CHAPTER 3: The Normal Distributions

Basic Practice of Statistics
7th Edition

Lecture PowerPoint 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 5)

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

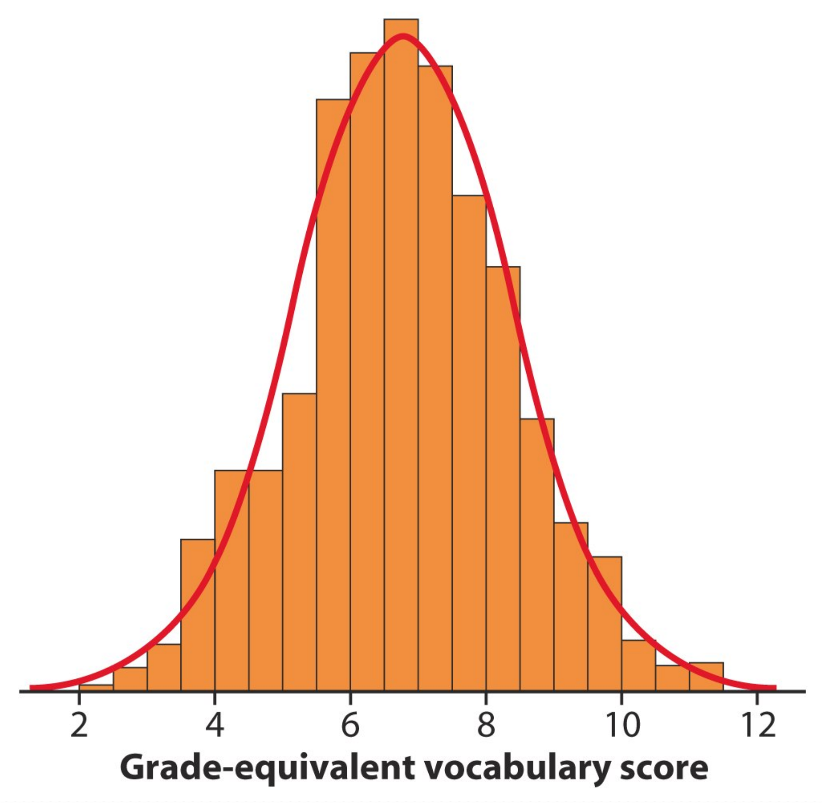
1. Always plot your data: Make a graph, usually a histogram or a stemplot.
2. Look for the overall pattern (shape, center, and variability) and for striking deviations, such as outliers.
3. Calculate a numerical summary to briefly describe center and variability.

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 5)

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



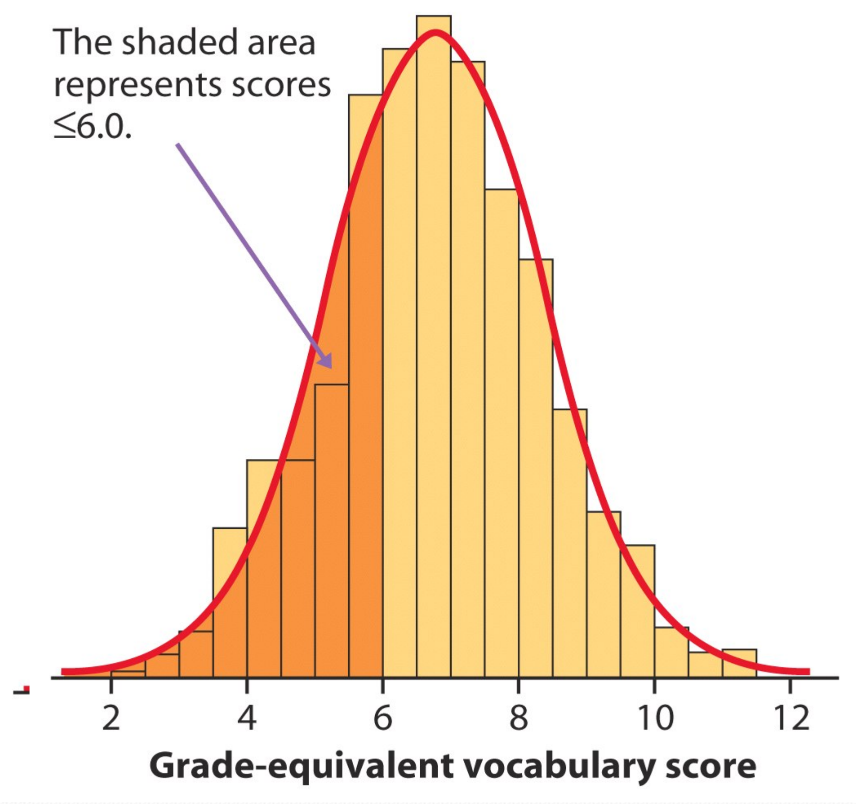
The smooth curve drawn over the histogram is a *mathematical "idialization"* for the distribution.

It is what the histogram "looks" like when we have LOTS of data.

Density Curves (3 of 5)

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.
This is what the proportion on the previous slide would equal to if we had LOTS of data.

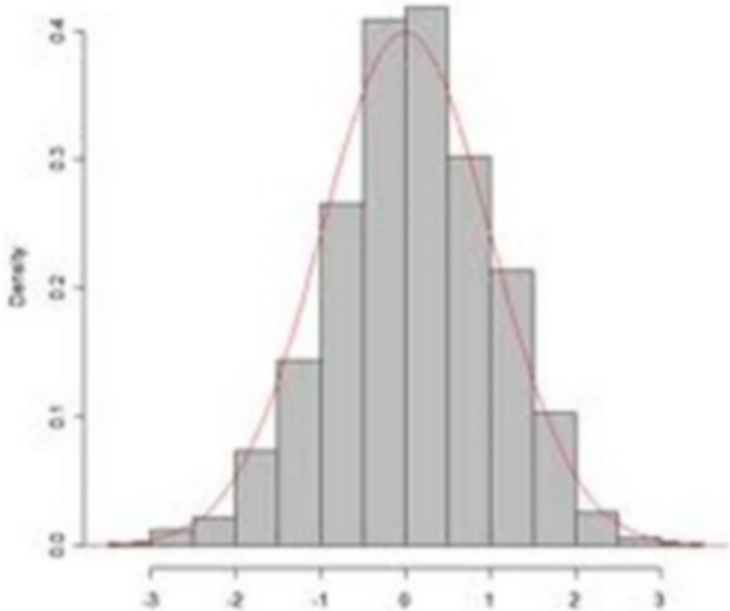
Now the area under the smooth curve to the left of 6.0 is shaded. If the scale is adjusted so that the total area under the curve is exactly 1, then this curve is called a **density curve.**

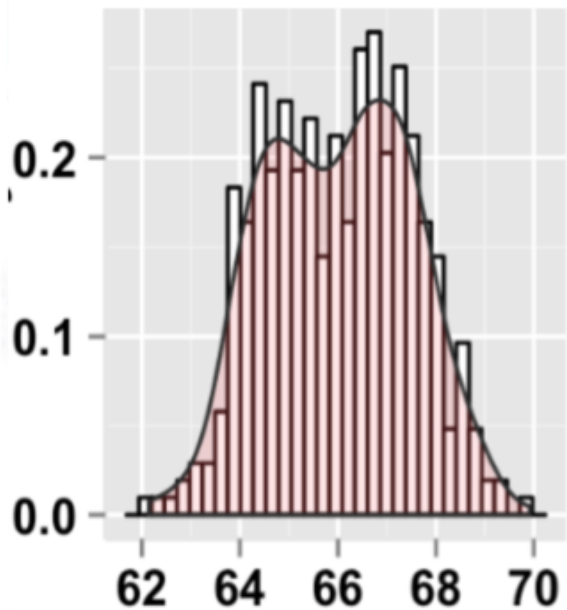


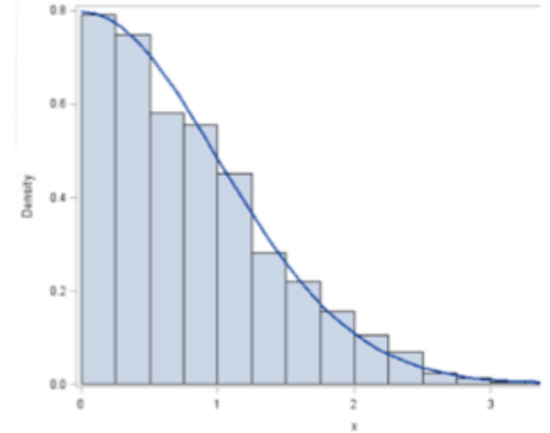
Density Curves (4 of 5)

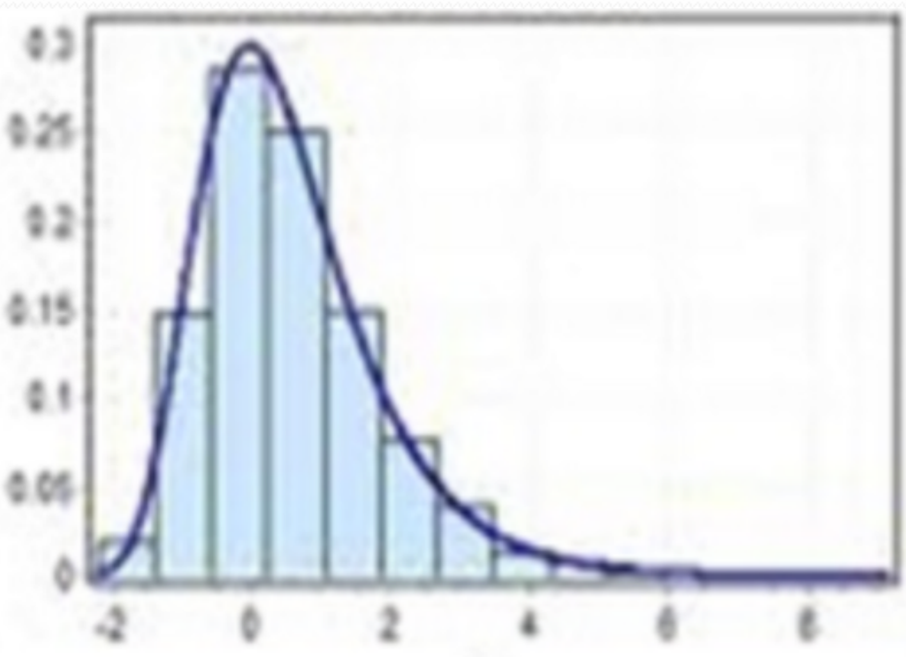
- A **density curve** is a curve that:
  - is always on or above the horizontal axis.
  - has an area of exactly 1 underneath it.
- A density curve describes the overall pattern of a distribution. The area under the curve and above any range of values on the horizontal axis is the proportion of all observations that fall in that range.

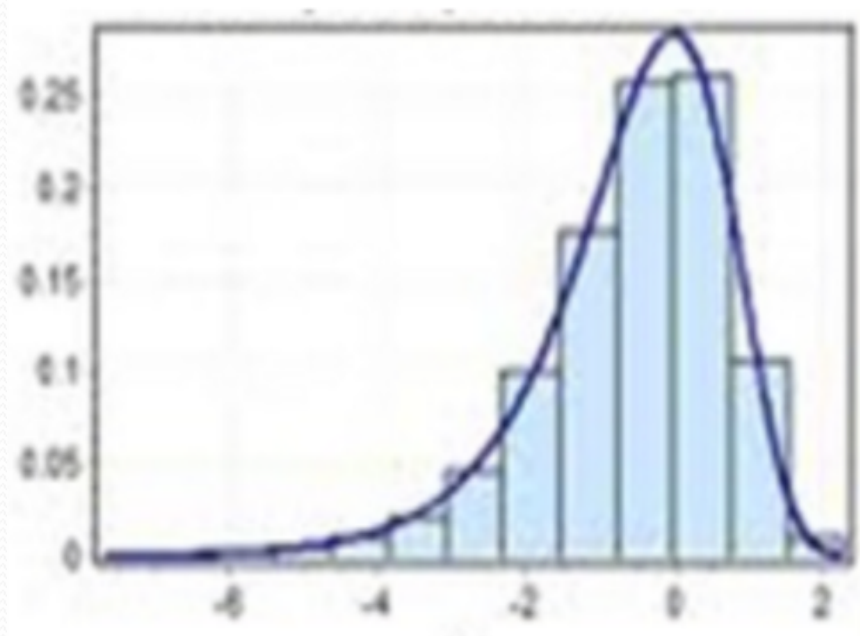
Caution: No set of real data is *exactly* described by a density curve!


Density
0.0
0.1
0.2
0.3
0.4
-3
-2
-1
0
1
2
3


0.0
0.1
0.2
62
64
66
68
70

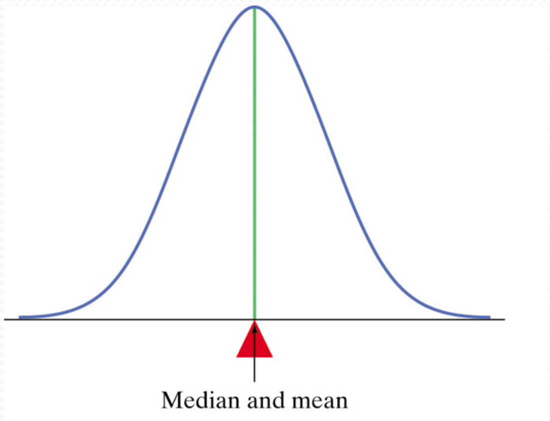

Density

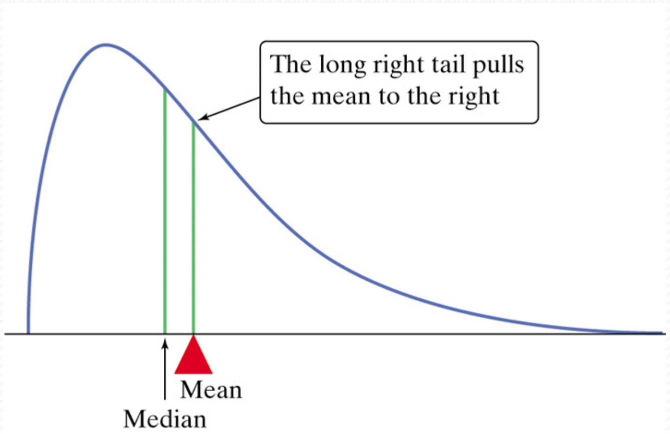

0
0.05
0.1
0.15
0.2
0.25
0.3
-2
0
2
4
6
8

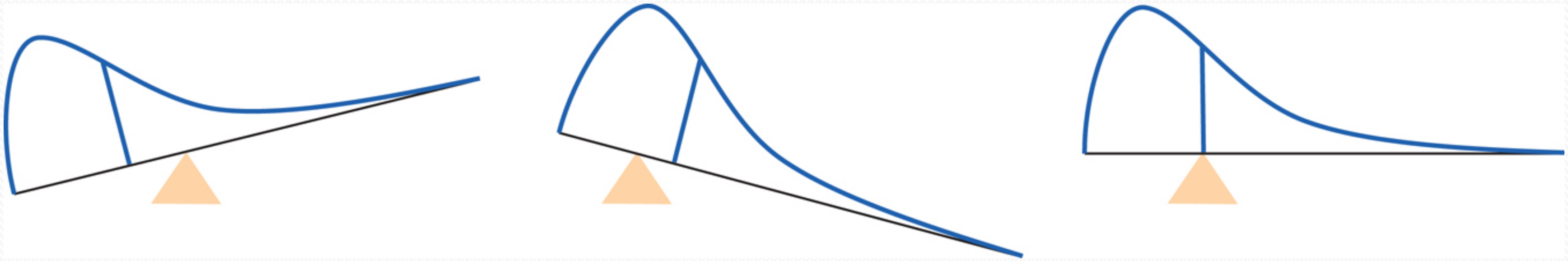

0
0.05
0.1
0.15
0.2
0.25
-6
-4
-2
0
2

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 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.







Density Curves

Which of the following is a false statement about density curves?

a) The median divides the area under the curve in half.
b) The mean is the balancing point of the density curve.
c) The mean and the median are always the same number.

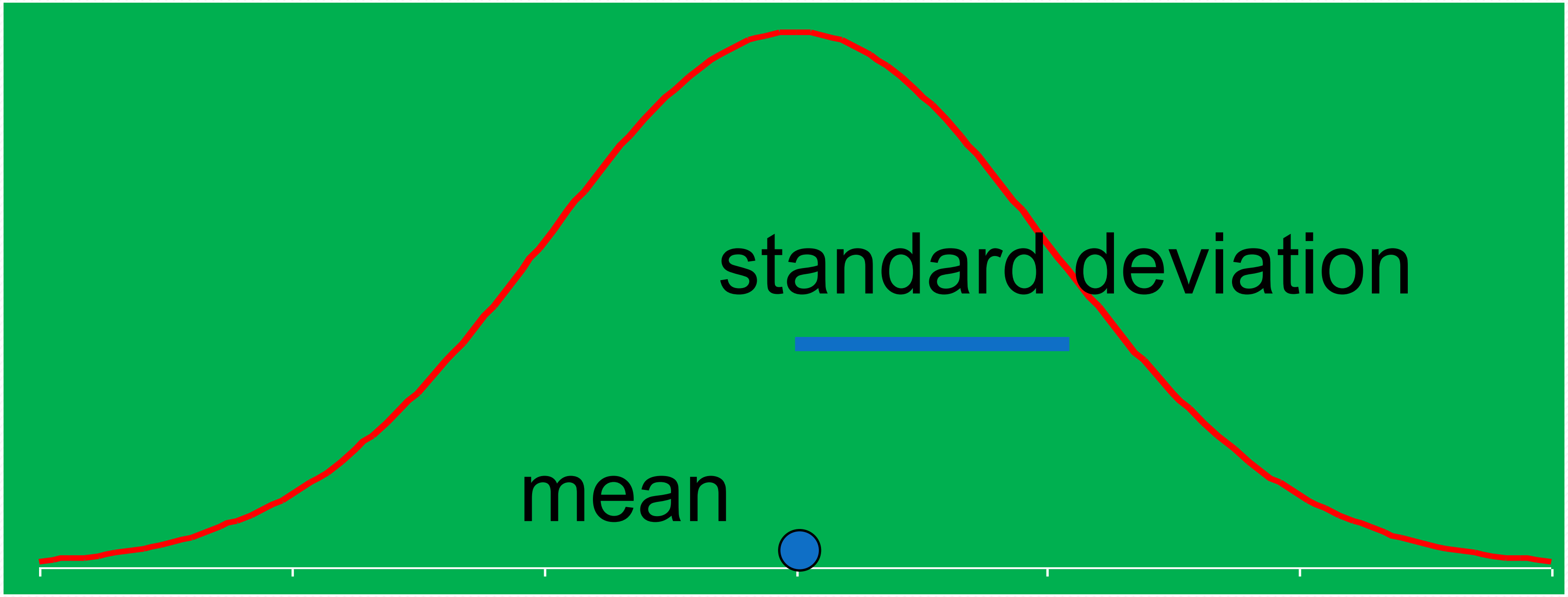
Density Curves (5 of 5)

- The mean and standard deviation computed from actual observations (data) are denoted by $\overline{x}$ and s, respectively.
- The mean and standard deviation of the actual distribution represented by the density curve are denoted by μ (“mu”) and σ (“sigma”), respectively.

Question

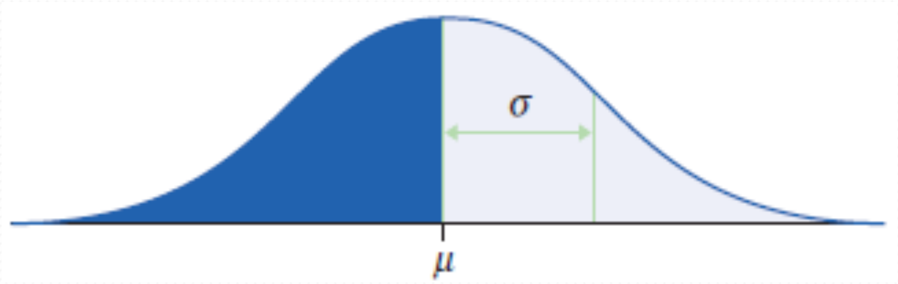
Data sets consisting of physical measurements (heights, weights, lengths of bones, and so on) for adults of the same species and sex tend to follow a similar pattern. The pattern is that most individuals are clumped around the average, with numbers decreasing the farther values are from the average in either direction. Describe what shape a histogram (or density curve) of such measurements would have.

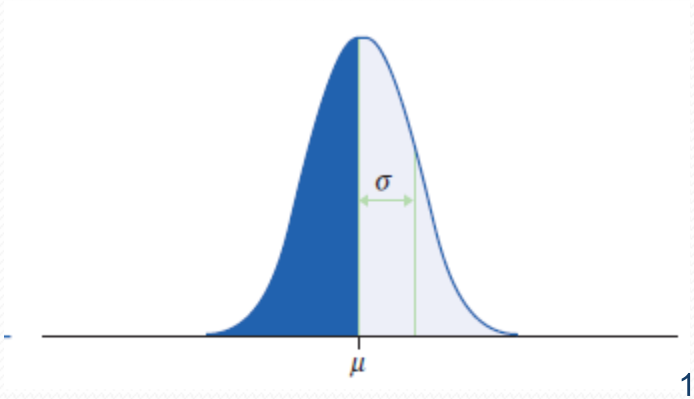
Bell-Shaped Curve: The Normal Distribution



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 described by giving its mean μ and standard deviation σ.





Normal Distributions (2 of 5)

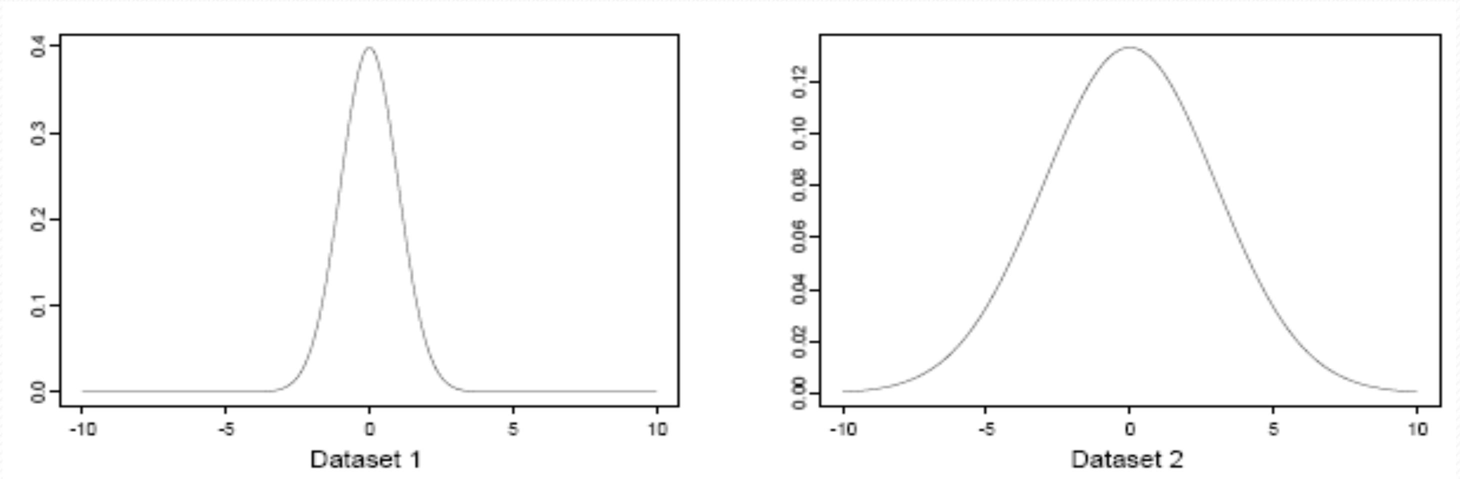
Unlike most distributions, any particular Normal distribution is completely specified by two numbers—its mean μ and standard deviation σ:

- The mean is located at the center of the symmetric curve and is the same as the median. Changing μ without changing σ moves the Normal curve along the horizontal axis without changing its variability.
- The standard deviation σ 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.

Density Curves

If you knew that $\mu = 0$ and $\sigma = 3$, which density curve applies?

Curve 1 Curve 2



a) Curve 1
b) Curve 2

Normal Distributions (3 of 5)

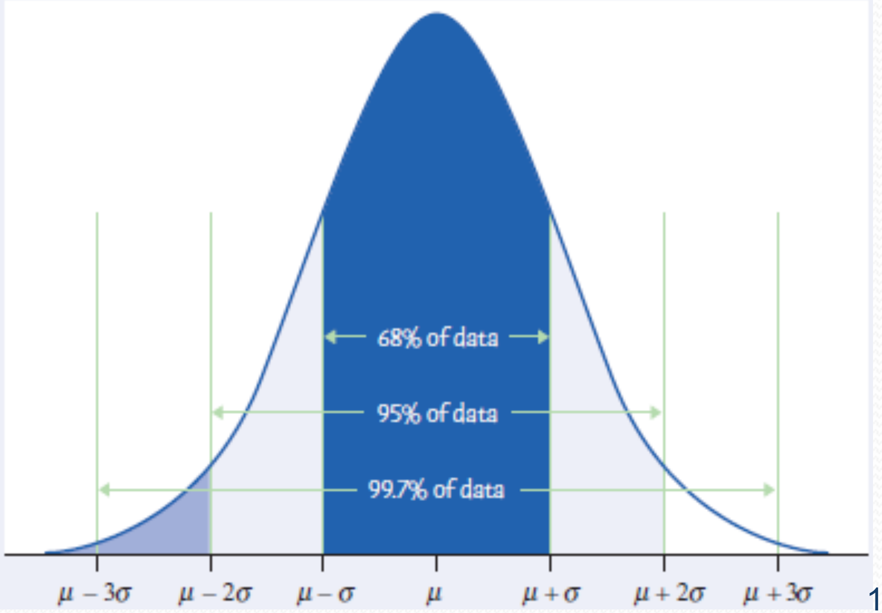
- A **Normal distribution** is described by a Normal density curve. Any particular Normal distribution is completely specified by two numbers: its mean μ and standard deviation σ.
- The mean of a Normal distribution is at the center of the symmetric Normal curve. The standard deviation is the distance from the center to the change-of-curvature points on either side.

Normal Distributions (4 of 5)

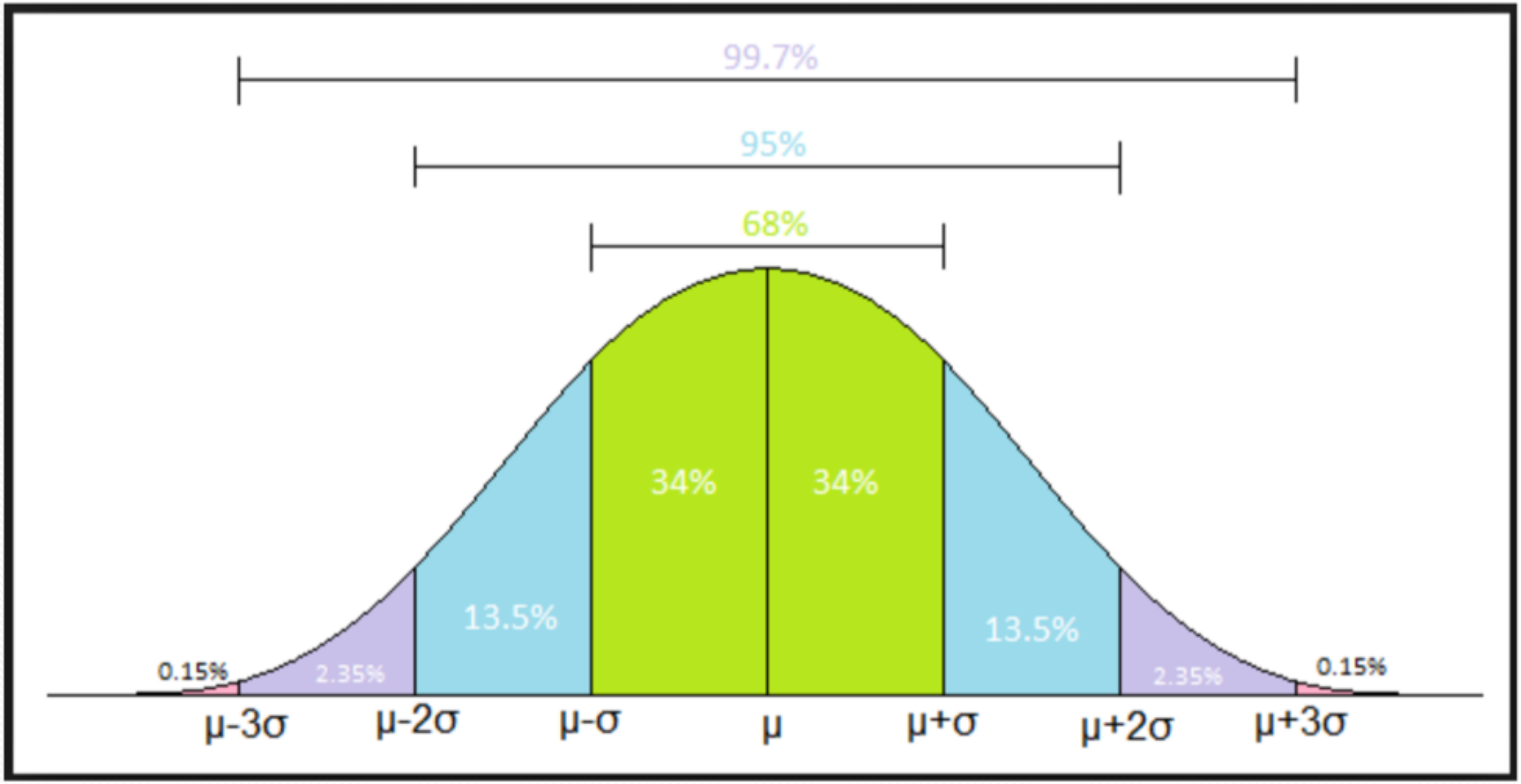
In the Normal distribution, with mean μ and standard deviation σ:

- approximately 68% of the observations fall within σ of μ.
- approximately 95% of the observations fall within 2σ of μ.
- approximately 99.7% of the observations fall within 3σ of μ.

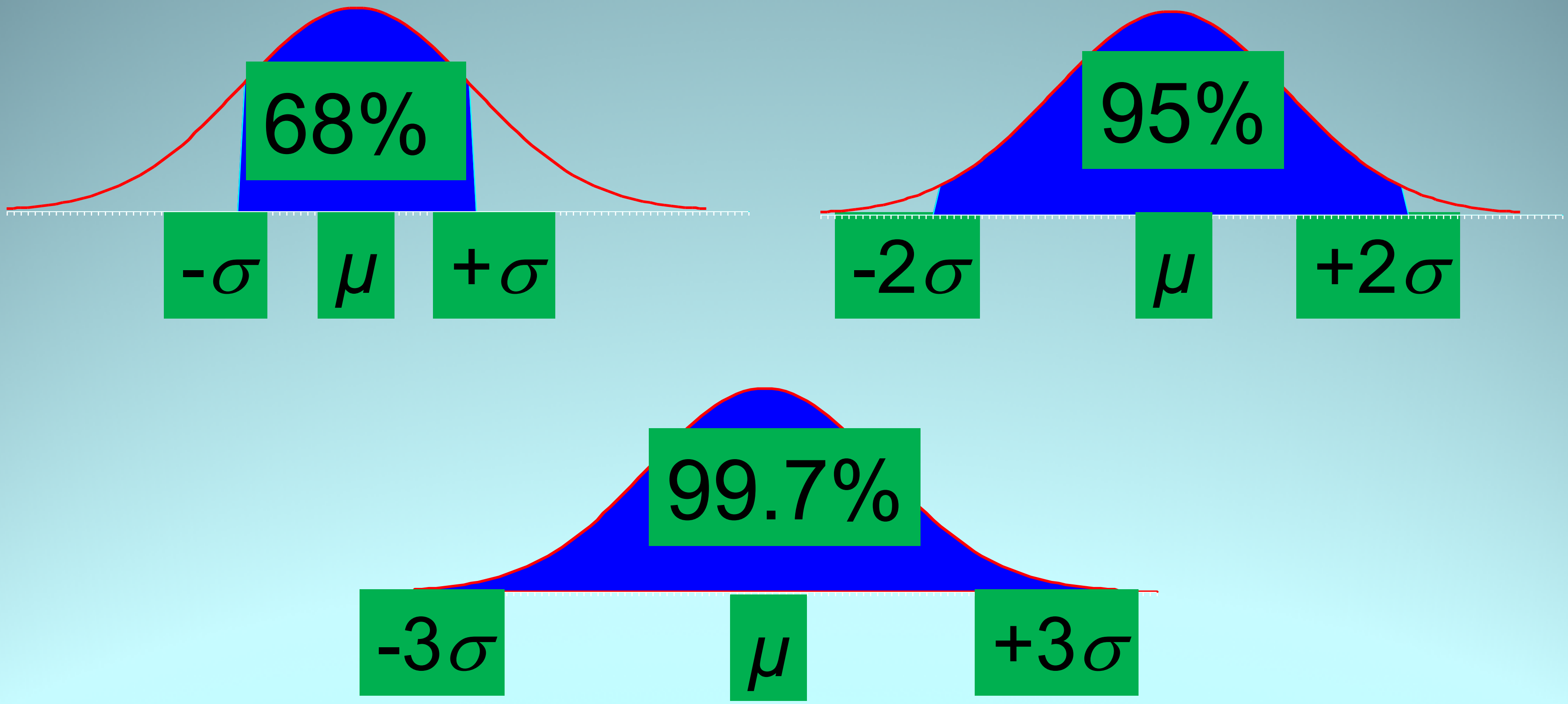
The 68–95–99.7 Rule



The 68–95–99.7 Rule



68-95-99.7 Rule for Any Normal Curve



Health and Nutrition Examination Study of 1976-1980

- Heights of adult men, aged 18-24
  - mean: 70.0 inches
  - standard deviation: 2.8 inches
  - heights follow a normal distribution, so we have that heights of men are N(70, 2.8).

Health and Nutrition Examination Study of 1976-1980

- 68-95-99.7 Rule for men's heights
  - ❖ 68% are between 67.2 and 72.8 inches

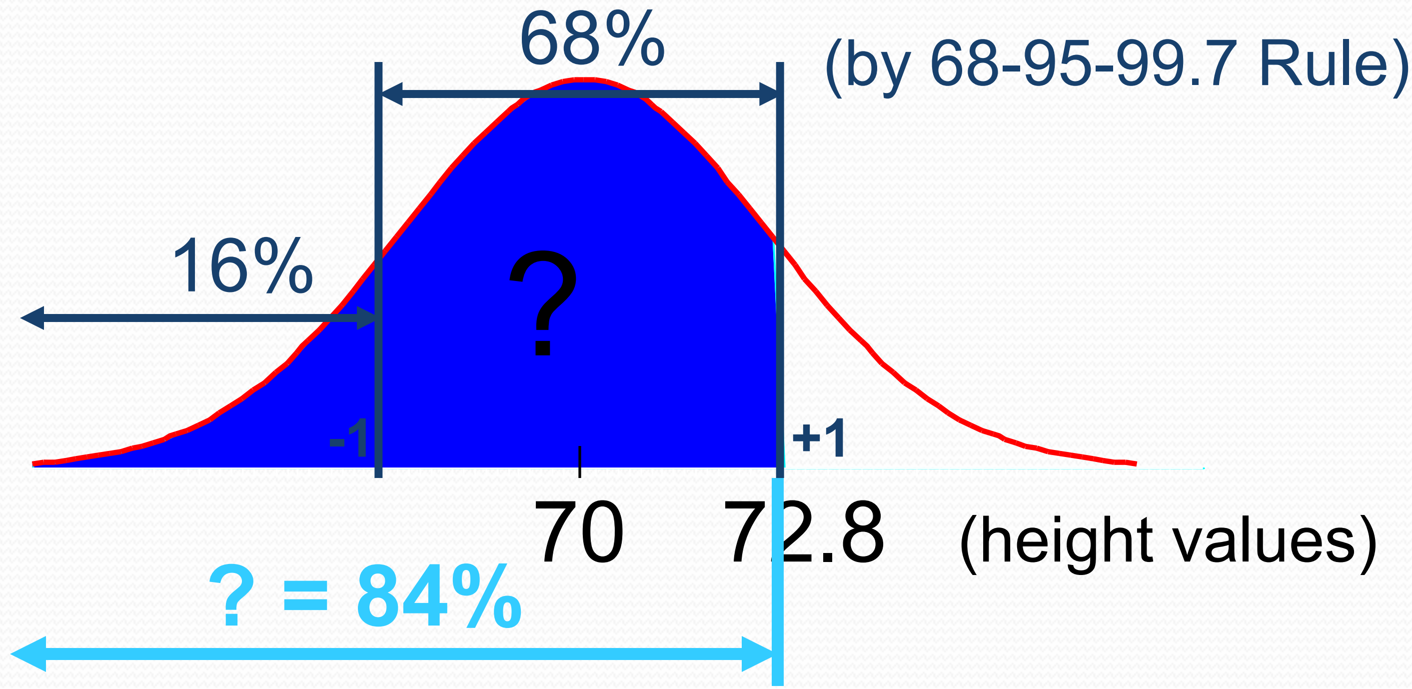
    $[\ \mu \pm \sigma = 70.0 \pm 2.8\ ]$
  - ❖ 95% are between 64.4 and 75.6 inches

    $[\ \mu \pm 2\sigma = 70.0 \pm 2(2.8) = 70.0 \pm 5.6\ ]$
  - ❖ 99.7% are between 61.6 and 78.4 inches

    $[\ \mu \pm 3\sigma = 70.0 \pm 3(2.8) = 70.0 \pm 8.4\ ]$

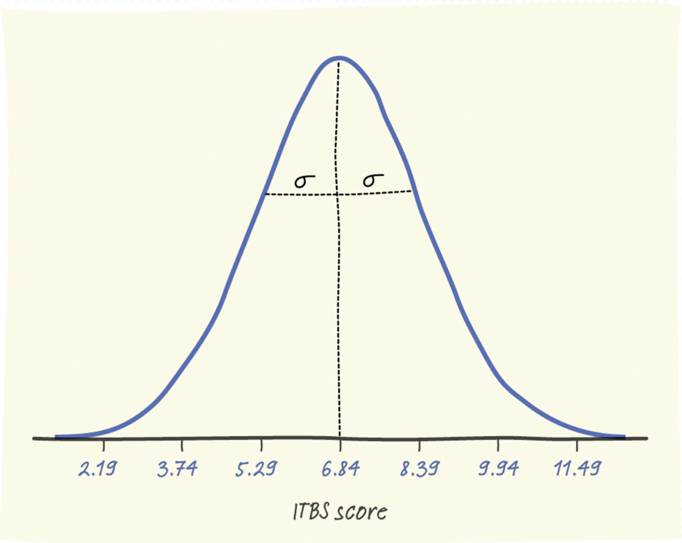
Health and Nutrition Examination Study of 1976-1980

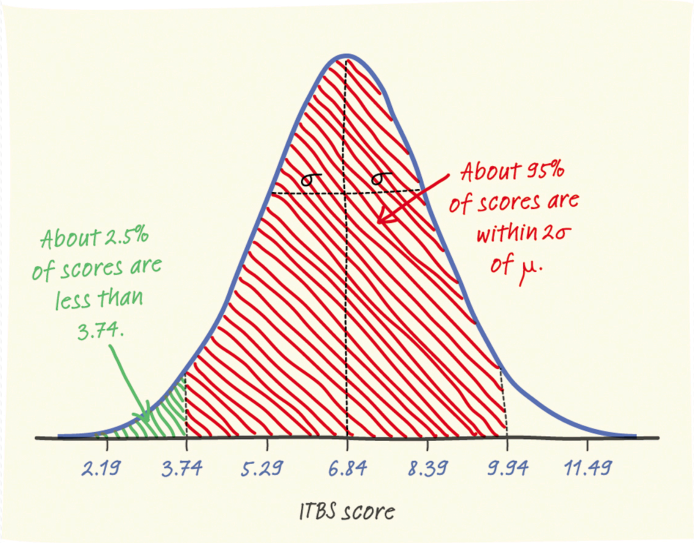
- What proportion of men are less than 72.8 inches tall?

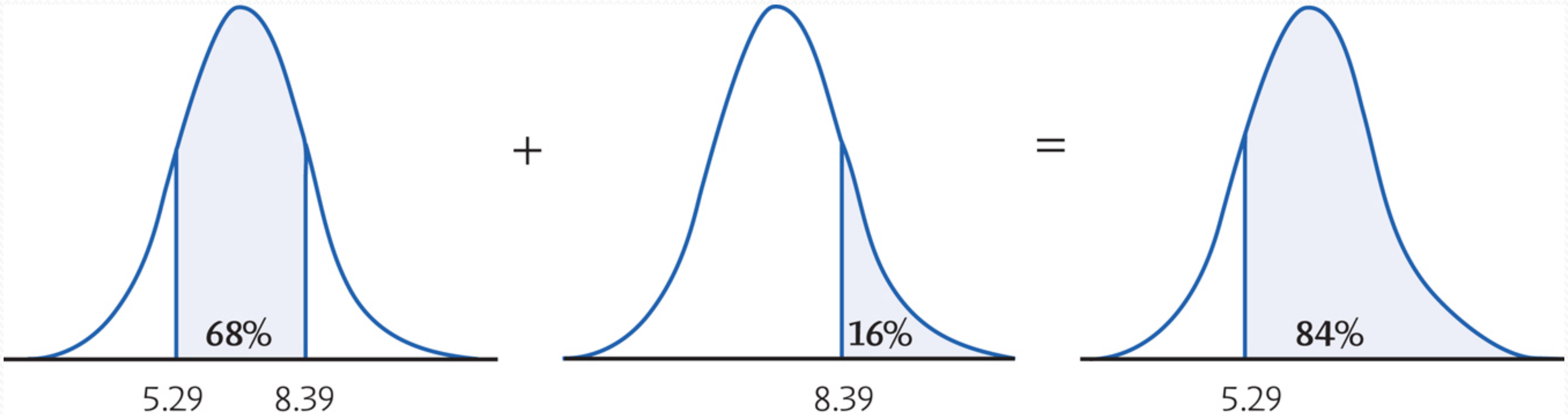


Example

- 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 is between 3.74 and 9.94?
- What percent of the scores is above 5.29?






68%
5.29
8.39
+
16%
8.39
=
84%
5.29

Standardizing (1 of 2)

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

STANDARDIZING AND *z*-SCORES

- If x is an observation from a distribution that has mean μ and standard deviation σ, 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 μ = 64.2 and σ = 2.8 inches. The standardized height is

$$z = \frac{\text{height} - 64.2}{2.8}$$

- 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

$$z = \frac{70 - 64.2}{2.8} = 2.07$$

  or 2.07 standard deviations above the mean.

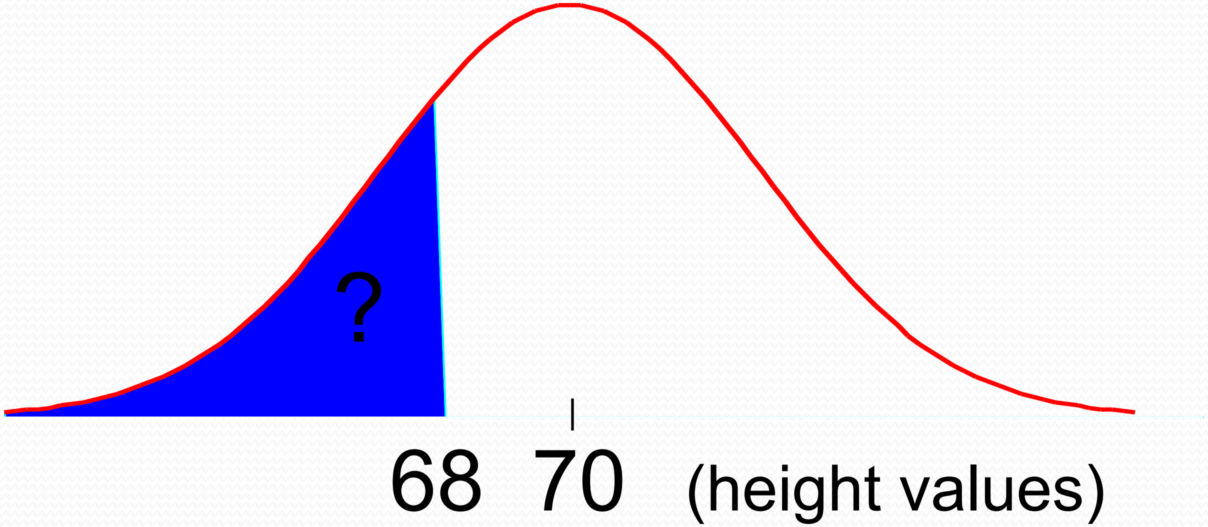
- Similarly, a woman 5 feet (60 inches) tall has standardized height

$$z = \frac{60 - 64.2}{2.8} = -1.50$$

  or 1.5 standard deviations *less than* the mean height.

Health and Nutrition Examination Study of 1976-1980

- What proportion of men are less than 68 inches tall?



How many standard deviations is 68 from 70?

Standardized Scores

- How many standard deviations is 68 from 70?
- ***standardized score*** =

  (observed value minus mean) / (std dev)

  [= (68 – 70) / 2.8 = −0.71]
- The value 68 is 0.71 standard deviations *below* the mean 70.

Standardized Scores

Jane is taking 1070-1. John is taking 1070-2.
Jane got 81 points. John got 76 points.
Question: Did Jane do slightly better?

Acount for difficulty: subtract class average.
Jane: 81-71=10; John: 76-56=20
Question: Did John do way better?

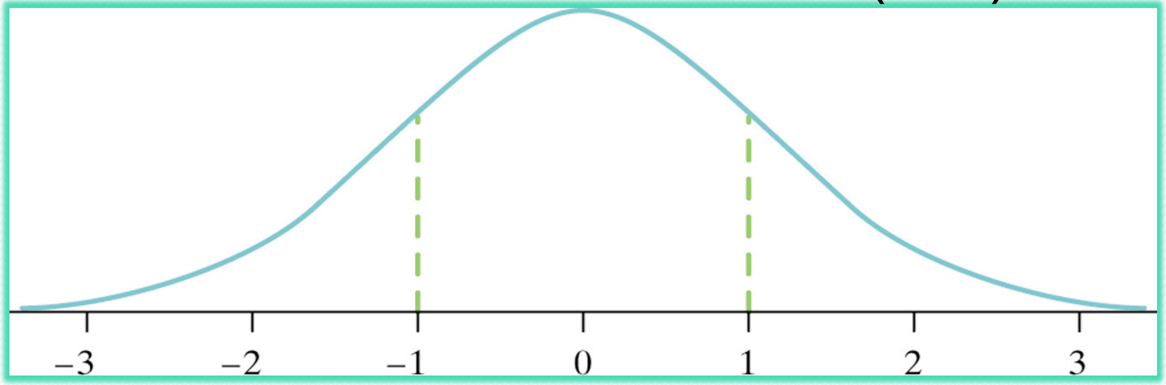
Acount for variability: divide by standard deviation.
Jane: (81-71)/2=5; John: (76-56)/10=2
Answer: Jane did way better!

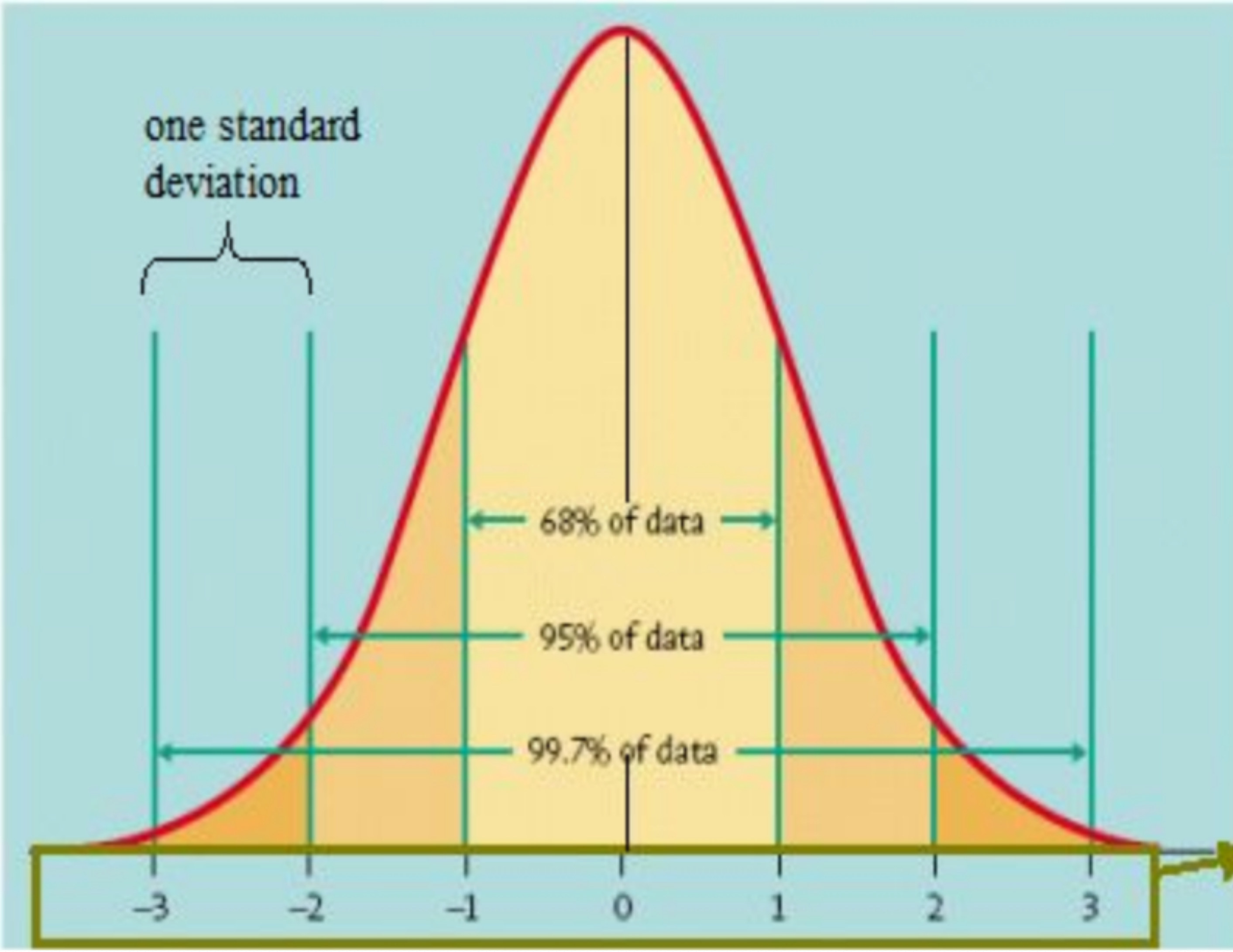
The Standard Normal Distribution

- The **standard Normal distribution** is the Normal distribution with mean 0 and standard deviation 1.
- Shown as ***N***(0,1)
- If a variable *x* has any Normal distribution *N*(μ, σ), with mean μ and standard deviation σ, then the standardized variable

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

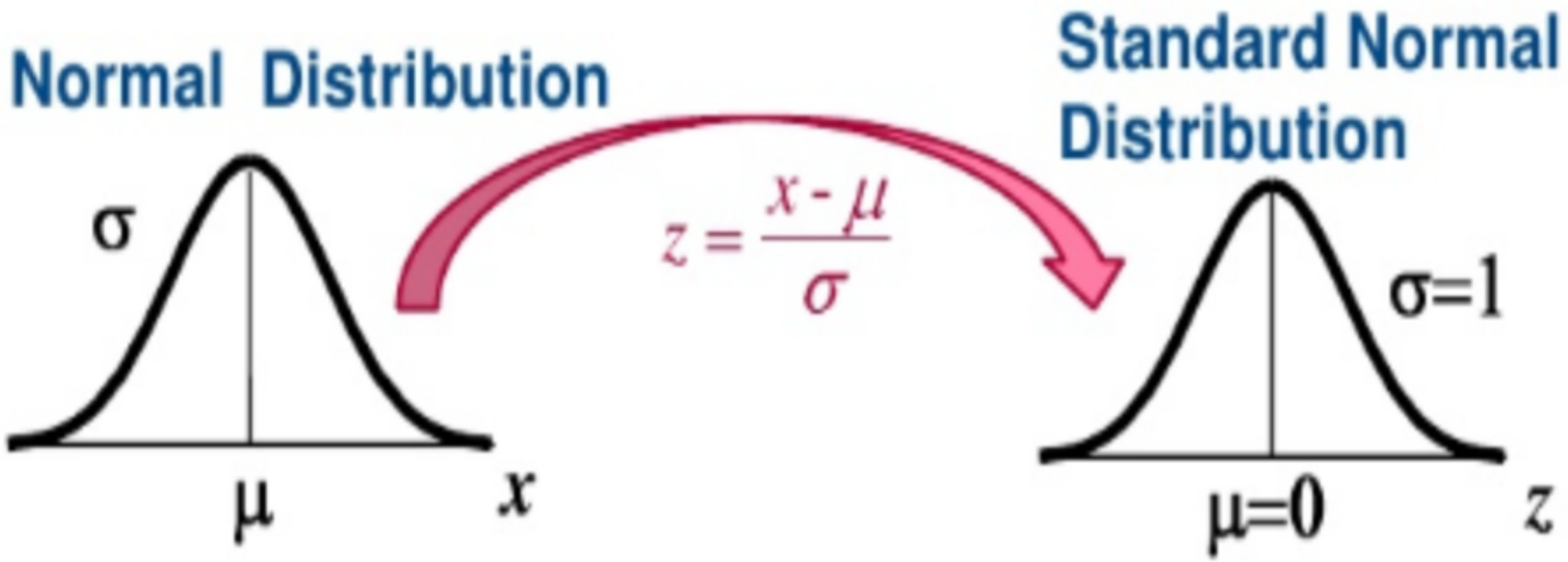
has the standard Normal distribution, *N*(0, 1).





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

Converting a normal distribution to The standard normal distribuion

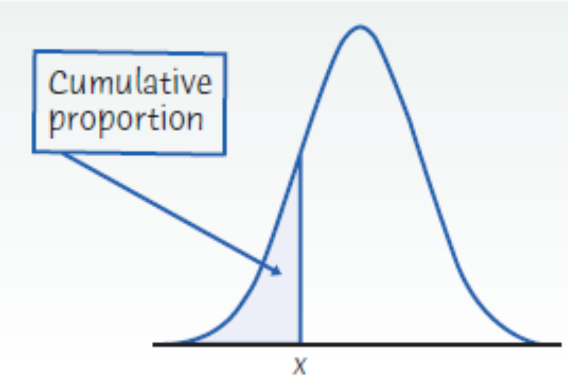


Exercise

IQ scores are approximately normal with a mean of 100 and a standard deviation of 16. Albert Einstein reportedly had an IQ of 160. Convert Einstein’s IQ score to a z score.

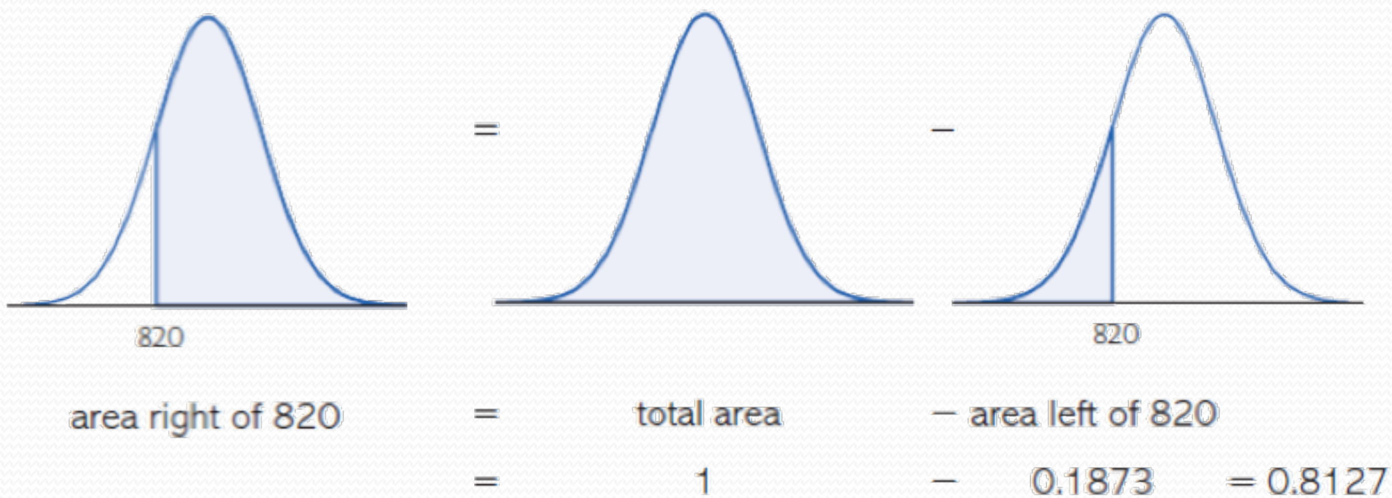
Cumulative Proportions

- Cumulative Proportion for x= Area to the left of x under the curve= Probability to the left of x
- A cumulative proportion for z corresponding x does **not change**.



Cumulative Proportions—Example

- Who qualifies for college sports?
- The combined scores of the almost 1.7 million high school seniors taking the SAT in 2013 were approximately Normal, with mean 1011 and standard deviation 216. What percent of high school seniors meet this SAT requirement of a combined score of 820 or better?
- Here is the calculation in a picture:



- So about 81% qualified for college sports.

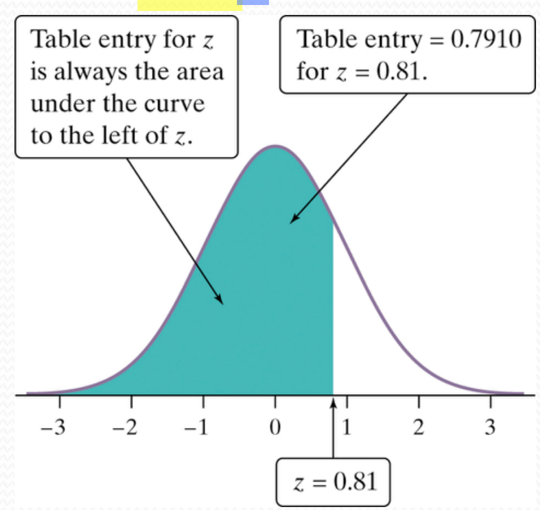
The Standard Normal Table

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 text, 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 0.81. We can use Table A:

Z	.00	.01	.02
0.7	.7580	[illegible]	.7642
0.8	.7881	.7910	.7939
0.9	.8159	.8186	.8212

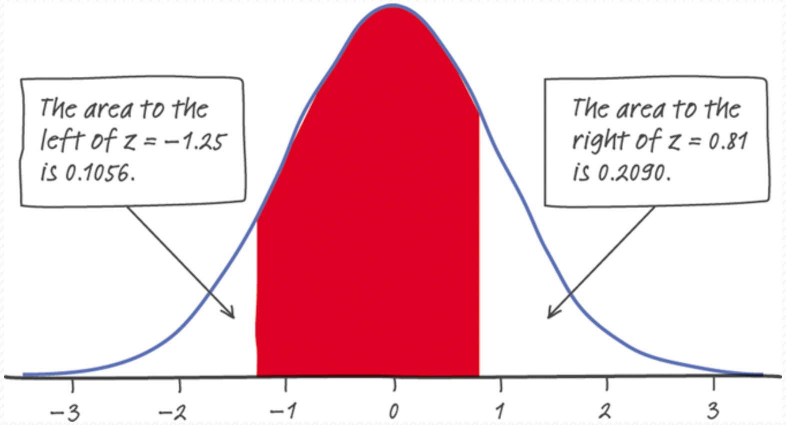
$P(z < 0.81) = 0.7910$



Normal Calculations (1 of 2)

Find the proportion of observations from the standard Normal distribution that are between –1.25 and 0.81.

Can you find the same proportion using a different approach?



$$1 - (0.1056 + 0.2090) = 1 - 0.3146$$

$$= \mathbf{0.6854}$$

Normal Calculations (2 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.

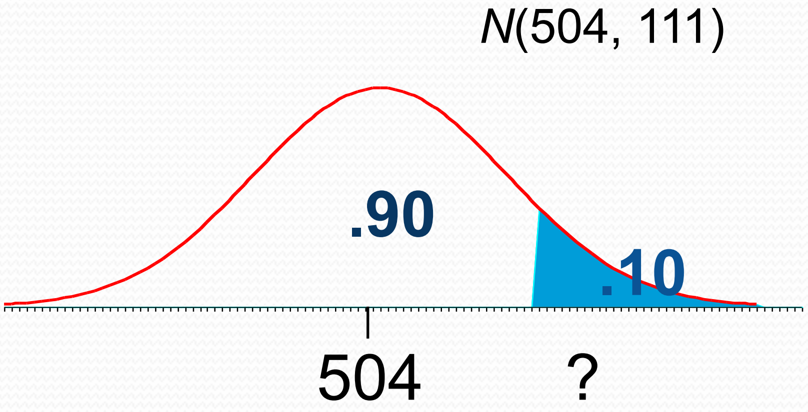
Problem

Scores of high school seniors on the SAT follow the Normal distribution with mean μ = 1026 and standard deviation σ = 209. What proportion of seniors score at least 820?

- **STEP1 :**Draw a **normal distribution** labeling mean and observation values and then **shade** area which the problem wants.
- **STEP2: Standardize** the observation values.
- **STEP3:**Use **z-Table**

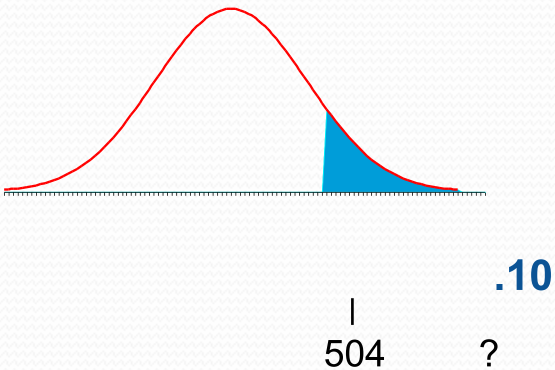
Normal Calculations (1 of 3)

- SAT reading scores for a recent year are distributed according to an $N(504, 111)$ 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 score has cumulative proportion 0.90 below it?



Normal Calculations (2 of 3)

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



- Look up the closest probability (closest to 0.10) in the table.
- Find the corresponding **standardized score**.
- The value you seek is ***that many standard deviations from the mean***.

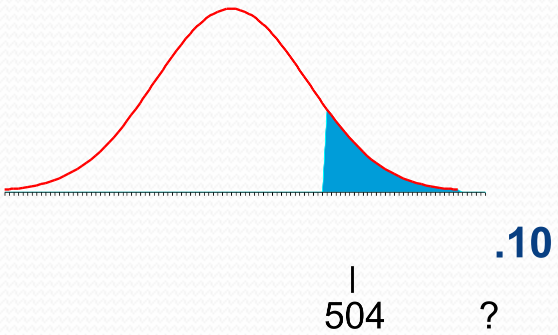
z	.07	.08	.09
1.1	.8790	.8[illegible]0	.8830
1.2	[illegible]	.8997	.8015
1.3	.8147	.8162	.8177

$$z = 1.28$$

Normal Calculations (3 of 3)

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*):

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

$$\begin{aligned} x &= 504 + z \cdot 111 \\ &= 504 + [(1.28) \cdot 111] \\ &= 504 + (142.08) = 646.08 \end{aligned}$$

A student would have to score at least 646.08 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.

Problem

High levels of cholesterol in the blood increase the risk of heart disease. For 14-year-old boys, the distribution of blood cholesterol is approximately Normal with mean μ = 170 milligrams of cholesterol per deciliter of blood (mg/dl) and standard deviation σ = 30 mg/dl.[8] What is the first quartile of the distribution of blood cholesterol?

- **STEP1:** Draw the **standard normal distribution** and then **shade** the area which the problem wants.
- **STEP2:**Use **z-Table** to find z-score corresponding given proportions.
- **STEP3:**Use $x = \mu + z\sigma$ to **convert** z-score to x.

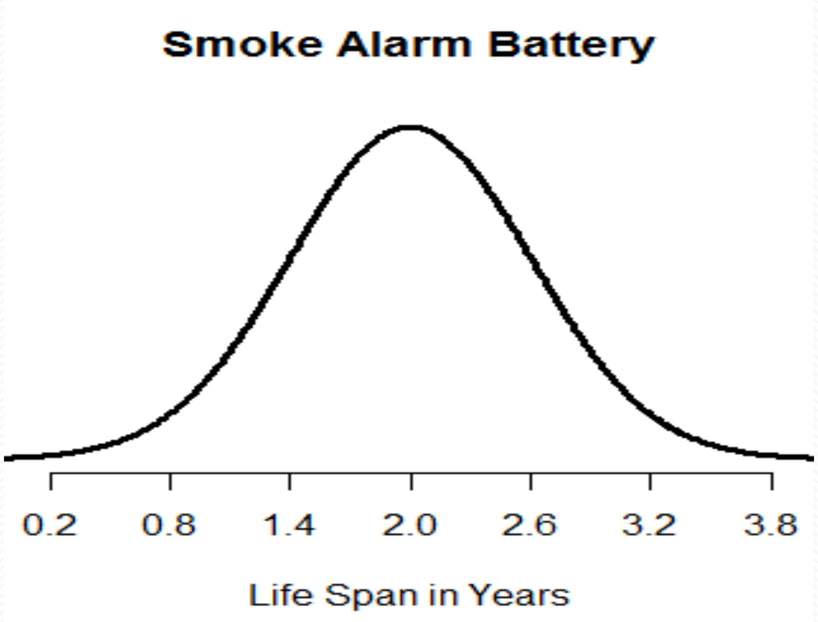
Finding Proportions (1 of 4)

What percent of values of a variable following the standard Normal distribution is between *z*-scores of 0 and 3?

a) ~33.2%
b) ~49.9%
c) ~90.7%
d) ~94.5%

Finding Proportions (2 of 4)

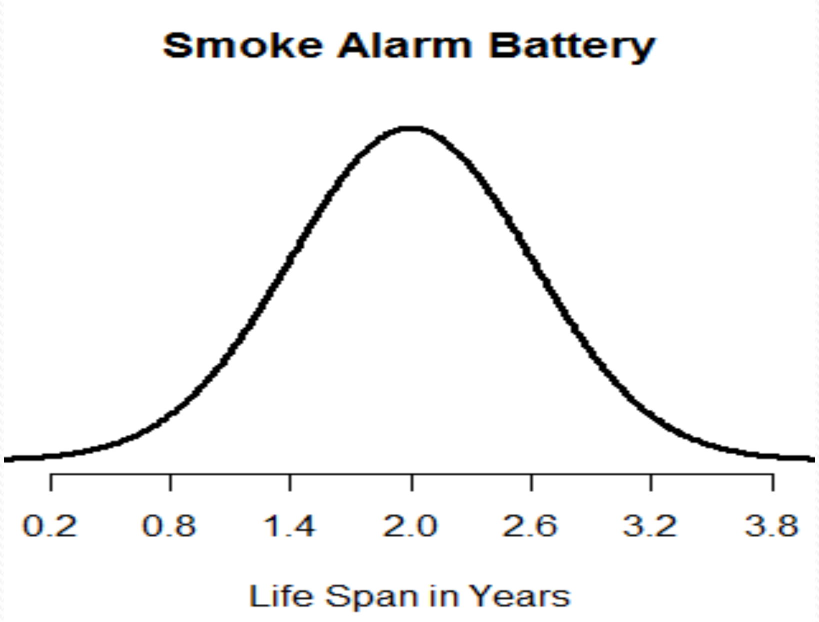
The battery in a certain smoke alarm has a life span that is Normally distributed, with a mean of 2 years and a standard deviation of 0.6 years. What proportion of smoke alarms will have a life span less than 1 year?



a) 0.167
b) -1.67
c) 0.953
d) 0.047

Finding Proportions (3 of 4)

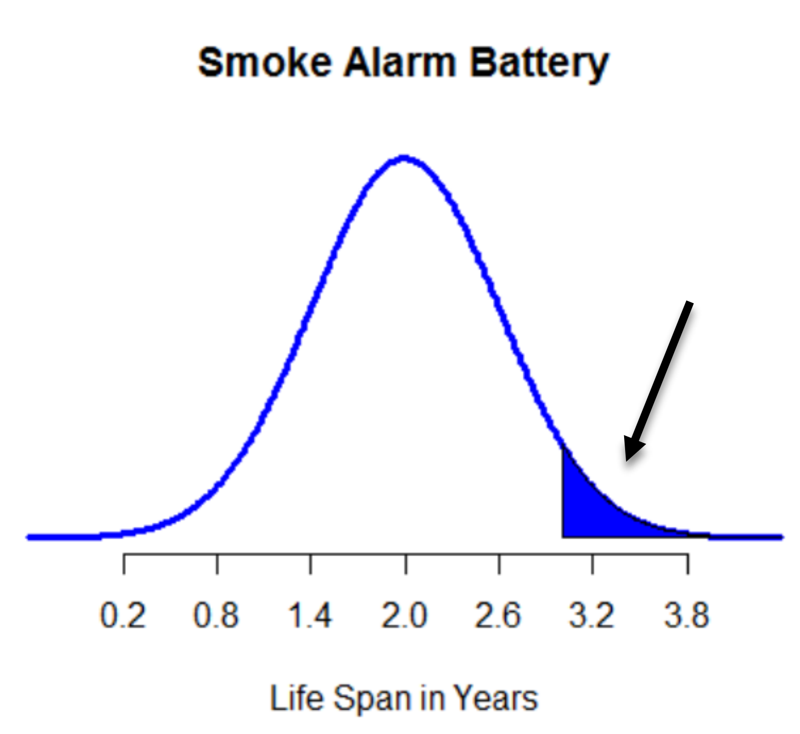
The battery in a certain smoke alarm has a life span that is Normally distributed, with a mean of 2 years and a standard deviation of 0.6 years. What proportion of smoke alarms will have a life span between 1 and 3 years?



a) 0.680
b) 0.950
c) 0.904
d) 1.0

Finding Proportions (4 of 4)

The battery in a certain smoke alarm has a life span that is Normally distributed, with a mean of 2 years and a standard deviation of 0.6 years. The graph shows the proportion of life spans above 3 years. Suppose you calculate the proportion above 3 to be 0.9522. Why is this the wrong answer?



a) 0.9522 is the proportion below 3 years.

b) According to the graph, the proportion should be smaller than 0.50.

c) You forgot to subtract the cumulative area 0.9522 from 1.

d) All of the above.

Finding Values (1 of 3)

Approximately, what *z*-score divides the lower 75% of the data from the upper 25%?

a) $z = 0.75$

b) $z = 0.675$

c) $z = -0.675$

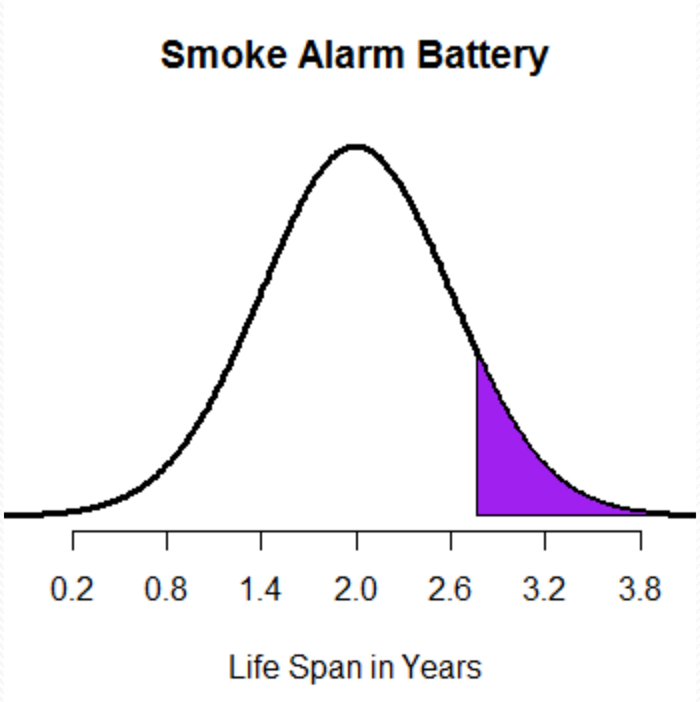
d) $z = -0.25$

e) None of the above.

Finding Values (2 of 3)

The battery in a certain smoke alarm has a life span that is Normally distributed, with a mean of 2 years and a standard deviation of 0.6 years. Only 10% of batteries last longer than ________.

a) 2.77 years
b) 2.91 years
c) 1.28 years
d) None of the above.



Finding Values (3 of 3)

Approximately 25% of full term babies weigh less than ________ grams, where μ = 3485 and σ = 425.

a) 2630
b) 4340
c) 3198
d) None of the above.