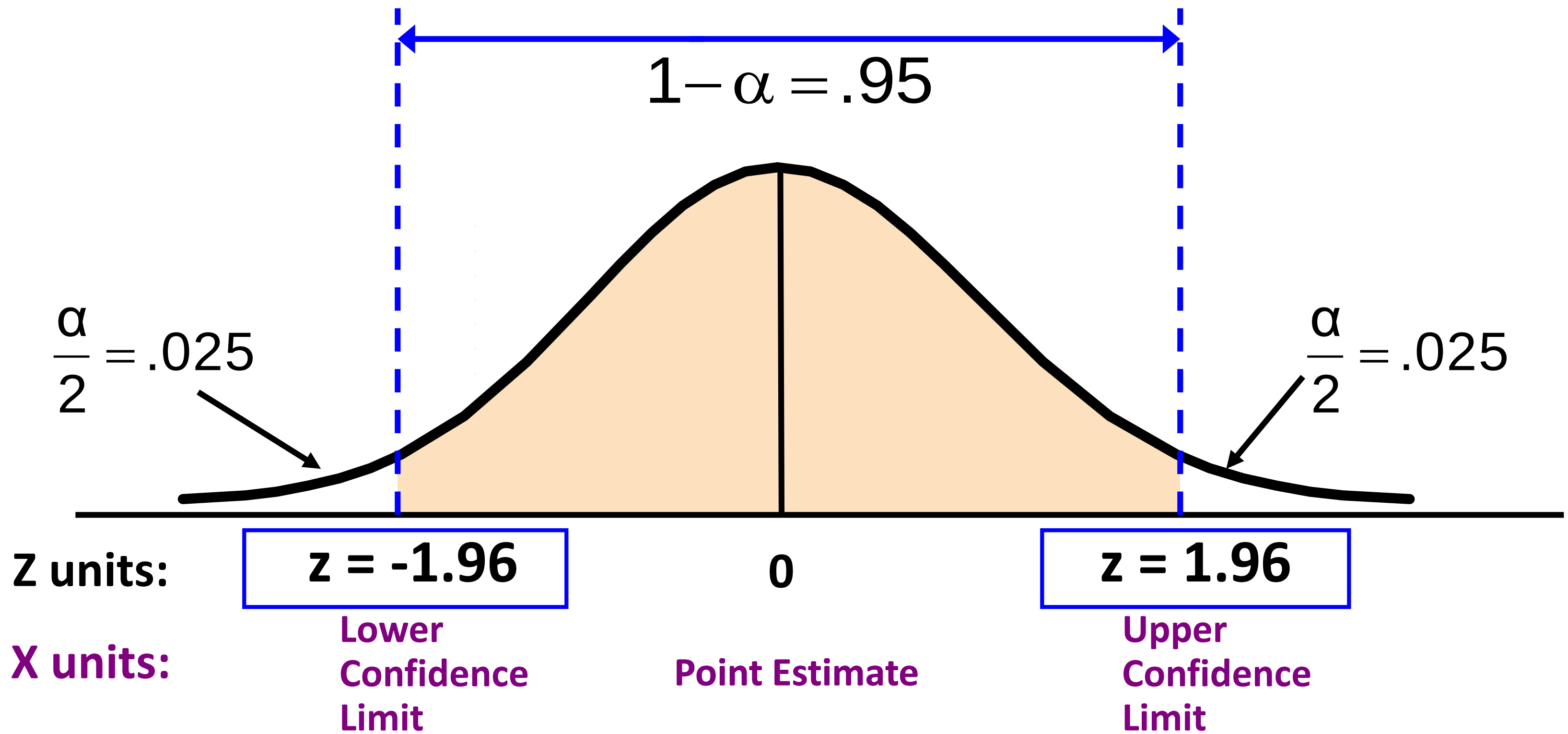
$$ME = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

The margin of error can be reduced if

- the population standard deviation can be reduced ($\sigma\downarrow$)
- The sample size is increased ($n\uparrow$)
- The confidence level is decreased, $(1 - \alpha) \downarrow$

Finding $z_{\alpha/2}$

- Consider a 95% confidence interval:



- Find $z_{.025} = \pm 1.96$ from the standard normal distribution table

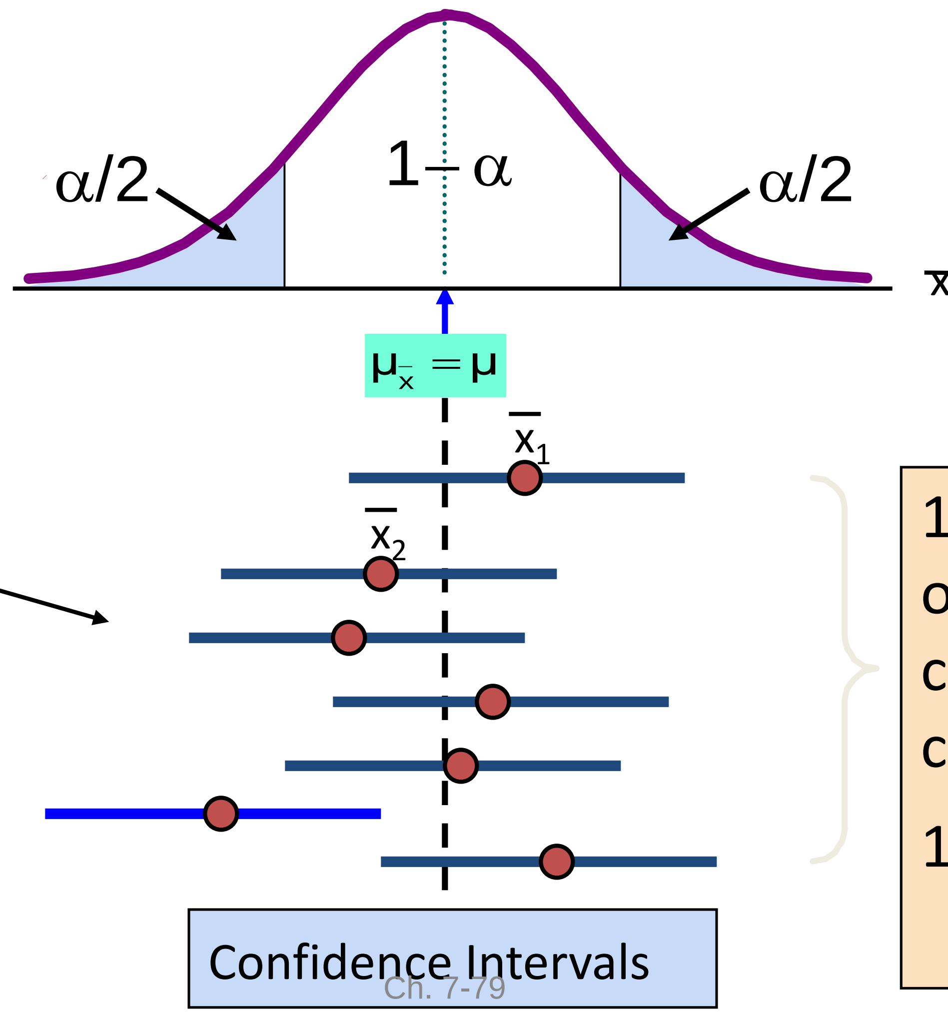
Common Levels of Confidence

- Commonly used confidence levels are 90%, 95%, 98%, and 99%

Confidence Level	***Confidence Coefficient,*** $1-\alpha$	$Z_{\alpha/2}$ ***value***
80%	**.80**	**1.28**
90%	**.90**	**1.645**
95%	**.95**	**1.96**
98%	**.98**	**2.33**
99%	**.99**	**2.58**
99.8%	**.998**	**3.08**
99.9%	**.999**	**3.27**

Intervals and Level of Confidence

Sampling Distribution of the Mean



Intervals extend from

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

to

$$UCL = \bar{X} + Z\frac{\sigma}{\sqrt{n}}$$

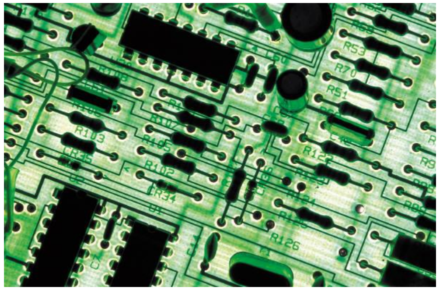
$100(1-\alpha)\%$ of intervals constructed contain μ;

$100(\alpha)\%$ do not.

Example

- A sample of 11 circuits from a large normal population has a mean resistance of 2.20 ohms. We know from past testing that the population standard deviation is 0.35 ohms.

- Determine a 95% confidence interval for the true mean resistance of the population.



Example

(continued)

- A sample of 11 circuits from a large normal population has a mean resistance of 2.20 ohms. We know from past testing that the population standard deviation is .35 ohms.

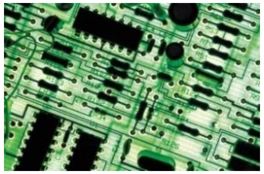
- Solution:

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$= 2.20 \pm 1.96\,(.35/\sqrt{11})$$

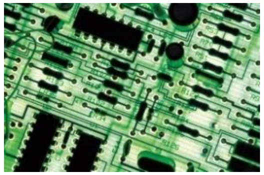
$$= 2.20 \pm .2068$$

$$\boxed{1.9932 < \mu < 2.4068}$$



Interpretation

- We are 95% confident that the true mean resistance is between 1.9932 and 2.4068 ohms
- Although the true mean may or may not be in this interval, 95% of intervals formed in this manner will contain the true mean



Confidence Intervals for the Population Variance

- **Goal:** Form a confidence interval for the population variance, σ^2

- The confidence interval is based on the sample variance, s^2
- Assumed: the population is normally distributed

Confidence Intervals for the Population Variance

(continued)

The random variable

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

follows a chi-square distribution with (n – 1) degrees of freedom

Where the chi-square value $\chi^2_{n-1,\alpha}$ denotes the number for which

$$\boxed{P(\chi^2_{n-1} > \chi^2_{n-1,\alpha}) = \alpha}$$

Confidence Intervals for the Population Variance

(continued)

The 100(1 - α)% confidence interval for the population variance is given by

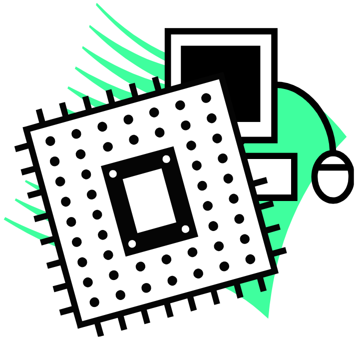
$$\mathrm{LCL} = \frac{(n-1)s^2}{\chi^2_{n-1,\,\alpha/2}}$$

$$\mathrm{UCL} = \frac{(n-1)s^2}{\chi^2_{n-1,\,1-\alpha/2}}$$

Example

You are testing the speed of a batch of computer processors. You collect the following data (in Mhz):

Sample size	**17**
Sample mean	**3004**
Sample std dev	**74**



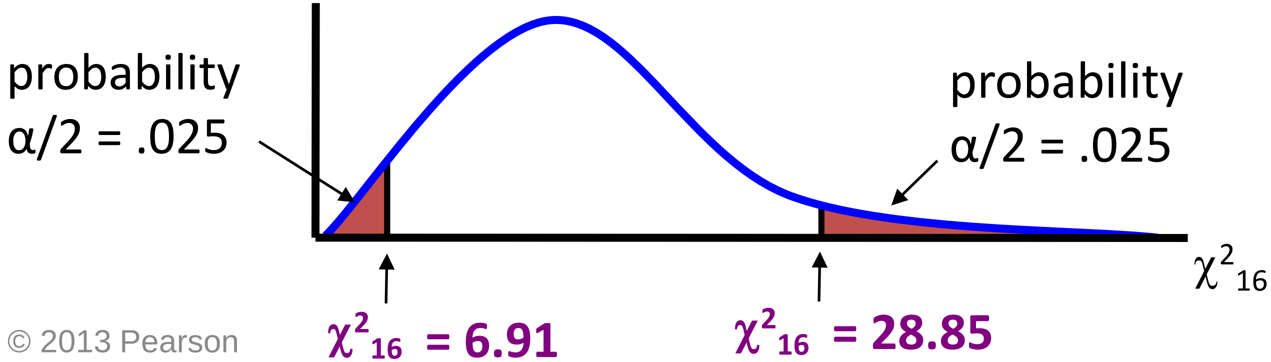
Assume the population is normal.
Determine the 95% confidence interval for σ_x^2

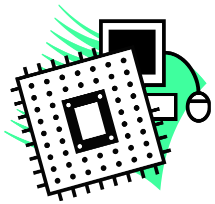
Finding the Chi-square Values

- n = 17 so the chi-square distribution has (n – 1) = 16 degrees of freedom
- α = 0.05, so use the the chi-square values with area 0.025 in each tail:

$$\chi^2_{n-1,\,\alpha/2} = \chi^2_{16\,,\,0.025} = 28.85$$

$$\chi^2_{n-1,\,1-\alpha/2} = \chi^2_{16\,,\,0.975} = 6.91$$





Calculating the Confidence Limits

- The 95% confidence interval is

$$\frac{(n-1)s^2}{\chi^2_{n-1,\,\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{n-1,\,1-\alpha/2}}$$

$$\frac{(17-1)(74)^2}{28.85} < \sigma^2 < \frac{(17-1)(74)^2}{6.91}$$

$$\boxed{3037 < \sigma^2 < 12680}$$

Converting to standard deviation, we are 95% confident that the population standard deviation of CPU speed is between 55.1 and 112.6 Mhz

7.6 Confidence Interval Estimation: Finite Populations

- If the sample size is more than 5% of the population size (and sampling is without replacement) then a finite population correction factor must be used when calculating the standard error

Finite Population Correction Factor

- Suppose sampling is without replacement and the sample size is large relative to the population size
- Assume the population size is large enough to apply the central limit theorem
- Apply the finite population correction factor when estimating the population variance

$$\text{finite population correction factor} = \frac{N-n}{N-1}$$

Estimating the Population Mean

- Let a simple random sample of size n be taken from a population of N members with mean μ
- The sample mean is an unbiased estimator of the population mean μ
- The point estimate is:

$$\boxed{\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i}$$

Finite Populations: Mean

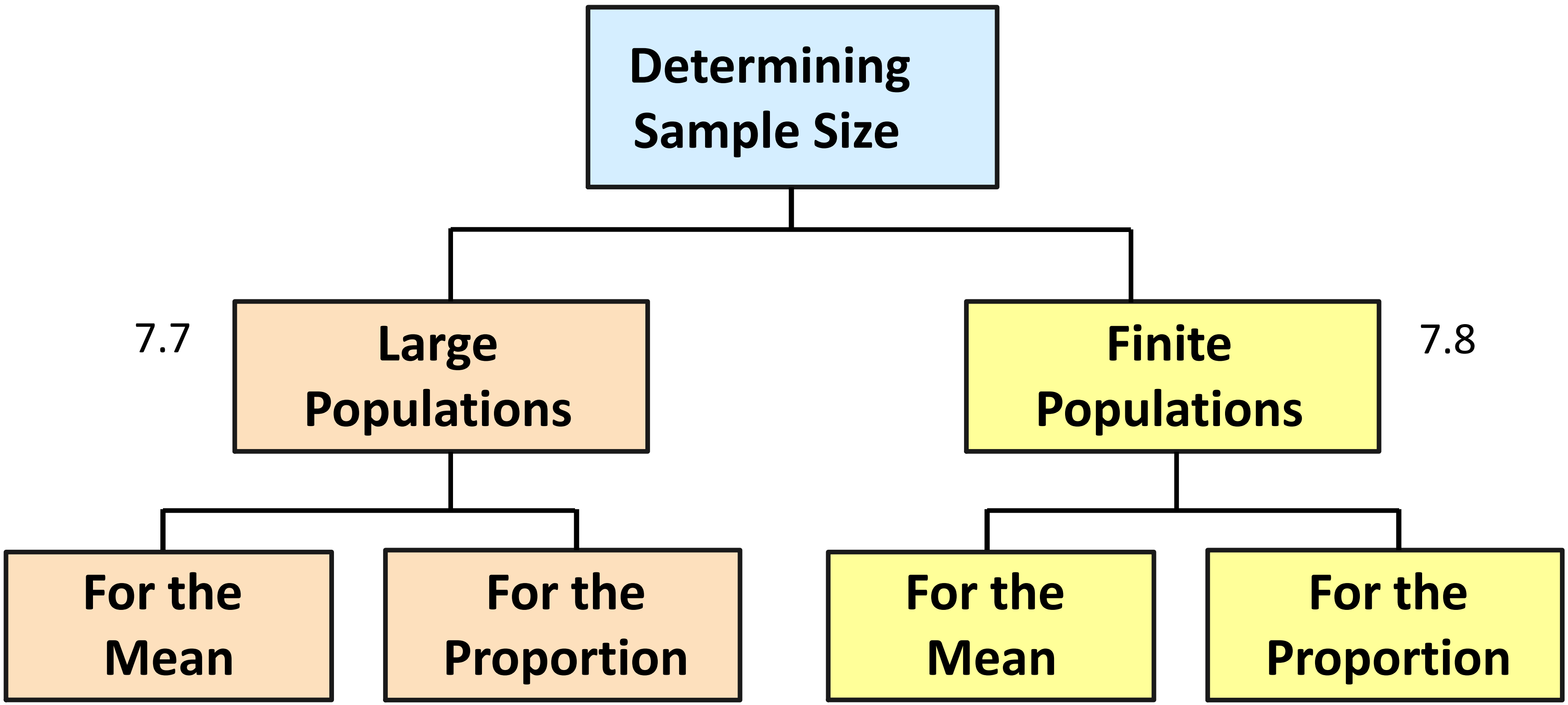
- If the sample size is more than 5% of the population size, an unbiased estimator for the variance of the sample mean is

$$\hat{\sigma}^2_{\bar{x}} = \frac{s^2}{n}\left(\frac{N-n}{N-1}\right)$$

- So the 100(1-α)% confidence interval for the population mean is

$$\bar{x} \pm t_{n-1,\alpha/2}\hat{\sigma}_{\bar{x}}$$

Sample-Size Determination



Sample-Size Determination: Large Populations

Large Populations

For the Mean

(Known population variance)

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Margin of Error (sampling error)

$$ME = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Sample Size Determination: Population Proportion

Large Populations

For the Proportion

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$ME = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Margin of Error (sampling error)

Sample Size Determination: Population Proportion

(continued)

Large Populations

For the Proportion

$$ME = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$\hat{p}(1-\hat{p})$ cannot be larger than 0.25, so max $\hat{p}$ = 0.5

Substitute 0.25 for $\hat{p}(1-\hat{p})$ and solve for n to get

$$n = \frac{0.25\, z_{\alpha/2}^2}{ME^2}$$

Sample Size Determination: Population Proportion

(continued)

- The sample and population proportions, $\hat{p}$ and P, are generally not known (since no sample has been taken yet)
- P(1 – P) = 0.25 generates the largest possible margin of error (so guarantees that the resulting sample size will meet the desired level of confidence)
- To determine the required sample size for the proportion, you must know:
  - The desired level of confidence ($1 - \alpha$), which determines the critical $z_{\alpha/2}$ value
  - The acceptable sampling error (margin of error), ME
  - Estimate P(1 – P) = 0.25

Required Sample Size Example: Population Proportion

How large a sample would be necessary to estimate the true proportion defective in a large population within ±3%, with 95% confidence?

Required Sample Size Example

(continued)

Solution:

For 95% confidence, use $z_{0.025} = 1.96$

ME = 0.03

Estimate P(1 – P) = 0.25

$$n = \frac{0.25\, z_{\alpha/2}^2}{ME^2} = \frac{(0.25)(1.96)^2}{(0.03)^2} = \boxed{1067.11}$$

So use n = 1068

Chapter Summary

- Introduced the concept of confidence intervals
- Discussed point estimates
- Developed confidence interval estimates
- Created confidence interval estimates for the mean (σ^2 known)
- Created confidence interval estimates for the proportion

Chapter Summary

(continued)

- Created confidence interval estimates for the variance of a normal population
- Applied the finite population correction factor to form confidence intervals when the sample size is not small relative to the population size
- Determined required sample size to meet confidence and margin of error requirements

Hypothesis Testing

Statistics for Business and Economics

8th Edition

Chapter 9

Hypothesis Testing:

Single Population

Chapter Goals

After completing this chapter, you should be able to:

- Formulate null and alternative hypotheses for applications involving
  - a single population mean from a normal distribution
  - a single population proportion (large samples)
  - the variance of a normal distribution
- Formulate a decision rule for testing a hypothesis
- Know how to use the critical value and *p*-value approaches to test the null hypothesis (for both mean and proportion problems)
- Define Type I and Type II errors and assess the power of a test
- Use the chi-square distribution for tests of the variance of a normal distribution

9.1 Concepts of Hypothesis Testing

- A hypothesis is a claim (assumption) about a population parameter:
  - population mean

Example: The mean monthly cell phone bill of this city is μ = \$52

  - population proportion

Example: The proportion of adults in this city with cell phones is P = .88

The Null Hypothesis, H_0

- States the assumption (numerical) to be tested

  Example: The average number of TV sets in U.S. Homes is equal to three $(H_0: \mu = 3)$

- Is always about a population parameter, not about a sample statistic

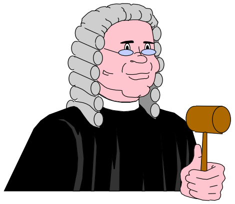
$$H_0: \mu = 3$$

$$\cancel{H_0: \bar{x} = 3}$$

The Null Hypothesis, H_0

(continued)

- Begin with the assumption that the null hypothesis is true
  - Similar to the notion of innocent until proven guilty
- Refers to the status quo
- Always contains "=" , "≤" or "≥" sign
- May or may not be rejected

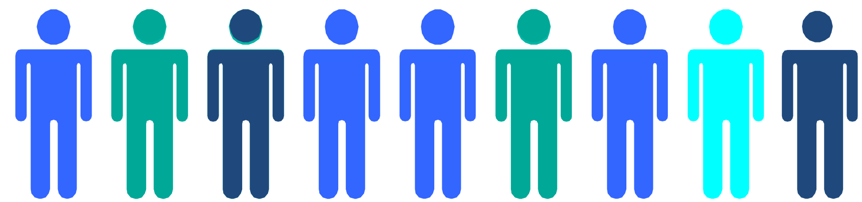


The Alternative Hypothesis, H_1

- Is the opposite of the null hypothesis
  - e.g., The average number of TV sets in U.S. homes is not equal to 3 ($H_1: \mu \neq 3$)
- Challenges the status quo
- Never contains the "=" , "≤" or "≥" sign
- May or may not be supported
- Is generally the hypothesis that the researcher is trying to support

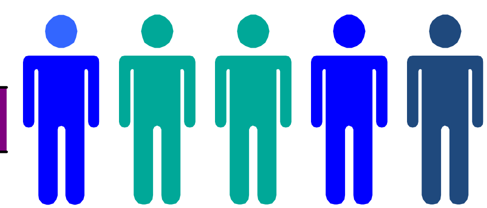
Hypothesis Testing Process

Claim: the population mean age is 50. (Null Hypothesis: $H_0: \mu = 50$)



Population

Now select a random sample



Sample

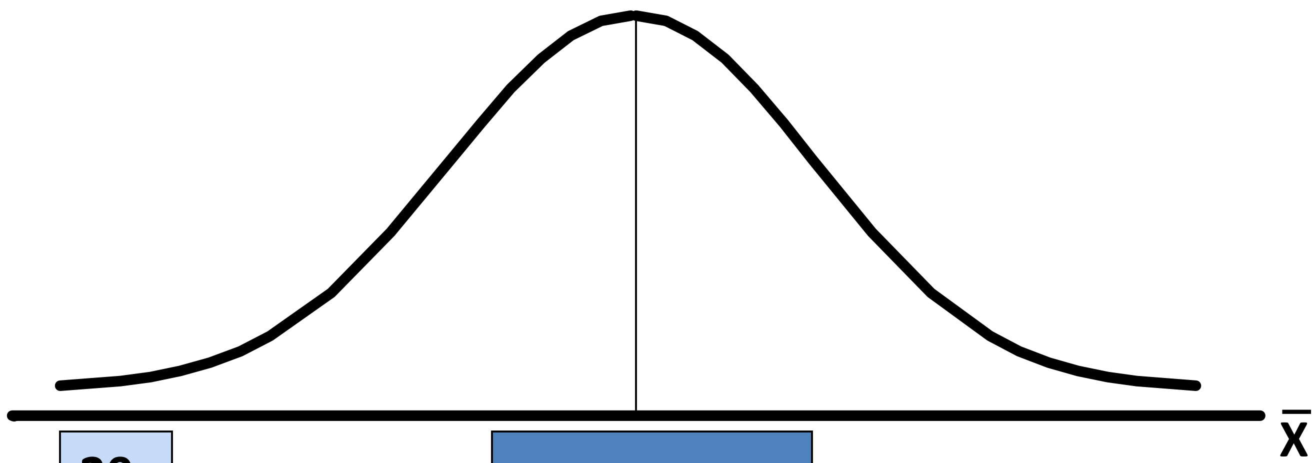
Suppose the sample mean age is 20: $\bar{x} = 20$

Is $\bar{x} = 20$ likely if $\mu = 50$?

If not likely, REJECT Null Hypothesis

Reason for Rejecting H_0

Sampling Distribution of $\bar{X}$



20

If it is unlikely that we would get a sample mean of this value ...

μ = 50

If H_0 is true

... if in fact this were the population mean...

... then we reject the null hypothesis that μ = 50.

Level of Significance, α

- **Defines the unlikely values of the sample statistic if the null hypothesis is true**
  - Defines rejection region of the sampling distribution
- Is designated by α , (level of significance)
  - Typical values are 0.01, 0.05, or 0.10
- Is selected by the researcher at the beginning
- Provides the critical value(s) of the test

Level of Significance and the Rejection Region

Level of significance = α

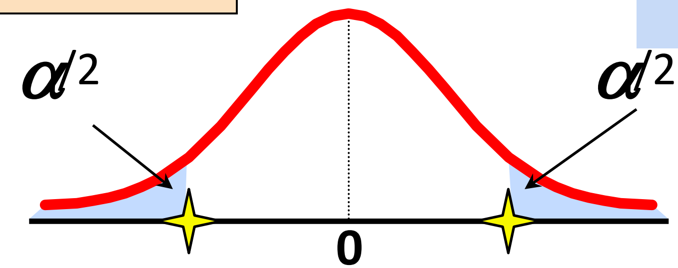
Represents critical value

Rejection region is shaded

$H_0: \mu = 3$

$H_1: \mu \neq 3$

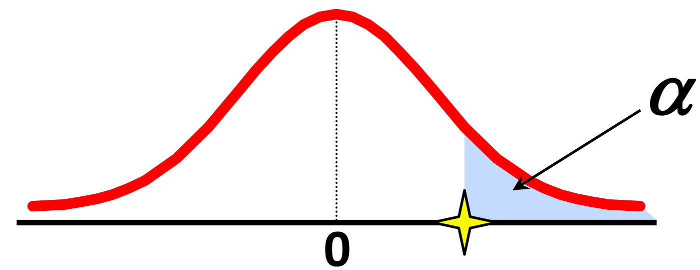
Two-tail test



$H_0: \mu \leq 3$

$H_1: \mu > 3$

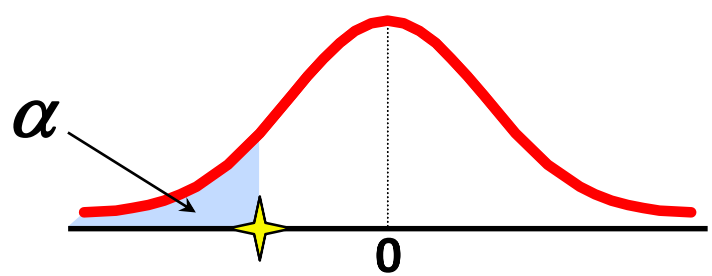
Upper-tail test



$H_0: \mu \geq 3$

$H_1: \mu < 3$

Lower-tail test



Errors in Making Decisions

- **Type I Error**
  - Reject a true null hypothesis
  - Considered a serious type of error

The probability of Type I Error is α

- Called level of significance of the test
- Set by researcher in advance

Errors in Making Decisions

(continued)

- **Type II Error**
  - Fail to reject a false null hypothesis

The probability of Type II Error is β

Outcomes and Probabilities

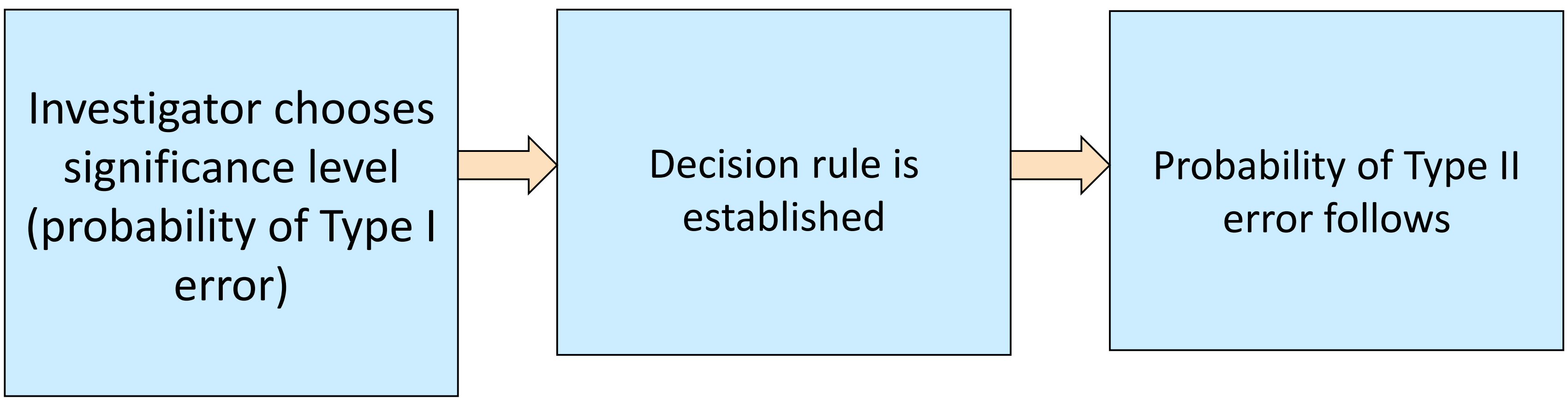
Possible Hypothesis Test Outcomes		
	Actual Situation	
Decision	H_0 True	H_0 False
Fail to Reject H_0	**Correct Decision** **$(1 - \alpha)$**	**Type II Error** **(β)**
Reject H_0	**Type I Error** **(α)**	**Correct Decision** **$(1 - \beta)$**

Key:
Outcome
(Probability)

$(1 - \beta)$ is called the power of the test

Consequences of Fixing the Significance Level of a Test

- Once the significance level α is chosen (generally less than 0.10), the probability of Type II error, β, can be found.



Type I & II Error Relationship

- Type I and Type II errors can not happen at the same time
  - Type I error can only occur if H_0 is true
  - Type II error can only occur if H_0 is false

If Type I error probability (α) ⬆, then

Type II error probability (β) ⬇

Factors Affecting Type II Error

- All else equal,
  - β $\Uparrow$ when the difference between hypothesized parameter and its true value $\Downarrow$
  - β $\Uparrow$ when α $\Downarrow$
  - β $\Uparrow$ when σ $\Uparrow$
  - β $\Uparrow$ when n $\Downarrow$

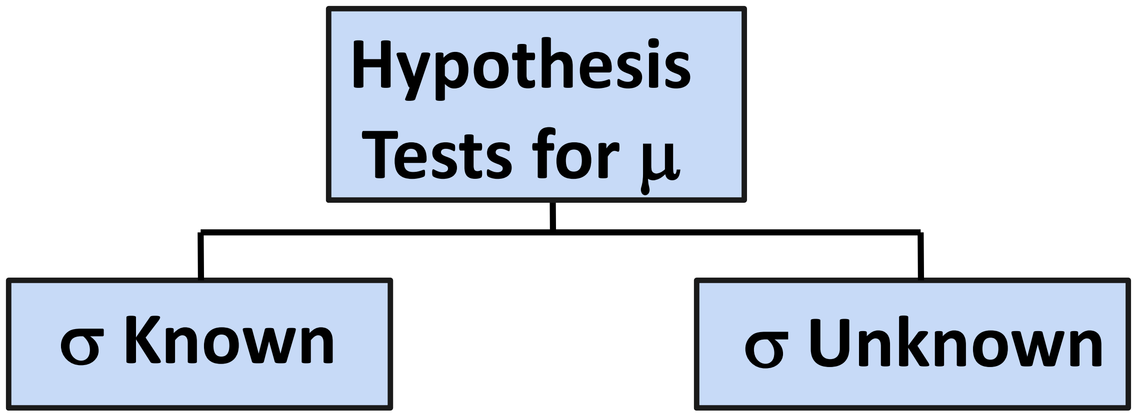
Power of the Test

- The power of a test is the probability of rejecting a null hypothesis that is false

- i.e., Power = P(Reject H_0 | H_1 is true)

  - Power of the test increases as the sample size increases

Hypothesis Tests for the Mean



9.2 Tests of the Mean of a Normal Distribution (σ Known)

- Convert sample result ($\bar{x}$) to a z value

Hypothesis Tests for μ	
σ Known	σ Unknown

Consider the test

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu > \mu_0$$

(Assume the population is normal)

The decision rule is:

$$\text{Reject } H_0 \text{ if } z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > z_\alpha$$

Decision Rule

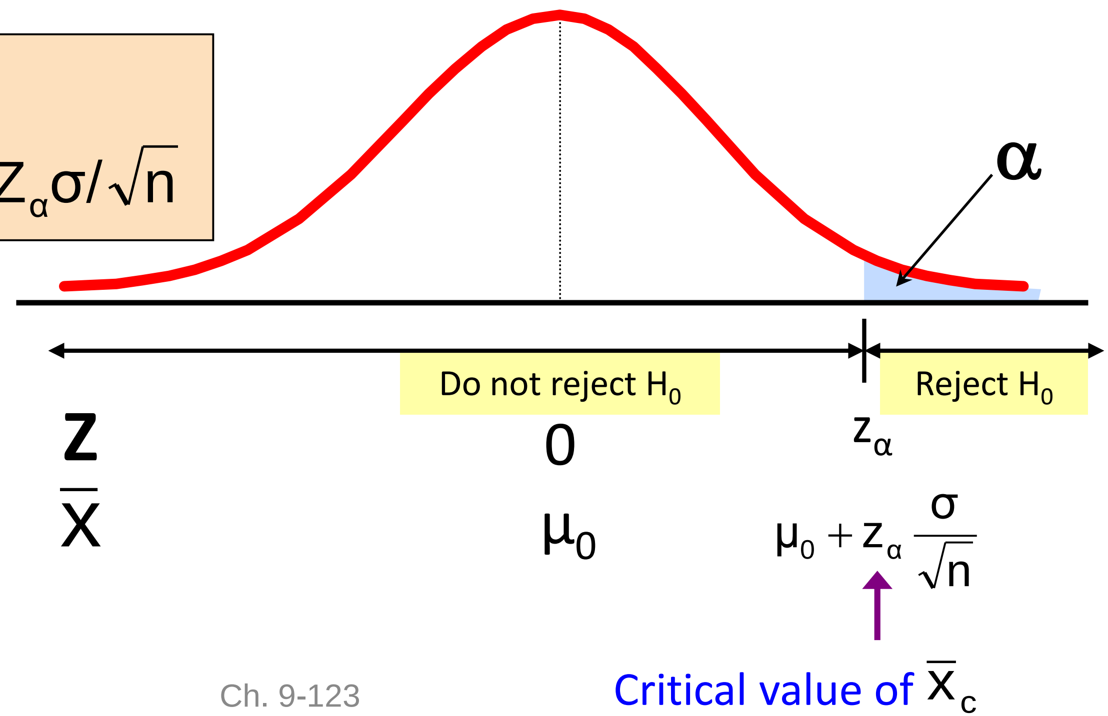
Reject H_0 if $z = \dfrac{\bar{x} - \mu_0}{\dfrac{\sigma}{\sqrt{n}}} > z_\alpha$

H_0: $\mu = \mu_0$

H_1: $\mu > \mu_0$

Alternate rule:

Reject H_0 if $\bar{x} > \mu_0 + Z_\alpha \sigma / \sqrt{n}$



p-Value

- *p*-value: Probability of obtaining a test statistic more extreme ($\leq$ or $\geq$) than the observed sample value given H_0 is true
  - Also called observed level of significance
  - Smallest value of α for which H_0 can be rejected

p-Value Approach to Testing

- Convert sample result (e.g., $\bar{x}$) to test statistic (e.g., z statistic)
- Obtain the *p*-value
  - For an upper tail test:

$$\begin{aligned} p\text{-value} &= P\left(z > \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}, \text{ given that } H_0 \text{ is true}\right) \\ &= P\left(z > \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \;\middle|\; \mu = \mu_0\right) \end{aligned}$$

- Decision rule: compare the *p*-value to α

  - If *p*-value $< \alpha$, reject H_0
  - If *p*-value $\geq \alpha$, do not reject H_0

Example: Upper-Tail Z Test for Mean (σ Known)

A phone industry manager thinks that customer monthly cell phone bill have increased, and now average over \$52 per month. The company wishes to test this claim. (Assume $\sigma = 10$ is known)

Form hypothesis test:

H_0: $\mu \leq 52$ the average is not over \$52 per month

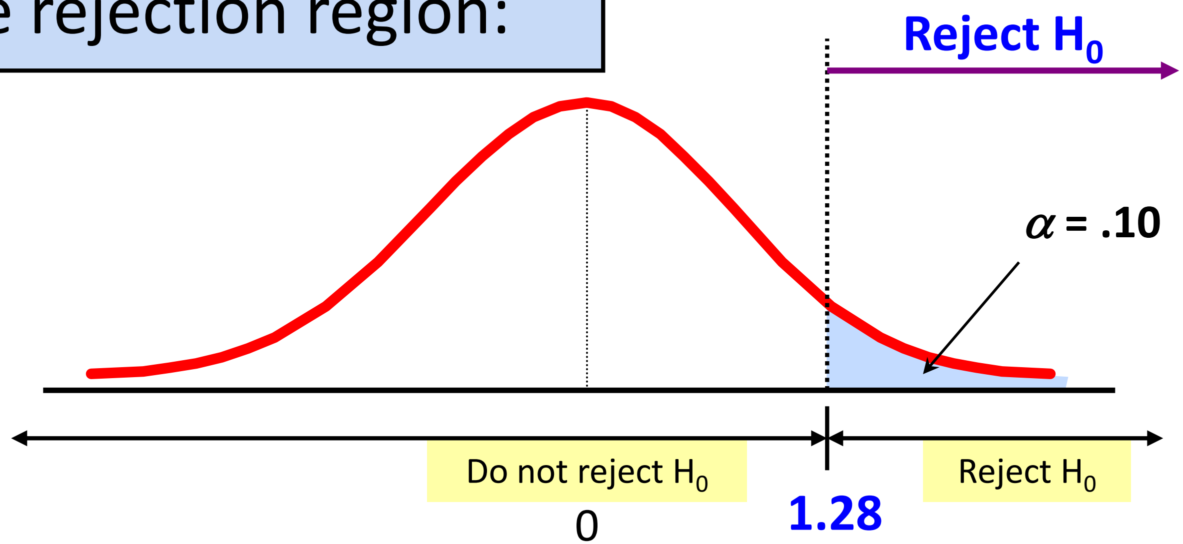
H_1: $\mu > 52$ the average is greater than \$52 per month (i.e., sufficient evidence exists to support the manager's claim)

Example: Find Rejection Region

(continued)

- Suppose that α = .10 is chosen for this test

Find the rejection region:



$$\text{Reject } H_0 \text{ if } \quad z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > 1.28$$

Example: Sample Results

(continued)

Obtain sample and compute the test statistic

Suppose a sample is taken with the following results: $n = 64$, $\bar{x} = 53.1$ ($\sigma = 10$ was assumed known)

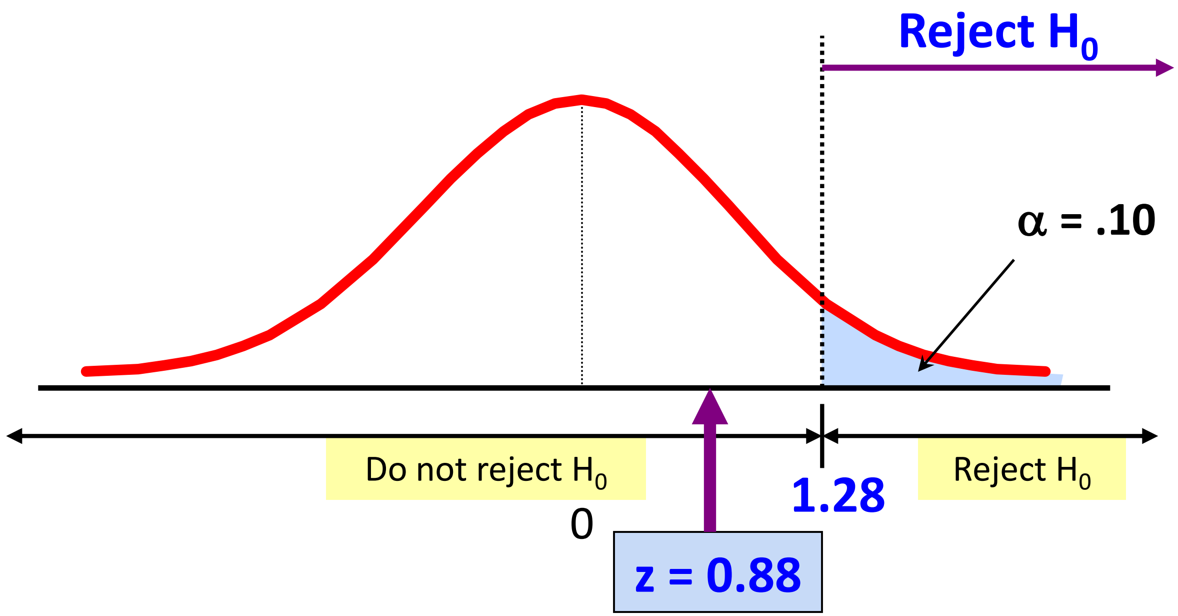
– Using the sample results,

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{53.1 - 52}{\frac{10}{\sqrt{64}}} = 0.88$$

Example: Decision

(continued)

Reach a decision and interpret the result:



Do not reject H_0 since z = 0.88 < 1.28

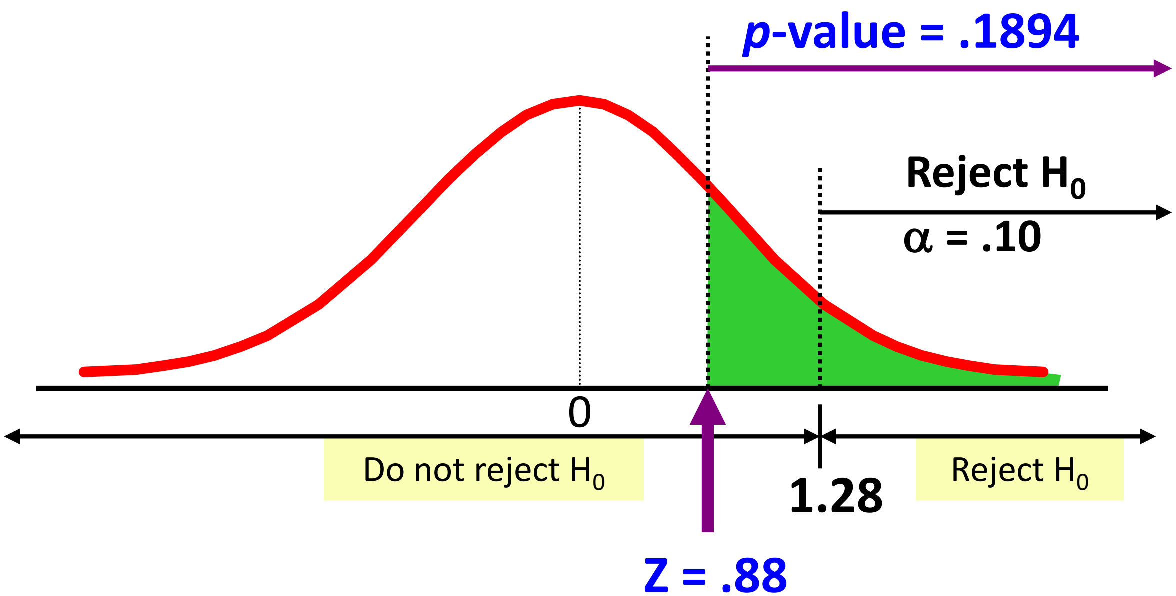
i.e.: there is not sufficient evidence that the mean bill is over $52

Example: p-Value Solution

(continued)

Calculate the *p*-value and compare to α

(assuming that μ = 52.0)



$$P(\bar{x} \geq 53.1 | \mu = 52.0)$$

$$= P\left(z \geq \frac{53.1 - 52.0}{10/\sqrt{64}} \right)$$

$$= P(z \geq 0.88) = 1 - .8106$$

$$= .1894$$

Do not reject H_0 since *p*-value = .1894 > α = .10

One-Tail Tests

- In many cases, the alternative hypothesis focuses on one particular direction

H_0: $\mu \le 3$

H_1: $\mu > 3$

This is an upper-tail test since the alternative hypothesis is focused on the upper tail above the mean of 3

H_0: $\mu \ge 3$

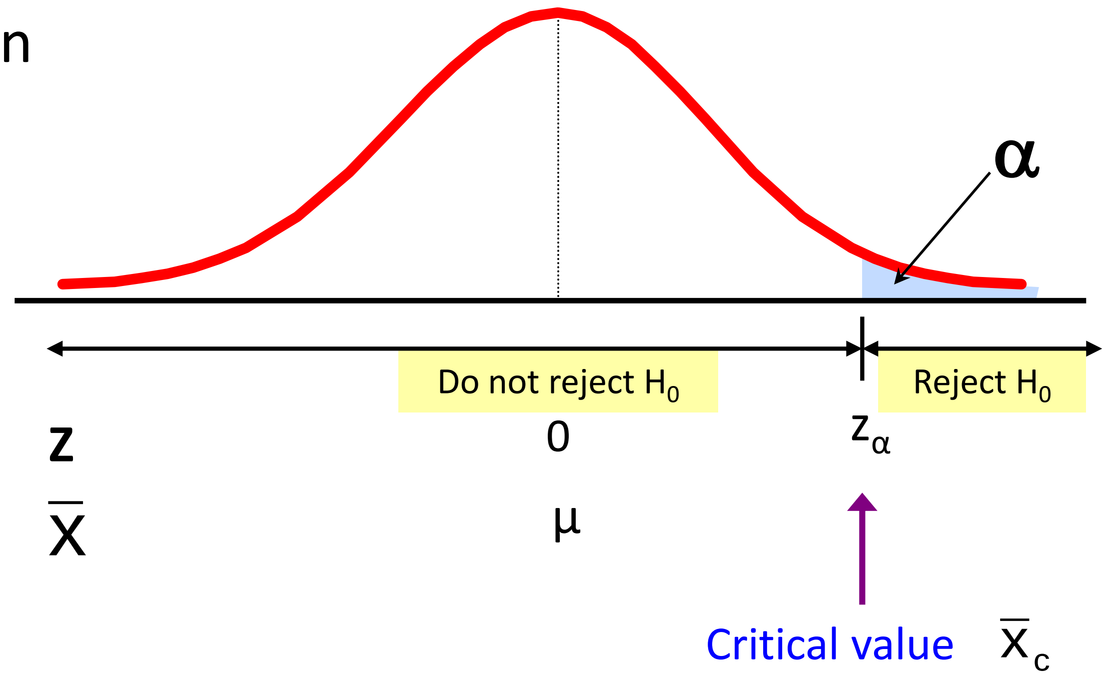
H_1: $\mu < 3$

This is a lower-tail test since the alternative hypothesis is focused on the lower tail below the mean of 3

Upper-Tail Tests

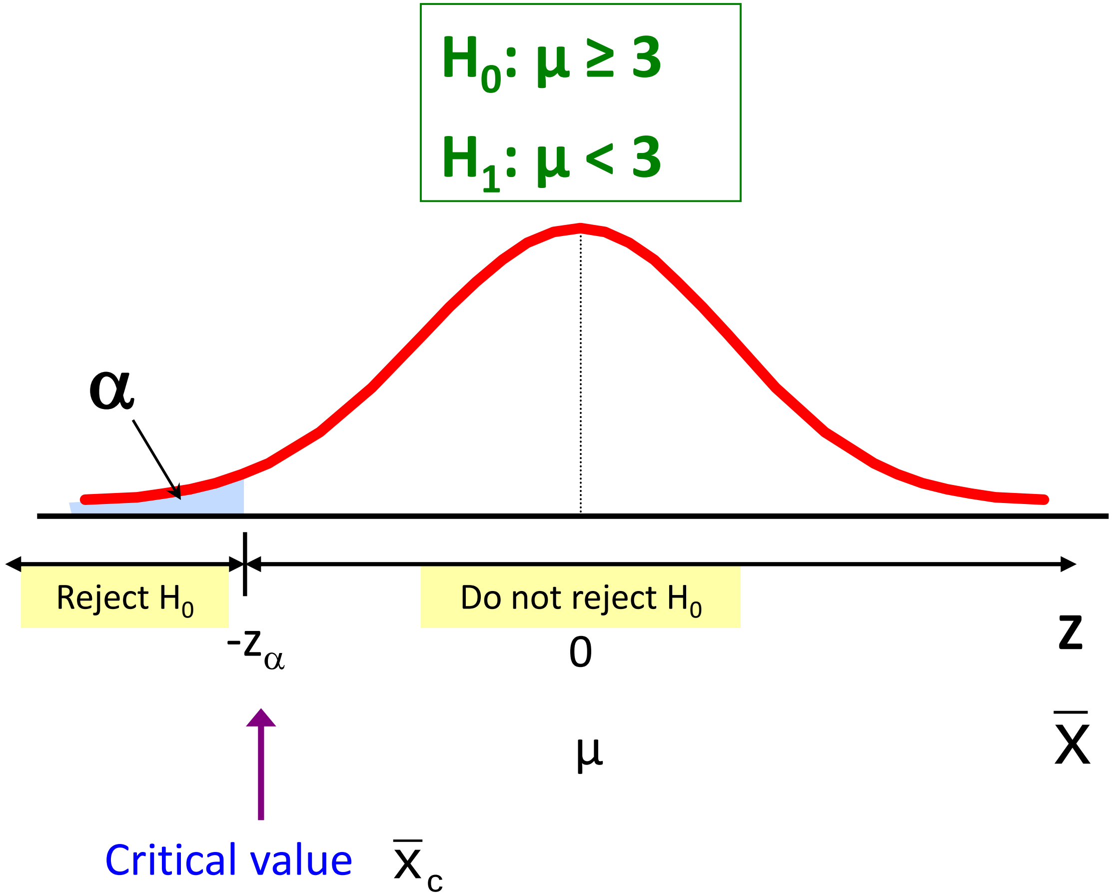
- There is only one critical value, since the rejection area is in only one tail

H_0: $\mu \leq 3$
H_1: $\mu > 3$



Lower-Tail Tests

- There is only one critical value, since the rejection area is in only one tail

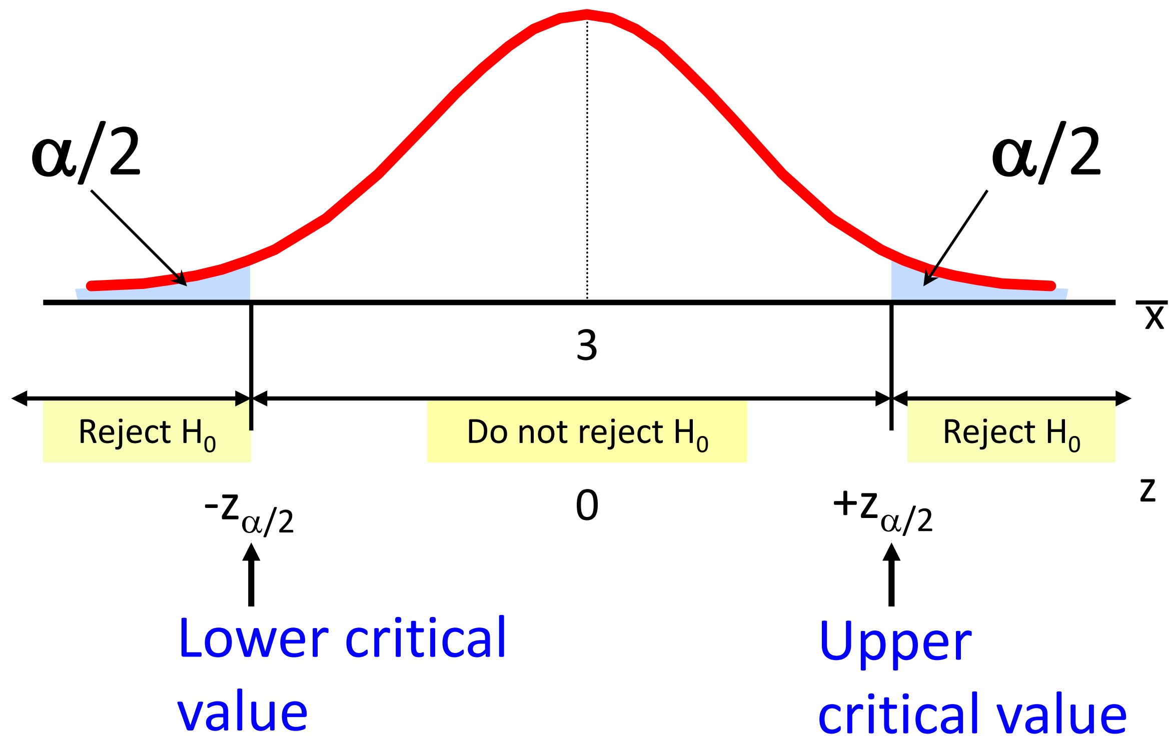


Two-Tail Tests

- In some settings, the alternative hypothesis does not specify a unique direction

H_0: $\mu = 3$ H_1: $\mu \neq 3$

- There are two critical values, defining the two regions of rejection



Hypothesis Testing Example

Test the claim that the true mean # of TV sets in US homes is equal to 3.
(Assume $\sigma = 0.8$)

- State the appropriate null and alternative hypotheses
  - H_0: $\mu = 3$, H_1: $\mu \neq 3$ (This is a two tailed test)
- Specify the desired level of significance
  - Suppose that $\alpha = .05$ is chosen for this test
- Choose a sample size
  - Suppose a sample of size $n = 100$ is selected

Hypothesis Testing Example

(continued)

- Determine the appropriate technique
  - σ is known so this is a z test
- Set up the critical values
  - For $\alpha = .05$ the critical z values are ±1.96
- Collect the data and compute the test statistic
  - Suppose the sample results are

    $n = 100$, $\bar{x} = 2.84$ ($\sigma = 0.8$ is assumed known)

So the test statistic is:

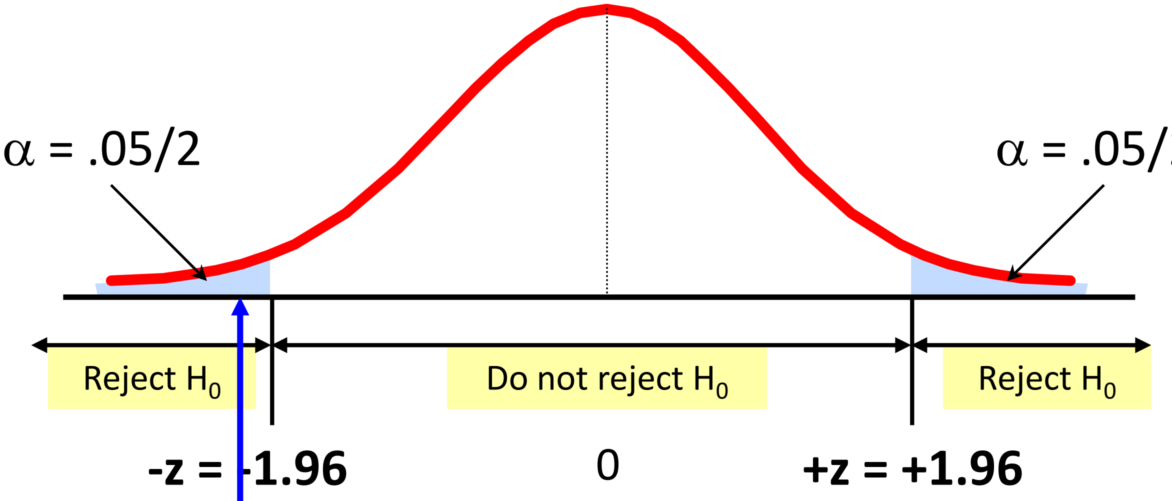
$$z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{2.84 - 3}{\frac{0.8}{\sqrt{100}}} = \frac{-.16}{.08} = -2.0$$

Hypothesis Testing Example

(continued)

- Is the test statistic in the rejection region?

Reject H_0 if $z < -1.96$ or $z > 1.96$; otherwise do not reject H_0

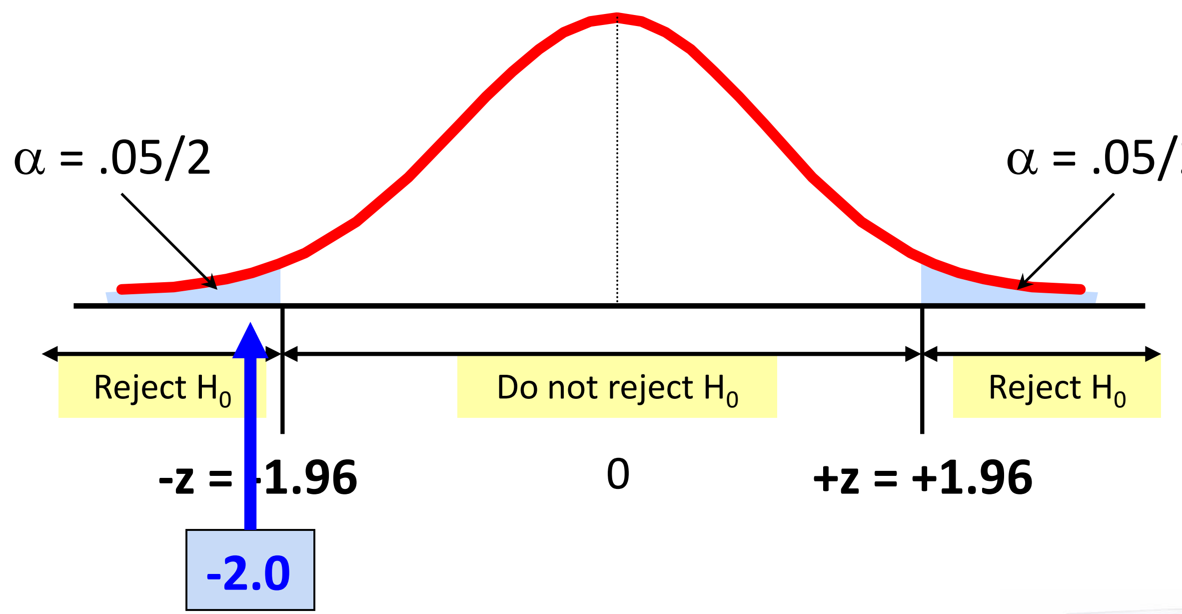


Here, $z = -2.0 < -1.96$, so the test statistic is in the rejection region

Hypothesis Testing Example

(continued)

- Reach a decision and interpret the result



Since $z = -2.0 < -1.96$, we <u>reject the null hypothesis</u> and conclude that there is sufficient evidence that the mean number of TVs in US homes is not equal to 3

Example: *p*-Value

- **Example:** How likely is it to see a sample mean of 2.84 (or something further from the mean, in either direction) if the true mean is μ = 3.0?

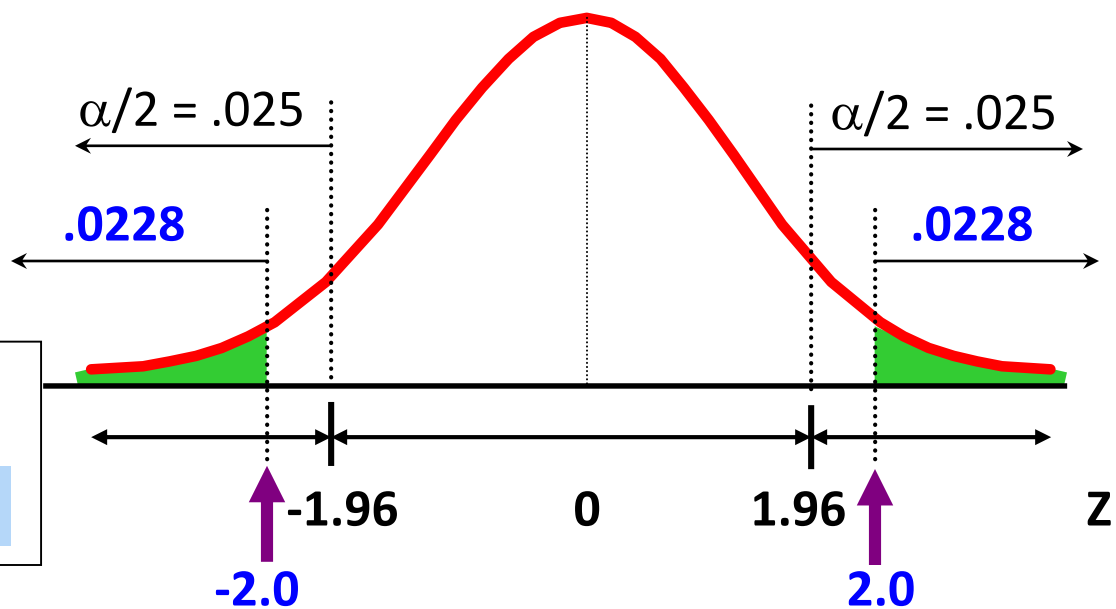
$\bar{x}$ = 2.84 is translated to a z score of z = -2.0

$$P(z < -2.0) = .0228$$

$$P(z > 2.0) = .0228$$

***p*-value**

= .0228 + .0228 = .0456



Example: p-Value

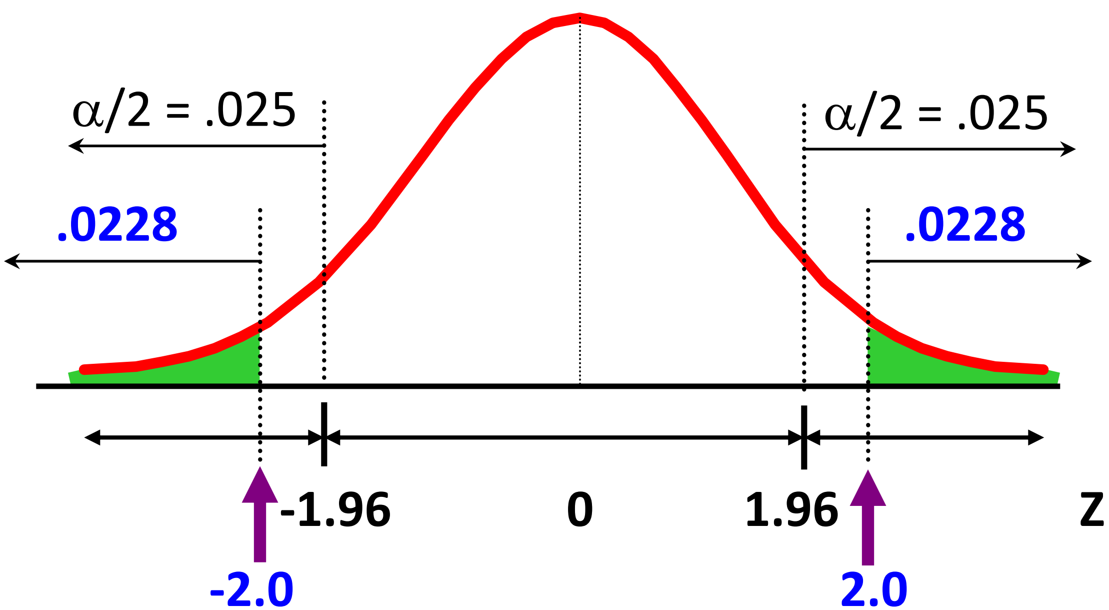
(continued)

- Compare the p-value with α
  - If p-value $<$ α , reject H_0
  - If p-value $\geq$ α , do not reject H_0

Here: p-value = .0456

α = .05

Since .0456 < .05, we reject the null hypothesis



9.3 Tests of the Mean of a Normal Population (σ Unknown)

- Convert sample result ($\bar{x}$) to a t test statistic

Hypothesis Tests for μ

σ Known

σ Unknown

Consider the test

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu > \mu_0$$

(Assume the population is normal)

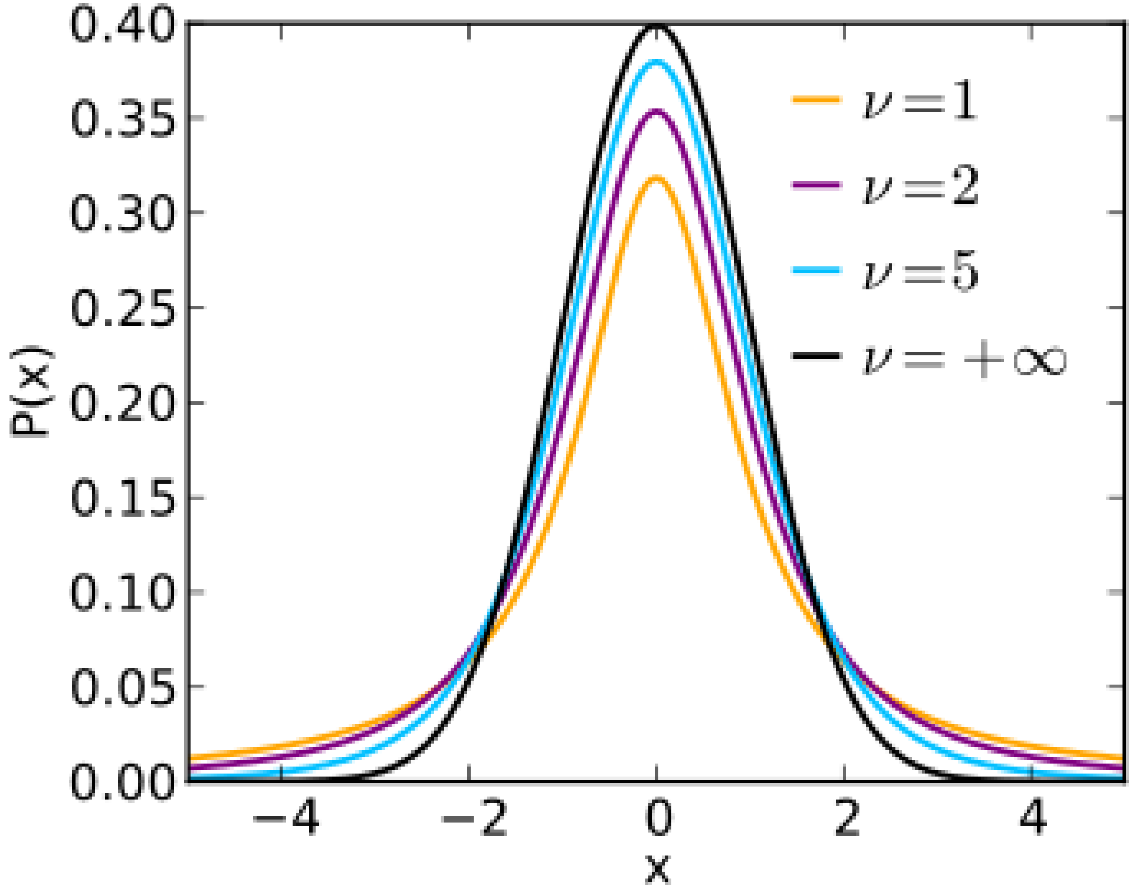
The decision rule is:

$$\text{Reject } H_0 \text{ if } \quad t = \frac{\bar{x} - \mu_0}{\dfrac{s}{\sqrt{n}}} > t_{n-1,\,\alpha}$$

Student's t-distribution

$$\overline{X} = \pm ? \times \frac{s}{\sqrt{n}}$$

$$t_{n-1,\alpha} > z_{\alpha}$$



Probability density function

https://en.wikipedia.org/

Tests of the Mean of a Normal Population (σ Unknown)

(continued)

- For a two-tailed test:

Consider the test

$$\boxed{\begin{aligned} &H_0 : \mu = \mu_0 \\ &H_1 : \mu \neq \mu_0 \end{aligned}}$$

(Assume the population is normal, and the population variance is unknown)

The decision rule is:

$$\text{Reject } H_0 \ \boxed{\text{if} \quad t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} < -t_{n\text{-}1,\,\alpha/2}} \ \text{or} \ \boxed{\text{if} \quad t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} > t_{n\text{-}1,\,\alpha/2}}$$

Example: Two-Tail Test (σ Unknown)

The average cost of a hotel room in Chicago is said to be \$168 per night. A random sample of 25 hotels resulted in

$\bar{x}$ = \$172.50 and

s = \$15.40. Test at the

α = 0.05 level.

(Assume the population distribution is normal)

H_0: μ = 168

H_1: $\mu \neq$ 168

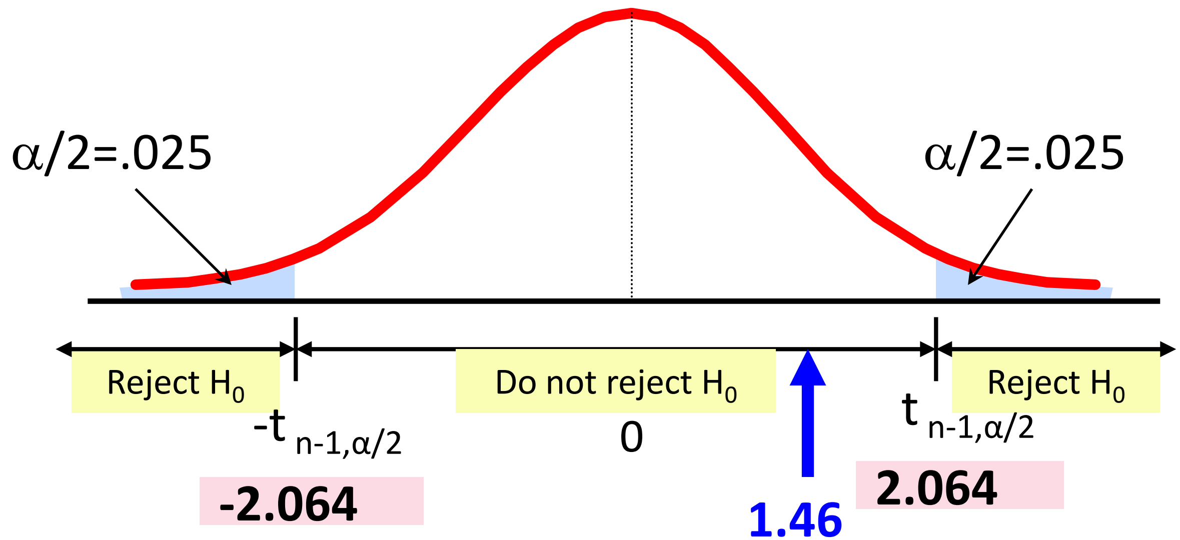
Example Solution: Two-Tail Test

H_0: $\mu = 168$

H_1: $\mu \neq 168$

- $\alpha = 0.05$
- $n = 25$
- σ is unknown, so use a **t statistic**
- **Critical Value:**

$t_{24\,,\,.025} = \pm\, 2.064$



$$t_{n-1} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{172.50 - 168}{\frac{15.40}{\sqrt{25}}} = 1.46$$

Do not reject H_0: not sufficient evidence that true mean cost is different than $168

9.6 Tests of the Variance of a Normal Distribution

- Goal: Test hypotheses about the population variance, σ^2 (e.g., H_0: $\sigma^2 = \sigma_0^2$)
  - If the population is normally distributed,

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

has a chi-square distribution with (n – 1) degrees of freedom

Tests of the Variance of a Normal Distribution

(continued)

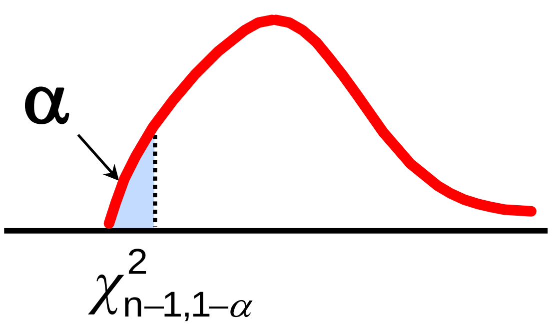
The test statistic for hypothesis tests about one population variance is

$$\boxed{\chi^2_{n-1} = \frac{(n-1)s^2}{\sigma_0^2}}$$

Decision Rules: Variance

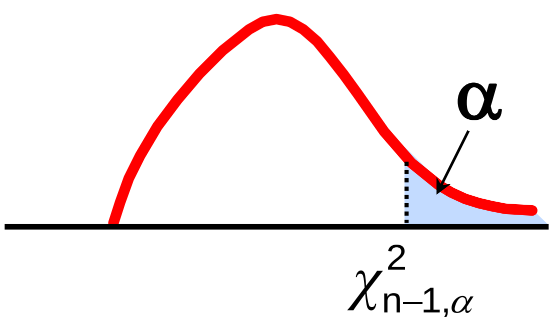
Population variance

Lower-tail test:	Upper-tail test:	Two-tail test:
H_0: $\sigma^2 \geq \sigma_0^2$ H_1: $\sigma^2 < \sigma_0^2$	H_0: $\sigma^2 \leq \sigma_0^2$ H_1: $\sigma^2 > \sigma_0^2$	H_0: $\sigma^2 = \sigma_0^2$ H_1: $\sigma^2 \neq \sigma_0^2$



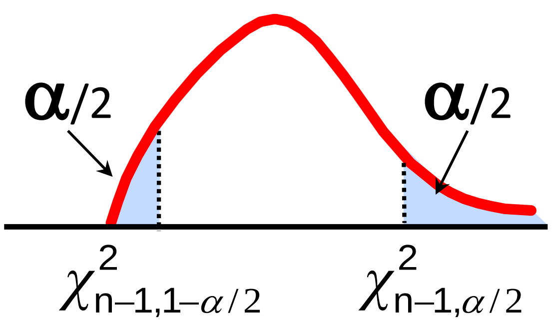
Reject H_0 if

$\chi^2_{n-1} < \chi^2_{n-1,1-\alpha}$



Reject H_0 if

$\chi^2_{n-1} > \chi^2_{n-1,\alpha}$



Reject H_0 if

$\chi^2_{n-1} > \chi^2_{n-1,\alpha/2}$

or $\chi^2_{n-1} < \chi^2_{n-1,1-\alpha/2}$

Overview of Matrices

An m x n matrix is a rectangular array of real numbers with m rows and n columns.

$$A = \begin{matrix} a11 & \dots & a1n \\ am1 & \dots & amn \end{matrix}$$

- **$A + B = (a_{ij}) + (b_{ij}) = (a_{ij} + b_{ij})$ Addition**
- **$\lambda A = A\lambda = (\lambda a_{ij})$**

Overview of Matrices (2)

An m x n matrix is a rectangular array of real numbers with m rows and n columns.

$$A = \begin{matrix} a11 & \ldots & a1n \\ am1 & \ldots & amn \end{matrix}$$

- **A + B = B + A**
- **(A + B) + C = A + (B + C)**
- **A + (-1) A = (0)**
- **(λ + μ) A = λ A + μ A**
- **λ(A + B) = λ A + λ B**
- **λ (μ A) = (λ μ) A**

Overview of Matrices

An m x n matrix is a rectangular array of real numbers with m rows and n columns.

$$A = \begin{matrix} a11 & \ldots & a1n \\ am1 & \ldots & amn \end{matrix}$$

- **A B = (a_{ij}) (b_{jk}) = $\sum_{j=1}^{m} a_{ij}$ bjk**
- **(AB)C = A(BC)**
- **A(B + C) = AB + AC**
- **(A + B) C = AC + BC**

Overview of Matrices

An m x n matrix is a rectangular array of real numbers with m rows and n columns.

A = $$\begin{matrix} \boldsymbol{a11} & \dots & \boldsymbol{a1n} \\ \boldsymbol{am1} & \dots & \boldsymbol{amn} \end{matrix}$$

- **A B = (a_{ij}) (b_{jk}) = $\sum_{j=1}^{m} a_{ij}$ bjk**
- **(AB)C = A(BC)**
- **A(B + C) = AB + AC**
- **(A + B) C = AC + BC**

Overview of Matrices

An m x n matrix with 1's on the diagonal and 0's elsewhere is called the identity matrix

$$I = \begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{matrix}$$

- **$AB = (a_{ij})(b_{jk}) = \sum_{j=1}^{m} a_{ij}$ bjk**
- **I A = A I = A**
- **Associated with any matrix A, is a matrix A^T, Transpose of A. If $A = (a_{ij})$, $A^T = (a_{ji})$**

Associated with any square matrix is a determinant | A |

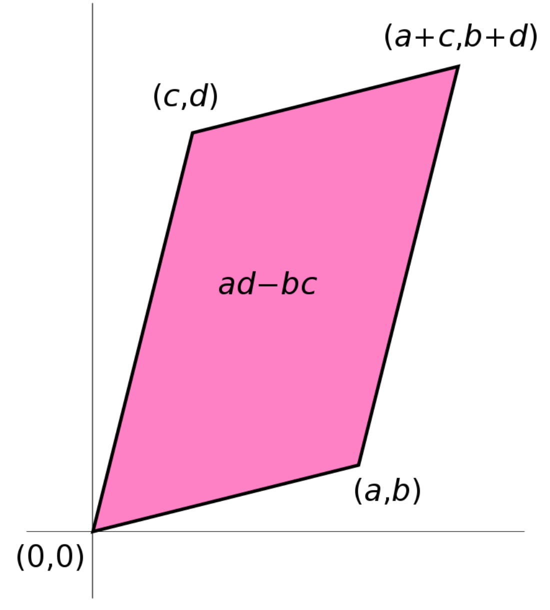
Overview of Matrices (source Wikipedia)

$$\det(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} A_{i,\sigma_i}.$$

Here the sum is computed over all permutations σ of the set {1, 2, ..., n}

1. $\det(I_n) = 1$ where I_n is the $n \times n$ identity matrix.
2. $\det(A^{\mathrm{T}}) = \det(A).$
3. $\det(A^{-1}) = \dfrac{1}{\det(A)}.$
4. For square matrices A and B of equal size,
$$\det(AB) = \det(A)\det(B).$$
5. $\det(cA) = c^n \det(A)$ for an $n \times n$ matrix.
6. If A is a triangular matrix, i.e. $a_{i,j} = 0$ whenever $i > j$ or, alternatively whenever $i < j$, then its determinant equals the product of the diagonal entries:

$$\det(A) = a_{1,1}a_{2,2}\cdots a_{n,n} = \prod_{i=1}^{n} a_{i,i}.$$



Overview of Matrices
Determinants
(source Wikipedia)

For a matrix equation

$$Ax = b$$

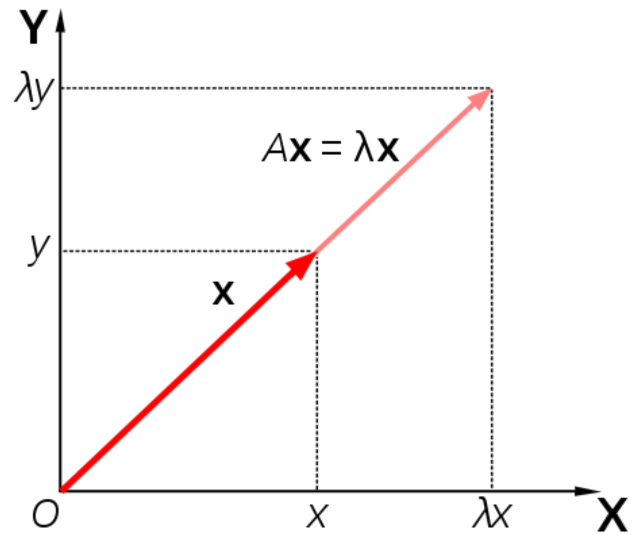
the solution is given by Cramer's rule:

$$x_i = \frac{\det(A_i)}{\det(A)} \qquad i = 1, \ldots, n$$

where A_i is the matrix formed by replacing the ith column of A by the column vector b.

Characteristic Equation

- Eigenvalues: Solution to Det(A –λI) = 0
- Eigenvectors: Solution to (A-λI)X = 0



STEVENS
INSTITUTE of TECHNOLOGY
School of Business

stevens.edu

Amir H Gandomi; PhD
Assistant Professor of Analytics & Information Systems
a.h.gandomi@stevens.edu