# The Basic Practice of Statistics
# Ninth Edition

David S. Moore William I. Notz

## Chapter 3
## The Normal Distributions

Lecture Slides

# In Chapter 3, we cover …

- Density curves
- Describing density curves
- Normal distributions
- The 68–95–99.7 rule
- The standard Normal distribution
- Finding Normal proportions
- Using the standard Normal table
- Finding a value given a proportion

# Density Curves (1 of 6)

We now have a toolbox of graphical and numerical methods for describing distributions. What is more, we have a clear strategy for exploring data on a single quantitative variable.

---

**EXPLORING A DISTRIBUTION**

1. Plot your data: usually a histogram, stemplot or boxplot.
2. Look for the overall pattern (shape, center, and variability) and for striking deviations, such as outliers.
3. Calculate a numerical summary to briefly describe center and variability.

---

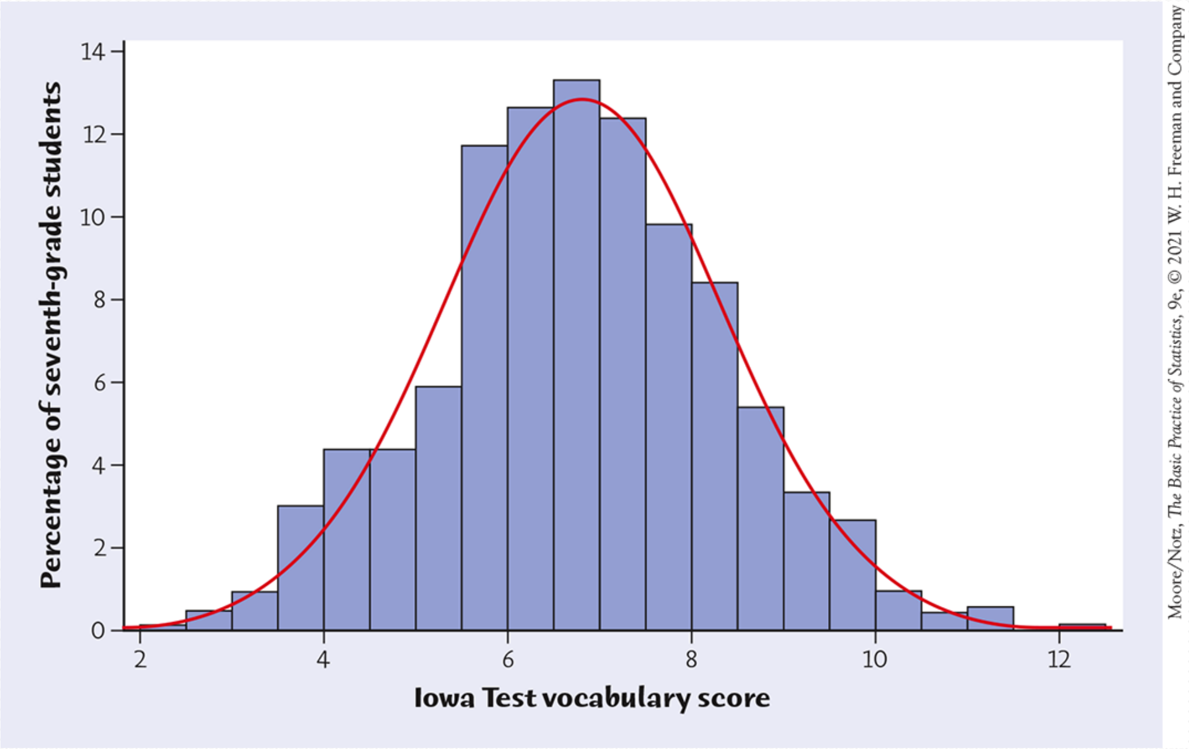
**Now we add one more step to this strategy:**

4. Sometimes the overall pattern of a large number of observations is so regular that we can describe it by a smooth curve.

# Density Curves (2 of 6)

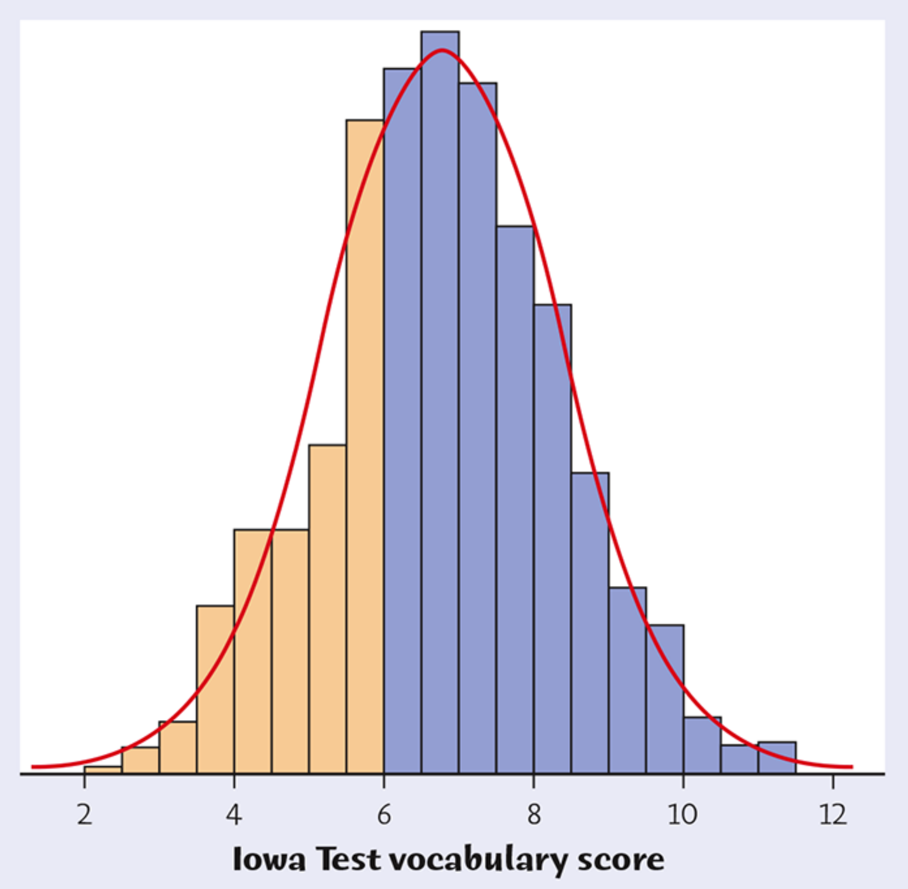
**Example:** Here is a histogram of the vocabulary scores of 947 seventh-graders.

The smooth curve closely approximates the tops of the histogram bars and provides a good description of the overall data pattern.



# Density Curves (3 of 6)

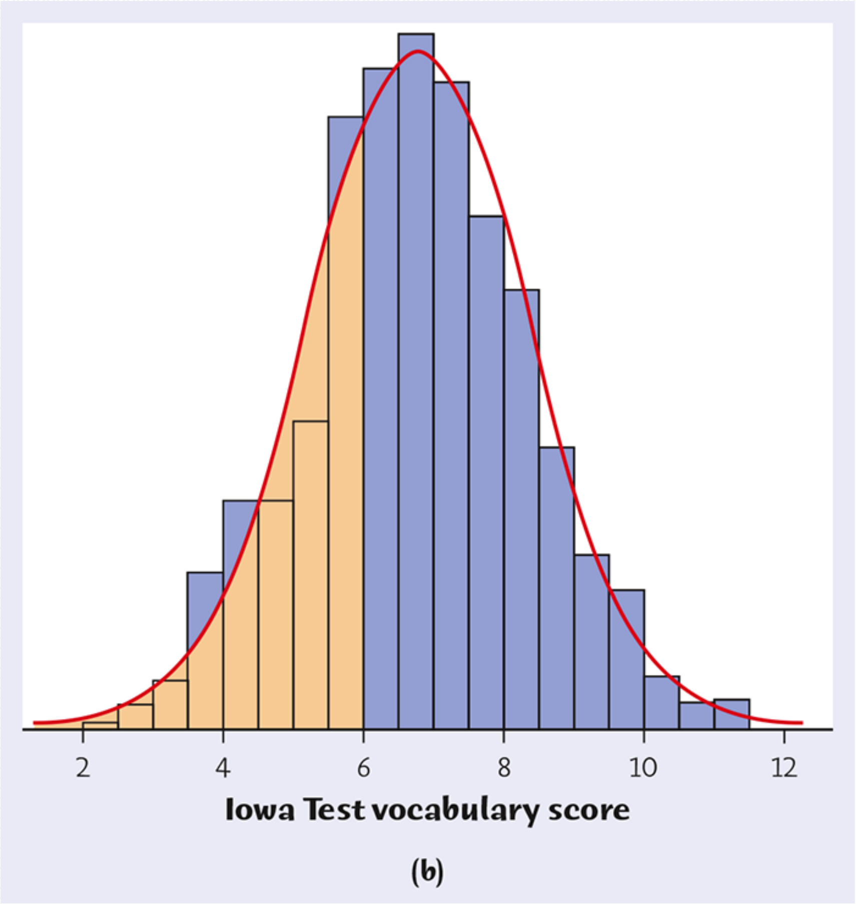
The areas of the shaded bars in this histogram represent the proportion of scores in the observed data that are less than or equal to 6.0. This proportion is equal to 0.303.




(a)

# Density Curves (4 of 6)

Now the area under the smooth curve to the left of 6.0 is shaded. If the scale is adjusted so that the total area under the curve is exactly 1, then this curve is called a **density curve.** The proportion of the area to the left of 6.0 is now equal to 0.293.



# Density Curves (5 of 6)

A **density curve** is a curve that:

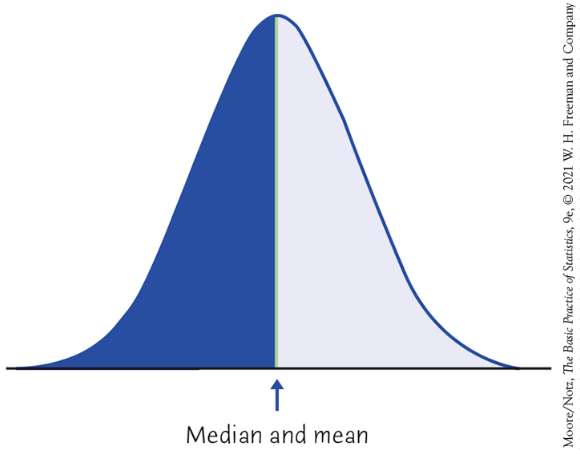
- is always on or above the horizontal axis.
- has an area of exactly 1 underneath it.

- A density curve describes the overall pattern of a distribution. The area under the curve and above any range of values on the horizontal axis is the proportion (or *probability*) of all observations that fall in that range.

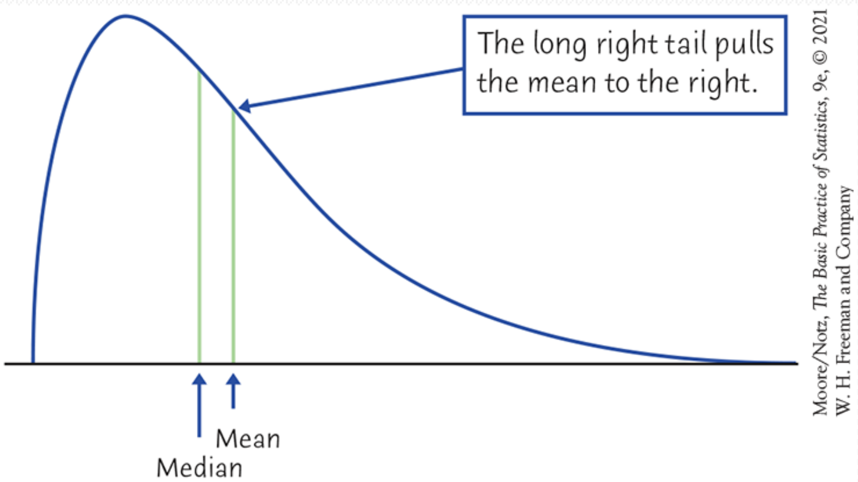
*Caution:* No set of real data is *exactly* described by a density curve. The curve is an idealized description that is easy to use and accurate enough for practical use.

# Describing Density Curves

- The **median** of a density curve is the equal-areas point, which divides the area under the curve in half.
- The **mean** of a density curve is the balance point, at which the curve would balance if it were made of solid material.
- The median and the mean are the same for a symmetric density curve—they both lie at the center of the curve. The mean of a skewed curve is pulled away from the median in the direction of the long tail.




(a)




(b)

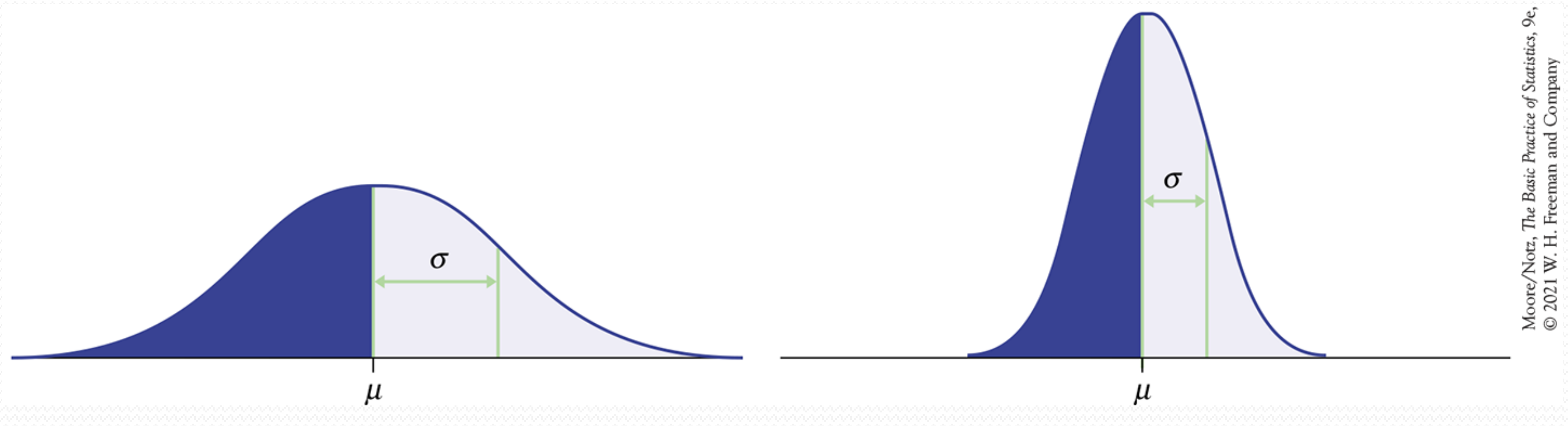
# Density Curves (6 of 6)

- The mean and standard deviation computed from a sample of observations (data) are denoted by $\bar{x}$ and $s$, respectively.

- The mean and standard deviation of the true population distribution represented by the density curve are denoted by $\mu$ ("mu") and $\sigma$ ("sigma"), respectively.

# Normal Distributions (1 of 5)

- One particularly important class of density curve comprises **Normal curves**, which describe **Normal distributions**.
- All Normal curves are symmetric, single-peaked, and bell-shaped.
- Any specific Normal curve is completely described by giving its mean $\mu$ and standard deviation $\sigma$.



- **NOTE:** Normal curves are also named **Gaussian** curves after the mathematician Carl Friedrich Gauss.

# Normal Distributions (2 of 5)

*Unlike most distributions,* any particular Normal distribution is completely specified by two numbers—its mean $\mu$ and standard deviation $\sigma$:

- The mean is located at the center of the symmetric curve and is the same as the median. Changing $\mu$ without changing $\sigma$ moves the Normal curve along the horizontal axis without changing its variability.

- The standard deviation $\sigma$ controls the variability of a Normal curve. When the standard deviation is larger, the area under the normal curve is less concentrated about the mean.

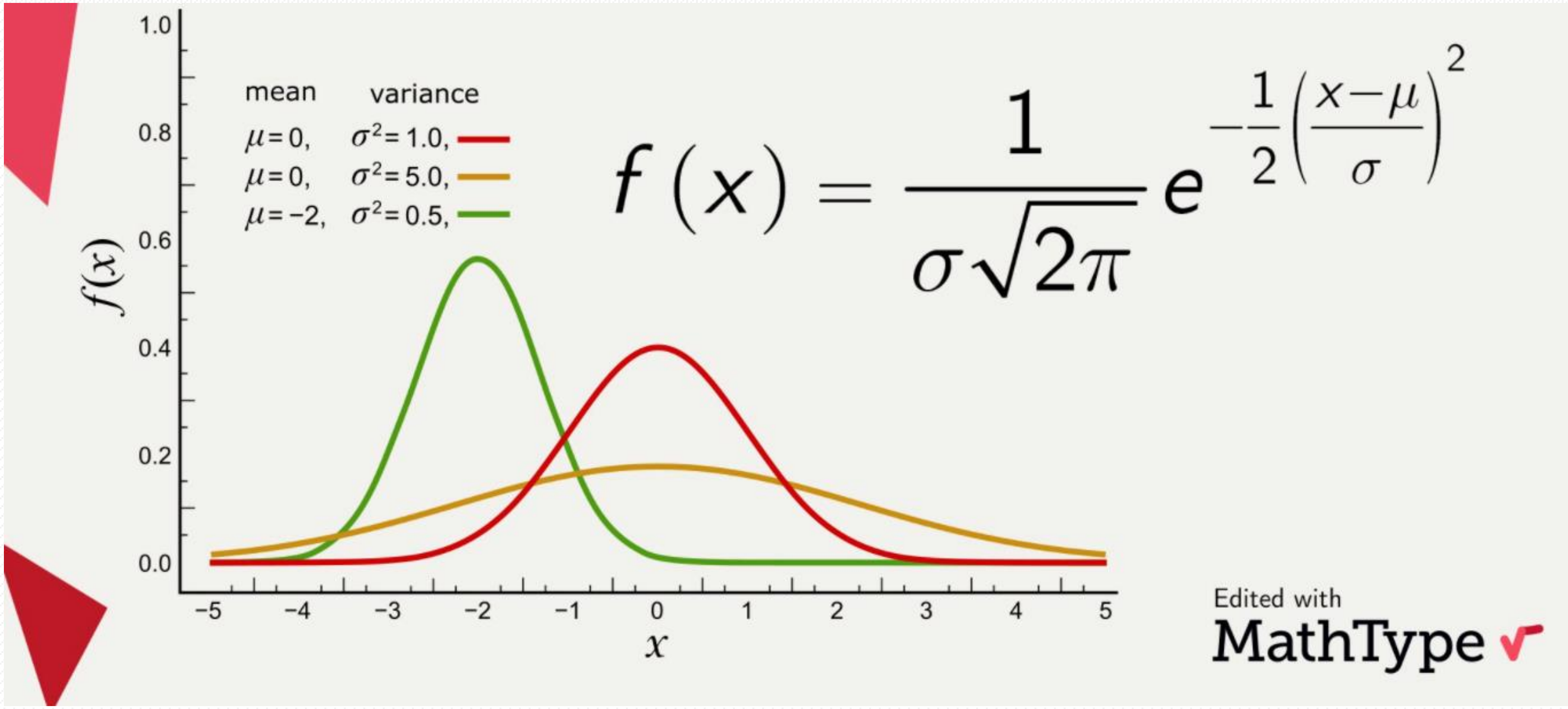
- The standard deviation is the distance from the center to the change-of-curvature points on either side.

# Normal Distributions (3 of 5)

- A **Normal distribution** is described by a Normal density curve. Any particular Normal distribution is completely specified by two numbers: its mean $\mu$ and standard deviation $\sigma$.

- The mean of a Normal distribution is at the center of the symmetric Normal curve. The standard deviation is the distance from the center to the change-of-curvature points on either side.

**(Bonus slide! Just for curiosity, you don't need to memorize this)**
The **Normal density curve** is a mathematical function given by:

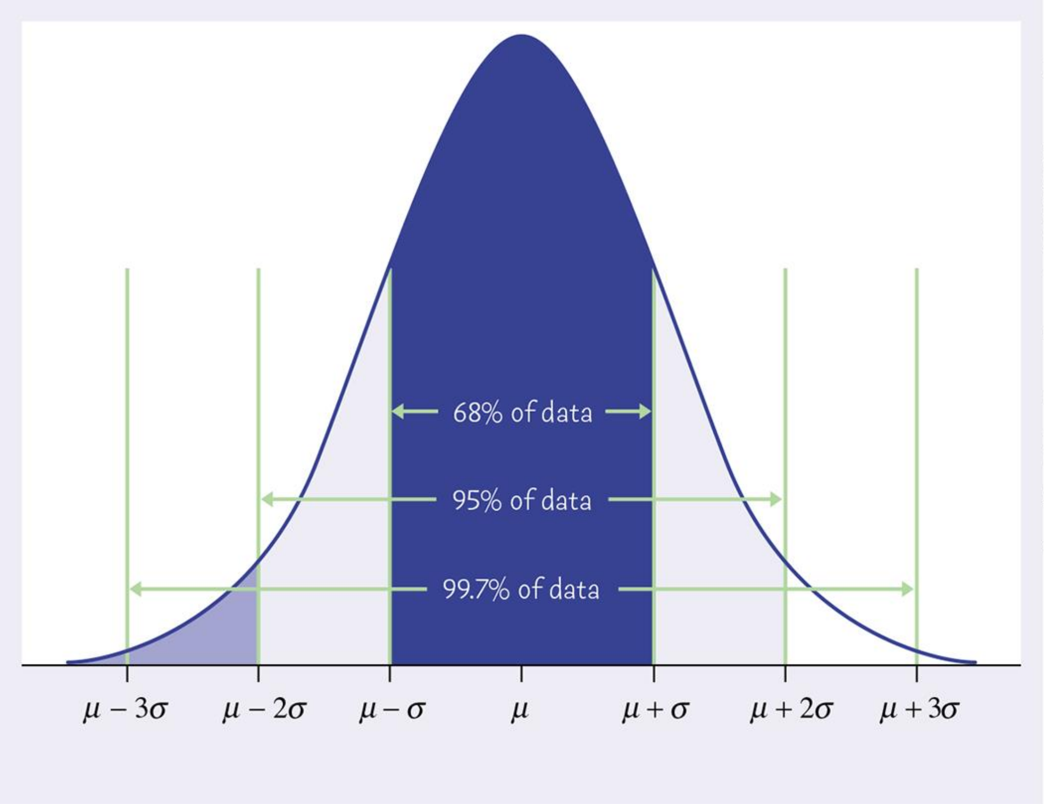


# Normal Distributions (4 of 5)

In the Normal distribution, with mean $\mu$ and standard deviation $\sigma$:

- approximately 68% of the observations fall within $\sigma$ of $\mu$.
- approximately 95% of the observations fall within $2\sigma$ of $\mu$.
- approximately 99.7% of the observations fall within $3\sigma$ of $\mu$.

## The 68–95–99.7 Rule

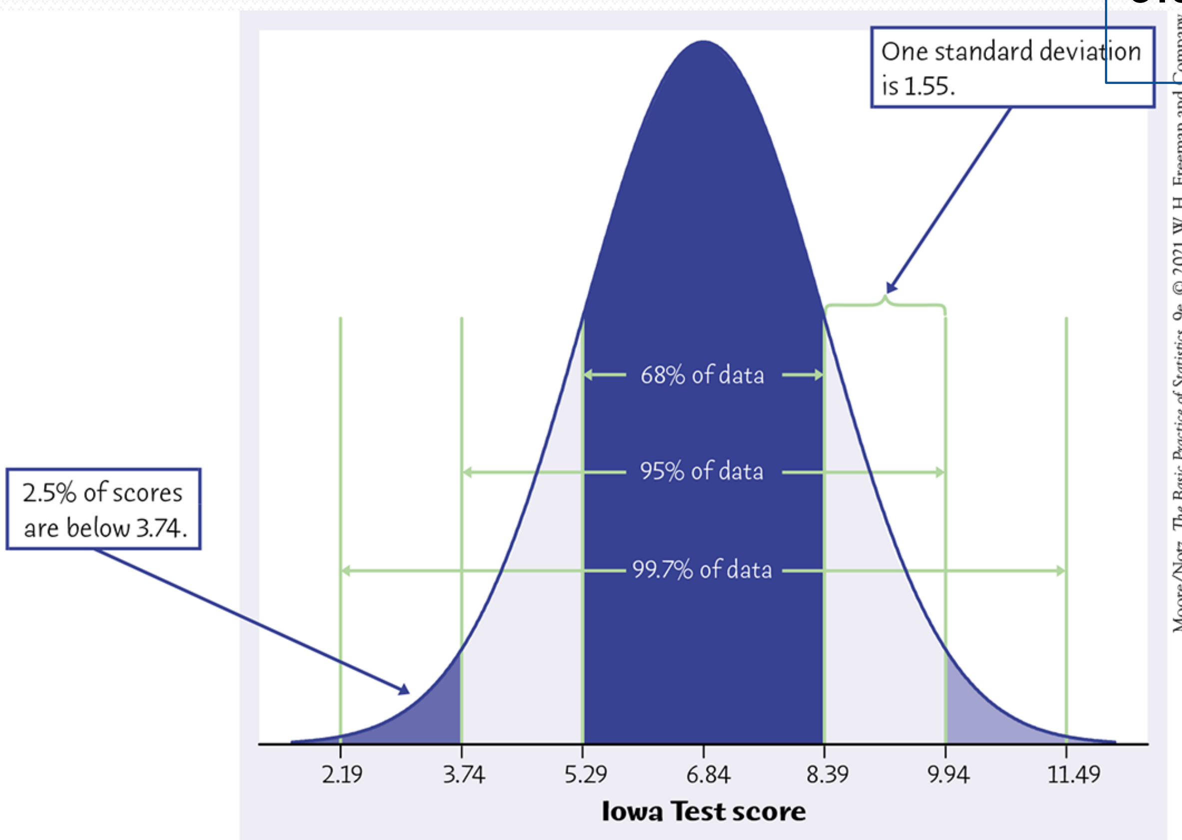


# Normal Distributions (5 of 5)

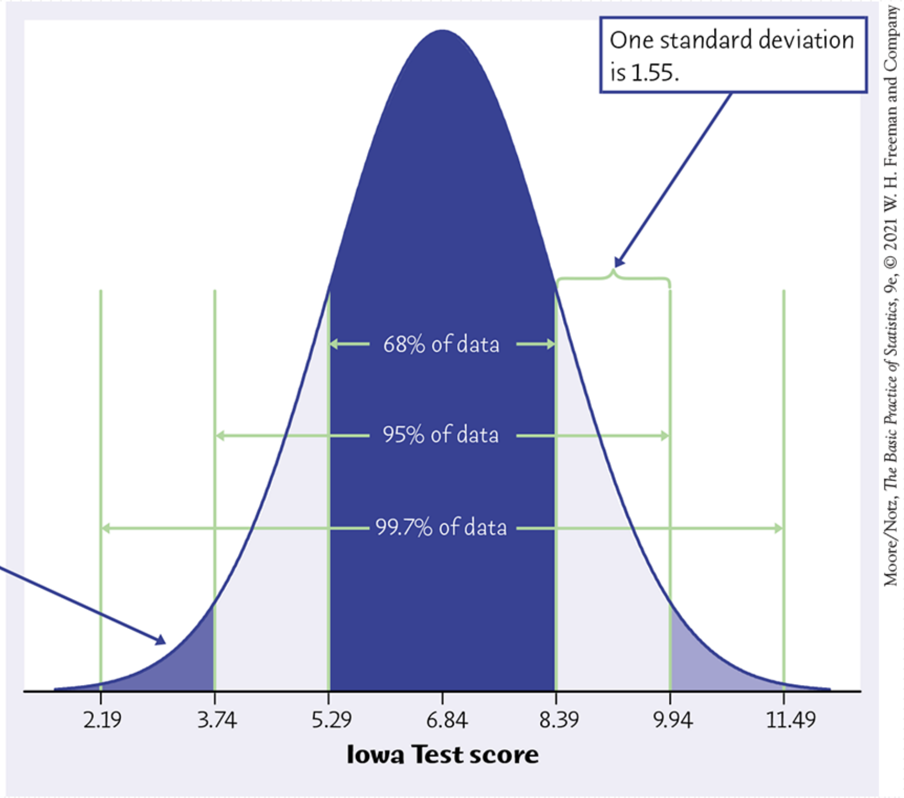
- The distribution of Iowa Test of Basic Skills (ITBS) vocabulary scores for seventh-grade students in Gary, Indiana, is close to Normal. Suppose the distribution is ***N*(6.84, 1.55).**
- Sketch the Normal density curve for this distribution.
- What percent of ITBS scores are between 3.74 and 9.94?
- What percent of the scores are above 5.29?

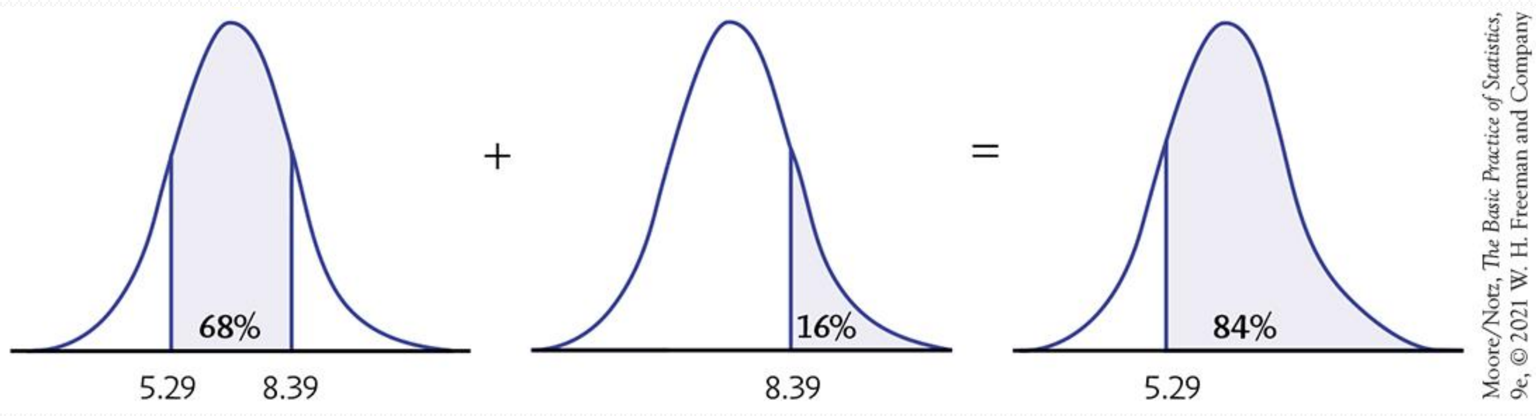
3.74 = mean - 2*sd
= 6.84 - 2*1.55

9.94 = mean + 2*sd
= 6.84 + 2*1.55



# Normal Distributions (5 of 5) – Cont.



# Standardizing (1 of 2)

- All Normal distributions are the same if we measure in units of size $\sigma$ from the mean $\mu$ as center.
- Changing to these units is called *standardizing*.

**STANDARDIZING AND *z*-SCORES**

- If $x$ is an observation from a distribution that has mean $\mu$ and standard deviation $\sigma$, then the **standardized value** of $x$ is

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

- A standardized value is often called a ***z*-score**.

# Standardizing (2 of 2)

**Example:**

- The heights of women aged 20 to 29 in the United States are approximately Normal, with $\mu$ = 64.1 and $\sigma$ = 3.7 inches. The standardized height $z$ can be calculated by:

$$z = \frac{\text{height} - 64.1}{3.7}$$

# Standardizing (2 of 2) – Cont.

**Example:**

$$z = \frac{\text{height} - 64.1}{3.7}$$

- A woman's standardized height is the number of standard deviations by which her height differs from the mean height of all women aged 20 to 29. A woman 70 inches tall, for example, has standardized height

$$z = \frac{70 - 64.1}{3.7} = 1.59$$

  or 1.59 standard deviations *above* the mean.

- Similarly, a woman 5 feet (60 inches) tall has standardized height

$$z = \frac{60 - 64.1}{3.7} = -1.11$$

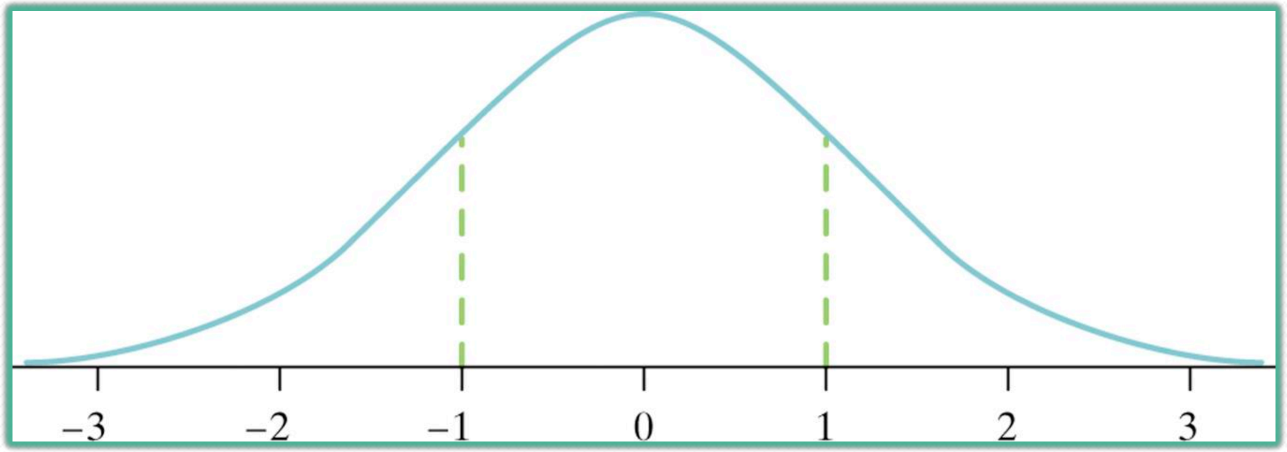
  or 1.11 standard deviations *less than* the mean height.

# The Standard Normal Distribution

- The **standard Normal distribution** is the Normal distribution with mean 0 and standard deviation 1.
- If a variable $x$ has any Normal distribution $N(\mu, \sigma)$, with mean $\mu$ and standard deviation $\sigma$, then the standardized variable

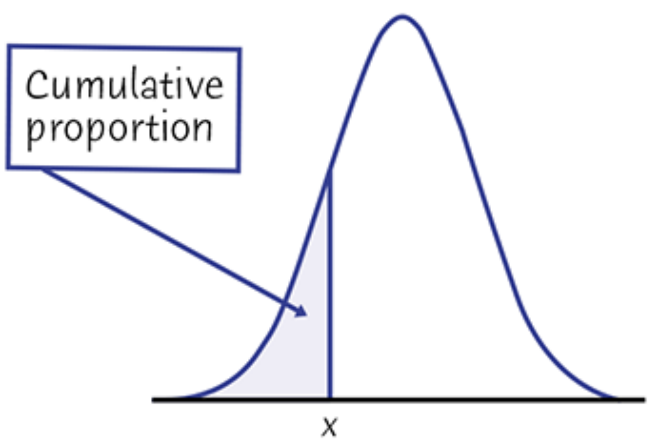
$$z = \frac{x - \mu}{\sigma}$$

has the standard Normal distribution, $N(0,1)$.



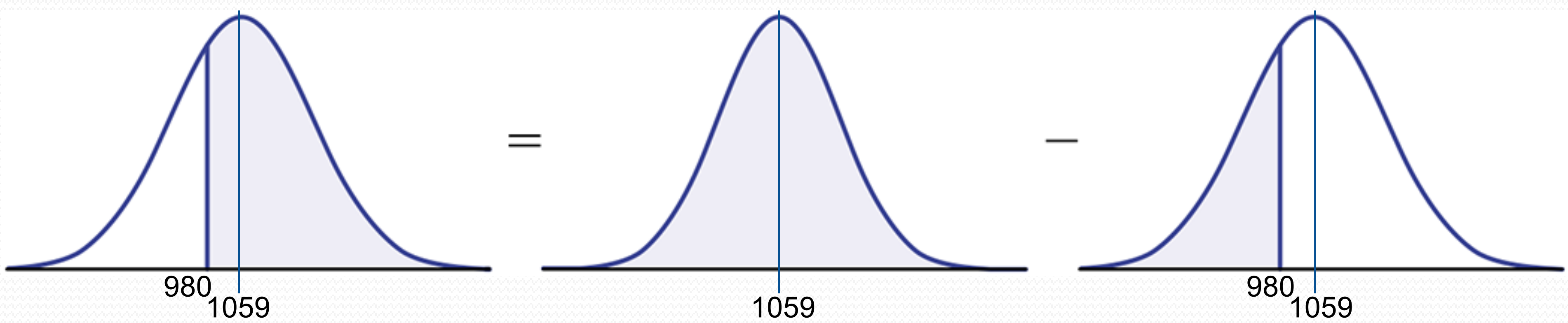
# Cumulative Proportions

The cumulative proportion for a value $x$ in a distribution is the proportion of observations in the distribution that are less than or equal to $x$.



# Cumulative Proportions—Example 3.5 (1 of 6)

- Who qualifies for college sports?
- The combined scores of the almost 2.2 million high school seniors taking the SAT in 2018 were approximately Normal, with mean 1059 and standard deviation 210. What percent of high school seniors meet this SAT requirement of a combined score of 980 or better?
- Here is the calculation in a picture:



# Cumulative Proportions—Example 3.5 (2 of 6)

- Software often uses terms such as "cumulative distribution" or "cumulative probability."
- What percent of high school seniors meet this SAT requirement of a combined score of 980 or better? First using *N(1059,210).*

```
In [3]: import numpy as np
        import pandas as pd
        import seaborn as sns
        import matplotlib.pyplot as plt
        from scipy.stats import norm

In [4]: norm.cdf(980, loc = 1059, scale = 210) ## this gives the proportion below 980

Out[4]: 0.35338764759507846

In [5]: 1 - norm.cdf(980, loc = 1059, scale = 210) ## this gives the desired proportion of students above 980

Out[5]: 0.6466123524049215
```

approx. 0.65 or 65%

# Cumulative Proportions—Example 3.5 (3 of 6)

- What percent of high school seniors meet this SAT requirement of a combined score of 980 or better? Now using z-score and *N(0,1)* – the standard normal distribution

```
In [10]: z = (980-1059)/210
         z

Out[10]: -0.3761904761904762

In [11]: norm.cdf(z)

Out[11]: 0.35338764759507846

In [12]: 1 - norm.cdf(z)

Out[12]: 0.6466123524049215
```

approx. 0.65 or 65%, same as previous slide

# Cumulative Proportions—Example 3.5 (4 of 6)

- What percent of high school seniors meet this SAT requirement of a combined score of 980 or better? Now using standard Normal table – **TABLE A at the end of the textbook**

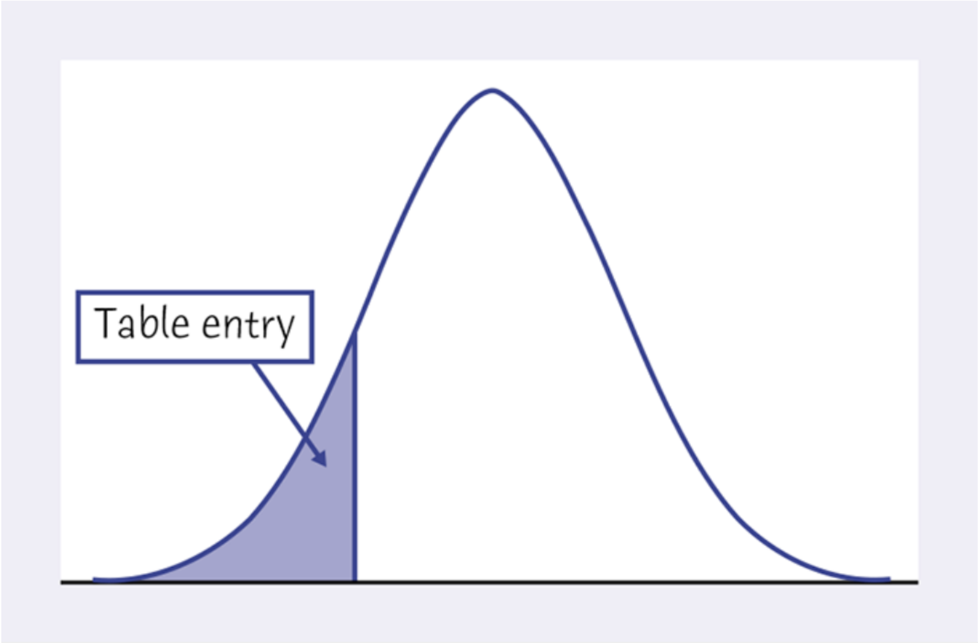





Table entry for *z* is the area under the standard Normal curve to the left of *z*.

**Table A Standard Normal cumulative proportions**

| z | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|---|---|---|---|---|---|---|---|---|---|---|
| −3.4 | .0003 | .0003 | .0003 | .0003 | .0003 | .0003 | .0003 | .0003 | .0003 | .0002 |
| −3.3 | .0005 | .0005 | .0005 | .0004 | .0004 | .0004 | .0004 | .0004 | .0004 | .0003 |

# Cumulative Proportions—Example 3.5 (5 of 6)

- What percent of high school seniors meet this SAT requirement of a combined score of 980 or better? Now using standard Normal table – **TABLE A at the end of the textbook**

**Table A Standard Normal cumulative proportions**

| z | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|---|---|---|---|---|---|---|---|---|---|---|
| −3.4 | .0003 | .0003 | .0003 | .0003 | .0003 | .0003 | .0003 | .0003 | .0003 | .0002 |
| −3.3 | .0005 | .0005 | .0005 | .0004 | .0004 | .0004 | .0004 | .0004 | .0004 | .0003 |
| −3.2 | .0007 | .0007 | .0006 | .0006 | .0006 | .0006 | .0006 | .0005 | .0005 | .0005 |
| −3.1 | .0010 | .0009 | .0009 | .0009 | .0008 | .0008 | .0008 | .0008 | .0007 | .0007 |
| −3.0 | .0013 | .0013 | .0013 | .0012 | .0012 | .0011 | .0011 | .0011 | .0010 | .0010 |
| −2.9 | .0019 | .0018 | .0018 | .0017 | .0016 | .0016 | .0015 | .0015 | .0014 | .0014 |
| −2.8 | .0026 | .0025 | .0024 | .0023 | .0023 | .0022 | .0021 | .0021 | .0020 | .0019 |
| −2.7 | .0035 | .0034 | .0033 | .0032 | .0031 | .0030 | .0029 | .0028 | .0027 | .0026 |
| −2.6 | .0047 | .0045 | .0044 | .0043 | .0041 | .0040 | .0039 | .0038 | .0037 | .0036 |
| −2.5 | .0062 | .0060 | .0059 | .0057 | .0055 | .0054 | .0052 | .0051 | .0049 | .0048 |
| −2.4 | .0082 | .0080 | .0078 | .0075 | .0073 | .0071 | .0069 | .0068 | .0066 | .0064 |
| −2.3 | .0107 | .0104 | .0102 | .0099 | .0096 | .0094 | .0091 | .0089 | .0087 | .0084 |
| −2.2 | .0139 | .0136 | .0132 | .0129 | .0125 | .0122 | .0119 | .0116 | .0113 | .0110 |
| −2.1 | .0179 | .0174 | .0170 | .0166 | .0162 | .0158 | .0154 | .0150 | .0146 | .0143 |

*Note: Table goes all the way from -3.4 to 3.4*

# Cumulative Proportions—Example 3.5 (6 of 6)

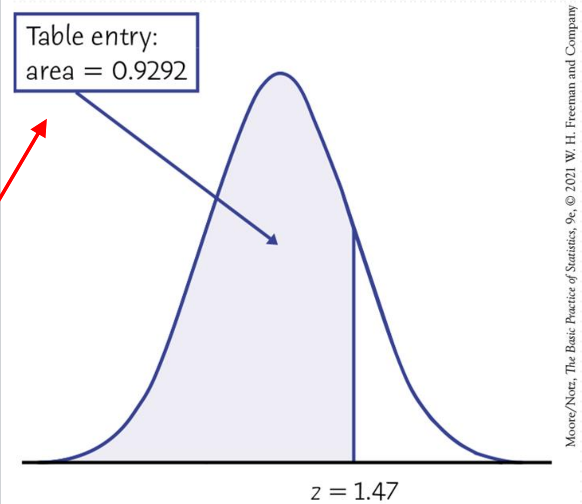
- What percent of high school seniors meet this SAT requirement of a combined score of 980 or better? Now using standard Normal table – **TABLE A in the end of the textbook**

| z | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|---|---|---|---|---|---|---|---|---|---|---|
| −0.4 | .3446 | .3409 | .3372 | .3336 | .3300 | .3264 | .3228 | .3192 | .3156 | .3121 |
| −0.3 | .3821 | .3783 | .3745 | .3707 | .3669 | .3632 | .3594 | .3557 | .3520 | .3483 |
| −0.2 | .4207 | .4168 | .4129 | .4090 | .4052 | .4013 | .3974 | .3936 | .3897 | .3859 |
| −0.1 | .4602 | .4562 | .4522 | .4483 | .4443 | .4404 | .4364 | .4325 | .4286 | .4247 |
| −0.0 | .5000 | .4960 | .4920 | .4880 | .4840 | .4801 | .4761 | .4721 | .4681 | .4641 |

- Recall that our z = (980-1059)/210 = 0.376 → approx. 0.38
- Proportion below z = 0.3520
- Proportion above z = 1 - 0.3520 = 0.648 → 0.65 or 65%

# Table A continues with positive z values

| z | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|---|---|---|---|---|---|---|---|---|---|---|
| **0.0** | .5000 | .5040 | .5080 | .5120 | .5160 | .5199 | .5239 | .5279 | .5319 | .5359 |
| **0.1** | .5398 | .5438 | .5478 | .5517 | .5557 | .5596 | .5636 | .5675 | .5714 | .5753 |
| **0.2** | .5793 | .5832 | .5871 | .5910 | .5948 | .5987 | .6026 | .6064 | .6103 | .6141 |
| **0.3** | .6179 | .6217 | .6255 | .6293 | .6331 | .6368 | .6406 | .6443 | .6480 | .6517 |
| **0.4** | .6554 | .6591 | .6628 | .6664 | .6700 | .6736 | .6772 | .6808 | .6844 | .6879 |
| **0.5** | .6915 | .6950 | .6985 | .7019 | .7054 | .7088 | .7123 | .7157 | .7190 | .7224 |
| **0.6** | .7257 | .7291 | .7324 | .7357 | .7389 | .7422 | .7454 | .7486 | .7517 | .7549 |
| **0.7** | .7580 | .7611 | .7642 | .7673 | .7704 | .7734 | .7764 | .7794 | .7823 | .7852 |
| **0.8** | .7881 | .7910 | .7939 | .7967 | .7995 | .8023 | .8051 | .8078 | .8106 | .8133 |
| **0.9** | .8159 | .8186 | .8212 | .8238 | .8264 | .8289 | .8315 | .8340 | .8365 | .8389 |
| **1.0** | .8413 | .8438 | .8461 | .8485 | .8508 | .8531 | .8554 | .8577 | .8599 | .8621 |
| **1.1** | .8643 | .8665 | .8686 | .8708 | .8729 | .8749 | .8770 | .8790 | .8810 | .8830 |
| **1.2** | .8849 | .8869 | .8888 | .8907 | .8925 | .8944 | .8962 | .8980 | .8997 | .9015 |
| **1.3** | .9032 | .9049 | .9066 | .9082 | .9099 | .9115 | .9131 | .9147 | .9162 | .9177 |
| **1.4** | .9192 | .9207 | .9222 | .9236 | .9251 | .9265 | .9279 | $\boxed{.9292}$ | .9306 | .9319 |



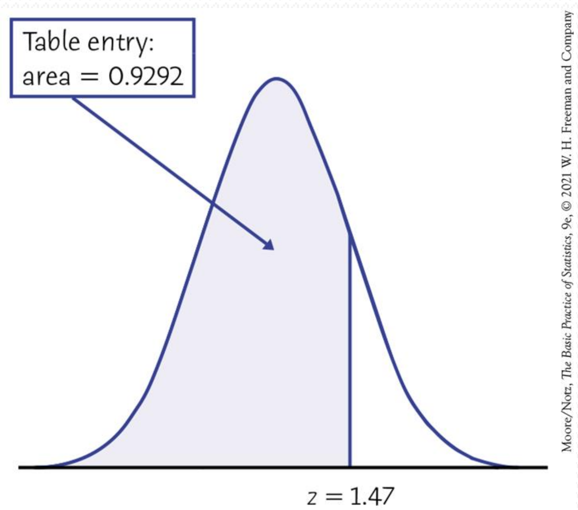
# The Standard Normal Table

Because all Normal distributions are the same once we standardize, we can find areas under any Normal curve from a single table: the standard normal table. Provided as **Table A in the textbook**, this is a table of areas under the standard Normal curve. The table entry for each value $z$ is the area under the curve to the left of $z$.

Suppose we want to find the proportion of observations from the standard Normal distribution that are less than 1.47. We can use Table A:

| **Z** | **.06** | **.07** | **.08** |
|---|---|---|---|
| **1.3** | .9131 | .9[illegible]7 | .9162 |
| **1.4** | .9279 | .9292 | .9306 |
| **1.5** | .9406 | .9418 | .9429 |

***P*(*z* < 1.47) = 0.9292**



# Normal Calculations (1 of 2)

**USING TABLE A TO FIND NORMAL PROPORTIONS**

**STEP 1: State the problem** in terms of the observed variable $x$. **Draw a picture** that shows the proportion you want in terms of cumulative proportions.

**STEP 2 : Standardize $x$** to restate the problem in terms of a standard Normal variable $z$.

**STEP 3 : Use Table A** and the fact that the total area under the curve is 1 to find the required area under the standard Normal curve.

# Normal Calculations (2 of 2)

- Repeating the SAT example, with approximately Normal distribution, with mean 1059 and standard deviation 210, use standardization to find what percent of high school seniors meet this SAT requirement of a combined score of 980 or better?
- **Standardize**:

$$\begin{aligned} x &\geq 980 \\ \frac{x-1059}{210} &\geq \frac{980-1059}{210} \\ z &\geq -0.38 \end{aligned}$$

- **Use the table**: the picture (from the first example where we solved this) said we needed the cumulative probability for $x = 980$. The previous step here showed this is the same as that of $z = -0.38$. So the area to the right of $-0.38$ or $1 - 0.3520 \sim 65\%$ qualified for college sports.

# Finding a Value Given a Proportion Example 3.9 (1 of 5)

- SAT reading scores for a recent year are distributed according to an $N(531, 104)$ distribution.
- How high must a student score in order to be in the top 10% of the distribution?
- We'll look at the same two approaches as we did for finding Normal proportions—using software (Python), and then using Table A.

# Finding a Value Given a Proportion Example 3.9 (2 of 5)

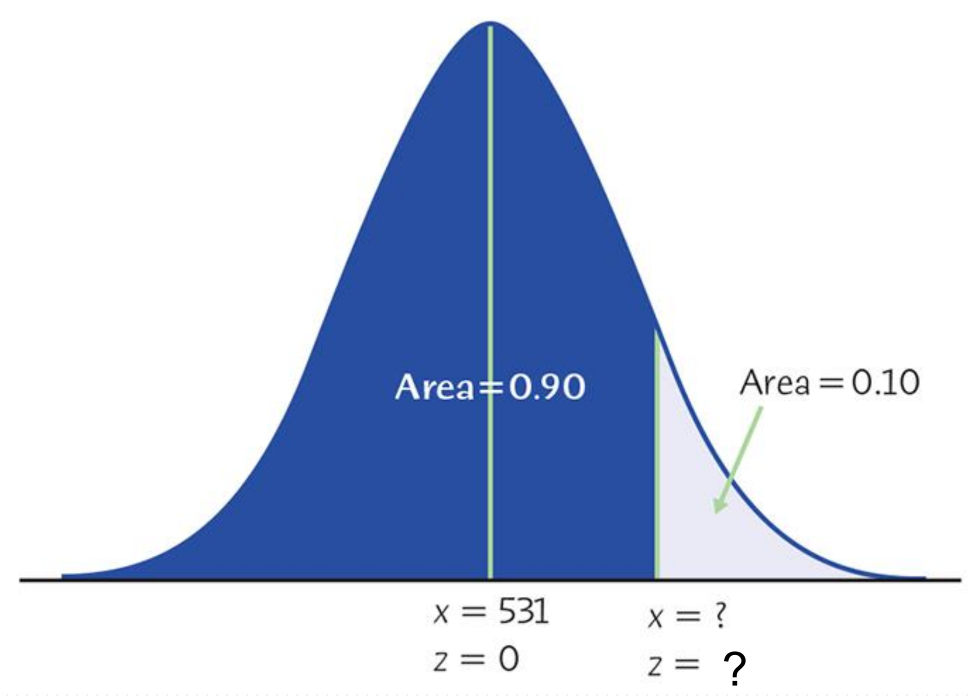
- Example: using Python
- How high must a student score in order to be in the top 10% of the distribution?
- We want to find the SAT score $x$ with area 0.1 to its right under the Normal curve with mean $\mu = 531$ and standard deviation $\sigma = 104$. That's the same as finding the SAT score $x$ with area 0.9 to its left.

```
In [1]: from scipy.stats import norm

In [2]: # this gives you the value x corresponding to a proportion of 0.9 to its left (lower tail)
        # which also corresponds to a proportion of 0.1 to its right
        norm.ppf(0.9,loc=531,scale=104)

Out[2]: 664.2813628166384
```

A student score must be higher than **664.28** to be in the top 10%

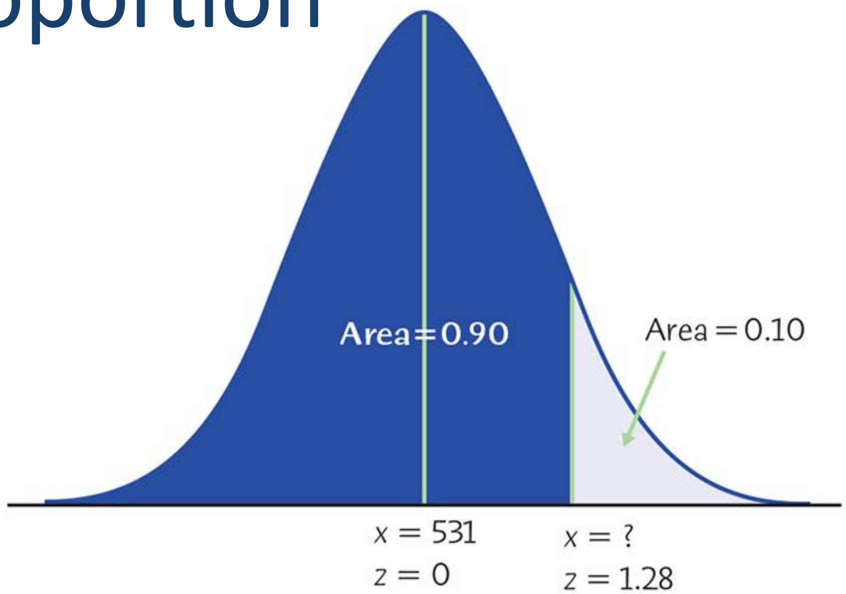
# Finding a Value Given a Proportion Example 3.9 (3 of 5)

- SAT reading scores for a recent year are distributed according to an *N*(531, 104) distribution.
- How high must a student score in order to be in the top 10% of the distribution?
- In order to use Table A, equivalently, what z-score has the cumulative proportion 0.90 below it?



# Finding a Value Given a Proportion Example 3.9 (4 of 5)

How high must a student score in order to be in the top 10% of the distribution?






- Look up the closest probability (closest to 0.90) in the table.
- Find the corresponding **standardized score**.
- The value you seek is ***that many standard deviations from the mean***.

| *z* | .07 | .08 | .09 |
|---|---|---|---|
| 1.1 | .8790 | .8[illegible]10 | .8830 |
| 1.2 | | .8997 | .8015 |
| 1.3 | .8147 | .8162 | .8177 |

$$z = 1.28$$

# Finding a Value Given a Proportion Example 3.9 (5 of 5)

How high must a student score in order to be in the top 10% of the distribution?

$$z = 1.28$$

We need to "unstandardize" the *z*-score to find the observed value (*x*):

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

$$x = 531 + z \cdot 104$$

$$x = 531 + [(1.28) \cdot 104]$$

$$x = 531 + (133.12) = 664.12$$

A student would have to score at least 664.12 to be in the top 10% of the distribution of SAT reading scores for this particular year.

# “Backward” Normal Calculations

USING **TABLE A** GIVEN A NORMAL PROPORTION

**STEP 1:** **State the problem** in terms of the given proportion. **Draw a picture** that shows the Normal value, $x$, that you want in relation to the cumulative proportion.

**STEP 2:** **Use Table A**, the fact that the total area under the curve is 1, and the given area under the standard Normal curve to find the corresponding $z$-value.

**STEP 3:** **Unstandardize** $z$ to solve the problem in terms of a non-standard Normal variable $x$.