Learning Outcomes

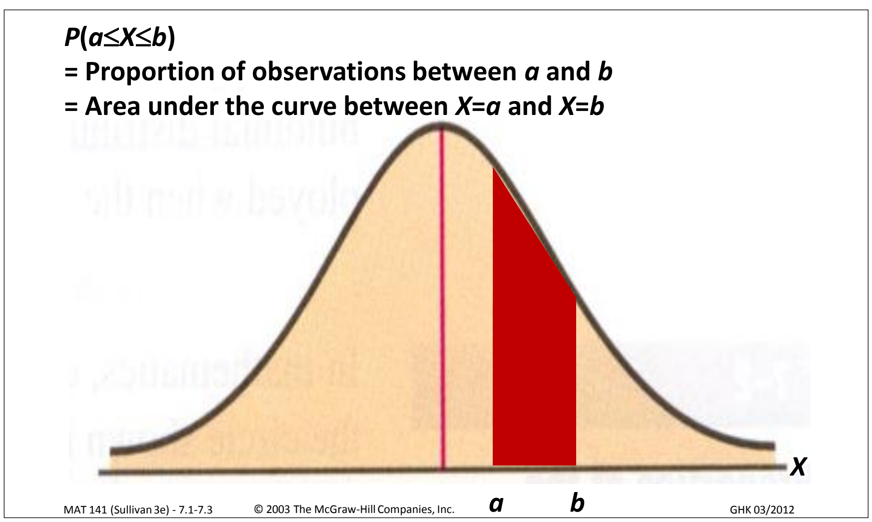
After we cover Section 7.2, you should be able to:

1. Describe what is meant by a z-score (without describing any arithmetic operations or writing a formula) and calculate the z-score for a given value of a random variable X.
2. Describe what is meant by a *standard normal distribution.*
3. Find areas (i.e., probabilities) under a (standard) normal distribution curve using Table V and/or calculator functions.
4. Given a normally distributed random variable X, use Table V and/or calculator functions to find:
   a. The value(s) of X that bound a given percentage of the observations (e.g., the middle 30%, the upper 10%).
   b. A given percentile of the random variable X.
   c. z_α , i.e., the z-score that is to the LEFT of a region whose area is α.
5. Draw a rough sketch of the (standard) normal distribution curve and shade the area associated with the probability being calculated. Use the Empirical Rule to draw a reasonably accurate graph (e.g., the curve will approach the horizontal axis at about 3 standard deviations on either side of the mean).
6. Apply these techniques to real world applications involving (approximately) normally distributed random variables.

Review: Interpreting the Area Under a Normal Probability Density Curve

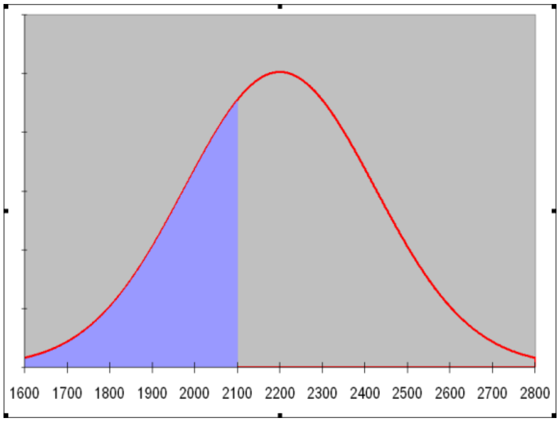
Probabilities (and proportions) are equivalent to areas under the probability density curve.

- This is true for *any* probability density curve, not just normal distributions!



At the end of Section 7.1 we started to look at the following example:

EXAMPLE: Giraffe Weights
Assume that giraffe weights are (approximately) normally distributed with a mean of 2200 lbs. and a standard deviation of 200 lbs.

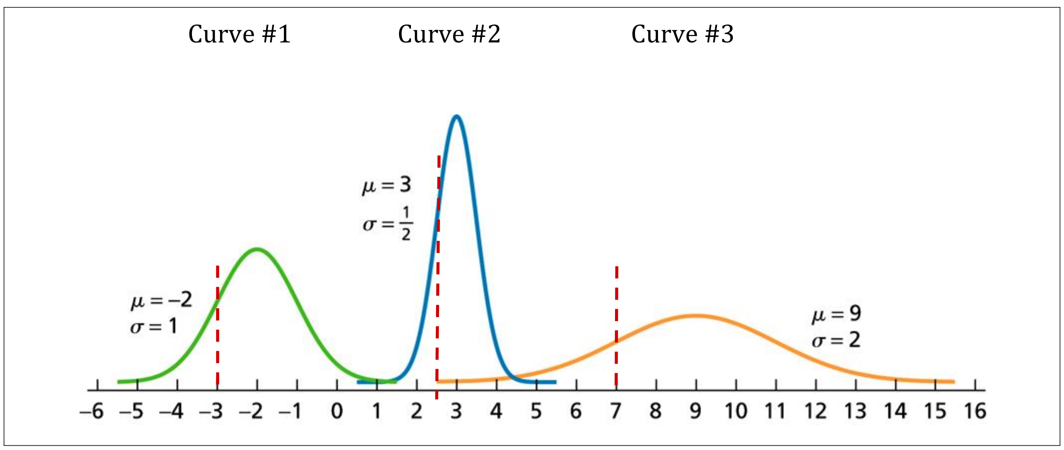


We learned that the shaded area represents:

- The probability that a randomly selected giraffe weighs less than 2100 lbs.
- The proportion of giraffes who weigh less than 2100 lbs.

In Section 7.2 we will learn how to calculate the shaded area.

Finding the Area Under a Normal Probability Density Curve



$Z =$ ____ ____ ____

The figure above shows probability density curves for three random variables on the same horizontal axis.

Suppose we want to find the following probabilities:

- For Curve #1: $P(X \leq -3)$
- For Curve #2: $P(X \leq 2.5)$
- For Curve #3: $P(X \leq 7)$

In order to find these probabilities, we need to calculate the areas under each curve to the left of the dashed line. We can calculate probabilities using a table or using calculator functions. Calculator functions are generally faster, more accurate and easier to use, but using the tables (at least at the beginning) may force you to learn more about the characteristics of the normal distribution!

Finding the Area Under a Normal Curve Using a Table

To find the area (probability) under a normal probability density curve, we will use a table where:

- You look up a value on the horizontal axis, and
- The table tells you the area under the curve to the left of that value.

However, each normal probability density curve is defined by the mean μ and standard deviation σ, so we would need a separate table for each curve. To see this more clearly, look at how $P(X \le 2.5)$ varies for each of the three curves on page 2. (Remember, each normal curve continues on the right to $+\infty$ and on the left to $-\infty$):

Curve #1	$P(X \le 2.5)$ is close to 1.	Almost 100% of the area under this curve is to the left of 2.5.
Curve #2	$P(X \le 2.5)$ is less than 0.5.	Less than 50% of the area under this curve is to the left of 2.5.
Curve #3	$P(X \le 2.5)$ is close to 0.	Practically none of the area under this curve is to the left of 2.5.

However, it would be impossible to construct a table for each normal curve because there are infinitely many normal curves!

We can avoid creating infinitely many tables by thinking about **<u>*z*-scores</u>** on the horizontal axis instead of the values of the random variable *X*. Recall (from Section 3.4) that the *z*-score describes the position of a data point *X* as the number of standard deviations from the mean:

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

EXAMPLE: Calculate *z*-scores
In Curve #3 (where $\mu = 9$ and $\sigma = 2$), find the *z*-score for $X = 7$.

By calculating the *z*-scores for values of *X*, we describe the values of the random variable relative to the mean and the standard deviation. For example, in Curve #3:

- 7 is _______ standard deviation(s) to the ____________ of the mean.

Therefore, if we had a table that showed areas under the curve relative to *z*-scores, we could look up the area to the left of *Z*= _____ to find the area to the left of *X*=7.

Finding the Area Under a Normal Curve Using a Table (cont'd)

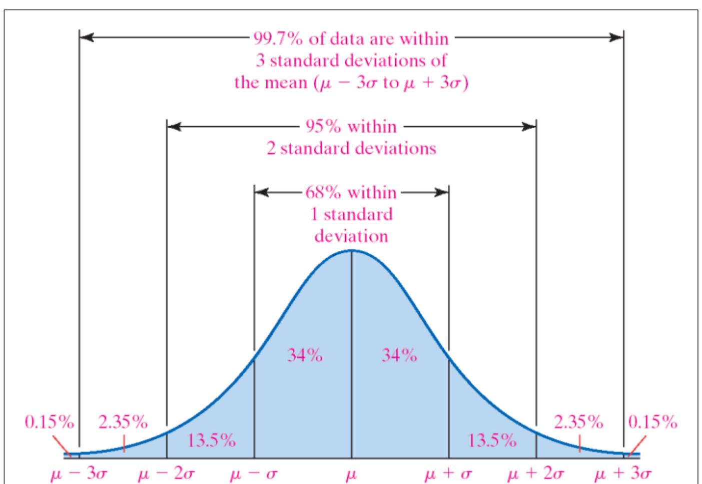
Why is it important to consider z-scores instead of values of the random variable X? In all three curves at the bottom of page 2, the dashed line is 1 standard deviation to the left of the mean. In other words, all three dashed lines represent a z-score of –1 and the **area to the left of Z= –1 is the same in all three curves** (and in *any* other normal distribution, for that matter).

Therefore, if we had just **one table** that told us the area to the left of Z= –1, we would know all three of the following areas:

- The area in Curve #1 to the left of –3
- The area in Curve #2 to the left of 2.5
- The area in Curve #3 to the left of 7

EXAMPLE: Finding the Area Under a Normal Distribution Curve
Use Table V to find the area to the left of Z= –1.

If you are not convinced that the area to the left of Z= –1 is the same for all three curves, recall the Empirical Rule from Section 3.2 (also known as the 68-95-99.7 Rule). The Empirical Rule tells us that in ***every*** normal distribution, approximately 68% of the observations lie within 1 standard deviation of the mean. This means that the area under the curve between Z= –1 and Z= 1 is approximately 0.68. Therefore the area outside of this region is 0.32. Because the normal distribution is symmetric, half that area (i.e., 0.16) lies to the left of Z= –1. Because the Empirical Rule applies to *every* normal distribution, the area to the left of Z= –1 must be the same in all three curves on page 2!

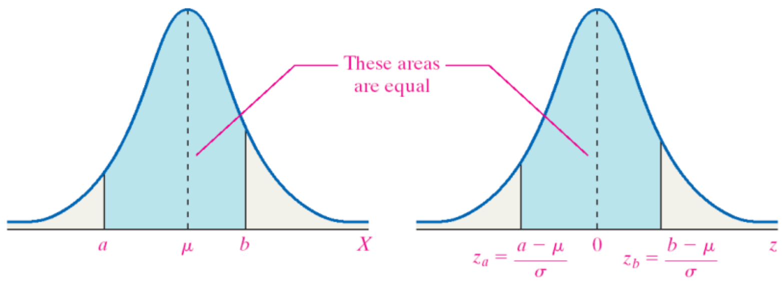


Source: Prentice Hall

<u>Finding the Area Under a Normal Curve Using a Table (cont'd)</u>

Calculating the area under the curve of a normal distribution

Converting to a Standard Normal Distribution



Normal distribution

μ, σ

Standard normal distribution

μ=0, σ=1

Notice that when we graph the probabilities associated with *z*-scores instead of probabilities of the random variable *X*, we only change the scale on the horizontal axis. The shape of the curve – and the shaded area under the curve – remain the same. Thus:

$$P(a \le X \le b) = P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right)$$

$$P(X \le a) = P\left(Z \le \frac{a-\mu}{\sigma}\right)$$

If *X* has a normal distribution, notice that the *z*-scores (shown above in the curve on the right) have a normal distribution with $\mu = 0$ and $\sigma = 1$. This normal distribution is called a ***<u>standard</u>* <u>normal distribution</u>**. The probability tables in the textbook are based on the standard normal distribution.

Any normal distribution can be converted to a standard normal distribution by converting *X*-values to *z*-scores. Therefore, **we only need one table based on *z*-scores** to find areas under *any* normal probability density curve.

<u>Summary</u>

- If we have a continuous random variable *X* with a normal distribution, the *z*-scores will have a *standard* normal distribution (i.e., $\mu = 0$ and $\sigma = 1$).
- To find $P(X \le a)$, use a standard normal distribution table (i.e., a table based on *z*-scores).

  Look up $P\left(Z \le \frac{a-\mu}{\sigma}\right)$ where $\frac{a-\mu}{\sigma}$ is the *z*-score associated with $X = a$.

EXAMPLE: Finding the Area Under a Normal Curve
In Curve #3 on page 2, use Table V to calculate:

- $P(X > 7)$

- $P(7 \le X \le 10.7)$

  STEP 1: Convert *X*-values to *z*-scores.

  STEP 2: Look up the *z*-score in the margins of the table. The body of the table shows the area to the left of the *z*-score.

  STEP 3: To compute the area between two *z*-scores, look up two areas and find the difference.

SUGGESTION: When calculating probabilities of normally distributed random variables, drawing a diagram may make it easier to see how to calculate the probability.

Finding the Area Under a Normal Curve Using Calculator Functions

Function In 2^{nd} > DISTR menu	**Description**	**Notes**
normalcdf(a, b)	Finds the area between $Z=a$ and $Z=b$ under a *standard* normal distribution curve.	Use $-$EE99 (i.e., -10^{99}) for $-\infty$ Use EE99 (i.e., 10^{99}) for $+\infty$
normalcdf (a, b, μ, σ)	Finds the area between $X=a$ and $X=b$ under a normal distribution curve with mean μ and standard deviation σ.	

CAUTION: Do NOT use the normalpdf function! This function measures the *height* of the curve, which we do not need to know.

EXAMPLE: Finding the Area Under a Normal Distribution Curve
Use calculator functions to find the following probabilities:

- Find $P(-1.0 \le Z \le 0.85)$.

- Find $P(7 \le X \le 10.7)$ in Curve #3.

- Find $P(Z \le -1.0)$

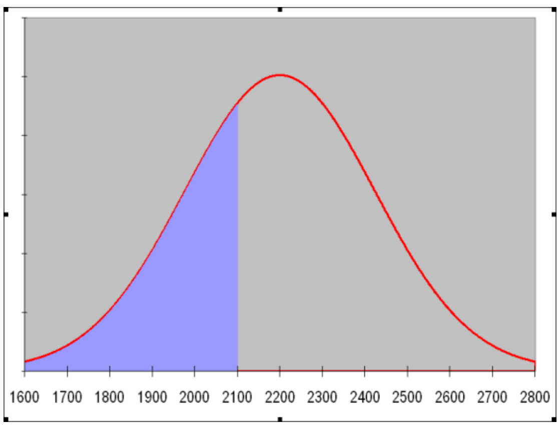
- Find $P(X \le 7)$ in Curve #3.

- Find $P(X > 7)$ in Curve #3.

Rounding (Using the Table or Calculator Functions)

- When using Table V, ROUND (*never truncate*) z-scores to two decimal places.
  - For example, if Z=1.63827, look up the area associated with Z=1.64.
- Round areas (probabilities) to four decimal places, unless told otherwise.
- Avoid reporting an area (probability) of 0 or 1 unless the events are truly *impossible* or *certain* events, respectively.
  - Use P<0.0001 or P>0.9999, respectively.
  - See P. 343 of the textbook ("Some Cautionary Thoughts") for a more detailed discussion on rounding.

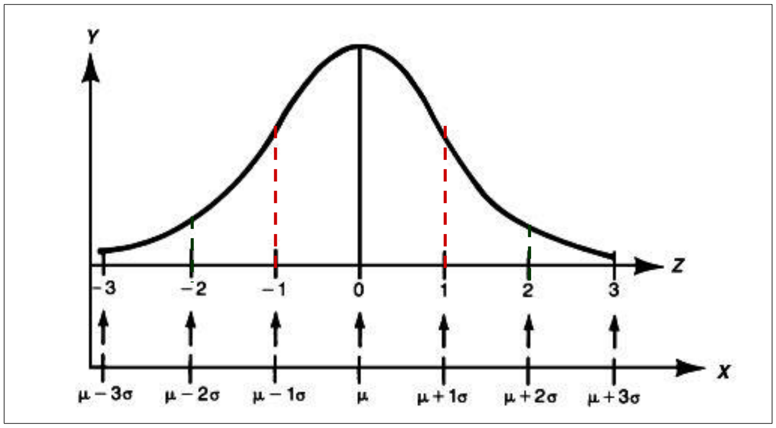
EXAMPLE: Giraffe Weights
Assume that giraffe weights are (approximately) normally distributed with a mean of 2200 lbs. and a standard deviation of 200 lbs.



- Find the probability that a randomly selected giraffe weighs less than 2100 lbs., i.e., find the proportion of giraffes that weigh less than 2100 lbs.

- Find the probability that a randomly selected giraffe weighs 2100 lbs. or more?

EXAMPLE: Normal Distributions and the Empirical Rule
Also known as the "68-95-99.7" rule



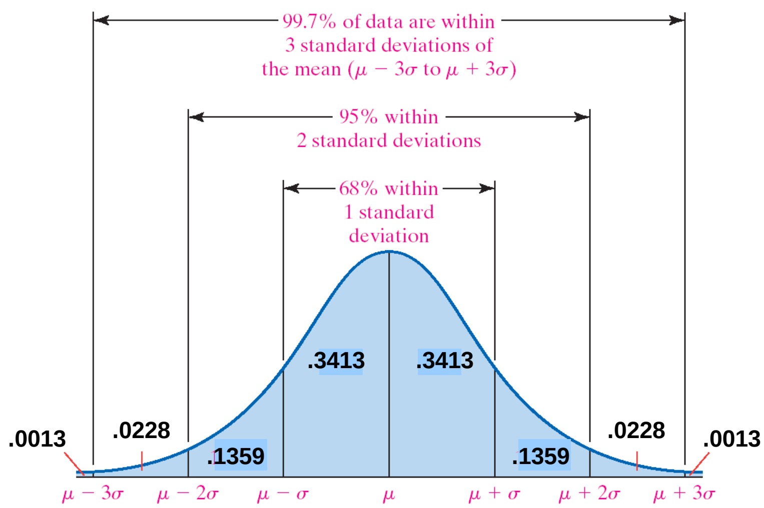
The following calculations show where the numbers in the Empirical Rule come from,

- $P(-1 \le Z \le 1)$

- $P(-2 \le Z \le 2)$

- $P(-3 \le Z \le 3)$

Normal Distributions and the Empirical Rule

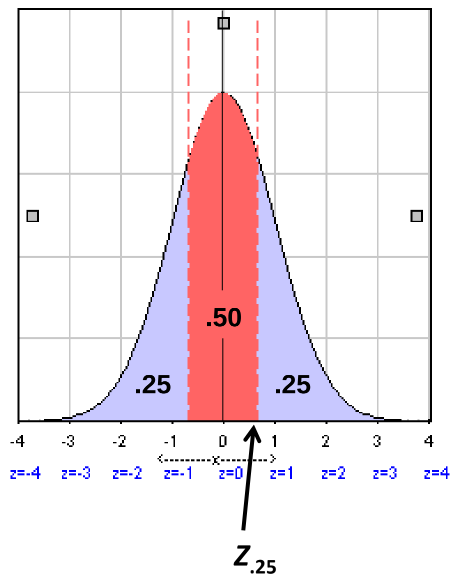


Source: Prentice Hall

Finding *z*-scores or *X*-values that Define Areas Under the Curve

In addition to using Table V or the graphing calculator to find areas under the normal distribution curve, we can use Table V or the calculator to find values along the horizontal axis that correspond to an area under the curve.

EXAMPLE: Finding *z*-scores that Define Areas Under the Curve.



- Find the *z*-scores that define the middle 50% of the distribution.
- What percentiles do these *z*-scores represent?

Notation: z_α refers to the *z*-score where the area to the *right* of the *z*-score is α.

For example, $z_{0.25}$ refers to the *z*-score that separates the *upper* 25% of the area from the lower 75% (see diagram above).

- $z_{0.25}$ is equivalent to P_{75}.

<u>Helpful Calculator Functions</u>

<table>
<tr><th>Function
In 2nd > DISTR menu</th><th>Description</th><th>Notes</th></tr>
<tr><td>invNorm(a)</td><td>Given a standard normal distribution (i.e., μ=0 and σ=1), this function finds the z-score such that the area to the left of the z-score is “a.”</td><td rowspan="2">The left boundary of the area is −∞.</td></tr>
<tr><td>invNorm(a, μ, σ)</td><td>Given a normal distribution with mean μ and standard deviation σ, this function finds the X-value such that the area to the left of the X-value is “a.”</td></tr>
</table>

EXAMPLE: Giraffe Weights
Assume that giraffe weights are (approximately) normally distributed with a mean of 2200 lbs. and a standard deviation of 200 lbs.

- What weights define the middle 60% of the giraffe population?
- What percentiles do these weights represent?

Summary of Calculator Functions for Normal Distributions

To find areas under a normal curve, given values of *X* (or *z*):

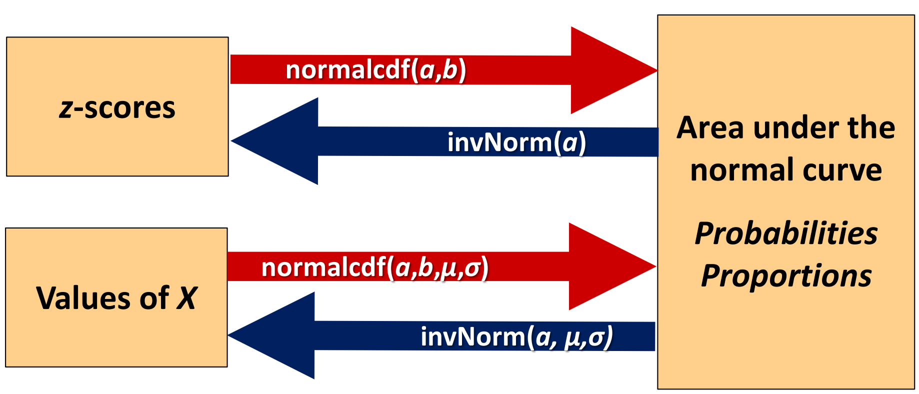
Function In 2nd > DISTR menu	**Description**	**Notes**
normalcdf(a, b)	Finds the area between $Z=a$ and $Z=b$ under a *standard* normal distribution curve.	Use –EE99 (i.e., -10^{99}) for $-\infty$ Use EE99 (i.e., 10^{99}) for $+\infty$
normalcdf (a, b, μ, σ)	Finds the area between $X=a$ and $X=b$ under a normal distribution curve with mean μ and standard deviation σ.	

To find values of *X* (or *z*) given the area under a normal curve:

Function In 2nd > DISTR menu	**Description**	**Notes**
invNorm(a)	Given a *standard* normal distribution (i.e., μ=0 and σ=1), this function finds the z-score such that the area to the *left* of the z-score is "a."	The left boundary of the area is $-\infty$.
invNorm(a, μ, σ)	Given a normal distribution with mean μ and standard deviation σ, this function finds the X-value such that the area to the *left* of the X-value is "a."	

Be sure to use the correct calculator function:

Function	**Input**	**Output**
normalcdf	z-scores or X-values	Area (i.e., probability, proportion) under a normal curve.
invNorm	Area (i.e., probability, proportion) under a normal curve.	z-scores or X-values



THIS PAGE IS INTENTIONALLY LEFT BLANK

Applications of Normal Distributions

In each exercise:

- Draw a rough diagram, including a "shaded area" under the curve.
- Solve the problem using Table V and/or the appropriate calculator function.

EXAMPLE: Army Helmets
The Army has found that it is best to order custom-made helmets for soldiers whose head circumference is at or above the 95th percentile. If head circumference is normally distributed with a mean of 22.8 inches and a standard deviation of 1.1 inches, what is the 95th percentile?

EXAMPLE: Cholesterol Levels
Cholesterol levels for US adults are normally distributed with a mean of 213 and a standard deviation of 48.4.

- What is the probability that a randomly selected adult will have "borderline high" cholesterol (defined as a total cholesterol level between 200 and 250)?

EXAMPLE: IQ Scores
If IQ scores are normally distributed with μ=100 and σ=16, then

- What is the probability that a randomly selected person will have an IQ greater than 110?
- What percentile is associated with an IQ of 110?

THIS PAGE IS INTENTIONALLY LEFT BLANK

__**Normal Distribution Curves for Application Problems**__

Army Helmets	Cholesterol Levels	IQ Scores