

Quantitative Literacy: Thinking Between the Lines

Crauder, Noell, Evans, Johnson

Chapter 6: Statistics

Chapter 6: Statistics

Lesson Plan

- ▶ Data summary and presentation: Boiling down the numbers
- ▶ The normal distribution: Why the bell curve?
- ▶ The statistics of polling: Can we believe the polls?
- ▶ Statistical inference and clinical trials: Effective drugs?



Chapter 6 Statistics

6.2 The normal distribution: Why the bell curve?

Learning Objectives:

- ▶ Understand why the normal distribution is so important.
 - ▶ The bell-shaped curve
 - ▶ Mean and standard deviation for the normal distribution
 - ▶ z-scores with Percentile scores
 - ▶ The Central Limit Theorem
 - ▶ Significance of apparently small deviations

Chapter 6 Statistics

6.2 The normal distribution: Why the bell curve?

- ▶ **The bell-shaped curve:** Figure 6.13 shows the distribution of heights of adult males in the United States. A graph shaped like this one resembles a bell—thus the *bell curve*. This bell-shaped graph is typical of normally distributed data.

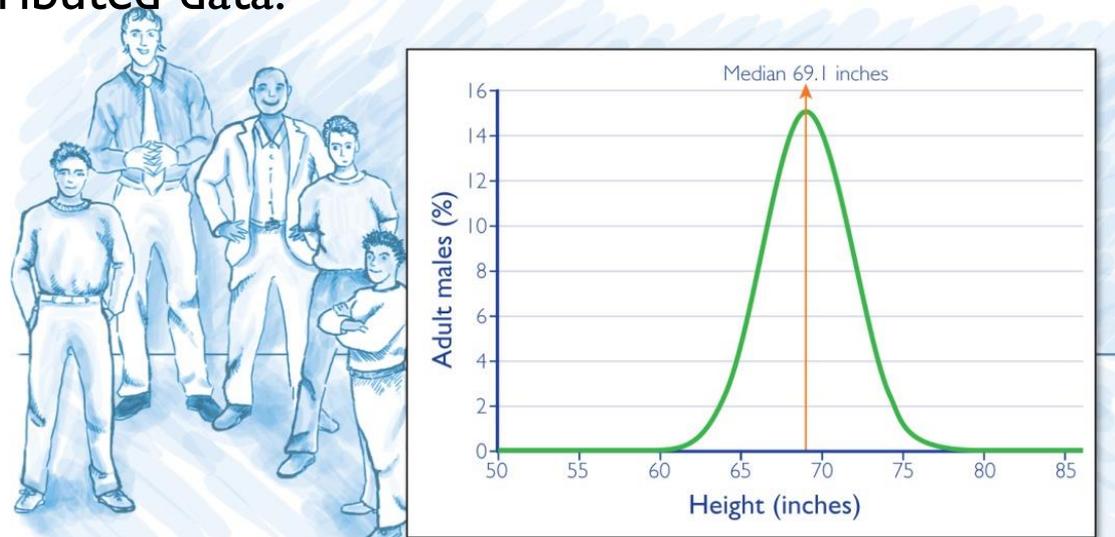


FIGURE 6.13 Heights of adult males are normally distributed.

- ▶ **The mean and median are the same:** For normally distributed data, the mean and median are the same. Figure 6.13 indicates that the median height of adult males is 69.1 inches. The average height of adult males is 69.1 inches.

Chapter 6 Statistics

6.2 The normal distribution: Why the bell curve?

- ▶ **Most data are clustered about the mean:** The vast majority of adult males are within a few inches of the mean.

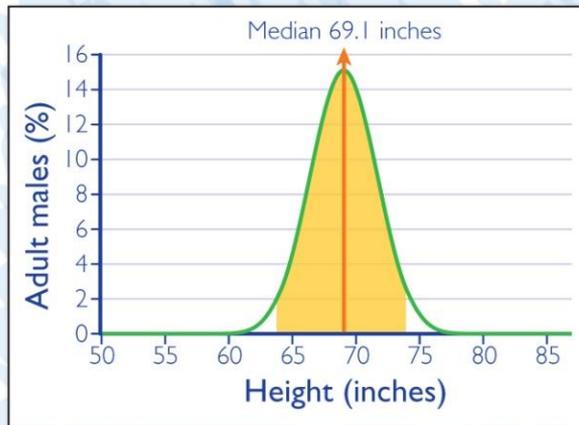


FIGURE 6.14 95% of adult males are within 5 inches of the median.

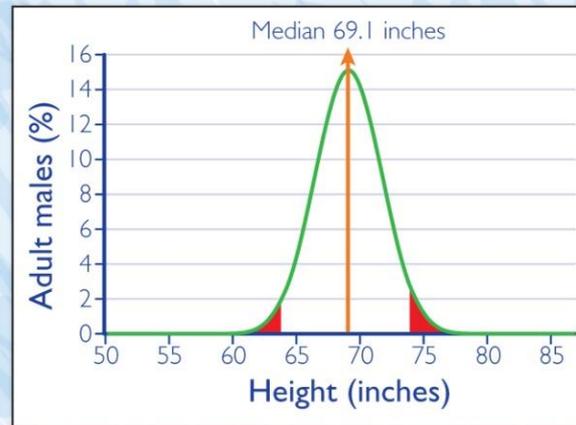


FIGURE 6.15 Relatively few men are very tall or very short.

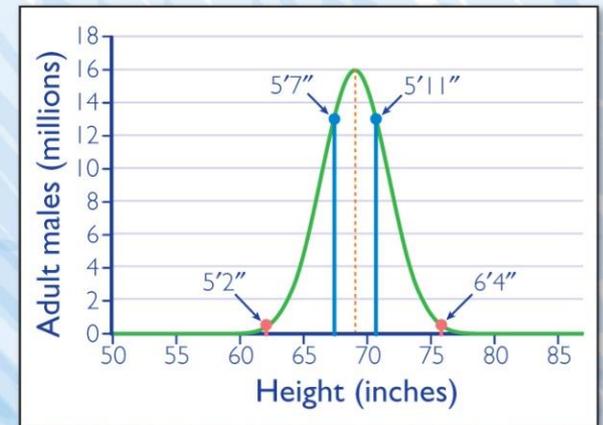


FIGURE 6.16 The bell curve is symmetric about the mean.

Chapter 6 Statistics

6.2 The normal distribution: Why the bell curve?

- ▶ **The bell curve is symmetric about the mean:** The curve to the left of the mean is a mirror image of the curve to the right of the mean. In terms of heights, there are about the same number of men 2 inches taller than the mean as there are men 2 inches shorter than the mean. This is illustrated in Figure 6.16.

- ▶ If data are **normally distributed**:
 1. Their graph is a bell-shaped curve.
 2. The mean and median are the same.
 3. Most of the data tend to be clustered relatively near the mean.
 4. The data are symmetrically distributed above and below the mean.

Chapter 6 Statistics

6.2 The normal distribution: Why the bell curve?

- ▶ **Example:** Figure 6.17 shows the distribution of IQ scores, and Figure 6.18 shows the percentage of American families and level of income. Which of these data sets appear to be normally distributed, and why?

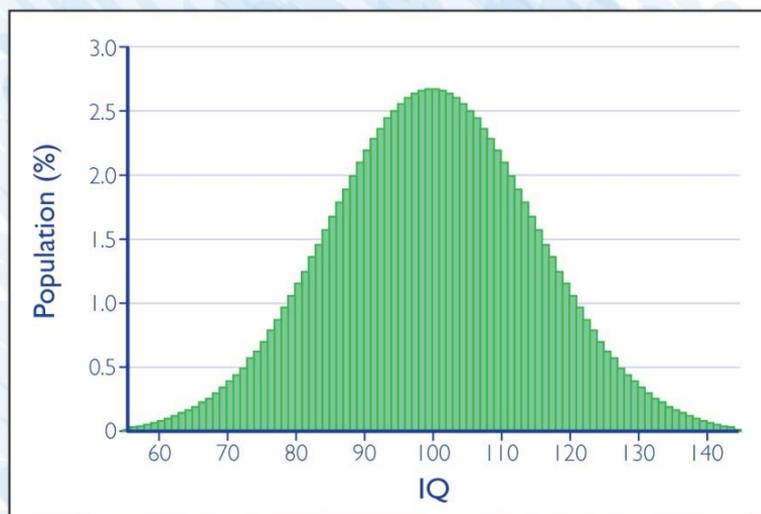


FIGURE 6.17 Percentage of population with given IQ score.

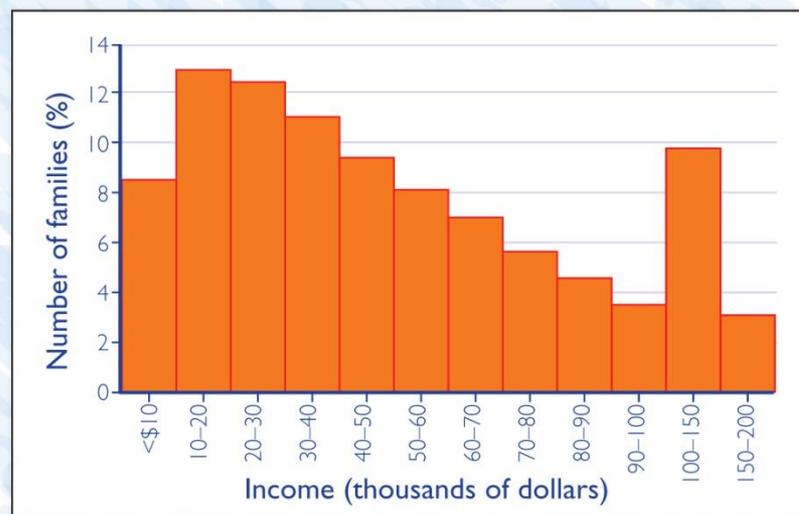


FIGURE 6.18 Percentage of families with given income.

Chapter 6 Statistics

6.2 The normal distribution: Why the bell curve?

- ▶ **Solution:** The IQ scores appear to be normally distributed because they are symmetric about the median score of 100, and most of the data relatively close to this value.
- ▶ Family incomes do not appear to be normally distributed because they are not symmetric. They are skewed toward the lower end of the scale, meaning there are many more families with low incomes than with high incomes.

Chapter 6 Statistics

6.2 The normal distribution: Why the bell curve?

- ▶ **Mean and standard deviation for the normal distribution:** A normal distribution, the mean and standard deviation completely determine the bell shape for the graph of the data.
- ▶ The mean determines the middle of the bell curve.
- ▶ The standard deviation determines how steep the curve is.
- ▶ A large standard deviation results in a very wide bell, and small standard deviation results in a thin, steep bell.

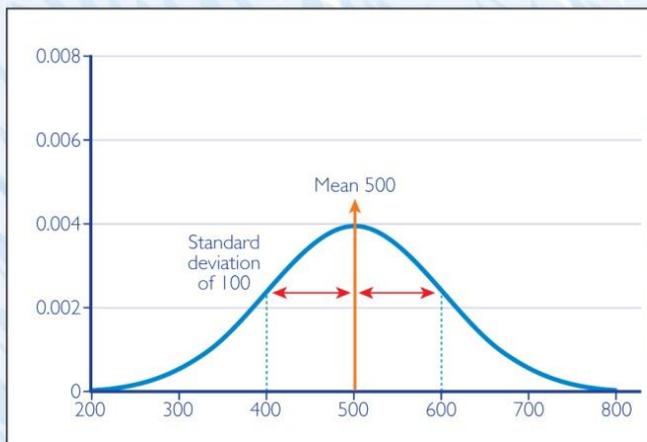


FIGURE 6.21 Normal curve with mean 500 and standard deviation 100.

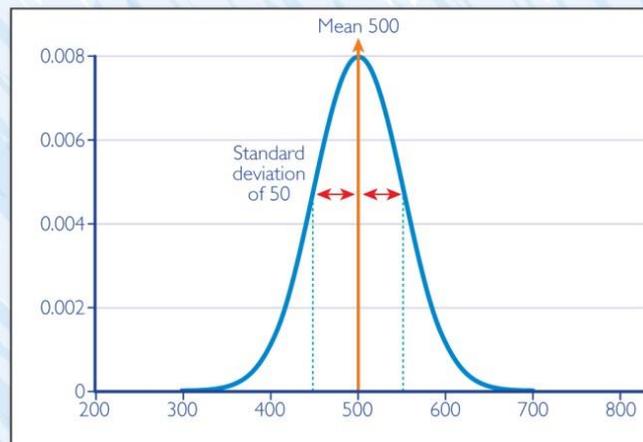


FIGURE 6.22 Normal curve with mean 500 and standard deviation 50.

Chapter 6 Statistics

6.2 The normal distribution: Why the bell curve?

Normal Data: 68-95-99.7% Rule

If a set of data is normally distributed:

- About 68% of the data lie within one standard deviation of the mean (34% within one standard deviation above the mean and 34% within one standard deviation below the mean). See Figure 6.23.
- About 95% of the data lie within two standard deviations of the mean (47.5% within two standard deviations above the mean and 47.5% within two standard deviations below the mean). See Figure 6.24.
- About 99.7% of the data lie within three standard deviations of the mean (49.85% within three standard deviations above the mean and 49.85% within three standard deviations below the mean). See Figure 6.25.

Chapter 6 Statistics

6.2 The normal distribution: Why the bell curve?

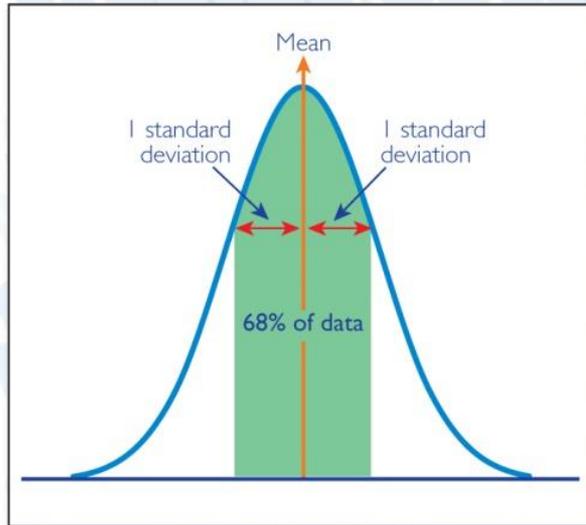


FIGURE 6.23 68% of data lie within one standard deviation of the mean.

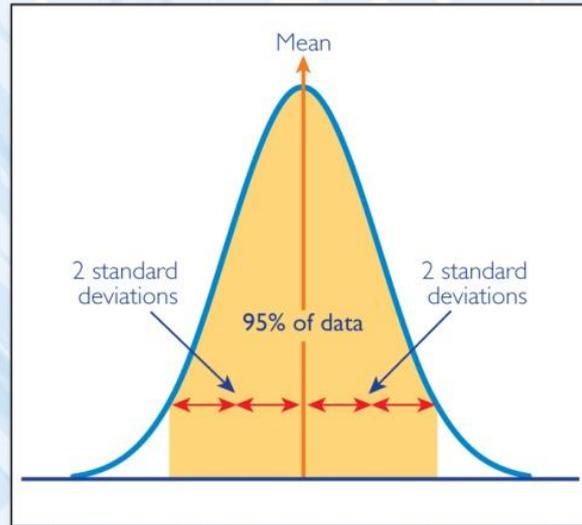


FIGURE 6.24 95% of data lie within two standard deviations of the mean.

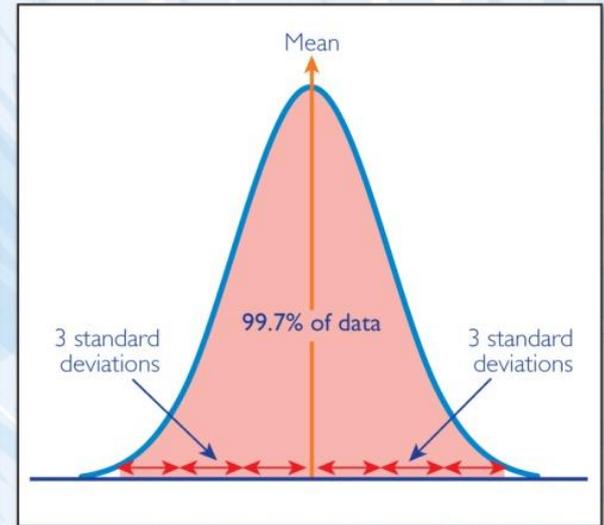


FIGURE 6.25 99.7% of the data lie within three standard deviations of the mean.

Chapter 6 Statistics

6.2 The normal distribution: Why the bell curve?

- ▶ **Example:** We noted earlier that adult male heights in the United States are normally distributed, with a mean of 69.1 inches. The standard deviation is 2.65 inches.

What does the 68-95-99.7% rule tell us about the heights of adult males?

- ▶ **Solution:** 68% of adult males are between
 $69.1 - 2.65 = 66.45$ inches (5 feet 6.45 inches) and
 $69.1 + 2.65 = 71.75$ inches (5 feet 11.75 inches) tall
- ▶ 95% are between
 $69.1 - (2 \times 2.65) = 63.8$ inches and
 $69.1 + (2 \times 2.65) = 74.4$ inches tall
- ▶ 99.7% are between
 $69.1 - (3 \times 2.65) = 61.15$ inches and
 $69.1 + (3 \times 2.65) = 77.05$ inches tall

Chapter 6 Statistics

6.2 The normal distribution: Why the bell curve?

- ▶ **Example:** The weights of apples in the fall harvest are normally distributed, with a mean weight of 200 grams and standard deviation of 12 grams. Figure 6.28 shows the weight distribution of 2000 apples. In a supply of 2000 apples, how many will weigh between 176 and 224 grams?

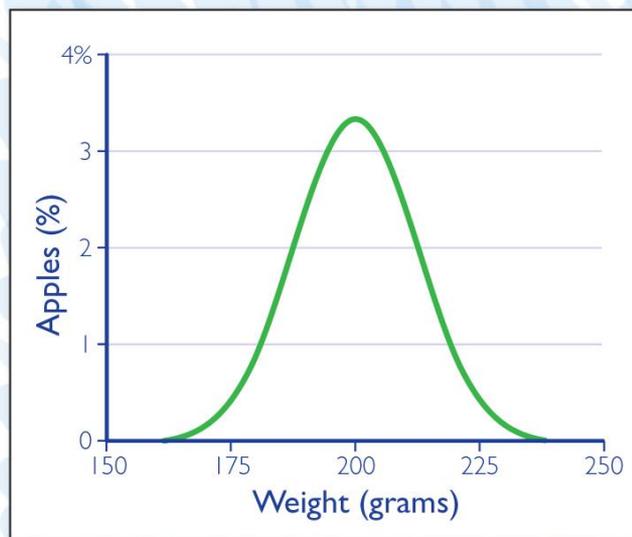


FIGURE 6.28 Apples with mean 200 grams and standard deviation 12 grams.

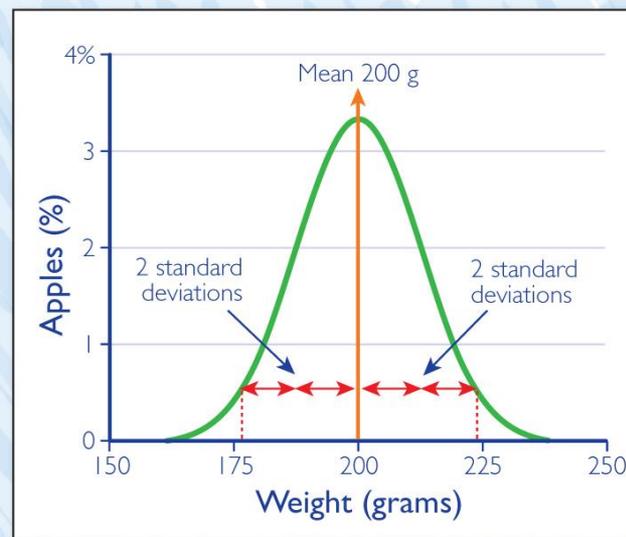


FIGURE 6.29 Apples between 176 and 224 grams.

Chapter 6 Statistics

6.2 The normal distribution: Why the bell curve?

▶ **Solution:**

- ▶ Apples weighing 176 grams are $200 - 176 = 24$ grams below the mean, and apples weighing 224 grams are $224 - 200 = 24$ grams above the mean.
- ▶ Now 24 grams represents $24/12 = 2$ standard deviations. So the weight range of 176 grams to 224 grams is within two standard deviations of the mean.
- ▶ Therefore, about 95% of data points will lie in this range. This means that about 95% of 2000, or 1900 apples, weigh between 176 and 224 grams.

Chapter 6 Statistics

6.2 The normal distribution: Why the bell curve?

In a normal distribution, the **z-score** or **standard score** for a data point is the number of standard deviations that point lies above or below the mean. For data points above the mean the z-score is positive, and for data points below the mean the z-score is negative.

$$z - \text{score} = (\text{Data point} - \text{Mean}) / \text{Standard deviation}$$

$$\text{Data point} = \text{Mean} + z - \text{score} \times \text{Standard deviation}$$

Chapter 6 Statistics

6.2 The normal distribution: Why the bell curve?

- ▶ **Example:** The weights of newborns in the United States are approximately normally distributed. The mean birthweight (for single births) is about 3332 grams (7 pounds, 5 ounces). The standard deviation is about 530 grams. Calculate the z-score for a newborn weighing 3700 grams (about 8 pounds, 2 ounces).
- ▶ **Solution:** A 3700-gram newborn is $3700 - 3332 = 368$ grams above the mean weight of 3332 grams. We divide by the number of grams in one standard deviation to find the z-score:

$$z - \text{score for 3700 grams} = \frac{368}{530} = 0.7$$

Chapter 6 Statistics

6.2 The normal distribution: Why the bell curve?

TABLE 6.2 Percentile from z-Score

z-score	Percentile	z-score	Percentile
-2.7	0.35	0.0	50.00
-2.6	0.47	0.1	53.98
-2.5	0.62	0.2	57.93
-2.4	0.82	0.3	61.79
-2.3	1.07	0.4	65.54
-2.2	1.39	0.5	69.15
-2.1	1.79	0.6	72.57
-2.0	2.28	0.7	75.80
-1.9	2.87	0.8	78.81
-1.8	3.59	0.9	81.59
-1.7	4.46	1.0	84.13
-1.6	5.48	1.1	86.43
-1.5	6.68	1.2	88.49
-1.4	8.08	1.3	90.32
-1.3	9.68	1.4	91.92
-1.2	11.51	1.5	93.32
-1.1	13.57	1.6	94.52
-1.0	15.87	1.7	95.54
-0.9	18.41	1.8	96.41
-0.8	21.19	1.9	97.13
-0.7	24.20	2.0	97.73
-0.6	27.43	2.1	98.21
-0.5	30.85	2.2	98.61
-0.4	34.46	2.3	98.93
-0.3	38.21	2.4	99.18
-0.2	42.07	2.5	99.38
-0.1	46.02	2.6	99.53
0.0	50.00	2.7	99.65

Chapter 6 Statistics

6.2 The normal distribution: Why the bell curve?

- ▶ The **percentile** for a number relative to a list of data is the percentage of data points that are less than or equal to that number.
- ▶ **Example:** The average length of illness for flu patients in a season is normally distributed, with a mean of 8 days and standard deviation of 0.9 day. What percentage of flu patients will be ill for more than 10 days?
- ▶ **Solution:** Ten days is 2 days above the mean of 8 days. This gives a z-score of $2/0.9$ or about 2.2. Table 6.2 gives a percentile of about 98.6% for this z-score. It means that about 98.6% of patients will recover in 10 days or less. Thus, only about $100\% - 98.6\% = 1.4\%$ will be ill for more than 10 days.

Chapter 6 Statistics

6.2 The normal distribution: Why the bell curve?

- ▶ **Example:** Recall from the previous Example that the weights of newborns in the United States are approximately normally distributed. The mean birthweight (for single births) is about 3332 grams (7 pounds, 5 ounces). The standard deviation is about 530 grams.
 1. What percentage of newborns weigh more than 8 pounds (3636.4 grams)?
 2. Low birthweight is a medical concern. The American Medical Association defines low birthweight to be 2500 grams (5 pounds, 8 ounces) or less. What percentage of newborns are classified as low-birthweight babies?

Chapter 6 Statistics

6.2 The normal distribution: Why the bell curve?

► **Solution:**

$$1. \quad z - \text{score} = \frac{3636.4 - \text{mean}}{\text{standard deviation}} = \frac{3636.4 - 3332}{530} = \frac{304.4}{530} = 0.6$$

Consulting Table 6.2, we find that this represents a percentile of about 72.6%.

This means that about 72.6% of newborns weigh 8 pounds or less. So, $100\% - 72.6\% = 27.4\%$ of newborns weigh more than 8 pounds.

$$2. \quad z - \text{score} = \frac{2500 - \text{mean}}{\text{standard deviation}} = \frac{3332 - 2500}{530} = \frac{832}{530} = 1.6$$

Table 6.2 shows a percentile of about 5.5% for a z-score of -1.6 . Hence, about 5.5% of newborns are classified as low-birthweight babies.

Chapter 6 Statistics

6.2 The normal distribution: Why the bell curve?

▶ **The Central Limit Theorem**

According to the **Central Limit Theorem**, percentages obtained by taking many samples of the same size from a population are approximately normally distributed.

- ▶ The mean $p\%$ of the normal distribution is the mean of the whole population.
- ▶ If the sample size is n , the standard deviation of the normal distribution is:

$$\text{Standard deviation} = \sigma = \sqrt{\frac{p(100 - p)}{n}} \text{ percentage points}$$

Here, p is a percentage, not a decimal.

Chapter 6 Statistics

6.2 The normal distribution: Why the bell curve?

- ▶ **Example:** For a certain disease, 30% of untreated patients can be expected to improve within a week. We observe a population of 50 patients and record the percentage who improve within a week. According to the Central Limit Theorem, the results of such a study will be approximately normally distributed.
 1. Find the mean and standard deviation for this normal distribution.
 2. Find the percentage of test groups of 50 patients in which more than 40% improve within a week.

Chapter 6 Statistics

6.2 The normal distribution: Why the bell curve?

► **Solution:**

1. $p = 30\%$, $n = 50$. A standard deviation of

$$\sigma = \sqrt{\frac{p(100-p)}{n}} = \sqrt{\frac{30(100-30)}{50}} = 6.5 \text{ percentage points}$$

2. The z-score for 40%:

$$z - \text{score} = \frac{40 - \text{mean}}{\text{standard deviation}} = \frac{40 - 30}{6.5} = \frac{10}{6.5} = 1.5$$

Table 6.2 gives a percentile of about 93.3%.

This means that in 93.3% of test groups, we expect that 40% or fewer will improve within a week.

Only $100\% - 93.3\% = 6.7\%$ of test groups will show more than 40% improving within a week.

Chapter 6 Statistics

6.2 The normal distribution: Why the bell curve?

- ▶ **Example:** Assume we know that 20% of Americans suffer from a certain type of allergy. Suppose we take a random sample of 100,000 Americans and record the percentage who suffer from this allergy.
 1. The Central Limit Theorem says that percentages from such surveys will be normally distributed. What is the mean of this distribution?
 2. What is the standard deviation of the normal distribution in part 1?
 3. Suppose we find that in a town of 100,000 people, 21% suffer from this allergy. Is this an unusual sample? What does the answer to such a question tell us about this town?

Chapter 6 Statistics

6.2 The normal distribution: Why the bell curve?

► **Solution:**

1. The mean is $p = 20\%$.
2. For a sample size of 100,000,

$$\sigma = \sqrt{\frac{p(100-p)}{n}} = \sqrt{\frac{20(100-20)}{100,000}} = 0.13 \text{ percentage point.}$$

3. Our sample of 21% is one percentage point larger than the mean of 20%.

$$z - \text{score} = \frac{21 - \text{mean}}{\text{standard deviation}} = \frac{21 - 20}{0.13} = \frac{1}{0.13} = 7.7$$

This score is far larger than any z-score in Table 6.2. There is almost no chance that in a randomly chosen sample of this size, 21% will suffer from this allergy. Thus, this is a truly anomalous sample: This town is not representative of the total population of Americans. Its allergy rate is highly unusual.

Chapter 6 Statistics: **Chapter Summary**

- ▶ **Data summary and presentation: Boiling down**
 - ▶ Four important measures in descriptive statistics:
mean, median, mode, and standard deviation
- ▶ **The normal distribution: Why the bell curve?**
 - ▶ A plot of normally distributed data: the *bell-shape* curve.
 - ▶ The z-score for a data point
 - ▶ The Central Limit Theorem

Chapter 6 Statistics: **Chapter Summary**

- ▶ **The statistics of polling:** Can we believe the polls?
 - ▶ Polling involves: a margin of error, a confidence level, and a confidence interval.
- ▶ **Statistical inference and clinical trials:** Effective drugs?
 - ▶ Statistical significance and p -values.
 - ▶ Positive correlated, negative correlated, uncorrelated or linearly correlated