Chapter 7.4 & 7.5

Sampling Distributions and the Central Limit Theorem

Learning Objectives

At the end of this lecture, the student should be able to:

- State the statistical notation for parameters and statistics for two measures of variation.
- Name one type of inference and describe it.
- Explain the difference between a frequency distribution and sampling distribution
- Describe the Central Limit Theorem in either words or formulas.
- Describe how to calculate standard error

Introduction

- Parameters, Statistics, and Inferences
- Introduction to Sampling Distribution
- Central Limit Theorem
- Finding Probabilities Regarding X-bar

Photograph by David Hawgood

Parameters, Statistics, and Inferences

Review and Overview

Reminder of Statistic and Parameter

- A *statistic* is a numerical measure describing a *sample*.

Statistic

Photographs by Che and Sandstein

Reminder of Statistic and Parameter

- A *statistic* is a numerical measure describing a *sample*.
- A *parameter* is a numerical measure describing a *population*.

Parameter

Statistic

Photographs by Che and Sandstein

Notation of Statistics and Parameters

Measure	Statistic	Parameter
Mean	$\bar{x}$ (x-bar)	μ (mu)
Variance	s^2	σ^2 (sigma squared)
Standard Deviation	s	σ (sigma)
Proportion	$\hat{p}$ (p-hat)	p

Inferences

Photograph by Tomasz Sienicki

Inferences

Photograph by Tomasz Sienicki

Types of Inferences

1. Estimation: we estimate the value of a population *parameter* using a *sample*

Types of Inferences

1. Estimation: we estimate the value of a population *parameter* using a *sample*

We will practice this in Chapter 8

Types of Inferences

1. Estimation: we estimate the value of a population *parameter* using a *sample*
2. Testing: we do a test to help us make a *decision* about a population parameter

Types of Inferences

1. <u>Estimation:</u> we estimate the value of a population *parameter* using a *sample*
2. <u>Testing:</u> we do a test to help us make a *decision* about a population parameter

Chapter 9

Types of Inferences

1. Estimation: we estimate the value of a population *parameter* using a *sample*
2. Testing: we do a test to help us make a *decision* about a population parameter
3. Regression: we make *predictions or forecasts* about a statistic

Types of Inferences

1. Estimation: we estimate the value of a population *parameter* using a *sample*
2. Testing: we do a test to help us make a *decision* about a population parameter
3. Regression: we make *predictions or forecasts* about a statistic

We already did this in Chapter 4.2

Types of Inferences

1. Estimation: we estimate the value of a population *parameter* using a *sample*
2. Testing: we do a test to help us make a *decision* about a population parameter
3. Regression: we make *predictions or forecasts* about a statistic

Requires understanding sampling distributions and the Central Limit Theorem

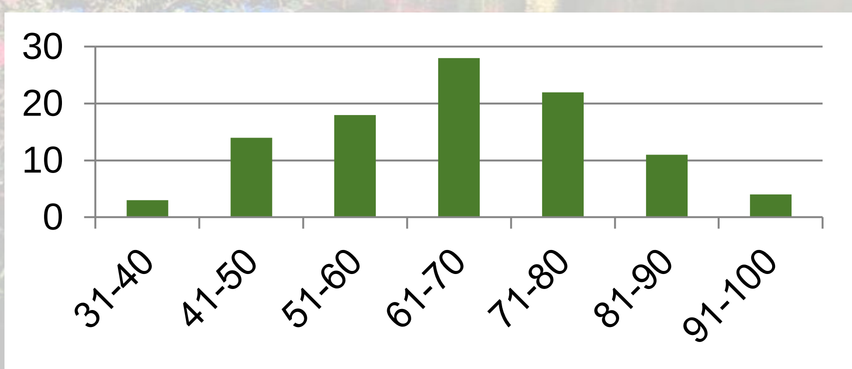
Introduction to Sampling Distribution

Different from Frequency Distribution

Frequency vs. Sampling Distribution

Frequency Distribution

1. Make a histogram of a quantitative variable.
2. Draw the shape and name the distribution.

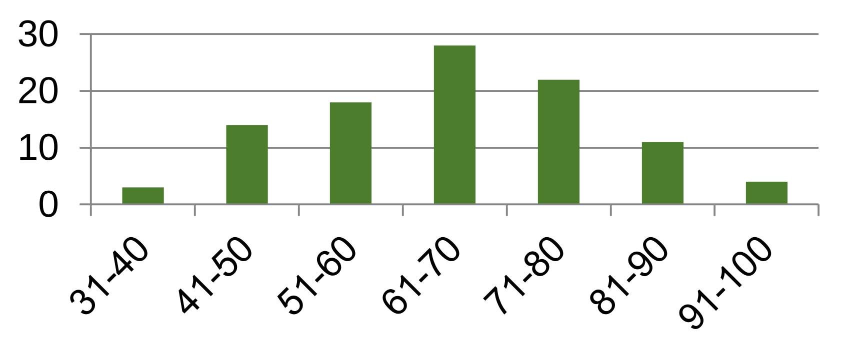


Sampling Distribution

Frequency vs. Sampling Distribution

Frequency Distribution

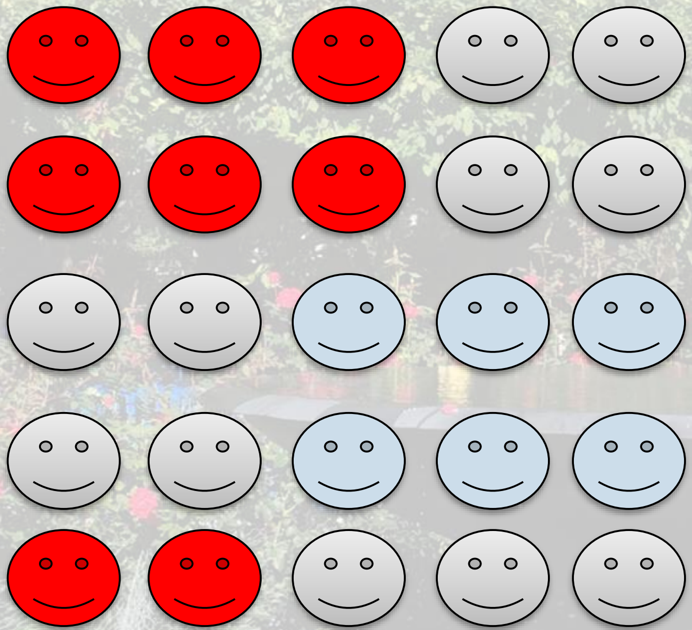
1. Make a histogram of a quantitative variable.
2. Draw the shape and name the distribution.



Sampling Distribution

1. Start with a population.
2. Decide on an n.
3. Take as many samples of n as possible from the population.
4. Make an x-bar for each sample.
5. Make a histogram of all the x-bars.

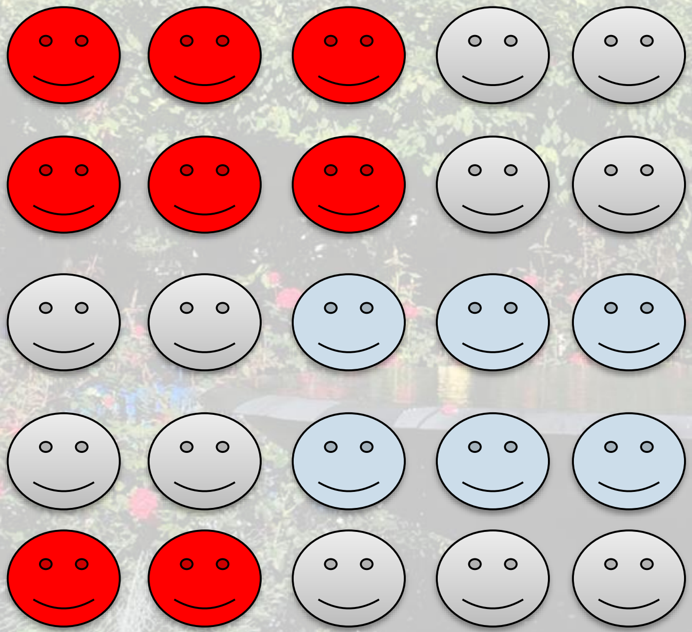
Explanation of Sampling Distribution



Sampling Distribution

1. Start with a population.
2. Decide on an n.
3. Take as many samples of n as possible from the population.
4. Make an x-bar for each sample.
5. Make a histogram of all the x-bars.

Explanation of Sampling Distribution

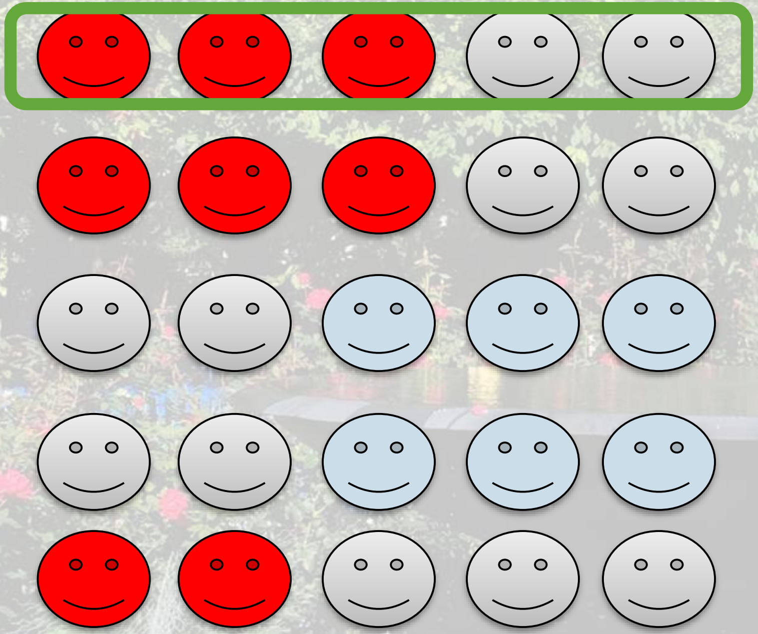


Sampling Distribution

1. Start with a population.
2. Decide on an n. **n=5**
3. Take as many samples of n as possible from the population.
4. Make an x-bar for each sample.
5. Make a histogram of all the x-bars.

Explanation of Sampling Distribution

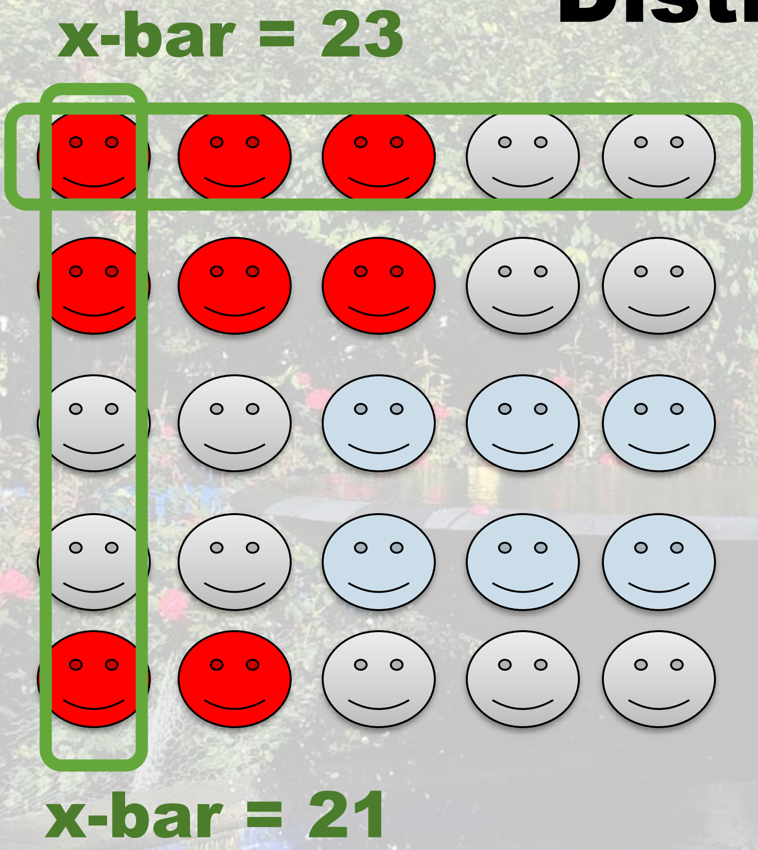
x-bar = 23



Sampling Distribution

1. Start with a population.
2. Decide on an n. **n=5**
3. Take as many samples of n as possible from the population.
4. Make an x-bar for each sample.
5. Make a histogram of all the x-bars.

Explanation of Sampling Distribution



Sampling Distribution

1. Start with a population.
2. Decide on an n. **n=5**
3. Take as many samples of n as possible from the population.
4. Make an x-bar for each sample.
5. Make a histogram of all the x-bars.

Explanation of Sampling Distribution

Sampling Distribution

1. Start with a population.
2. Decide on an n. **n=5**
3. Take as many samples of n as possible from the population.
4. Make an x-bar for each sample.
5. Make a histogram of all the x-bars.

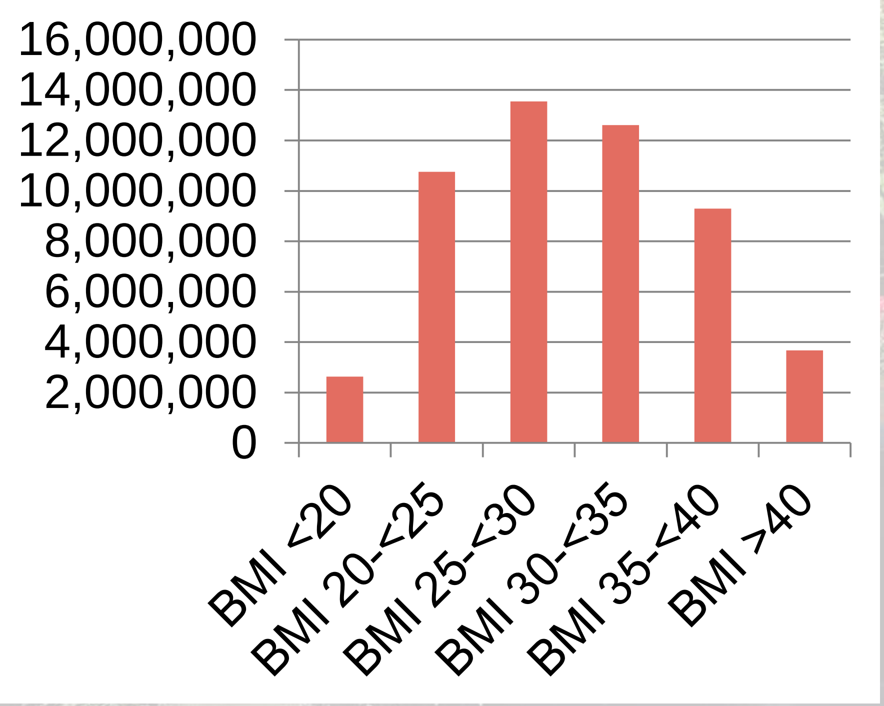
Explanation of Sampling Distribution

Class Limits of BMI	Frequency of x-bars
BMI <20	2,626,094
BMI 20-<25	10,758,762
BMI 25-<30	13,554,687
BMI 30-<35	12,605,250
BMI 35-<40	9,300,551
BMI >40	3,676,531
Total	52,521,875

Sampling Distribution

1. Start with a population.
2. Decide on an n. **n=5**
3. Take as many samples of n as possible from the population.
4. Make an x-bar for each sample.
5. Make a histogram of all the x-bars.

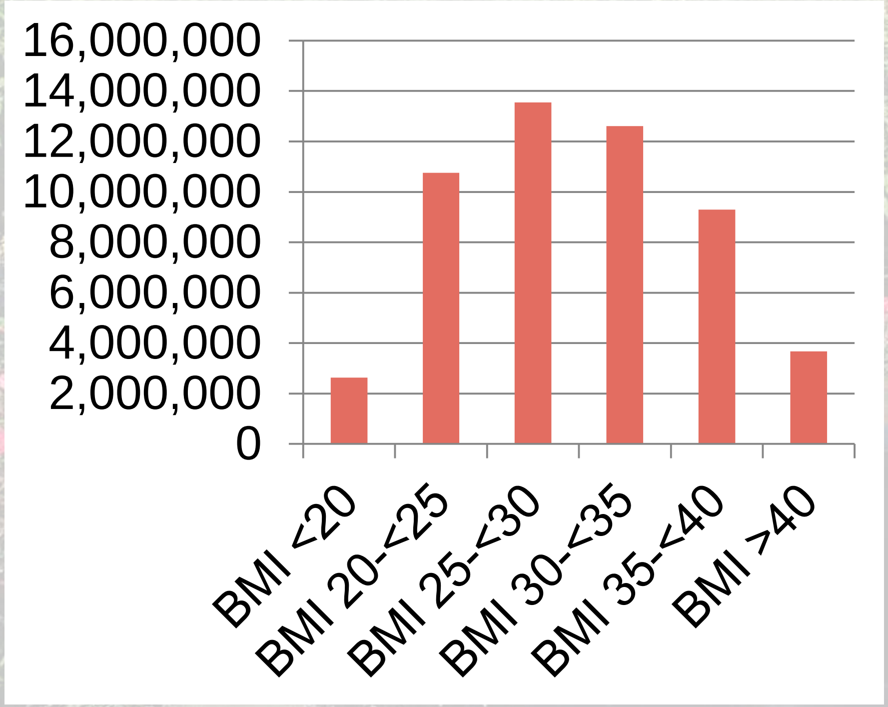
Explanation of Sampling Distribution



Sampling Distribution

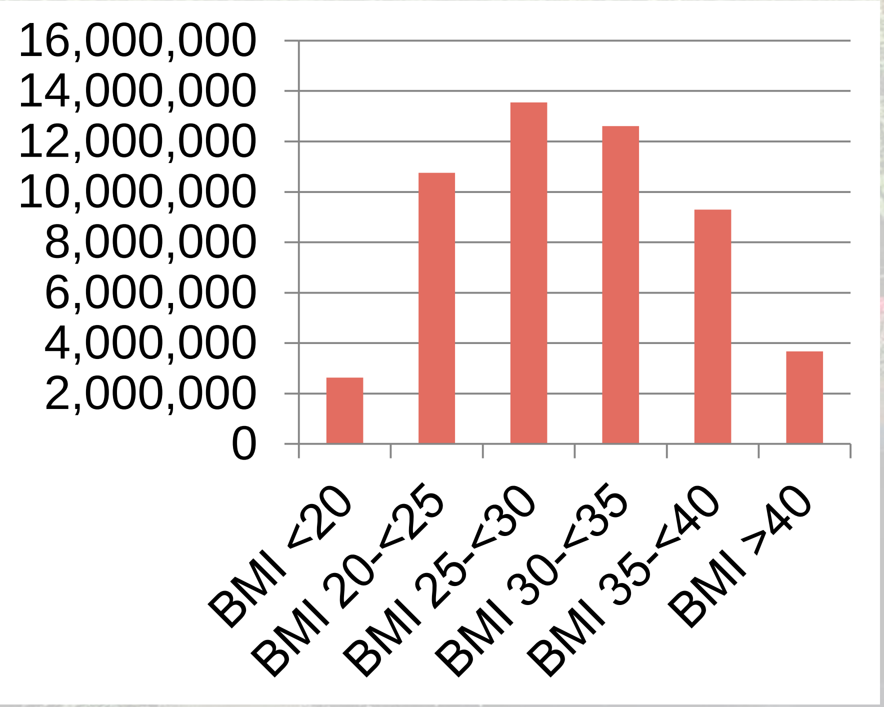
1. Start with a population.
2. Decide on an n. **n=5**
3. Take as many samples of n as possible from the population.
4. Make an x-bar for each sample.
5. Make a histogram of all the x-bars.

Definition of a Sampling Distribution



- A sampling distribution is a probability distribution of a *sample statistic* based *on all possible simple random samples* of the *same size* from the *same population*.

Definition of a Sampling Distribution



- A sampling distribution is a probability distribution of a *sample statistic* based *on all possible simple random samples* of the *same size* from the *same population*.
- In the next section, we will talk about the Central Limit Theorem, which is a proof that shows how we can use a sampling distribution for inference.

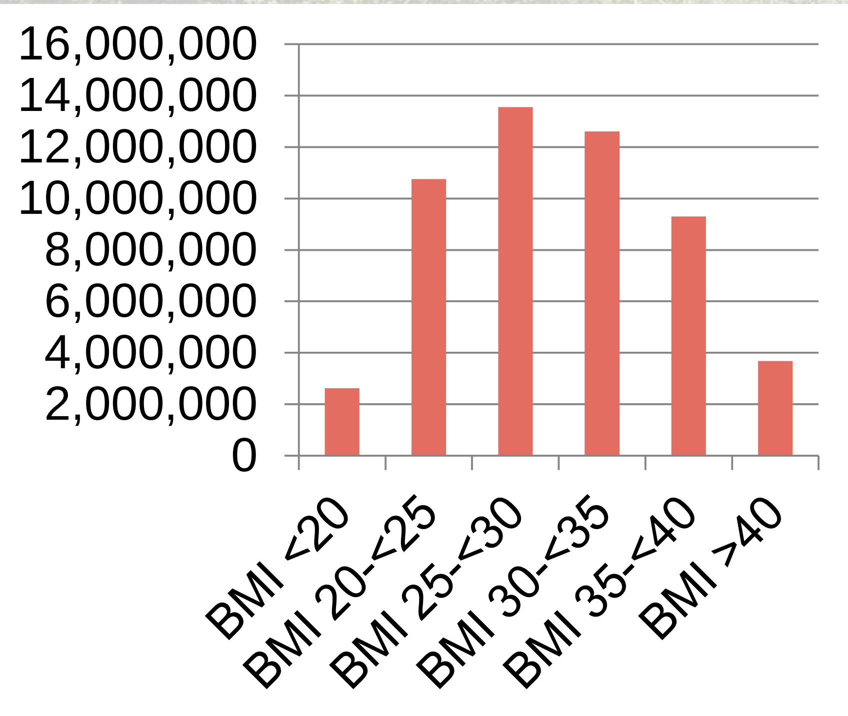
Central Limit Theorem

Using it for Statistical Inference

Central Limit Theorem: In Words

For any normal distribution:

1. The sampling distribution (the distributions of x-bars from all possible samples) is also a normal distribution
2. The mean of the x-bars is actually μ
3. The standard deviation of the x-bars is actually σ/√n



Central Limit Theorem: In Formulas

$\mu_{\text{x-bars}} = \mu$

$\sigma_{\text{x-bars}} = \sigma/\sqrt{n}$

$$z = \frac{\text{x-bar} - \mu_{\text{x-bars}}}{\sigma_{\text{x-bars}}} = \frac{\text{x-bar} - \mu}{\sigma/\sqrt{n}}$$

….where

- n is the sample size
- μ is the mean of the x distribution (population mean), and
- σ is the standard deviation of the x distribution (population standard deviation)

Central Limit Theorem: In Formulas

$\mu_{x\text{-bars}} = \mu$

$\sigma_{x\text{-bars}} = \sigma/\sqrt{n}$

$$z = \frac{\text{x-bar} - \mu_{x\text{-bars}}}{\sigma_{x\text{-bars}}} = \frac{\text{x-bar} - \mu}{\boxed{\sigma/\sqrt{n}}}$$

….where

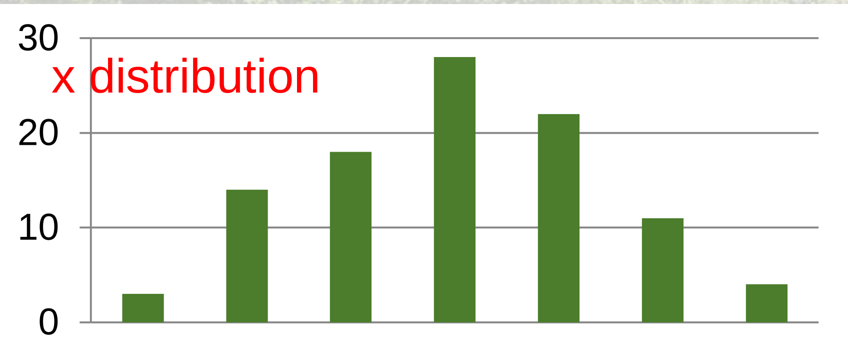
- n is the sample size
  - Note: Only works when $n \geq 30$!
- μ is the mean of the x distribution (population mean), and
- σ is the standard deviation of the x distribution (population standard deviation)

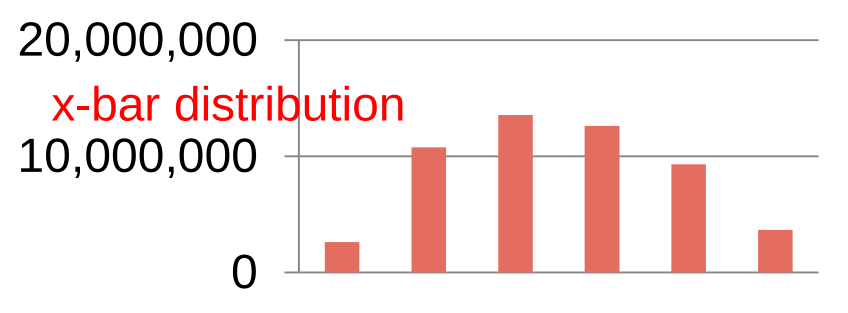
The standard error is the standard deviation of the sampling distribution.

For the x-bar sampling distribution, standard error (SE) = $\sigma/\sqrt{n}$

Central Limit Theorem

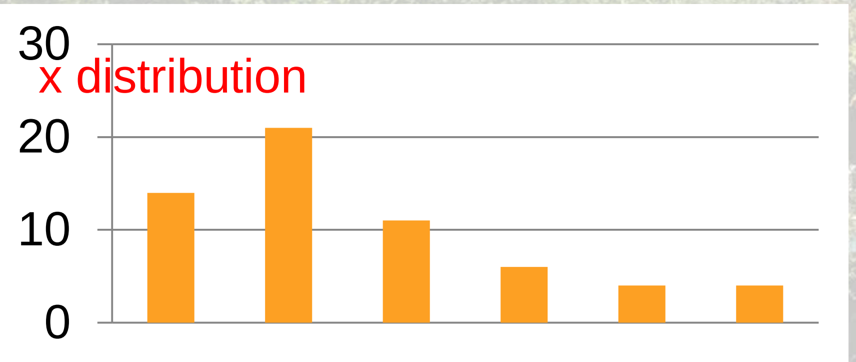
1. If the distribution of x is normal, then the distribution of x-bar is also normal.

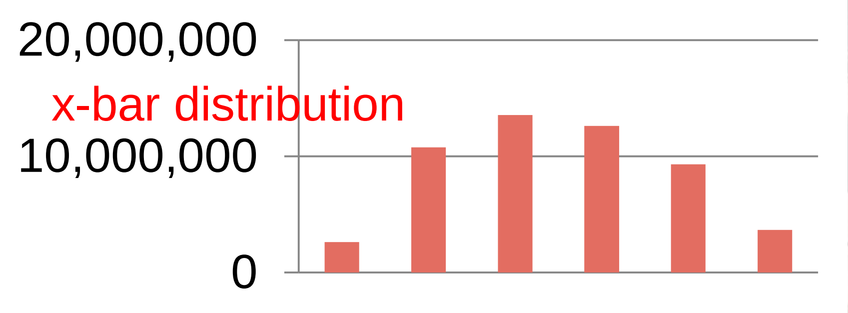




Central Limit Theorem

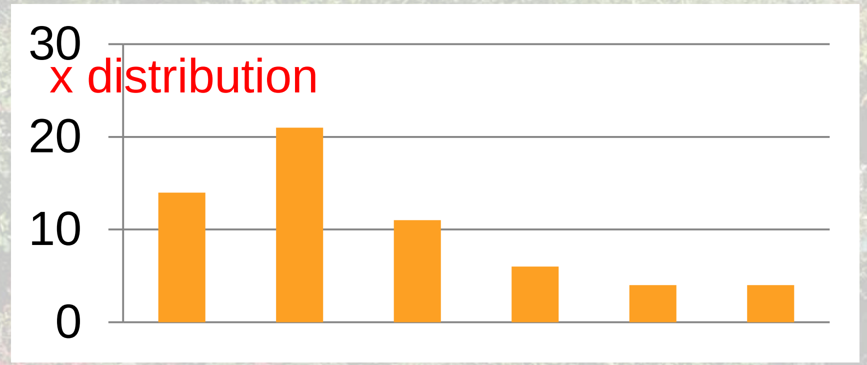
1. If the distribution of x is normal, then the distribution of x-bar is also normal.
2. Even if the distribution of x is NOT normal, as long as n≥30, the Central Limit Theorem says that the x-bar distribution is approximately normal.

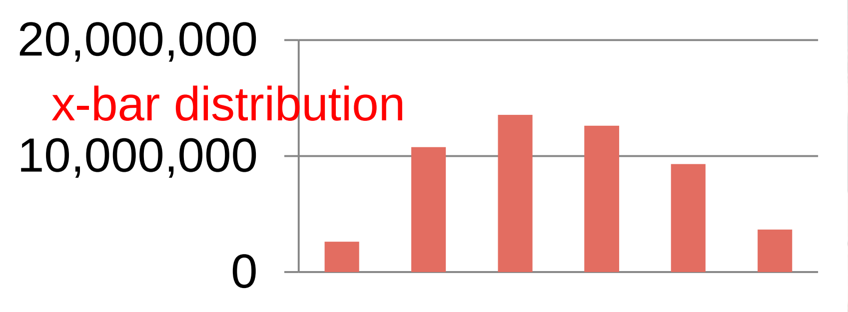




Central Limit Theorem

1. If the distribution of x is normal, then the distribution of x-bar is also normal.
2. Even if the distribution of x is NOT normal, as long as n≥30, the Central Limit Theorem says that the x-bar distribution is approximately normal.





A sample statistic is considered ***unbiased*** if the mean of its sampling distribution equals the parameter being estimated.

Finding Probabilities Regarding X-bar

Applying the Central Limit Theorem

What Are We Doing?

Chapter 7.1-7.3

- We had a normally distributed x.
- We had a μ and a σ.
- We want to find the probability of selecting a value (from the population) above or below a value of x, so we use the z-score and z-table for probabilities.
- We used this formula:

$$z = \frac{x - \mu}{\sigma}$$

What Are We Doing?

Chapter 7.1-7.3

- We had a normally distributed x.
- We had a μ and a σ.
- We want to find the probability of selecting a value (from the population) above or below a value of x, so we use the z-score and z-table for probabilities.
- We used this formula: $z = \dfrac{x - \mu}{\sigma}$

Chapter 7.4-7.5

- We have a normally distributed x.
- We have a μ and a σ.
- We want to find the probability of selecting a *sample* n (from the population) with a mean value *(x-bar)* above or below a value of x-bar, so we use the z-score and z-table for probabilities.

We will use this formula: $z = \dfrac{\text{x-bar} - \mu}{\sigma/\sqrt{n}}$

What Are We Doing?

Chapter 7.1-7.3

- We had a normally distributed x.
- We had a μ and a σ.
- We want to find the probability of selecting a value (from the population) above or below a value of x, so we use the z-score and z-table for probabilities.
- We used this formula:

$$z = \frac{x - \mu}{\sigma}$$

Chapter 7.4-7.5

- We have a normally distributed x.
- We have a μ and a σ.
- We want to find the probability of selecting a *sample* n (from the population) with a mean value *(x-bar)* above or below a value of x-bar, so we use the z-score and z-table for probabilities.

We will use this formula:

$$z = \frac{\text{x-bar} - \mu}{\sigma/\sqrt{n}}$$

What Are We Doing?

Chapter 7.1-7.3

- We had a normally distributed x.
- We had a μ and a σ.
- We want to find the probability of selecting a value (from the population) above or below a value of x, so we use the z-score and z-table for probabilities.
- We used this formula:

$$z = \frac{x - \mu}{\sigma}$$

Chapter 7.4-7.5

- We have a normally distributed x.
- We have a μ and a σ.
- We want to find the probability of selecting a *sample* n (from the population) with a mean value *(x-bar)* above or below a value of x-bar, so we use the z-score and z-table for probabilities.

We will use this formula: $z = \frac{\text{x-bar} - \mu}{\sigma/\sqrt{n}}$

Standard Error (SE)

How to Find Probabilities Regarding X-Bar

1. Convert x-bar to a z-score using the following formula:

$$z = \frac{\text{x-bar} - \mu}{\sigma/\sqrt{n}}$$

2. Look up the probability for the z-score in the z-table (like in Chapters 7.2-7.3, only this is about x-bar).

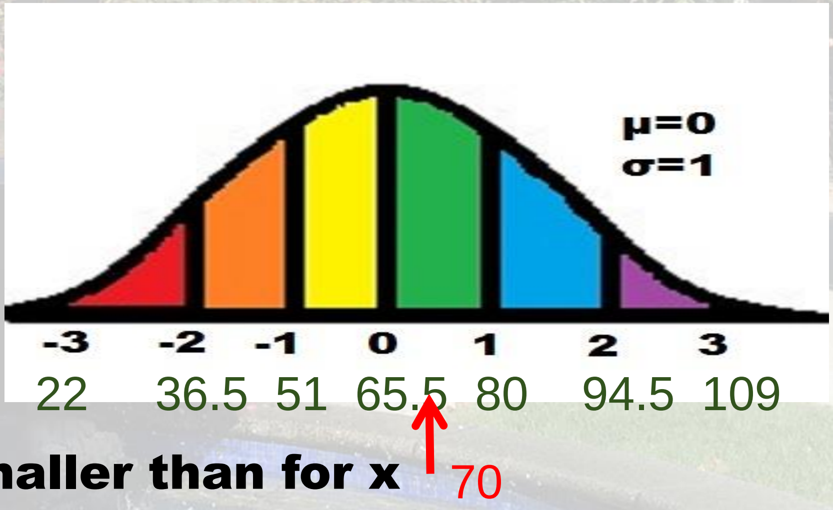
Remember the Students?

$$z = \frac{\text{x-bar} - \mu}{\sigma/\sqrt{n}}$$

μ = 65.5
σ = 14.5

x-bar = 70
n=49

- Assume the 100-student class is a population.
- Now I have to pick an n
  - Let's pick 49.
- To pass the class, students have to get at least 70, which is a C.
- *Question:* What is the probability of me selecting a sample of 49 students with an x-bar greater than 70?



Probability for x-bar will be smaller than for x

Remember the Students?

- Assume the 100-student class is a population.
- Now I have to pick an n
  - Let's pick 49.
- To pass the class, students have to get at least 70, which is a C.
- *Question:* What is the probability of me selecting a sample of 49 students with an x-bar greater than 70?

$$z = \frac{\text{x-bar} - \mu}{\sigma/\sqrt{n}}$$

$$\sigma/\sqrt{n} = SE$$

$$z = \frac{\text{x-bar} - \mu}{SE}$$

μ = 65.5
σ = 14.5

x-bar = 70
n=49

Remember the Students?

- Assume the 100-student class is a population.
- Now I have to pick an n
  - Let's pick 49.
- To pass the class, students have to get at least 70, which is a C.
- *Question:* What is the probability of me selecting a sample of 49 students with an x-bar greater than 70?

$$z = \frac{\text{x-bar} - \mu}{\sigma/\sqrt{n}}$$

$$\sigma/\sqrt{n} = SE$$

$$z = \frac{\text{x-bar} - \mu}{SE}$$

$$SE = \sigma/\sqrt{n} = 14.5/\sqrt{49} = 2.1$$

μ = 65.5
σ = 14.5

x-bar = 70
n=49

Remember the Students?

- Assume the 100-student class is a population.
- Now I have to pick an n
  - Let's pick 49.
- To pass the class, students have to get at least 70, which is a C.
- *Question:* What is the probability of me selecting a sample of 49 students with an x-bar greater than 70?

$$z = \frac{\text{x-bar} - \mu}{\sigma/\sqrt{n}}$$

$$\sigma/\sqrt{n} = SE$$

$$z = \frac{\text{x-bar} - \mu}{SE}$$

μ = 65.5
σ = 14.5

x-bar = 70

n=49

$$SE = \sigma/\sqrt{n} = 14.5/\sqrt{49} = 2.1$$

$$z = (70-65.5)/2.1 = 2.17$$

Remember the Students?

- Assume the 100-student class is a population.
- Now I have to pick an n
  - Let's pick 49.
- To pass the class, students have to get at least 70, which is a C.
- *Question:* What is the probability of me selecting a sample of 49 students with an x-bar greater than 70?

$$z = (70-65.5)/2.1 = 2.17$$

Remember the Students?

- Assume the 100-student class is a population.
- Now I have to pick an n
  - Let’s pick 49.
- To pass the class, students have to get at least 70, which is a C.
- *Question:* What is the probability of me selecting a sample of 49 students with an x-bar greater than 70?

$z = (70-65.5)/2.1 = 2.17$

Use -2.17*

$p = 0.0150$

Remember the Students?

- Assume the 100-student class is a population.
- Now I have to pick an n
    - Let's pick 49.
- To pass the class, students have to get at least 70, which is a C.
- *Question:* What is the probability of me selecting a sample of 49 students with an x-bar greater than 70?

$z = (70-65.5)/2.1 = 2.17$

Use -2.17*

$p = 0.0150$

Probability is 0.0150, or 1.5%

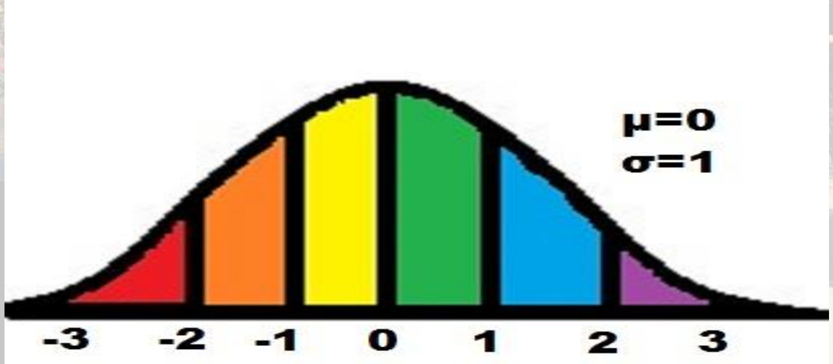
Remember the Students?

- Assume the 100-student class is a population.
- Now I have to pick an n
  - Let's pick 36.
- *Question:* What is the probability of me selecting a sample of 36 students with an x-bar between 60 and 65?

$$z = \frac{\text{x-bar} - \mu}{\sigma/\sqrt{n}}$$

μ = 65.5
σ = 14.5

x-bar1 = 60
x-bar2 = 65
n=36



22 36.5 51 65.5 80 94.5 109

Probability for x-bar will be smaller than for x

Remember the Students?

- Assume the 100-student class is a population.
- Now I have to pick an n
  - Let's pick 36.
- *Question:* What is the probability of me selecting a sample of 36 students with an x-bar between 60 and 65?

$$z = \frac{\text{x-bar} - \mu}{\sigma/\sqrt{n}}$$

$$\sigma/\sqrt{n} = SE$$

$$z = \frac{\text{x-bar} - \mu}{SE}$$

μ = 65.5
σ = 14.5

x-bar1 = 60
x-bar2 = 65
n=36

Remember the Students?

- Assume the 100-student class is a population.
- Now I have to pick an n
  - Let's pick 36.
- *Question:* What is the probability of me selecting a sample of 36 students with an x-bar between 60 and 65?

$$z = \frac{\text{x-bar} - \mu}{\sigma/\sqrt{n}}$$

$$\sigma/\sqrt{n} = SE$$

$$z = \frac{\text{x-bar} - \mu}{SE}$$

$$SE = \sigma/\sqrt{n} = 14.5/\sqrt{36} = 2.4$$

μ = 65.5
σ = 14.5

x-bar1 = 60
x-bar2 = 65
n=36

Remember the Students?

- Assume the 100-student class is a population.
- Now I have to pick an n
  - Let's pick 36.
- *Question:* What is the probability of me selecting a sample of 36 students with an x-bar between 60 and 65?

$$z = \frac{\text{x-bar} - \mu}{\sigma/\sqrt{n}}$$

$\sigma/\sqrt{n}$ = SE

$$z = \frac{\text{x-bar} - \mu}{SE}$$

μ = 65.5
σ = 14.5

x-bar1 = 60
x-bar2 = 65
n=36

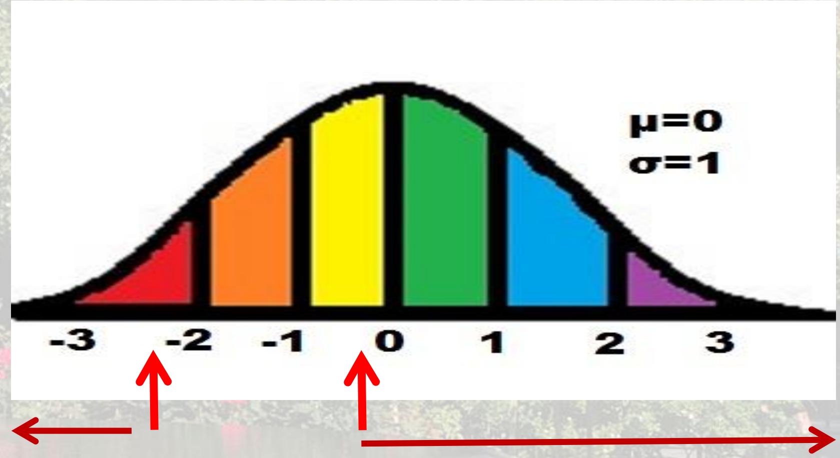
SE = $\sigma/\sqrt{n}$ = $14.5/\sqrt{36}$ = 2.4

z1 = (60-65.5)/2.4 = -2.28

z2 = (65-65.5)/2.4 = -0.21

Remember the Students?

- Assume the 100-student class is a population.
- Now I have to pick an n
  - Let’s pick 36.
- *Question:* What is the probability of me selecting a sample of 36 students with an x-bar between 60 and 65?

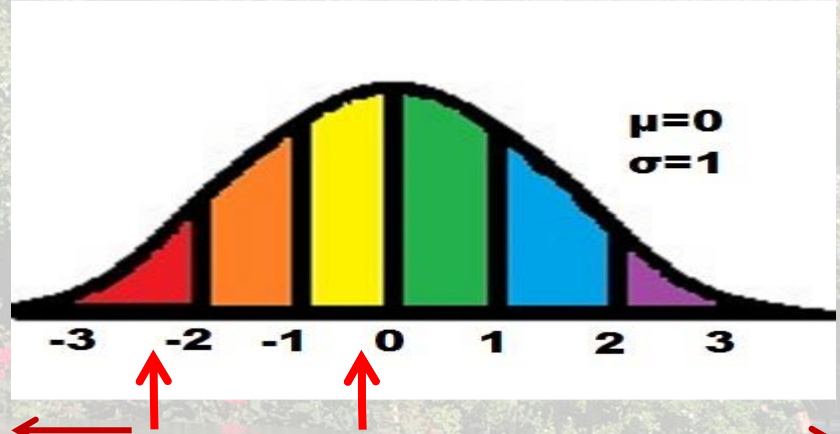


z1 = (60-65.5)/2.4 = -2.28

z2 = (65-65.5)/2.4 = -0.21

Remember the Students?

- Assume the 100-student class is a population.
- Now I have to pick an n
  - Let's pick 36.
- *Question:* What is the probability of me selecting a sample of 36 students with an x-bar between 60 and 65?



z1 = (60-65.5)/2.4 = -2.28

p1 = 0.0113

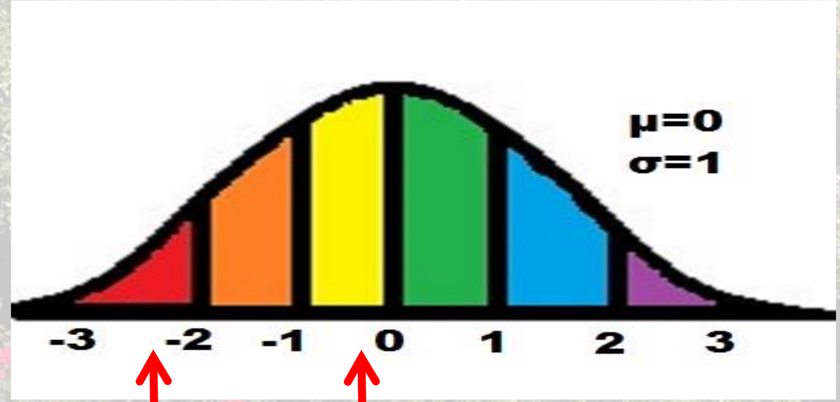
z2 = (65-65.5)/2.4 = -0.21

p2 = 0.5832

***Use 0.21 ***

Remember the Students?

- Assume the 100-student class is a population.
- Now I have to pick an n
  - Let's pick 36.
- *Question:* What is the probability of me selecting a sample of 36 students with an x-bar between 60 and 65?



z1 = (60-65.5)/2.4 = -2.28

p1 = 0.0113

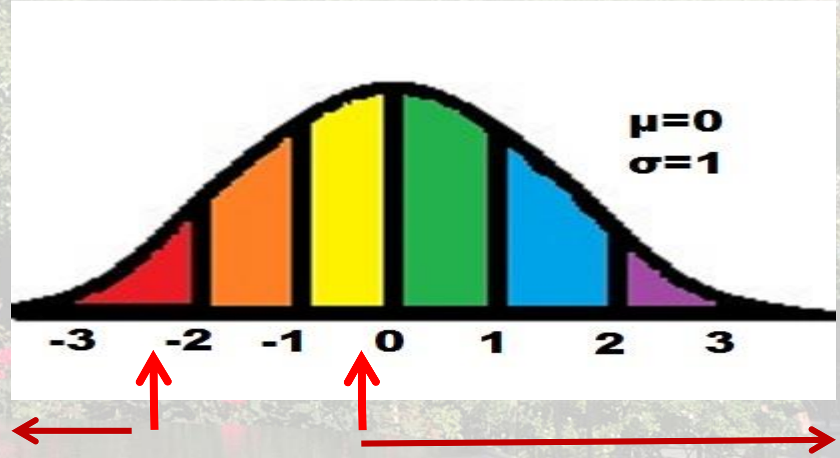
z2 = (65-65.5)/2.4 = -0.21

p2 = 0.5832

1 – 0.0113 – 0.5832 = 0.4055

Remember the Students?

- Assume the 100-student class is a population.
- Now I have to pick an n
  - Let's pick 36.
- *Question:* What is the probability of me selecting a sample of 36 students with an x-bar between 60 and 65?



z1 = (60-65.5)/2.4 = -2.28

p1 = 0.0113

z2 = (65-65.5)/2.4 = -0.21

p2 = 0.5832

The probability is 0.4055 or 41%

Conclusion

- Reviewed parameters and statistics, and discussed inferences
- Description of sampling distribution
- Presented Central Limit Theorem
- Examples of finding probabilities regarding x-bar

Photograph by Steve Cadman