Normal Distribution and Parameter Estimation

Learning Objectives

In this chapter, you learn:

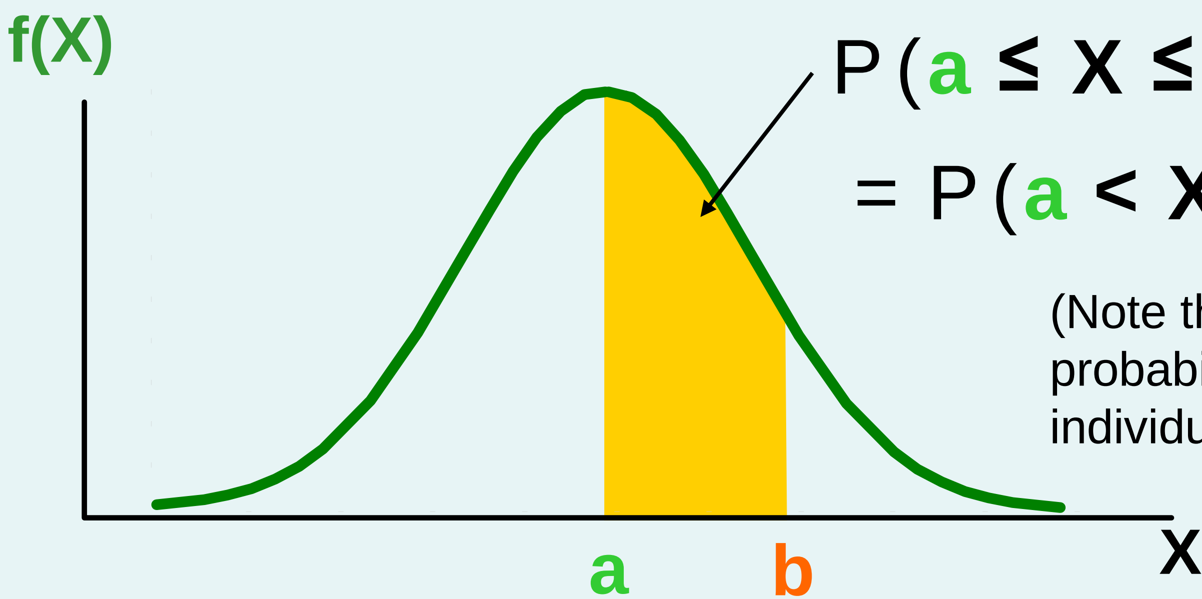
- How continuous distributions are different from discrete distributions.
- How to compute probabilities from the normal distribution
- How to use the normal distribution to solve practical problems

Continuous Probability Distributions

- A continuous random variable is a variable that can assume any value on an interval
  - thickness of an item, weight, length
  - time required to complete a task
  - temperature
  - height, in inches
  - miles per gallon

Probability Density

- A function f(x) ≥ 0 that shows the more likely and less likely intervals of variable X.
- P(a ≤ X ≤ b) = area under f(x) from a to b

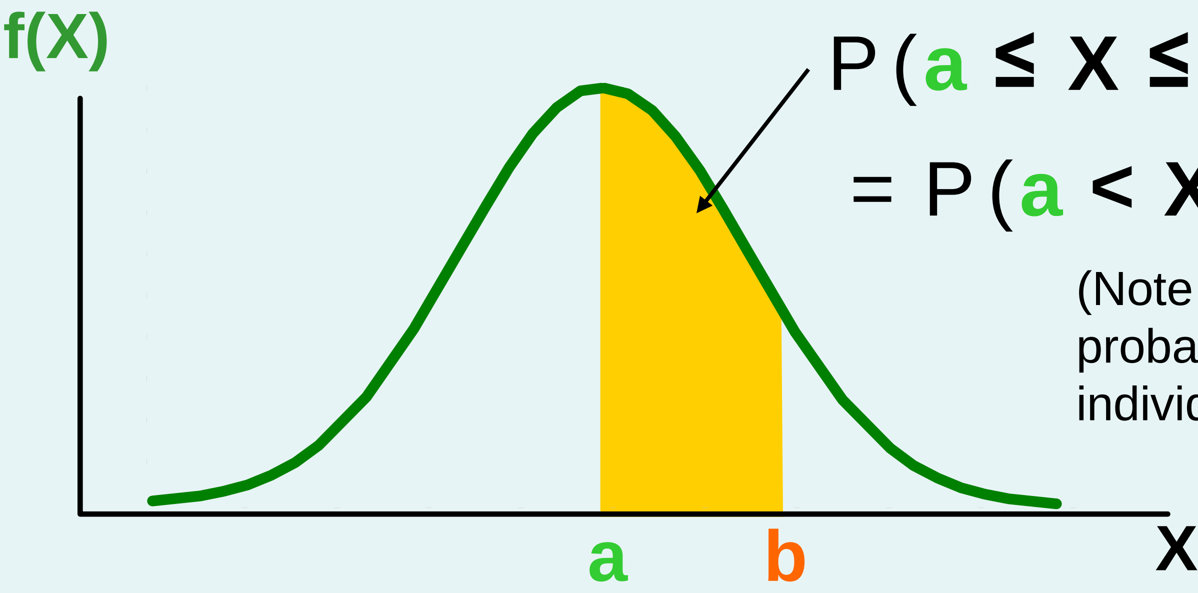


$P(a \le X \le b)$

$= P(a < X < b)$

(Note that the probability of any individual value is zero)

Probability is the area under the density curve



$P(a \le X \le b)$

$= P(a < X < b)$

(Note that the probability of any individual value is zero)

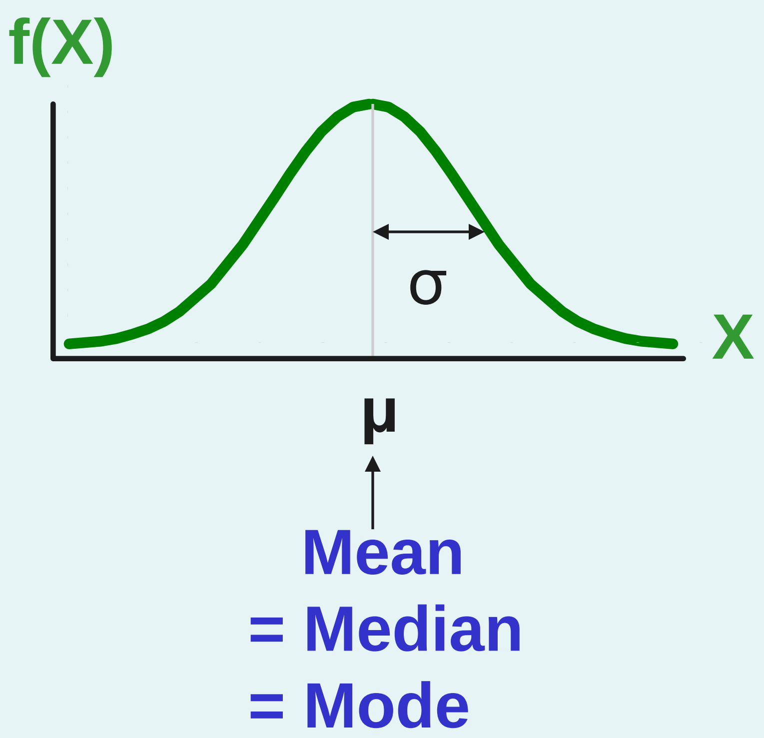
The Normal Distribution

- 'Bell Shaped'
- Symmetrical
- Mean, Median and Mode are Equal

Location is determined by the mean, μ

Spread is determined by the standard deviation, σ

The random variable has an infinite theoretical range:
+ ∞ to − ∞



The Normal Distribution Density Function

- The formula for the normal probability density function is

$$f(X) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{(X-\mu)}{\sigma}\right)^2}$$

Where e = the mathematical constant approximated by 2.71828

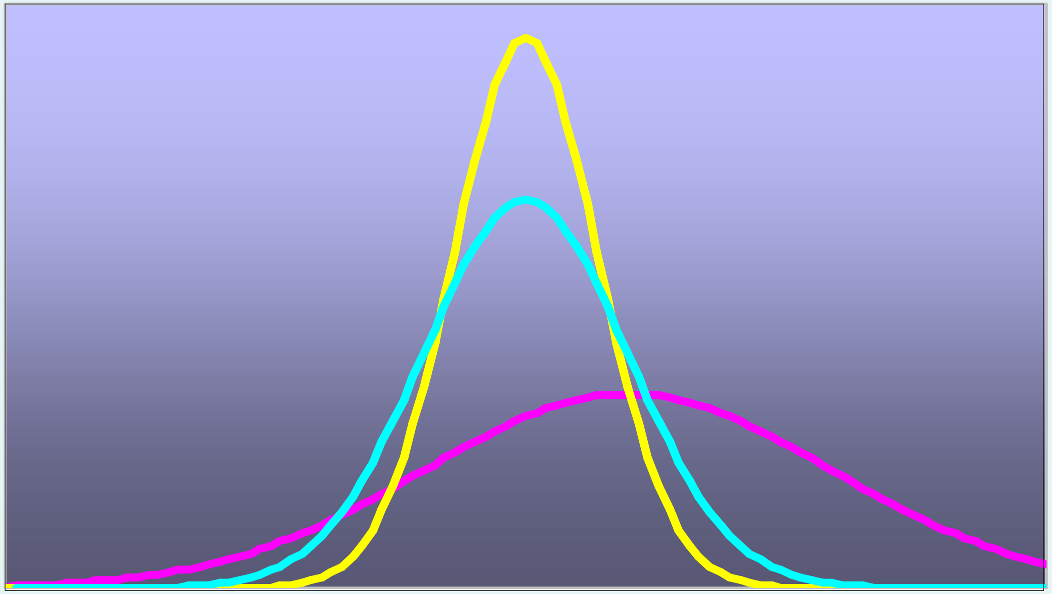
π = the mathematical constant approximated by 3.14159

μ = the population mean

σ = the population standard deviation

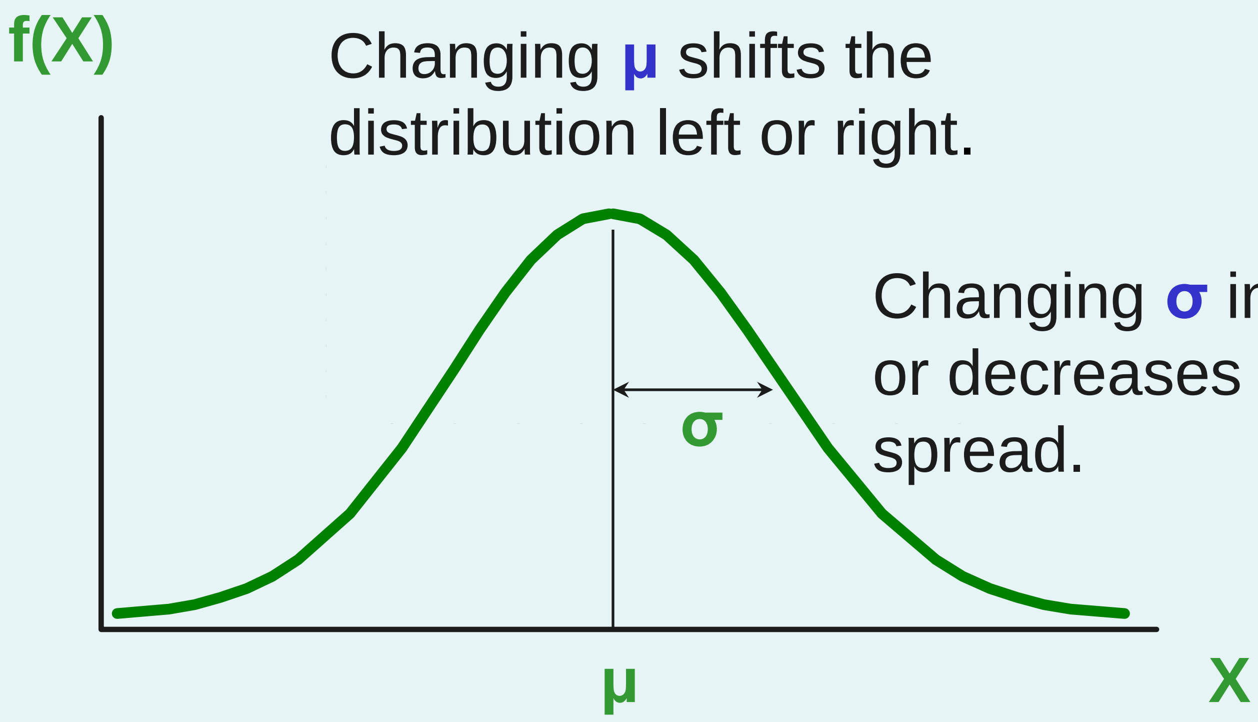
X = any value of the continuous variable

Many Normal Distributions



By varying the parameters μ and σ, we obtain different normal distributions

The Normal Distribution Shape



Changing μ shifts the distribution left or right.

Changing σ increases or decreases the spread.

The Standardized Normal

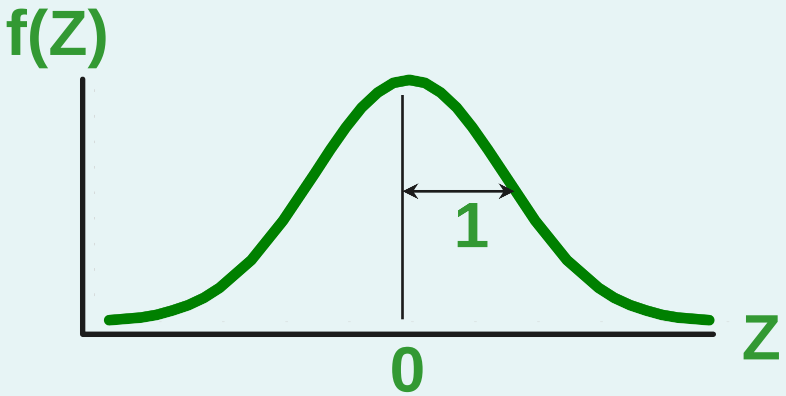
- Any normal distribution (with any mean μ and standard deviation σ) can be transformed into the standardized normal distribution (Z)

$$Z = \frac{X - \mu}{\sigma}$$

- The standardized normal distribution (Z) has a mean of 0 and a standard deviation of 1

The Standardized Normal Distribution

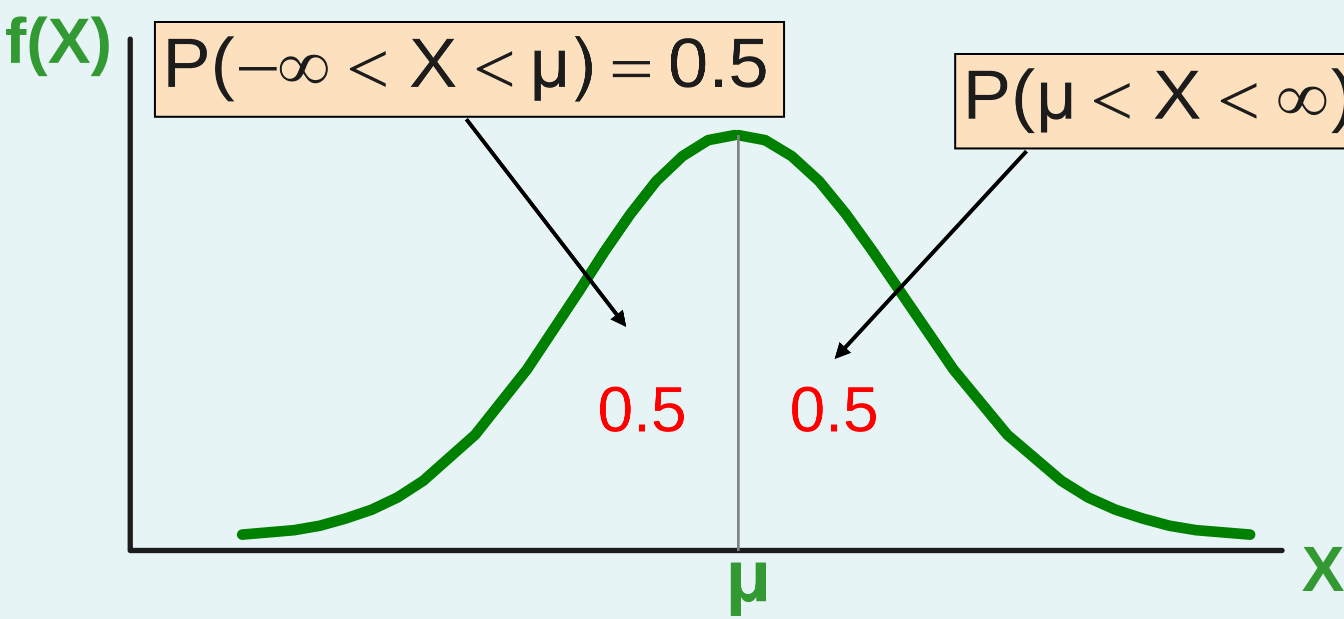
- Also known as the “Z” distribution
- Mean is 0
- Standard Deviation is 1



Values above the mean have positive Z-values, values below the mean have negative Z-values

Probability as Area Under the Curve

The total area under the curve is 1.0, and the curve is symmetric, so half is above the mean, half is below



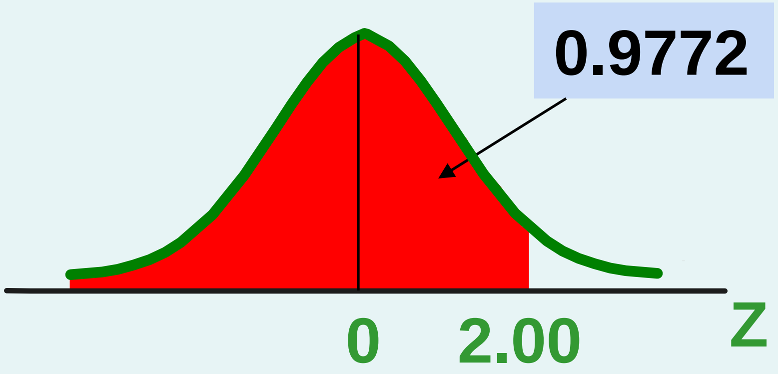
$$P(-\infty < X < \infty) = 1.0$$

The Standardized Normal Table

- The Cumulative Standardized Normal table in the textbook (Appendix table E.2) gives the probability less than a desired value of Z (i.e., from negative infinity to Z)

Example:

$P(Z < 2.00) = 0.9772$



The Standardized Normal Table

(continued)

The column gives the value of Z to the second decimal point

Z	0.00	0.01	0.02 …
0.0			
0.1			
:			
2.0	.9772		

The row shows the value of Z to the first decimal point

The value within the table gives the probability from $Z = -\infty$ up to the desired Z value

$P(Z < 2.00) = 0.9772$

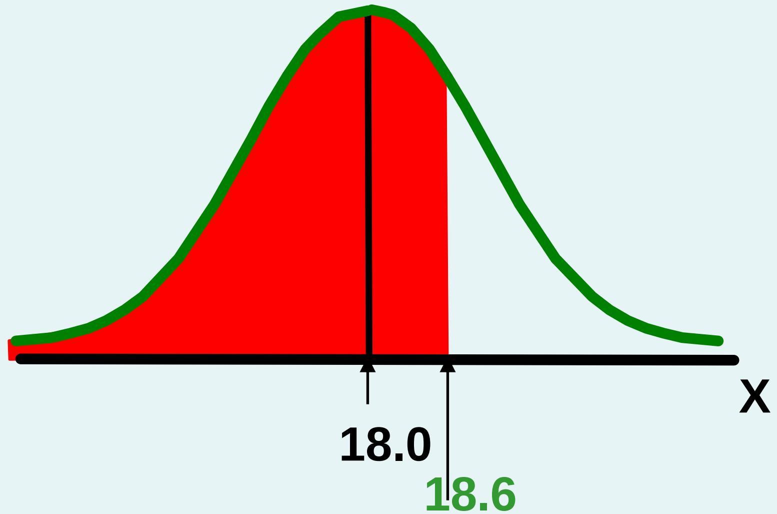
General Procedure for Finding Normal Probabilities

To find $P(a < X < b)$ when X is distributed normally:

- Draw the normal curve for the problem in terms of X
- Standardize X by computing Z
- Use the Standardized Normal Table for Z

Example

- Let X represent the time it takes (in seconds) to download an image file from the internet.
- Suppose X is normal with a mean of 18.0 seconds and a standard deviation of 5.0 seconds. Find $P(X < 18.6)$

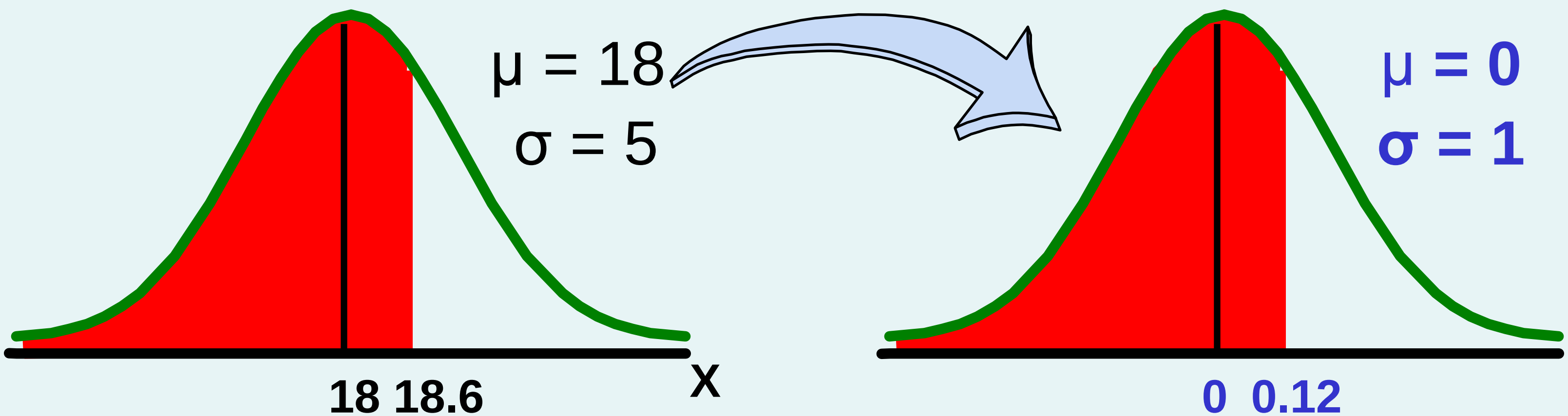


Finding Normal Probabilities

(continued)

- Let X represent the time it takes, in seconds to download an image file from the internet.
- Suppose X is normal with a mean of 18.0 seconds and a standard deviation of 5.0 seconds. Find P(X < 18.6)

$$Z = \frac{X - \mu}{\sigma} = \frac{18.6 - 18.0}{5.0} = 0.12$$



P(X < 18.6)   P(Z < 0.12)

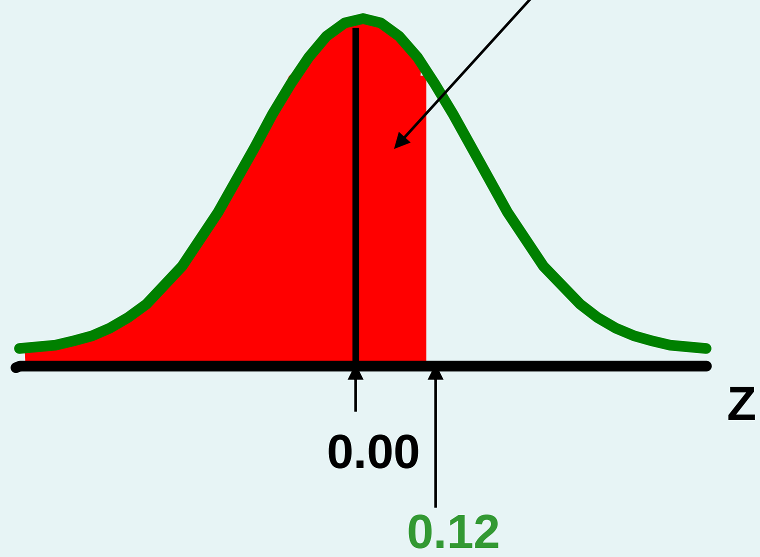
Solution: Finding P(Z < 0.12)

Standardized Normal Probability Table (Portion)

Z	.00	.01	.02
0.0	.5000	.5040	.5080
0.1	.5398	.5438	.5478
0.2	.5793	.5832	.5871
0.3	.6179	.6217	.6255

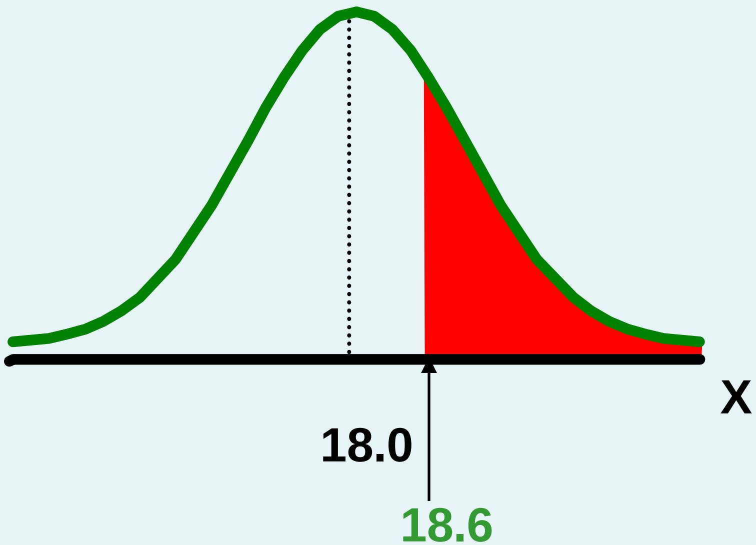
P(X < 18.6)
= P(Z < 0.12)

0.5478



Finding Normal Upper Tail Probabilities

- Suppose X is normal with mean 18.0 and standard deviation 5.0.
- Now Find P(X > 18.6)



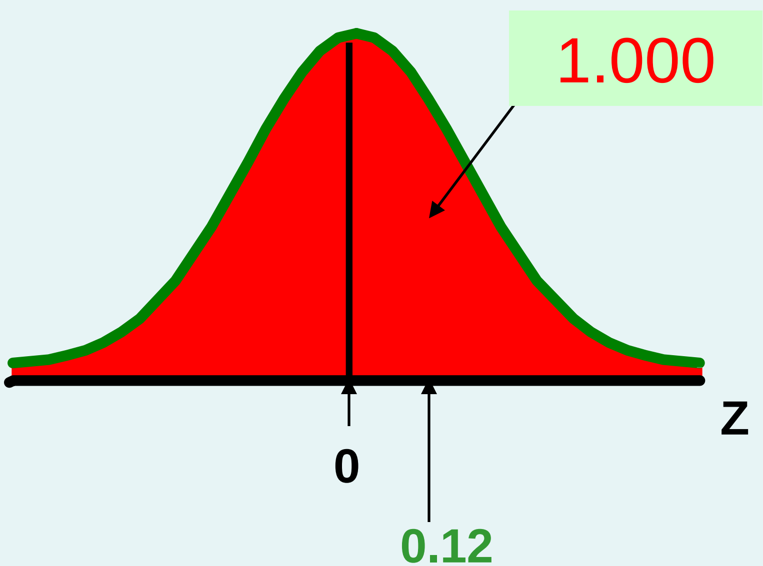
Finding Normal Upper Tail Probabilities

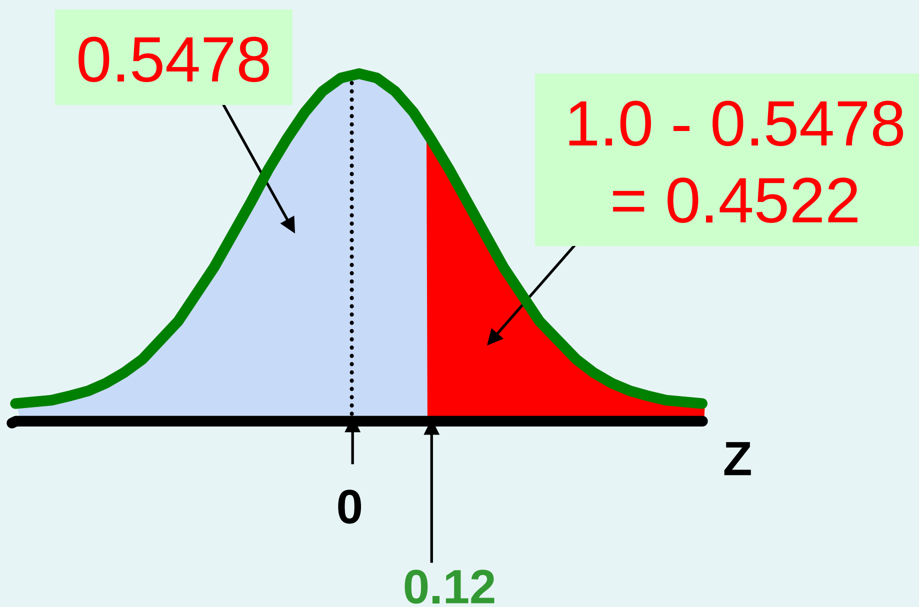
(continued)

- Now Find P(X > 18.6) by the complement rule…

$$P(X > 18.6) = P(Z > 0.12) = 1.0 - P(Z \le 0.12)$$

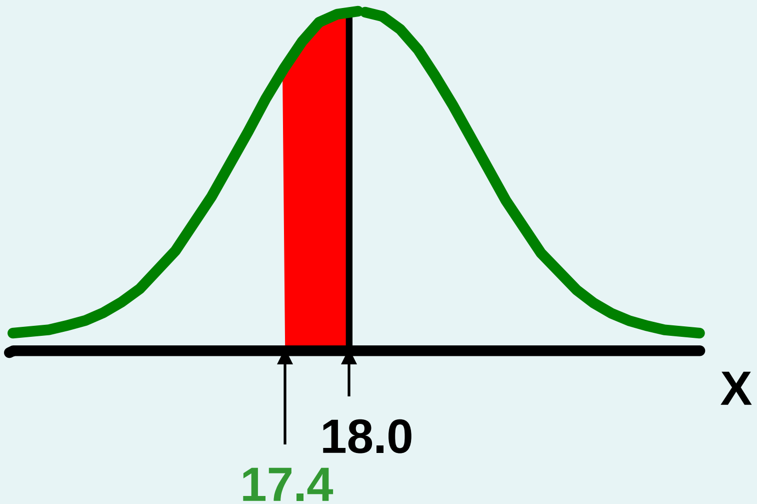
$$= 1.0 - 0.5478 = \boxed{0.4522}$$





Probabilities in the Lower Tail

- Suppose X is normal with mean 18.0 and standard deviation 5.0.
- Find $P(17.4 < X < 18)$



Probabilities in the Lower Tail

(continued)

Find P(17.4 < X < 18)…

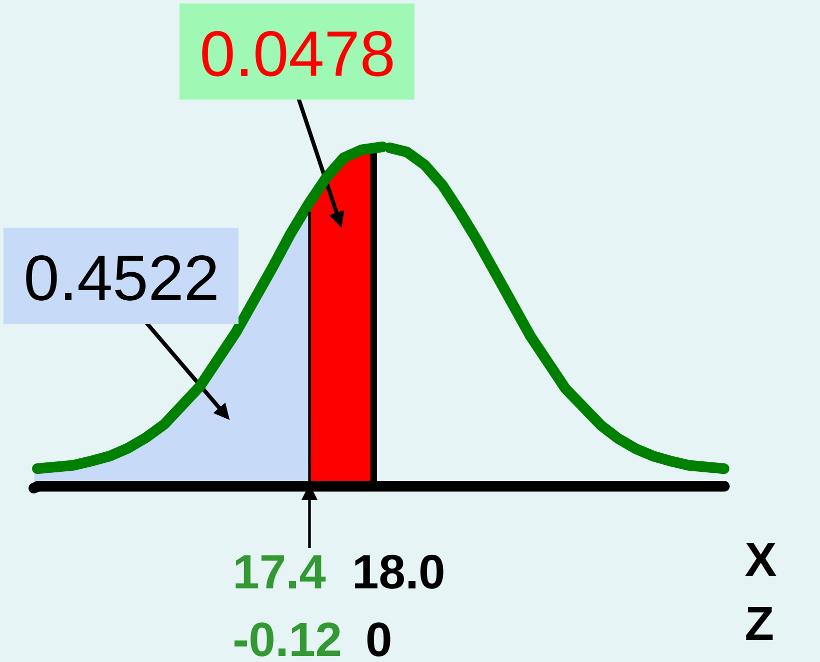
$P(17.4 < X < 18)$

$= P(-0.12 < Z < 0)$

$= P(Z < 0) - P(Z \le -0.12)$

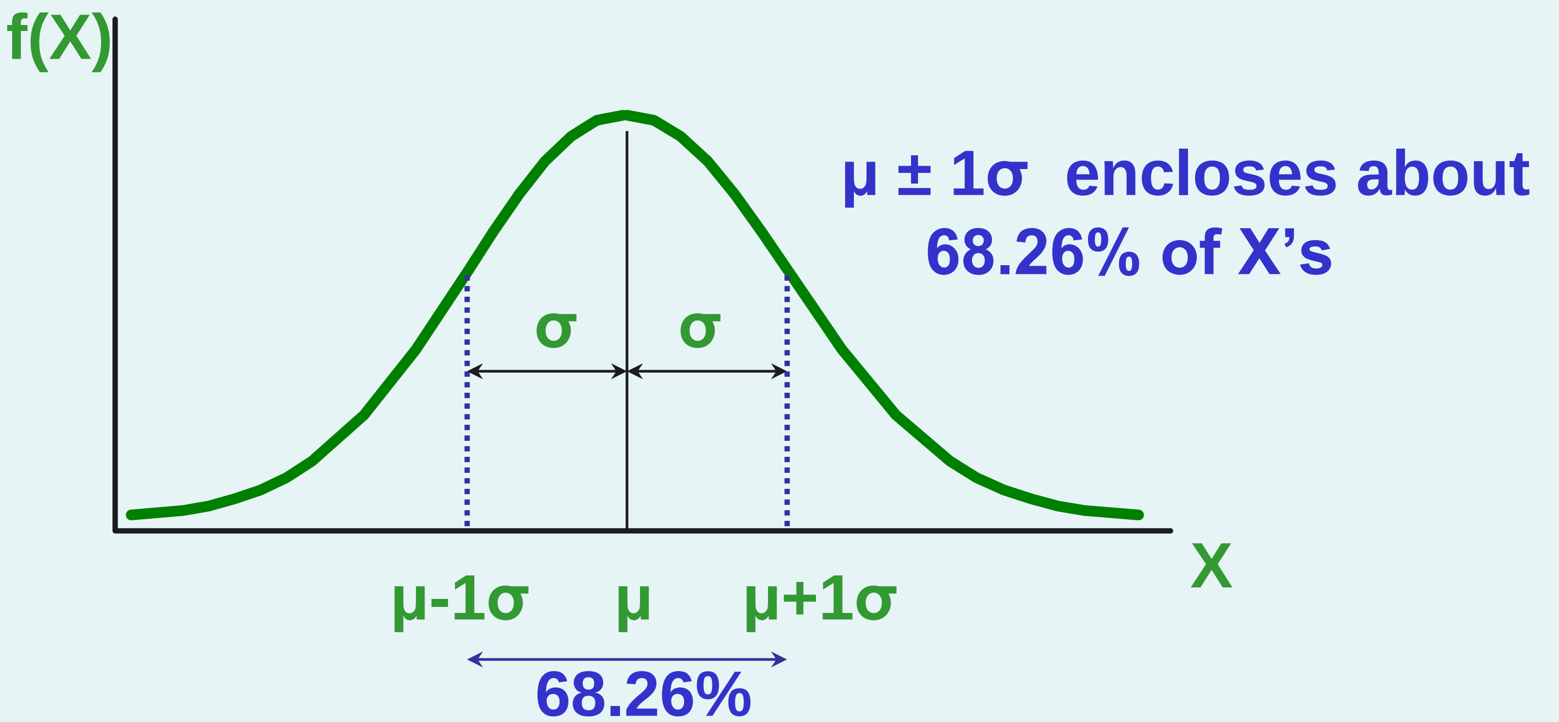
$= 0.5000 - 0.4522 = \boxed{0.0478}$

The Normal distribution is symmetric, so this probability is the same as $P(0 < Z < 0.12)$



"Empirical" Rules

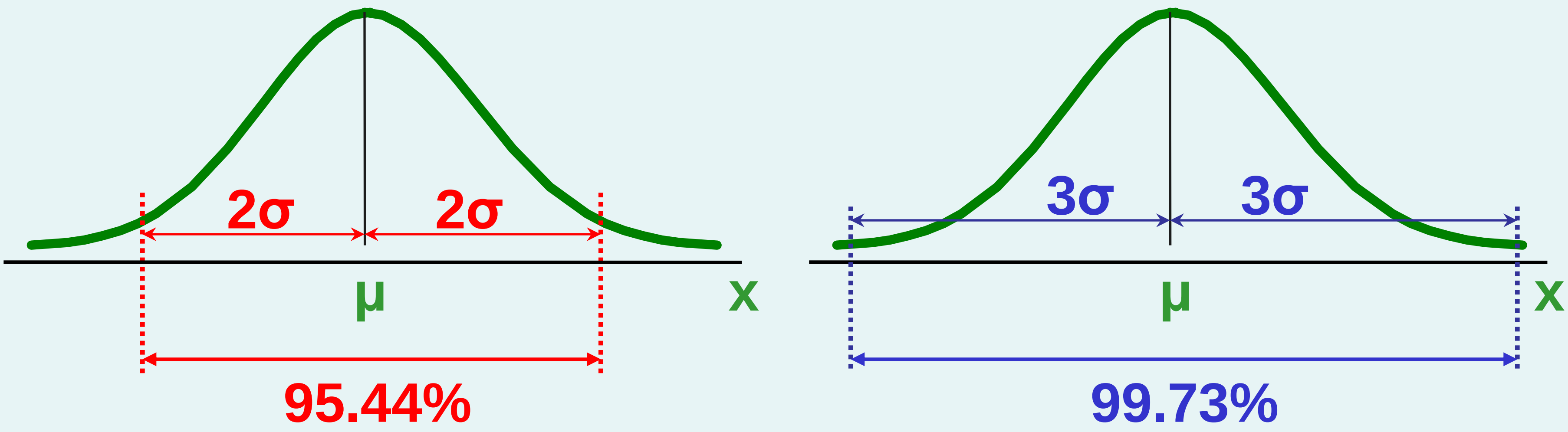
What can we say about the distribution of values around the mean? For any normal distribution:



The Empirical Rule

(continued)

- $\mu \pm 2\sigma$ covers about 95% of X's
- $\mu \pm 3\sigma$ covers about 99.7% of X's



Given a Normal Probability Find the X Value

- Steps to find the X value for a known probability:
  1. Find the Z value for the known probability
  2. Convert to X units using the formula:

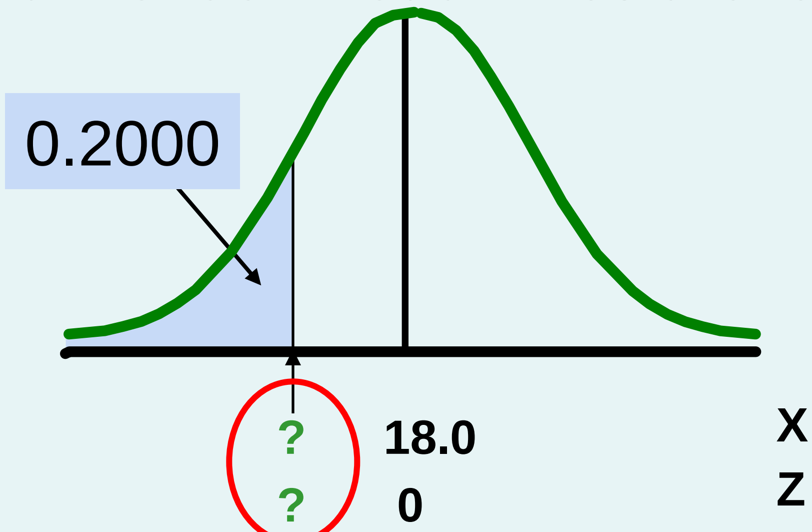
$$\boxed{X = \mu + Z\sigma}$$

Finding the X value for a Known Probability

(continued)

Example:

- Let X represent the time it takes (in seconds) to download an image file from the internet.
- Suppose X is normal with mean 18.0 and standard deviation 5.0
- Find X such that 20% of download times are less than X.



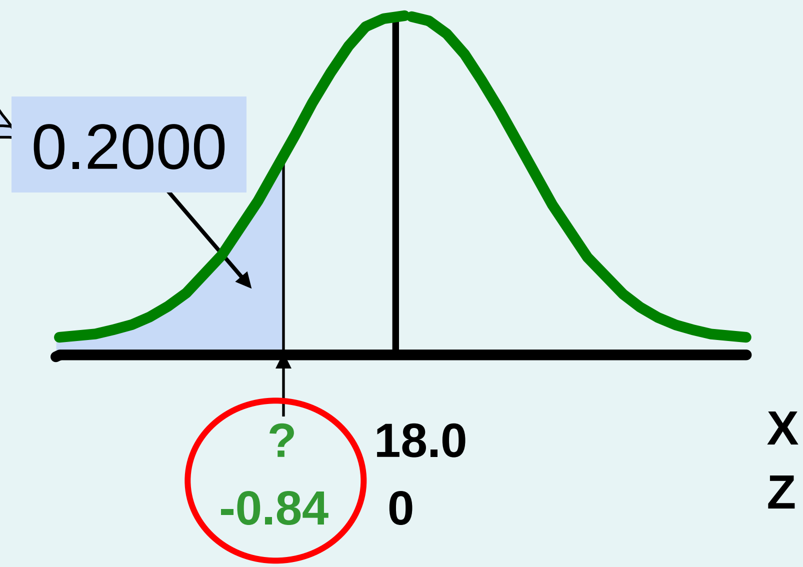
Find the Z value for 20% in the Lower Tail

1. Find the Z value for the known probability

Standardized Normal Probability Table (Portion)

Z	…	.03	.04	.05
-0.9	…	.1762	.1736	.1711
-0.8	…	.2033	.2005	.1977
-0.7	…	.2327	.2296	.2266

- 20% area in the lower tail is consistent with a Z value of -0.84



Finding the X value

2. Convert to X units using the formula:

$$
\begin{aligned}
X &= \mu + Z\sigma \\
&= 18.0 + (-0.84)5.0 \\
&= 13.8
\end{aligned}
$$

So 20% of the values from a distribution with mean 18.0 and standard deviation 5.0 are less than 13.80

Parameter Estimation and Sampling Distributions

Learning Objectives

In this chapter, you learn:

- Estimation of means, variances, proportions
- The concept of a sampling distribution
- To compute probabilities about the sample mean and the sample proportion
- How to use the *Central Limit Theorem*

Sampling Distributions

- A sampling distribution is a distribution of a statistic computed from a sample of size *n*.

- A sample is random, collected from a population. Hence, all statistics computed from it are *random variables*.

Example

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

Example

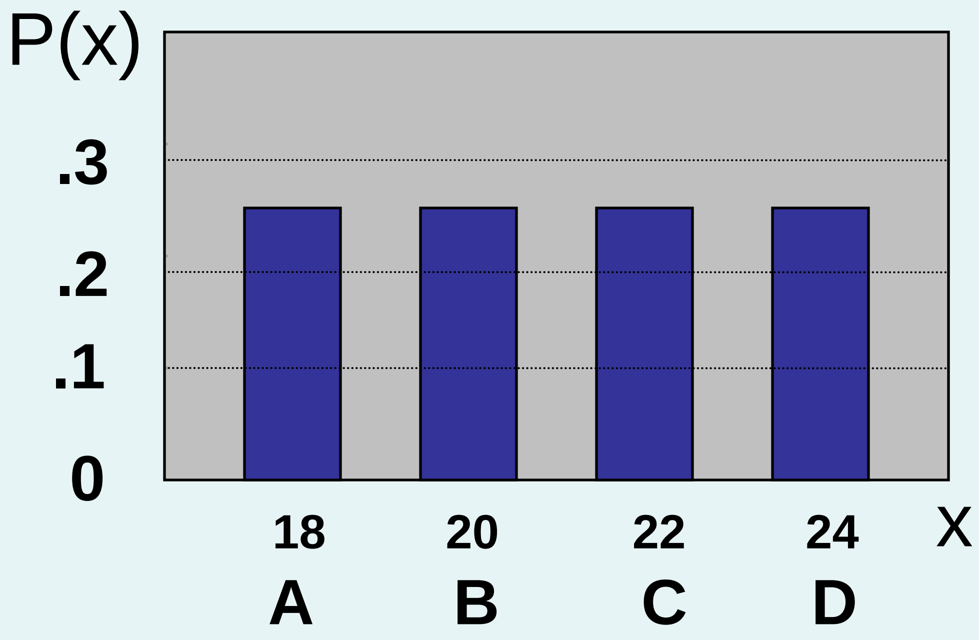
(continued)

Population parameters:

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

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

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



Example

(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

Example

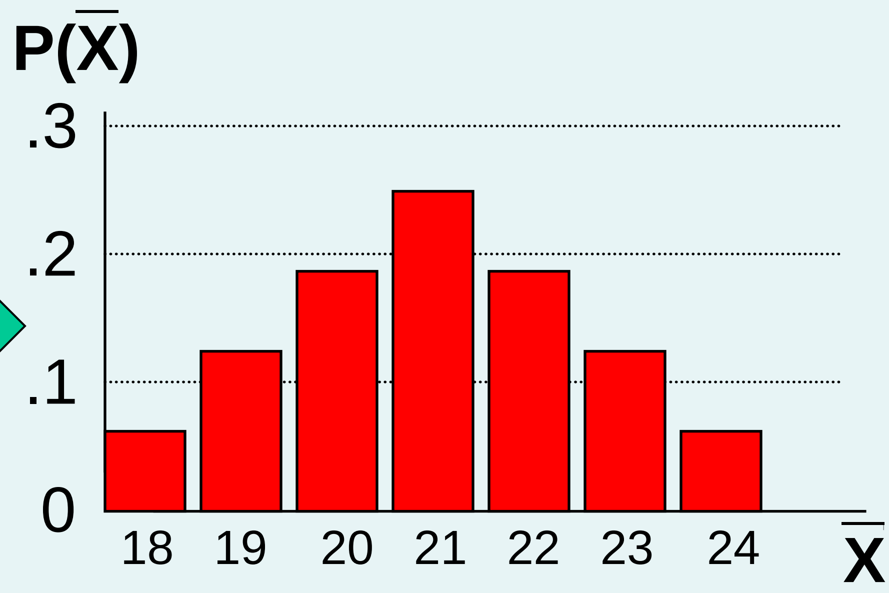
(continued)

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



Sampling Distribution of the Sample Mean

Sample mean has the following mean and standard deviation:

$$\mu_{\overline{X}} = E(\overline{X}) = \mu = \text{population mean}$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{\text{population standard deviation}}{\sqrt{\text{sample size}}}$$

Sample mean is unbiased because $E(\overline{X}) = \mu$

Its standard deviation is also called the standard error of the sample mean. It decreases as the sample size increases.

Sample Mean for a Normal Population

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

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

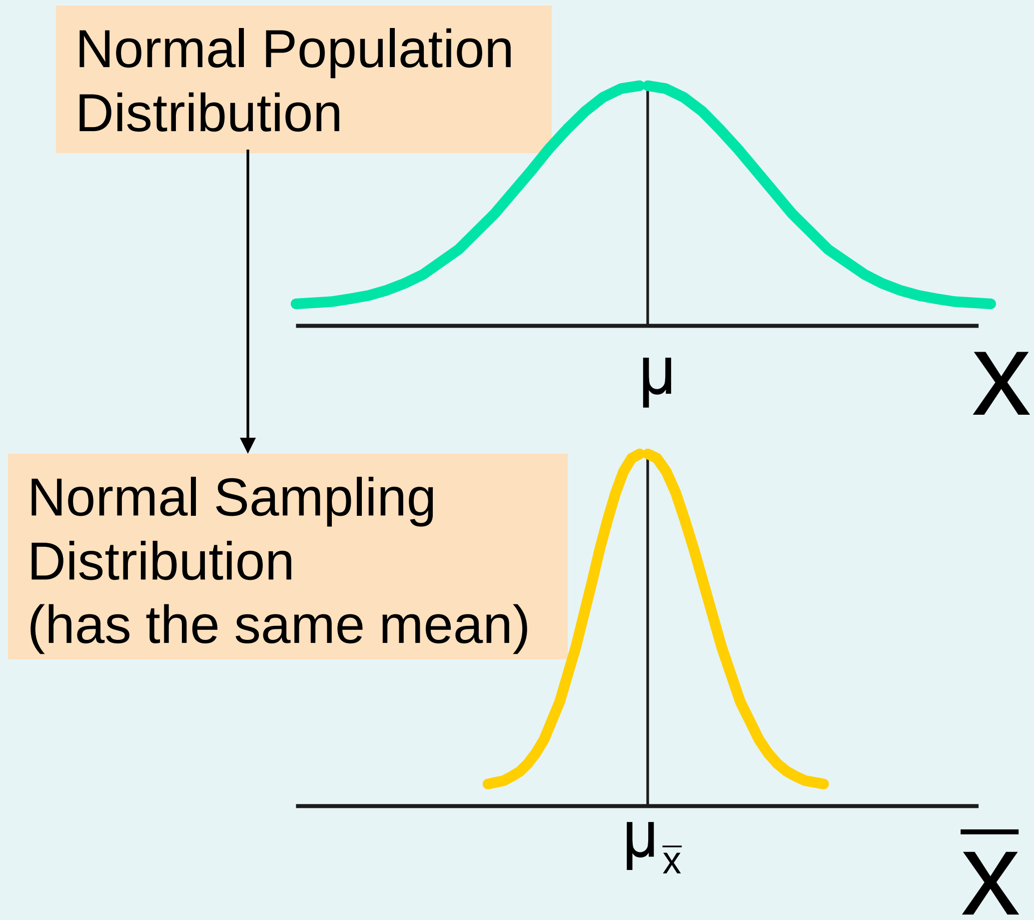
and

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

Sampling Distribution Properties

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

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

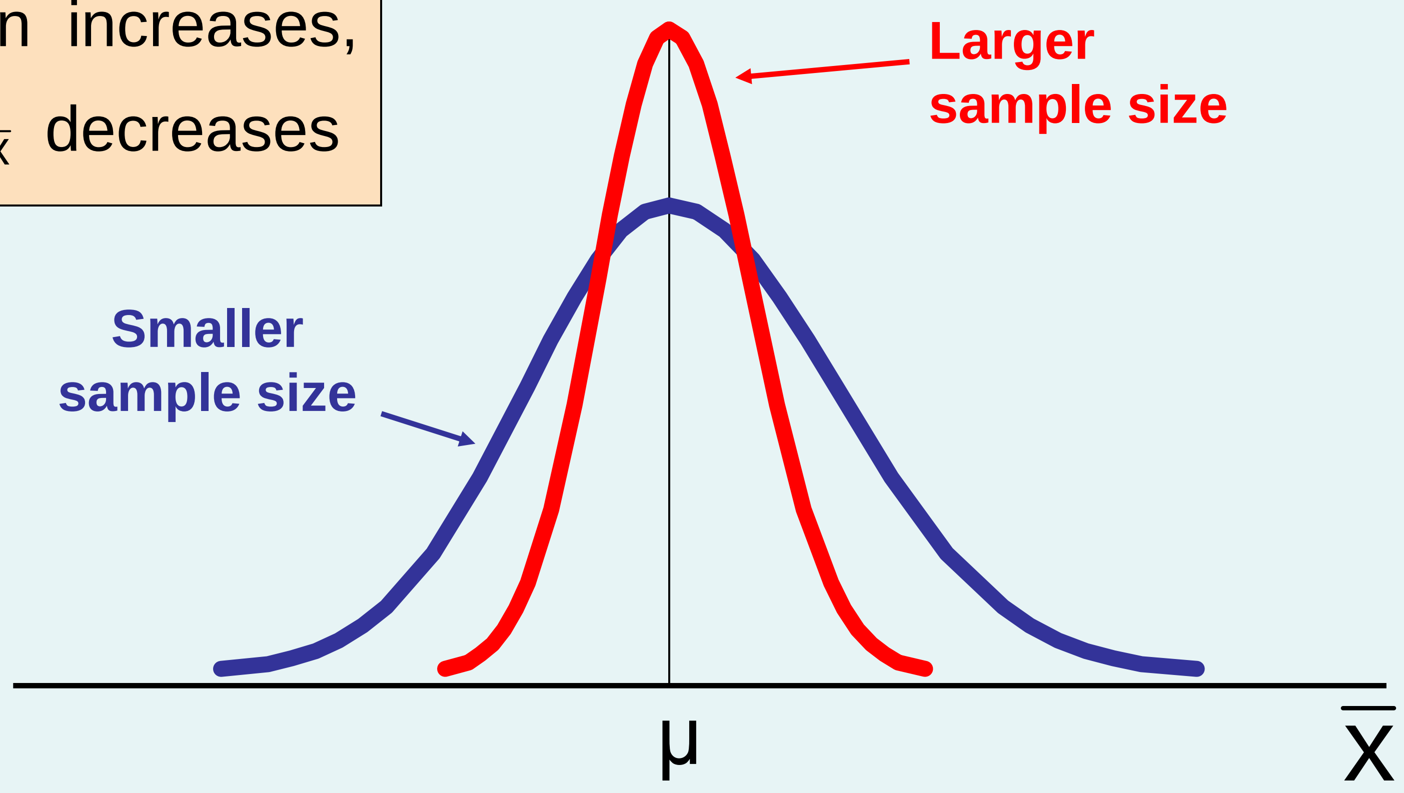


Sampling Distribution Properties

(continued)

As n increases,

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

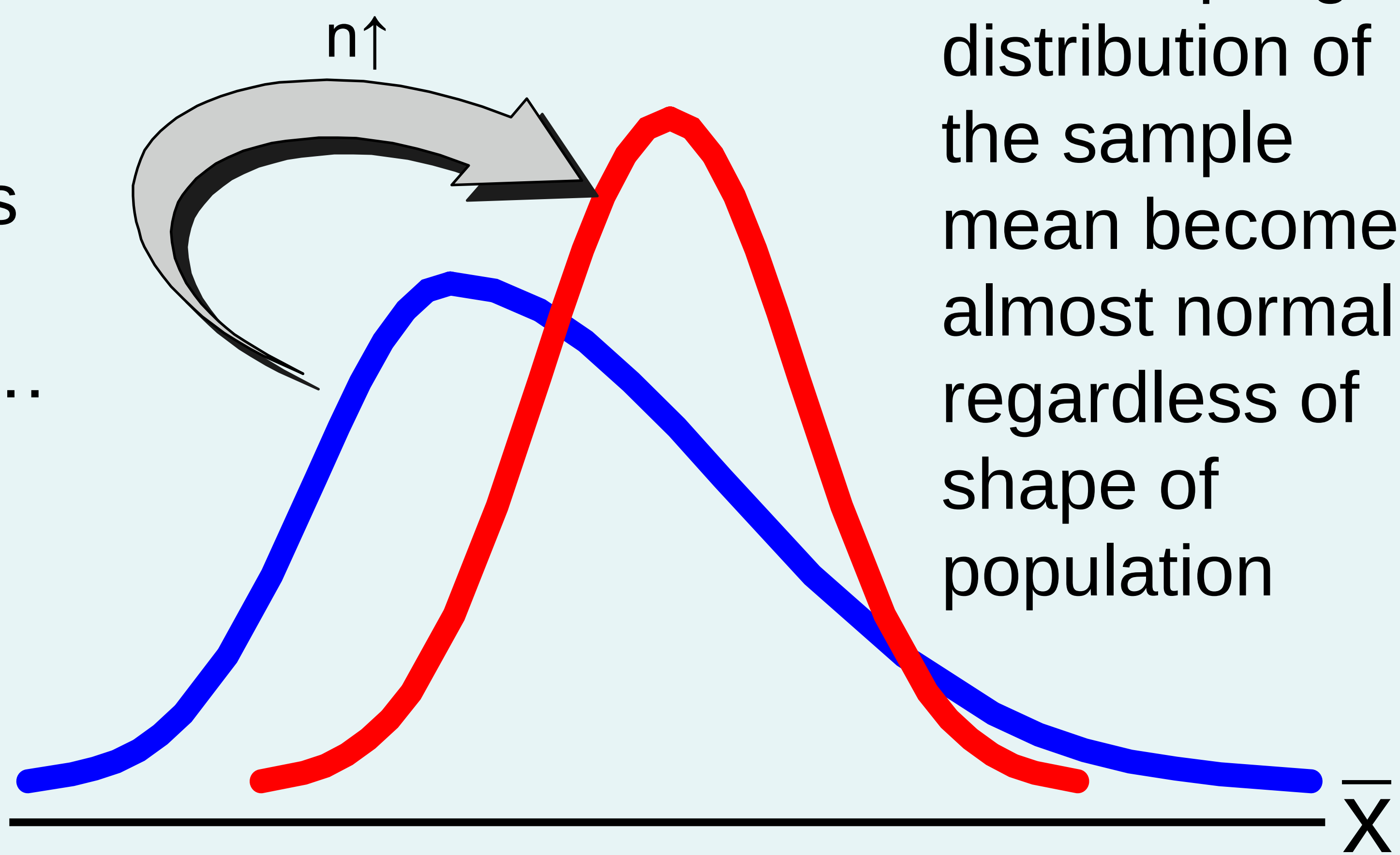


Sample Mean for **non**-Normal Populations

- Central Limit Theorem:
  - Even if the population is not normal,
  - …sample means are approximately normal as long as the sample size is large enough.

Central Limit Theorem

As the sample size gets large enough…



the sampling distribution of the sample mean becomes almost normal regardless of shape of population

Sample Mean if the Population is **not** Normal

(continued)

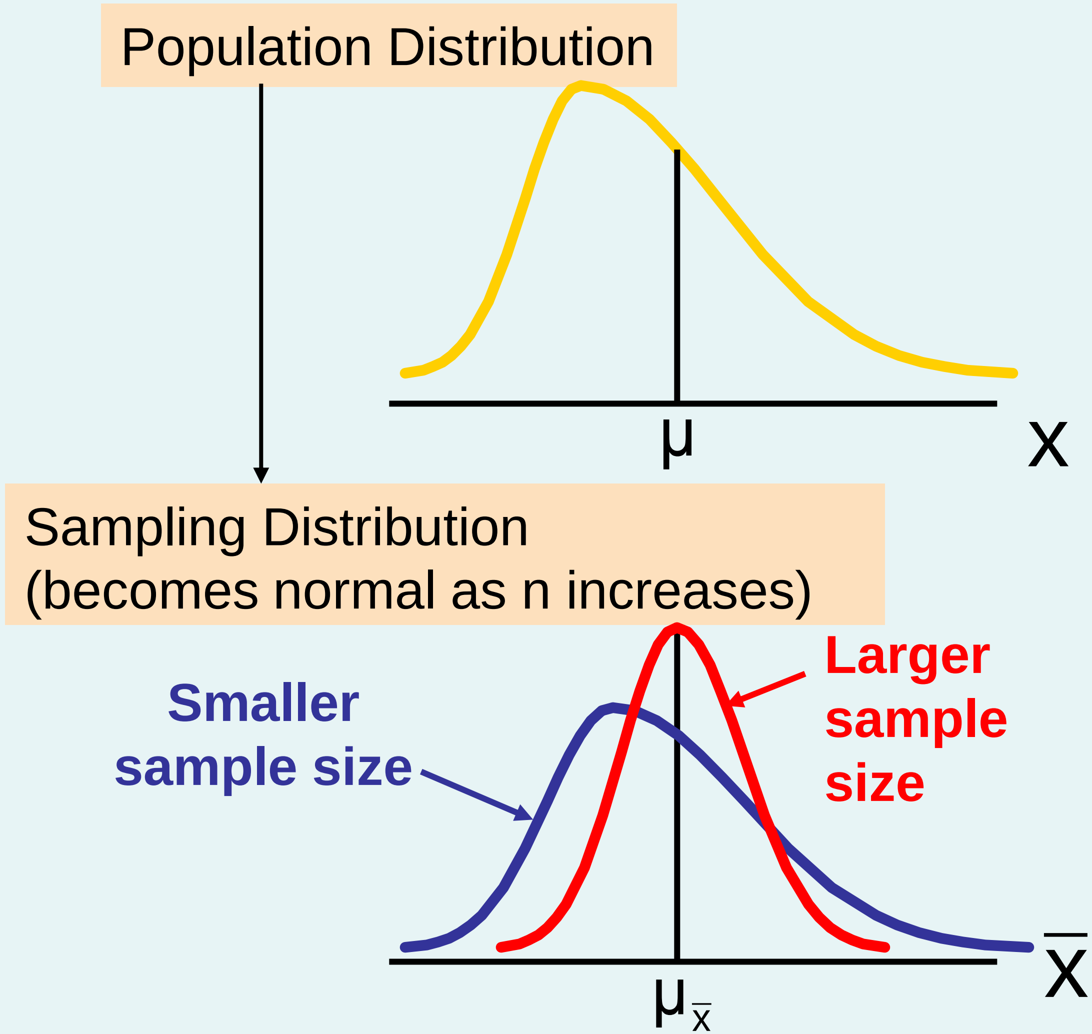
Sampling distribution properties:

Central Tendency

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

Variation

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



How Large is Large Enough?

- For most distributions, n > 30 will give a sampling distribution that is nearly normal
- For fairly symmetric distributions, n > 15
- For normal population distributions, the sampling distribution of the mean is <u>always</u> normally distributed

Z-value for Sampling Distribution of the Mean

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

$$Z = \frac{(\overline{X} - \mu_{\overline{X}})}{\sigma_{\overline{X}}} = \frac{(\overline{X} - \mu)}{\sigma / \sqrt{n}}$$

where:
$\overline{X}$ = sample mean
μ = population mean
σ = population standard deviation
n = sample size

Example

- Suppose a 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?

Example

(continued)

Solution:

- Even if the population is not normally distributed, the central limit theorem can be used (n > 30)
- … so the sampling distribution of $\bar{X}$ is approximately normal
- … with mean $\mu_{\bar{X}} = 8$
- …and standard deviation $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{36}} = 0.5$

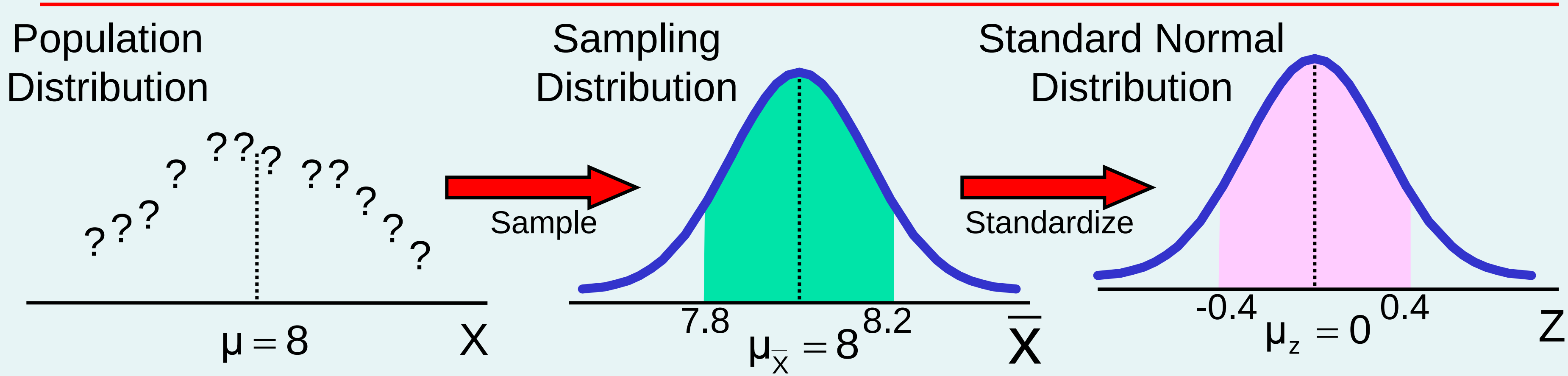
Example

(continued)

Solution (continued):

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

$$= P(-0.4 < Z < 0.4) = 0.6554 - 0.3446 = \boxed{0.3108}$$



Chapter Summary

In this chapter we discussed

- Continuous random variables
- Normal distribution
- Sampling distributions
- The sampling distribution of the mean
  - For normal populations
  - Using the Central Limit Theorem
- Calculating probabilities using sampling distributions