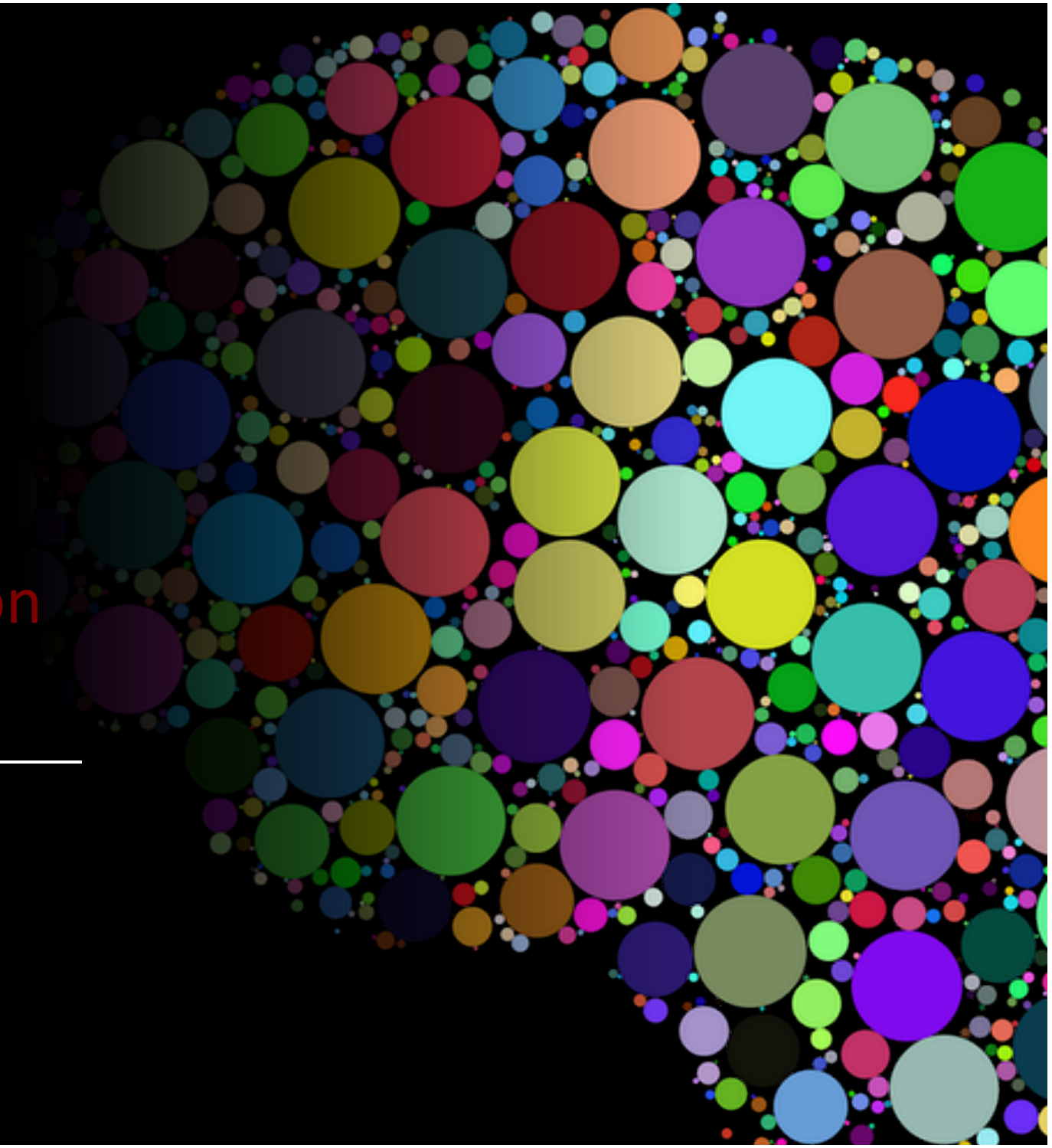
The Normal Curve, Standardization and z Scores

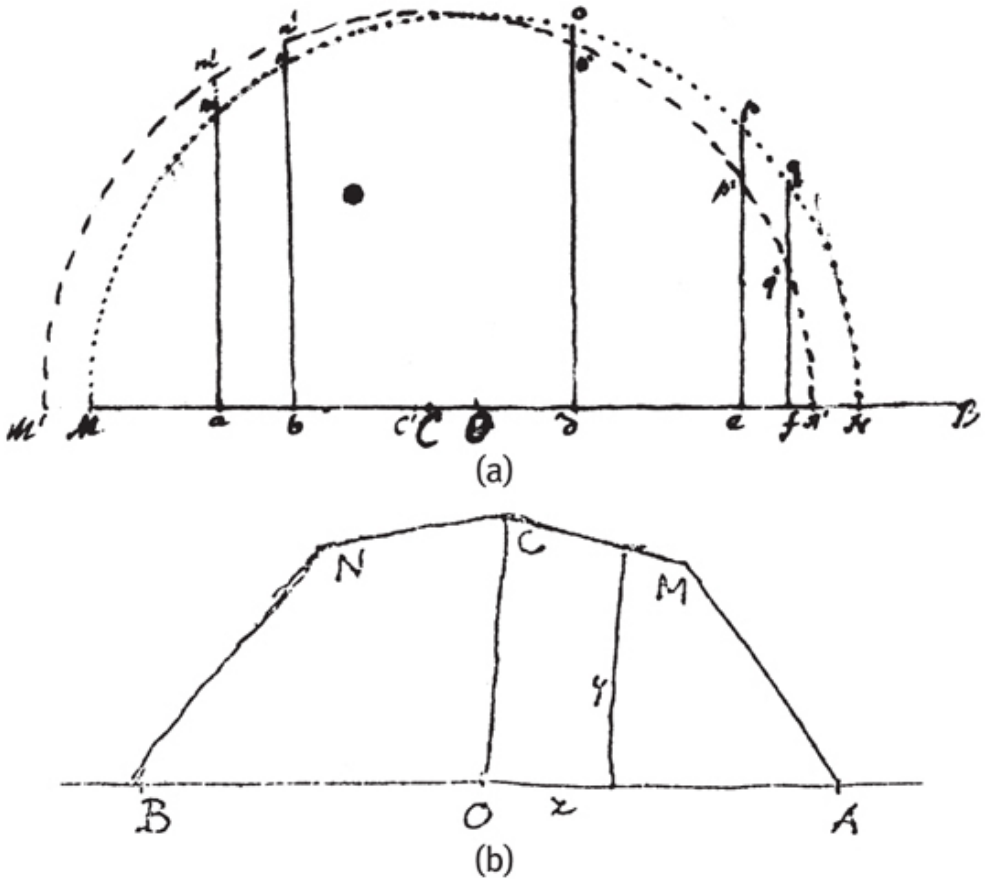
Chapter 6



Freakanomics!

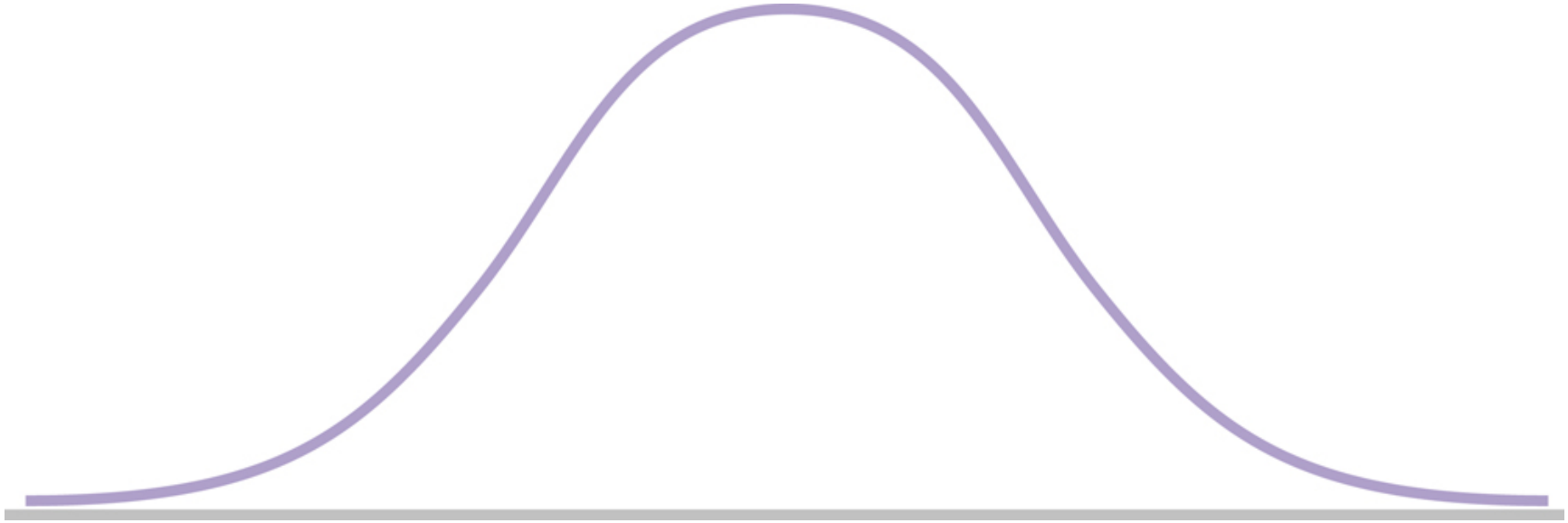
- Go go go!

The Bell Curve is Born (1769)



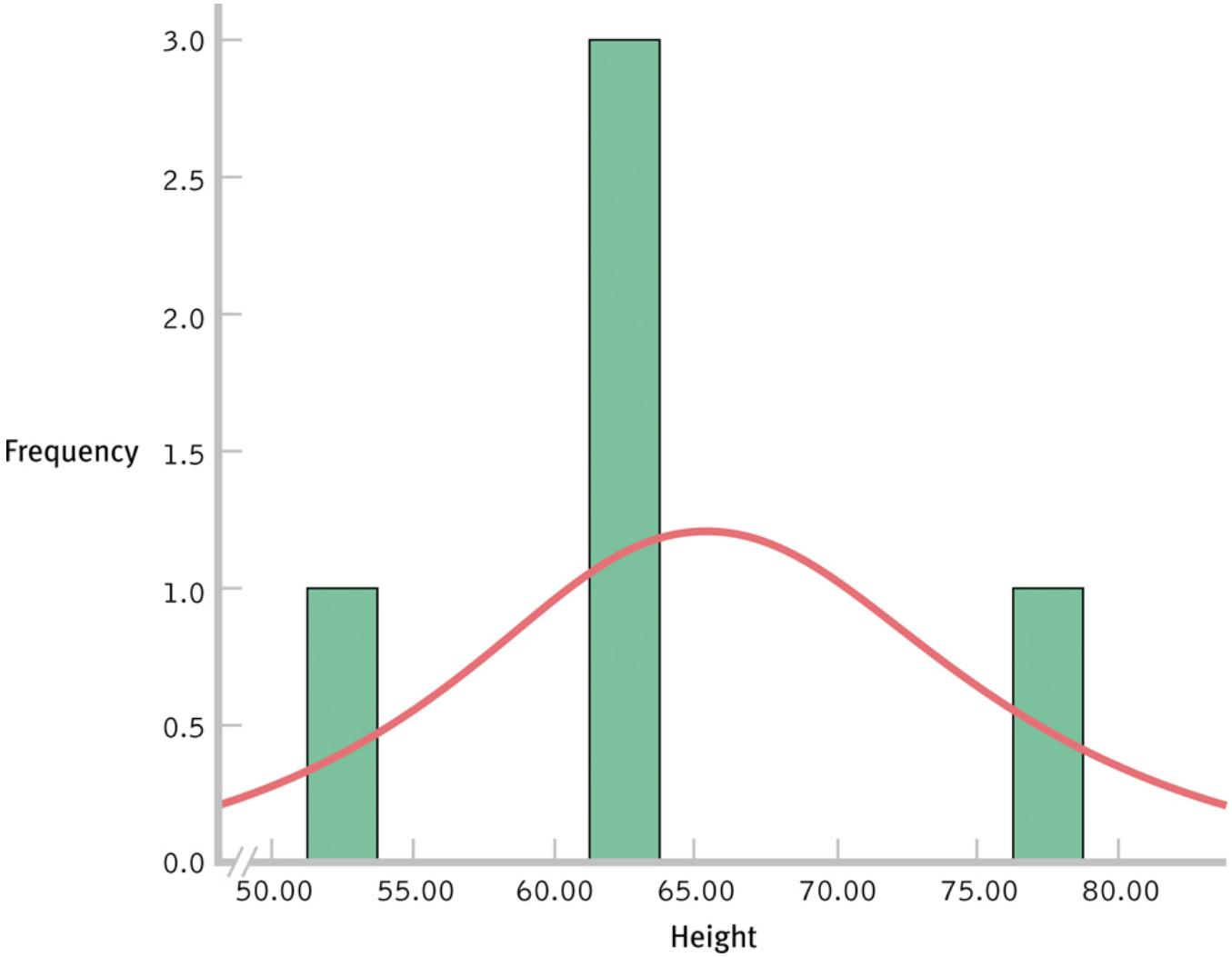
De Moivre – Bernoulli – De Morgan

A Modern Normal Curve

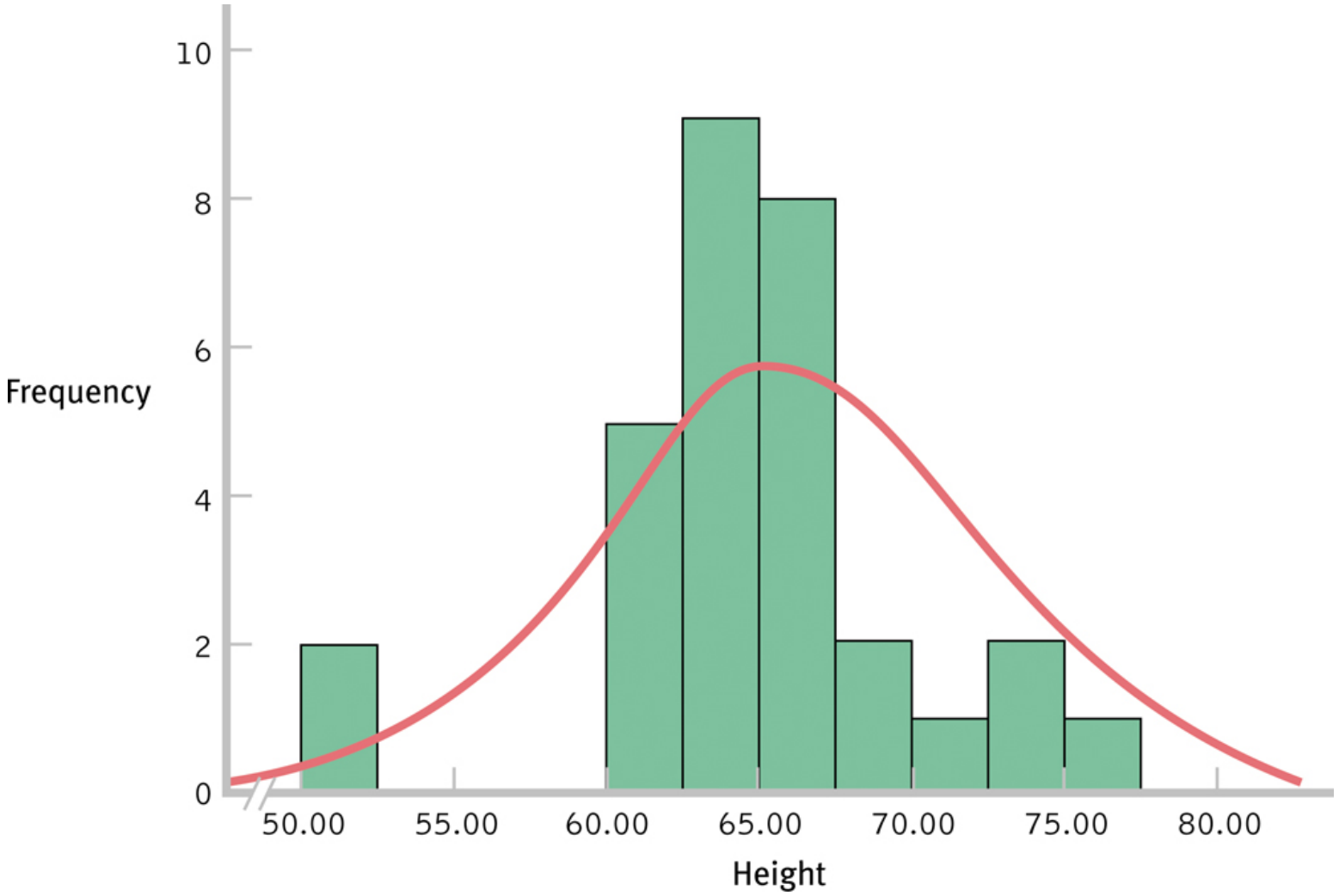


Remember: unimodal, symmetric

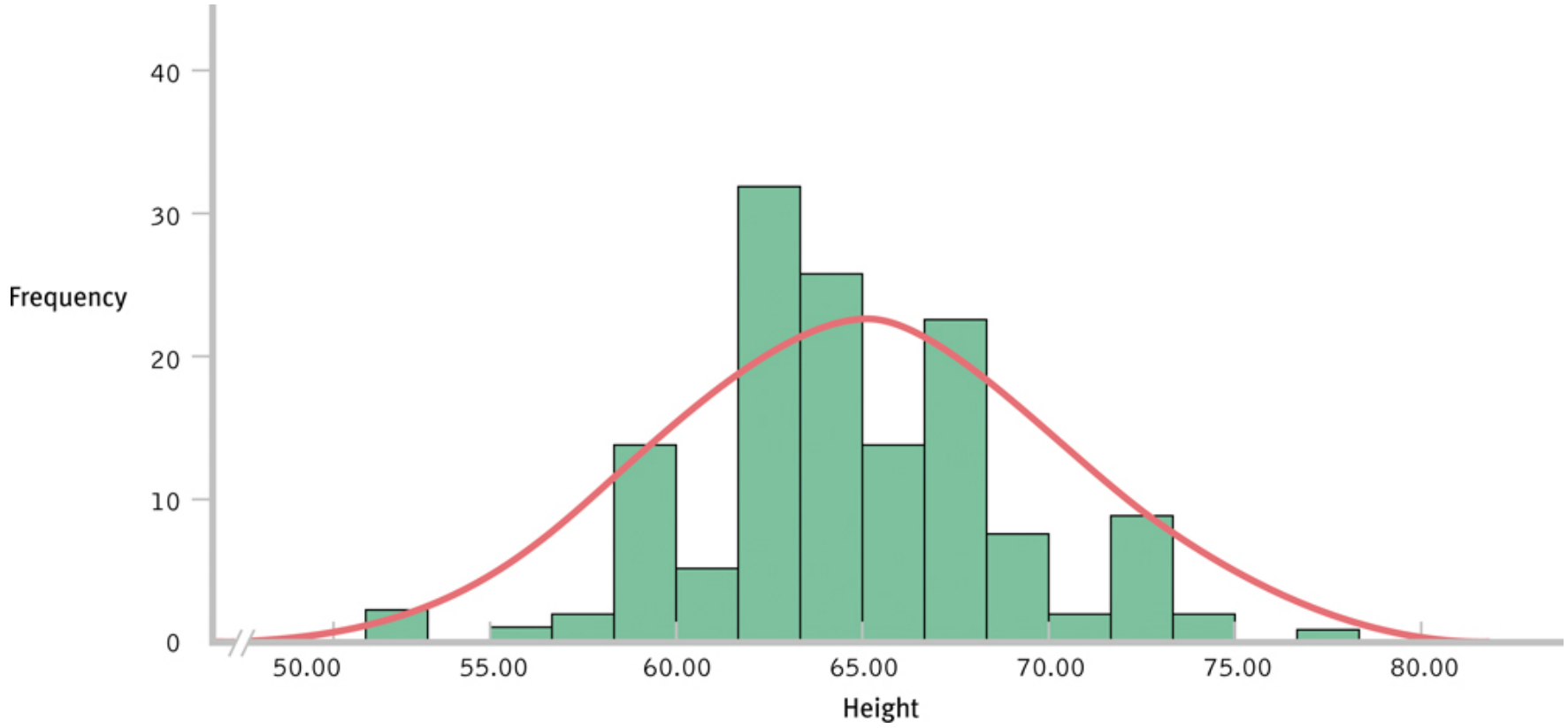
Development of a Normal Curve: Sample of 5



Development of a Normal Curve: Sample of 30



Development of a Normal Curve: Sample of 140



Central Limit Theorem

- As the sample size increases, the shape of the distribution becomes more like the normal curve.
- Can you think of variables that might be normally distributed?
  - Think about it: Can nominal (categorical) variables be normally distributed?

Standardization, z Scores, and the Normal Curve

- Let's say we wanted to compare our student scores on the old GRE (800 point scale) to the new GRE (170 point scale)
- Standardization: allows comparisons by creating a common shared distribution
  - Also allows us to create percentiles (*p*-values!)

Standardization, z Scores, and the Normal Curve

- Normal curve = standardized
  - z distribution (draw it)
  - z scores
    - Comparing z scores
  - Percentiles are *p* values.
    - Different ways to think about *p*.

Standardization, z Scores, and the Normal Curve

- Z-distribution – normal distribution of standardized scores
- Also called standard normal distribution

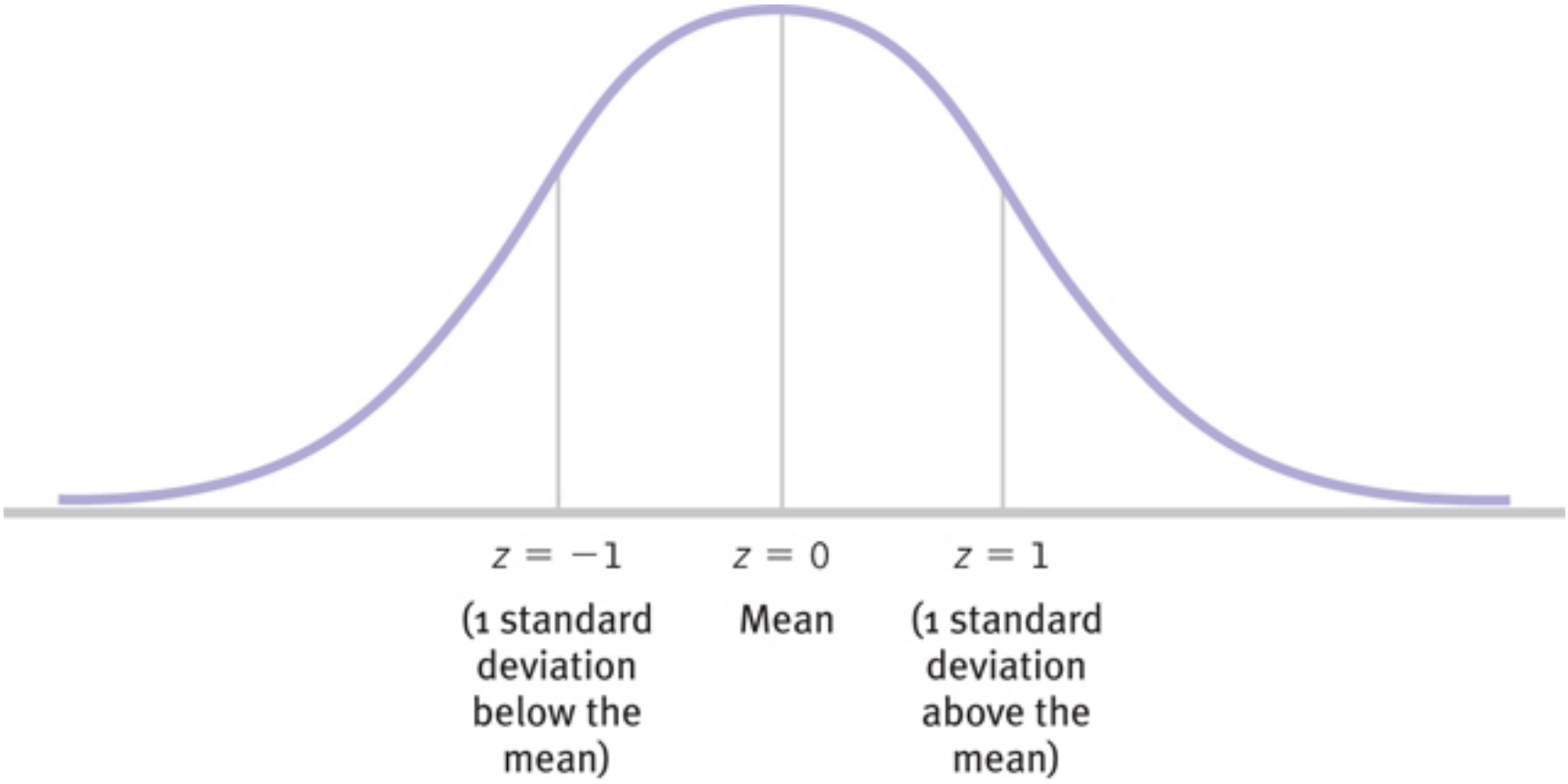
Standardization, z Scores, and the Normal Curve

- So what are z-scores?
  - Number of standard deviations away from the mean of a particular score
  - Can be positive or negative
    - Positive = above mean
    - Negative = below mean

$$z = \frac{(X - \mu)}{\sigma}$$

Tip! Make yourself a symbols chart!

The *z* Distribution



Standardization, z Scores, and the Normal Curve

- Z-distribution
  - Mean = 0
  - Standard deviation = 1

Examples

- Be sure you can do the following:
  1. Find a z score
  2. Find a raw score (x)
  3. Compare scores
  4. Find a percent above
  5. Find a percent below
  6. Find a percent between
  7. Given percent find a z
  8. Given percent find a raw score

Transforming Raw Scores to z Scores

- Step 1: Subtract the mean of the population from the raw score
- Step 2: Divide by the standard deviation of the population

$$z = \frac{(X - \mu)}{\sigma}$$

Example 1

Transforming z Scores into Raw Scores

- Step 1: Multiply the z score by the standard deviation of the population
- Step 2: Add the mean of the population to this product

$$X = z\sigma + \mu$$

Example 2

Using z Scores to Make Comparisons

- If you know your score on an exam, and a friend's score on an exam, you can convert to z scores to determine who did better and by how much.

- z scores are standardized, so they can be compared!

Example 3

Comparing Apples and Oranges

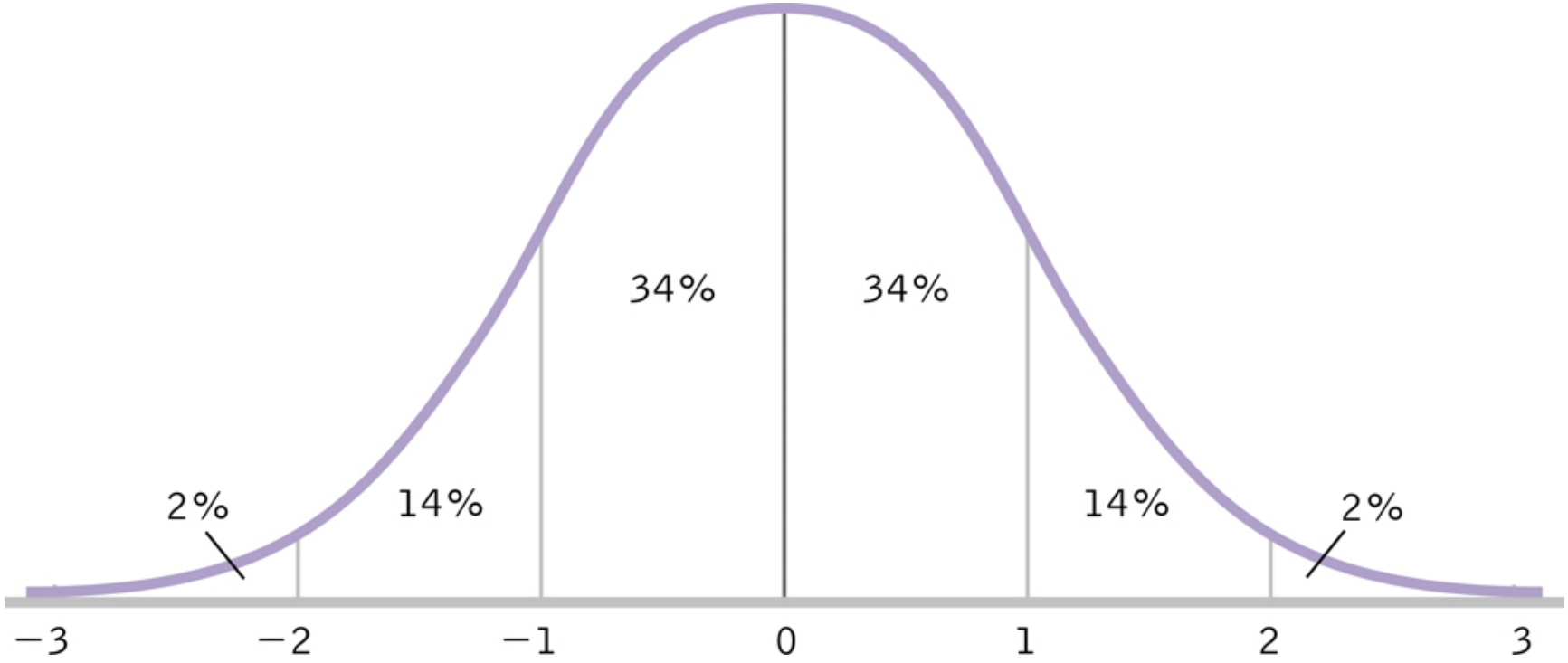
- If we can standardize the raw scores on two different scales, converting both scores to z scores, we can then compare the scores directly.

Example 3

Transforming *z* Scores into Percentiles

- *z* scores tell you where a value fits into a normal distribution.
- Based on the normal distribution, there are rules about where scores with a *z* value will fall, and how it will relate to a percentile rank.
- You can use the area under the normal curve to calculate percentiles for any score.

The Normal Curve and Percentages



Called the 34-14 rule

TABLE 7-1. Excerpt from the *z* Table

The *z* table provides the percentage of scores between the mean and a given *z* value. The full table includes positive *z* statistics from 0.00 to 4.50. The negative *z* statistics are not included because all we have to do is change the sign from positive to negative. The percentage between the mean and a positive *z* statistic is identical to the percentage between the mean and the negative version of that *z* statistic. Remember, the normal curve is symmetric: one side always mirrors the other.

z	% Between Mean and *z*
.	.
.	.
.	.
0.97	33.40
0.98	33.65
0.99	33.89
1.00	34.13
1.01	34.38
1.02	34.61
.	.
.	.
.	.

Remember *R*

- Only the positive numbers are on the table
  - The z distribution is normal, so we don't need the negatives (it's symmetric).
- However, tables are dumb when we have a program that will calculate for us!

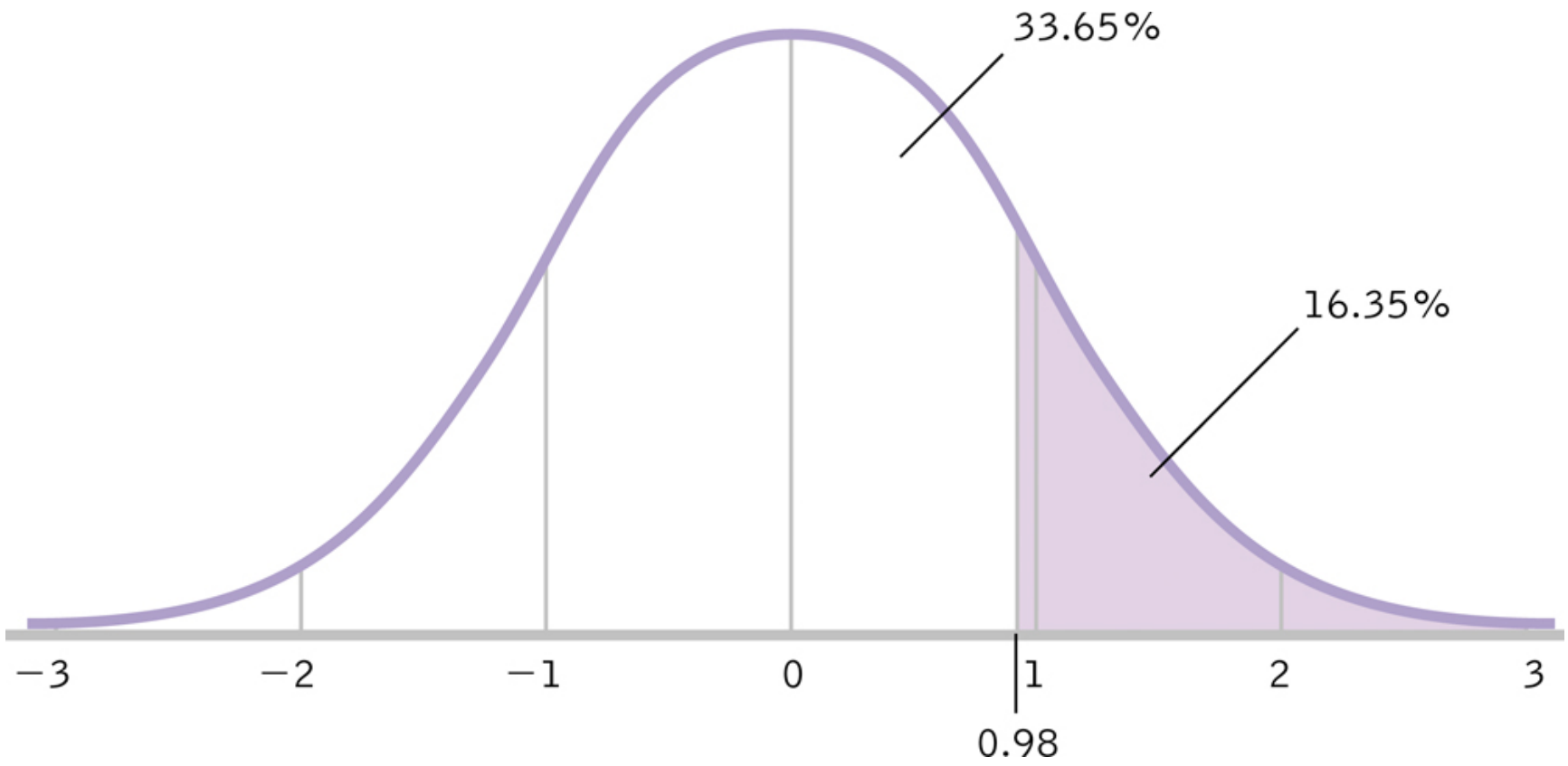
Sketching the Normal Curve

- The benefits of sketching the normal curve:
  - Stays clear in memory; minimizes errors
  - Practical reference
  - Condenses the information
  - Allows you to make sure the *R* information you are getting seems right.

R Curves

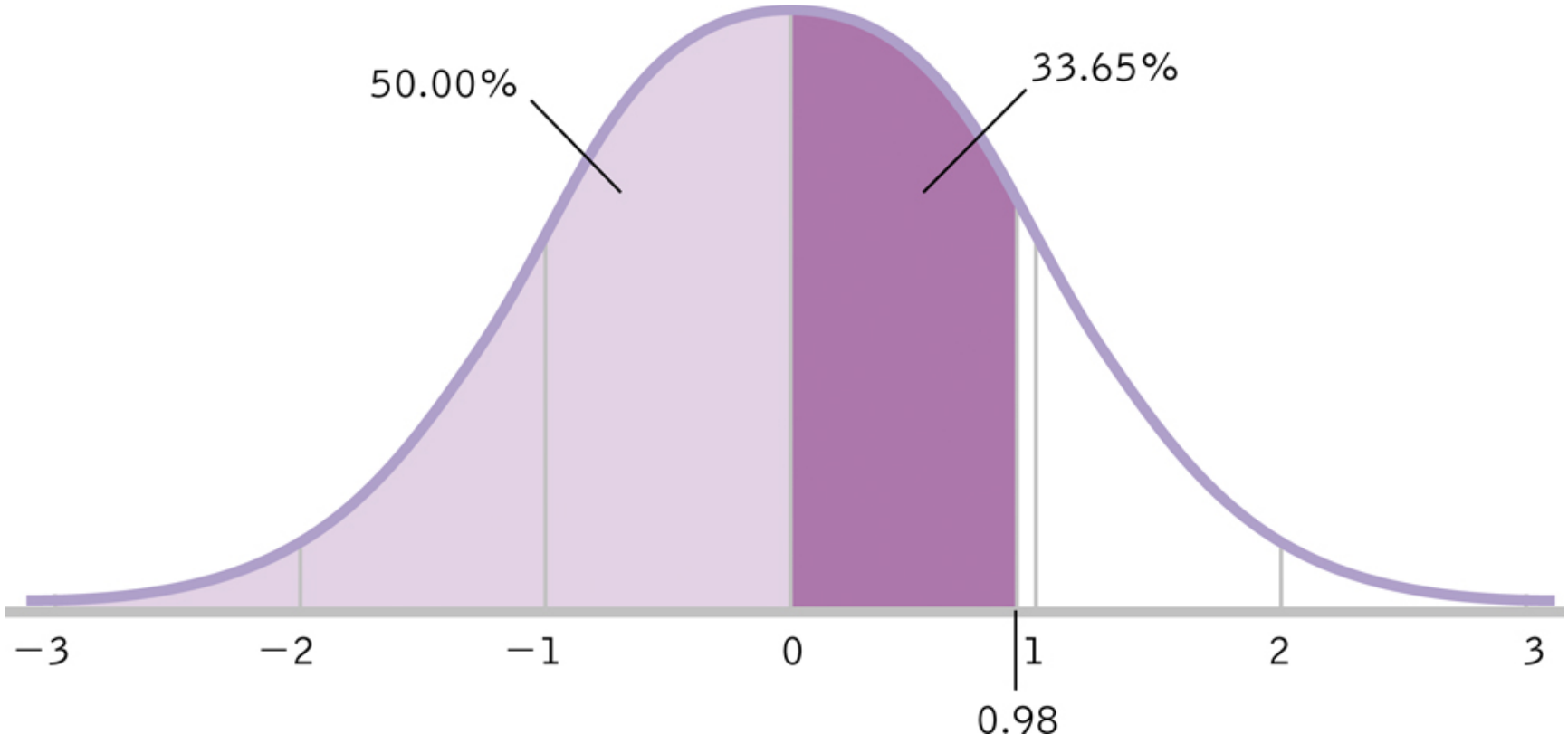
- To get a *p* value from a z score:
  - pnorm(*z*, lower.tail = *F*)
  - lower.tail depends on what you want (options are T or F)

Calculating the Percentage Above a Positive *z* Score



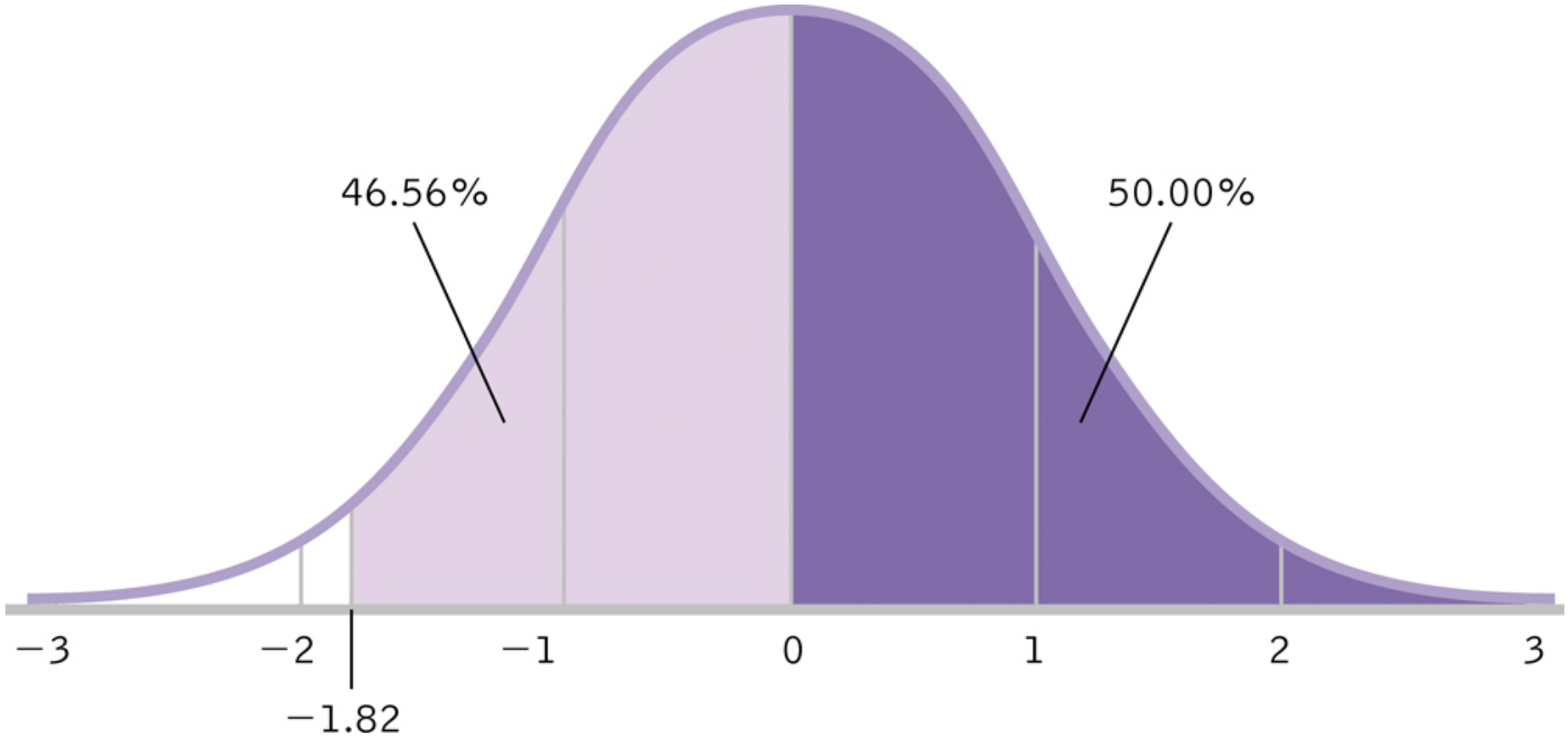
Example 4

Calculating the Percentile for a Positive *z* Score



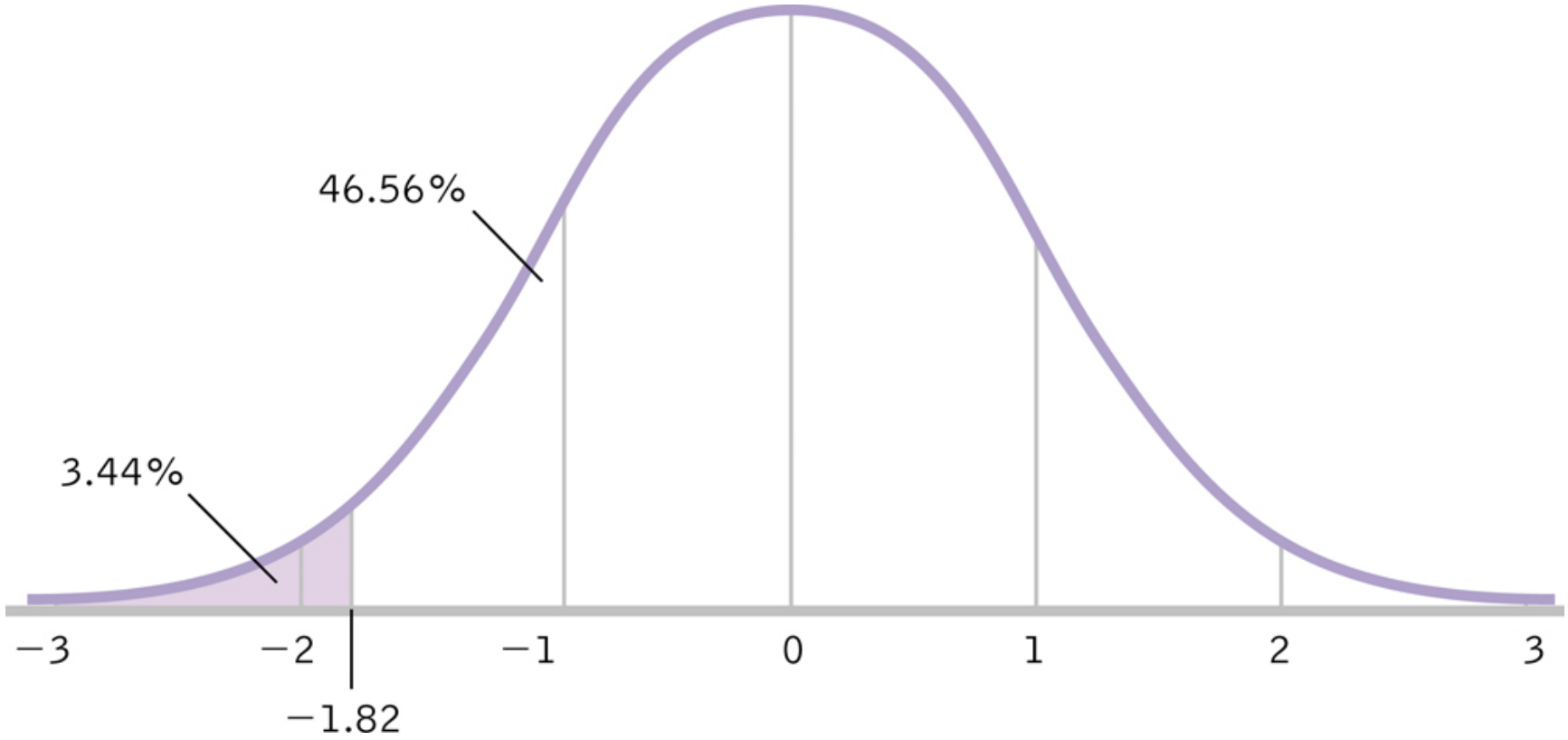
Example 5

Calculating the Percentage Above a Negative *z* Score



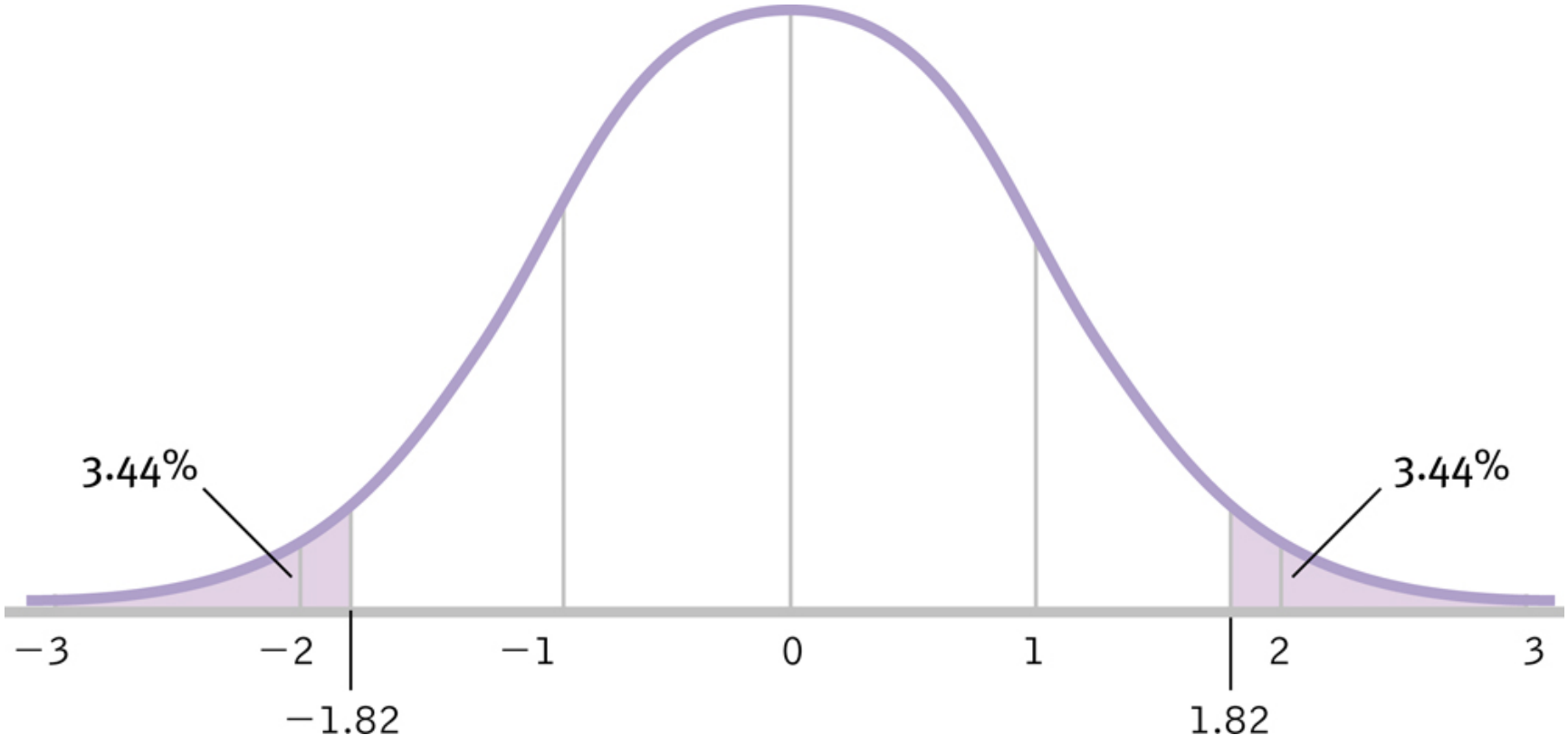
Example 4

Calculating the Percentile for a Negative *z* Score



Example 5

Calculating the Percentage at Least as Extreme as Our *z* Score

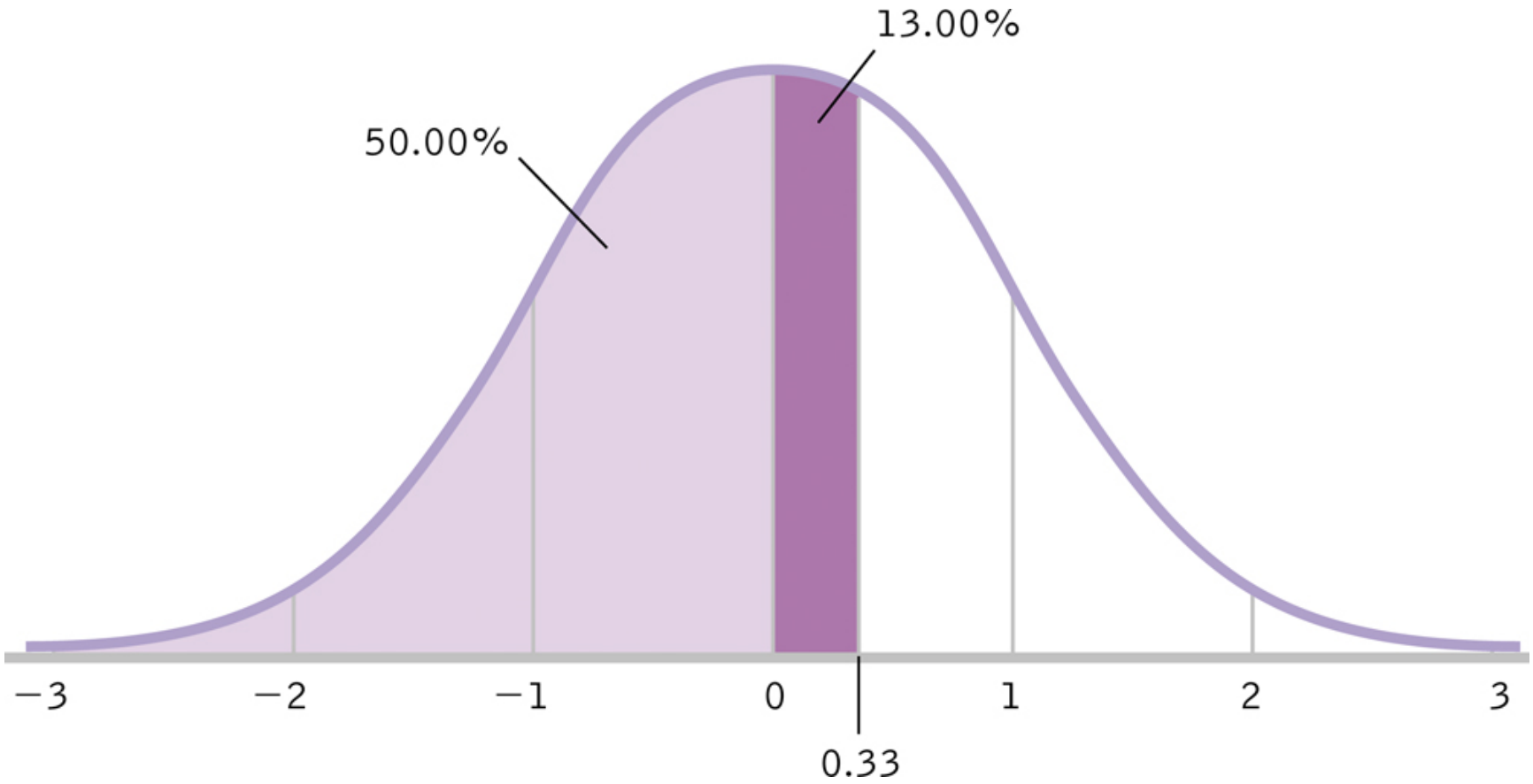


Example 6 (ish)

R Curves

- To get a z score from a *p* value:
- qnorm(*p*, lower.tail = *F*)
  - lower.tail depends on what you want (options are T or F)

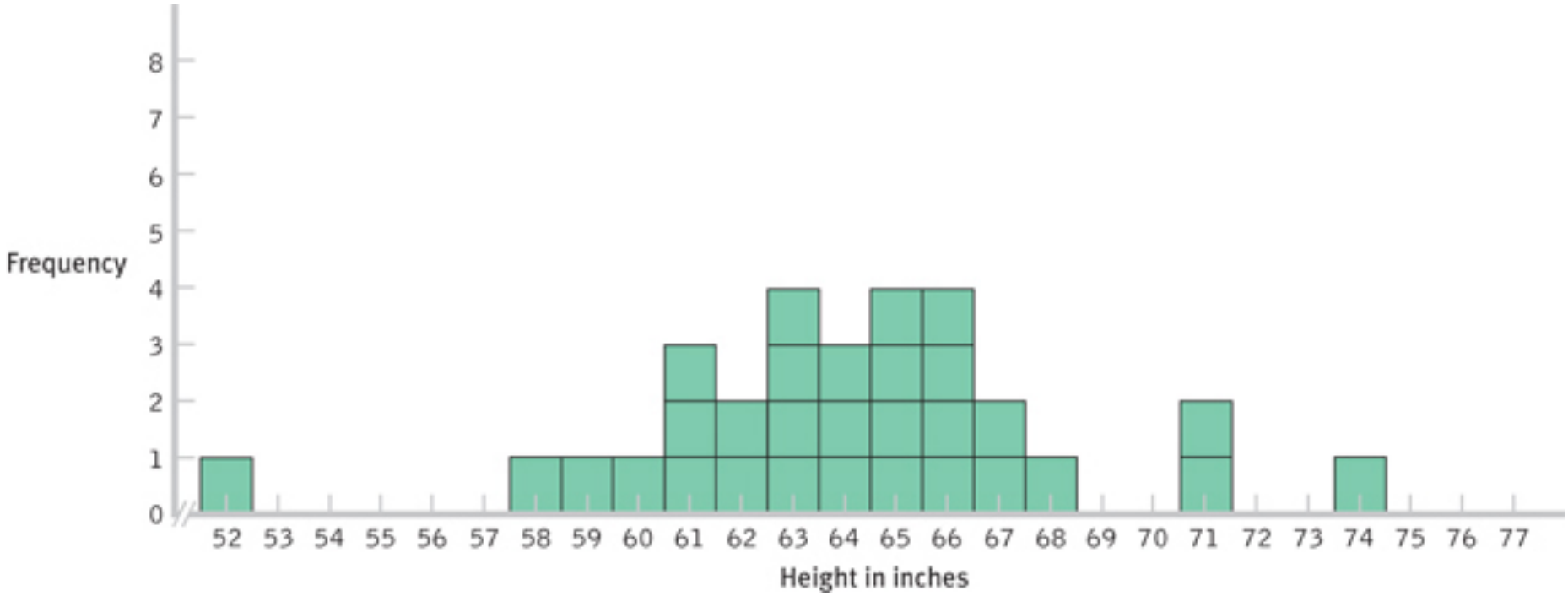
Calculating a Score from a Percentile



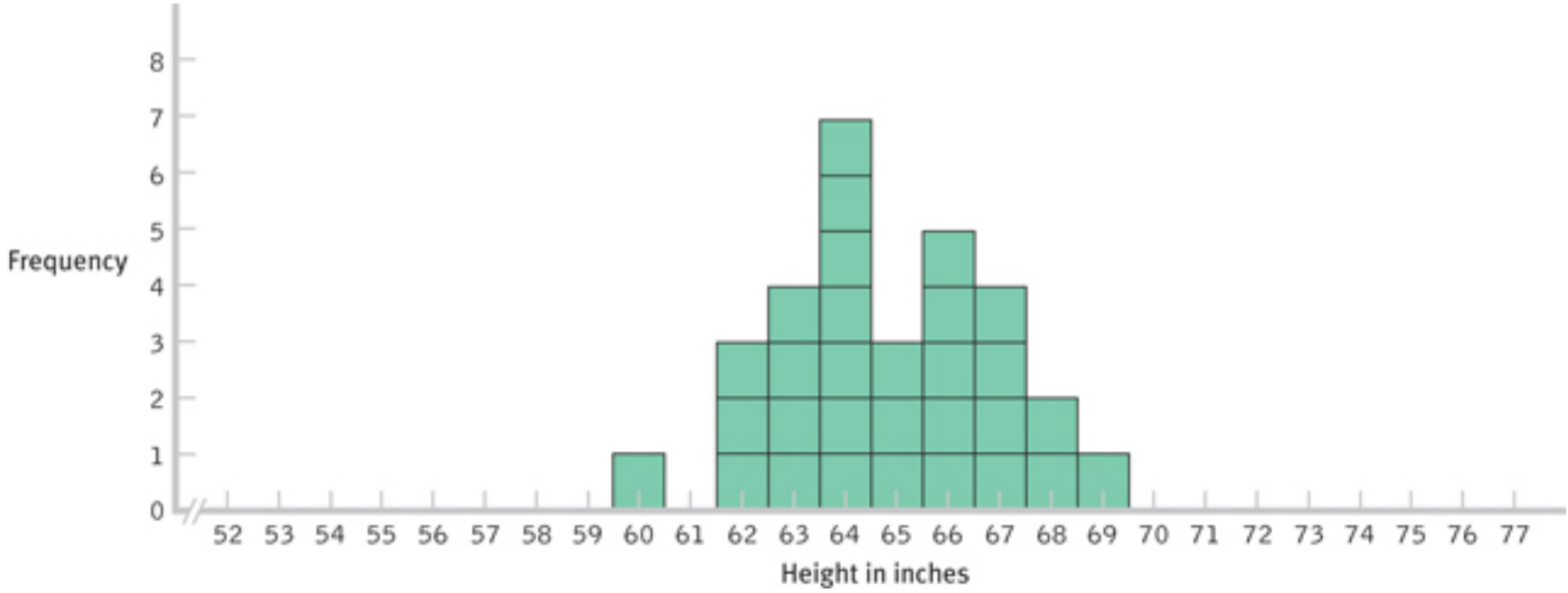
Example 7 and 8

Creating a Distribution of Scores

These distributions were obtained by drawing from the same population.

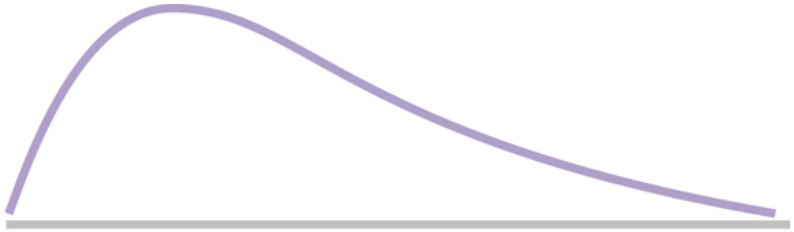


Creating a Distribution of Means



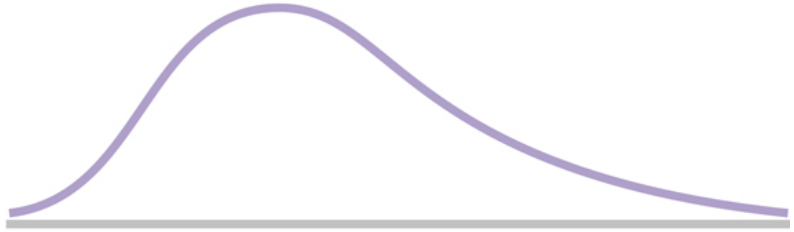
The Mathematical Magic of Large Samples

A severely skewed distribution of scores in a population



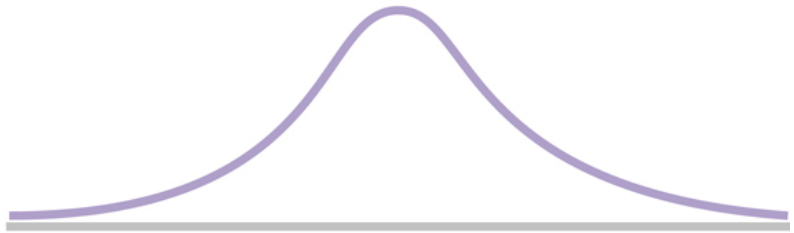
(a)

A less severely skewed distribution of means using samples of 2 from the same population



(b)

A normal distribution of means using samples of 25 from the same population



(c)

Distribution Bunnies!

- https://youtu.be/jvoxEYmQHNM

The Central Limit Theorem

- Distribution of sample means is normally distributed even when the population from which it was drawn is not normal!
- A distribution of means is less variable than a distribution of individual scores.
  - (meaning SD is smaller, but we don't call it SD)

Most of statistics is based on making beer better.
Which is why it's awesome!

Distribution of Means

- Mean of the distribution tends to be the mean of the population.
- Standard deviation of the distribution tends to be less than the standard deviation of the population.
  - The standard error: standard deviation of the distribution of means

$$\sigma_M = \frac{\sigma}{\sqrt{N}}$$

Using the Appropriate Measure of Spread

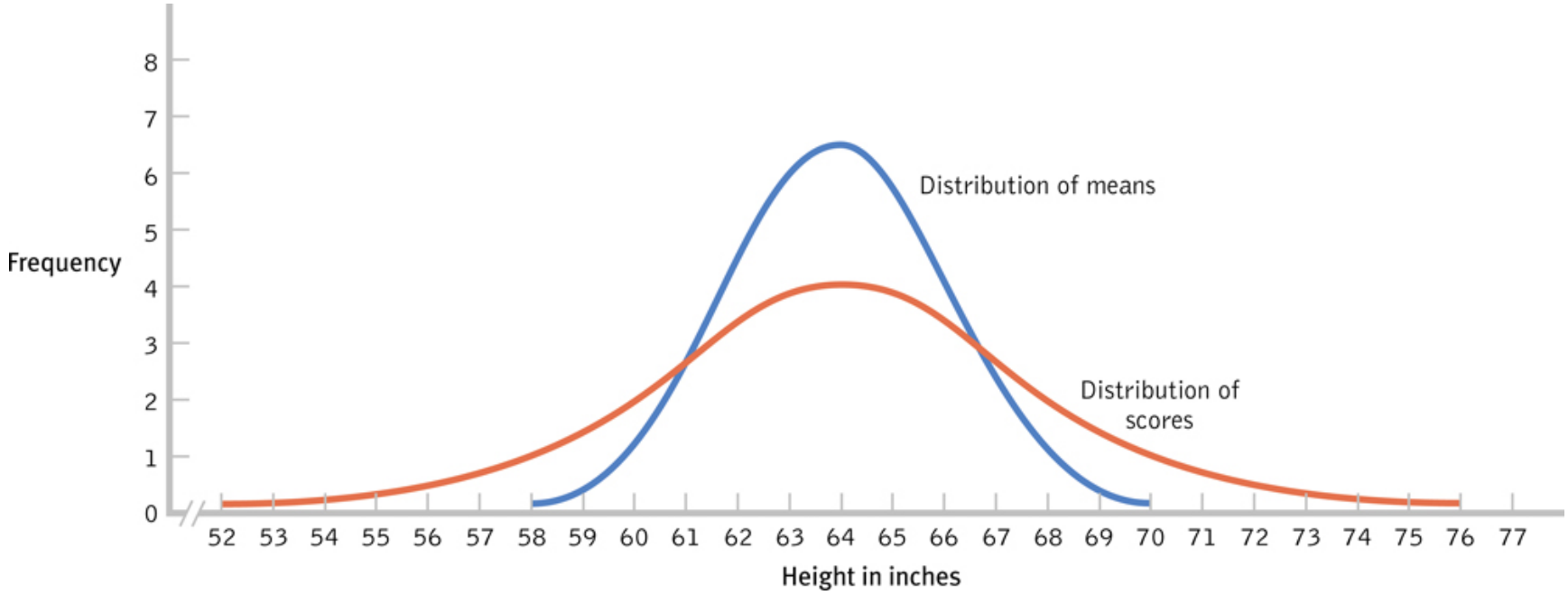


TABLE 6-2. PARAMETERS FOR DISTRIBUTIONS OF SCORES VERSUS MEANS

When we determine the parameters of a distribution, we must consider whether the distribution is composed of means or scores.

DISTRIBUTION	SYMBOL FOR MEAN	SYMBOL FOR SPREAD	NAME FOR SPREAD
Scores	μ	σ	Standard deviation
Means	μ_M	σ_M	Standard error

Z statistic for Distribution of Means

- When you use a distribution of means, you tweak how you calculate z!
- Calculation of percentages stays the same.

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

The Normal Curve and Catching Cheaters

- This pattern is an indication that researchers might be manipulating their analyses to push their z statistics beyond the cutoffs.

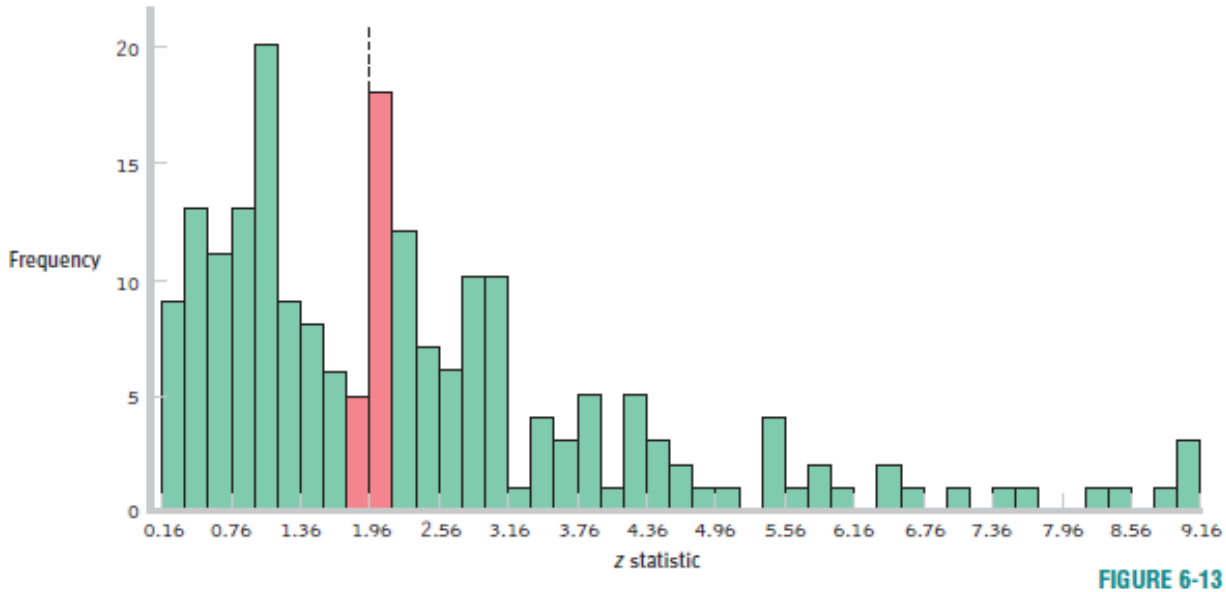


FIGURE 6-13
Identifying Cheaters