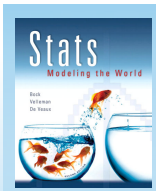


## Chapter 6

### The Standard Deviation as a Ruler and the Normal Model



PEARSON Addison Wesley  
Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley

### The Standard Deviation as a Ruler

- The trick in comparing very different-looking values is to use standard deviations as our rulers.
- The standard deviation tells us how the whole collection of values varies, so it's a natural ruler for comparing an individual to a group.
- As the most common measure of variation, the standard deviation plays a crucial role in how we look at data.

Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley  
**Slide 6-3**

### Standardizing with z-scores

- We compare individual data values to their mean, relative to their standard deviation using the following formula:

$$z = \frac{(y - \bar{y})}{s}$$

- We call the resulting values **standardized values**, denoted as **z**. They can also be called **z-scores**.

Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley  
**Slide 6-4**

### Standardizing with z-scores (cont.)

- Standardized values have no units.
- z-scores measure the distance of each data value from the mean in standard deviations.
- A negative z-score tells us that the data value is *below* the mean, while a positive z-score tells us that the data value is *above* the mean.

Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley  
**Slide 6-5**

### Benefits of Standardizing

- Standardized values have been converted from their original units to the standard statistical unit of *standard deviations from the mean*.
- Thus, we can compare values that are measured on different scales, with different units, or from different populations.

Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley  
**Slide 6-6**

## Shifting Data

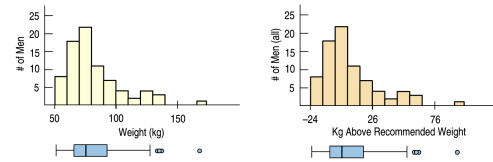
- Shifting data:
  - Adding (or subtracting) a *constant* amount to each value just adds (or subtracts) the same constant to (from) the mean. This is true for the median and other measures of position too.
  - In general, adding a constant to every data value adds the same constant to measures of center and percentiles, but leaves measures of spread unchanged.

Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley

Slide 6-7

## Shifting Data (cont.)

- The following histograms show a **shift** from men's actual weights to kilograms above recommended weight:



Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley

Slide 6-8

## Rescaling Data

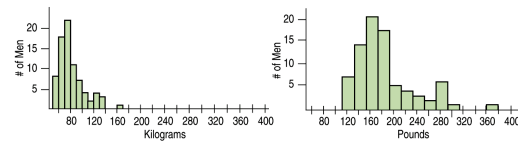
- Rescaling data:
  - When we divide or multiply all the data values by any constant value, both measures of location (e.g., mean and median) and measures of spread (e.g., range, IQR, standard deviation) are divided and multiplied by the same value.

Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley

Slide 6-9

## Rescaling Data (cont.)

- The men's weight data set measured weights in kilograms. If we want to think about these weights in pounds, we would **rescale** the data:



Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley

Slide 6-10

## Back to z-scores

- Standardizing data into z-scores **shifts** the data by subtracting the mean and **rescales** the values by dividing by their standard deviation.
  - Standardizing into z-scores does not change the **shape** of the distribution.
  - Standardizing into z-scores changes the **center** by making the mean 0.
  - Standardizing into z-scores changes the **spread** by making the standard deviation 1.

Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley

Slide 6-11

## When Is a z-score Big?

- A z-score gives us an indication of how unusual a value is because it tells us how far it is from the mean.
- Remember that a negative z-score tells us that the data value is *below* the mean, while a positive z-score tells us that the data value is *above* the mean.
- The larger a z-score is (negative or positive), the more unusual it is.

Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley

Slide 6-12

### When Is a z-score Big? (cont.)

- There is no universal standard for z-scores, but there is a model that shows up over and over in Statistics.
- This model is called the **Normal model** (You may have heard of “bell-shaped curves.”).
- Normal models are appropriate for distributions whose shapes are unimodal and roughly symmetric.
- These distributions provide a measure of how extreme a z-score is.

Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley

Slide 6- 13

### When Is a z-score Big? (cont.)

- There is a Normal model for every possible combination of mean and standard deviation.
  - We write  $N(\mu, \sigma)$  to represent a Normal model with a mean of  $\mu$  and a standard deviation of  $\sigma$ .
- We use Greek letters because *this* mean and standard deviation do not come from data—they are numbers (called **parameters**) that specify the model.

Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley

Slide 6- 14

### When Is a z-score Big? (cont.)

- Summaries of data, like the sample mean and standard deviation, are written with Latin letters. Such summaries of data are called **statistics**.
- When we standardize Normal data, we still call the standardized value a **z-score**, and we write

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

Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley

Slide 6- 15

### When Is a z-score Big? (cont.)

- Once we have standardized, we need only one model:
  - The  $N(0, 1)$  model is called the **standard Normal model** (or the **standard Normal distribution**).
- Be careful—don't use a Normal model for just any data set, since standardizing does not change the shape of the distribution.

Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley

Slide 6- 16

### When Is a z-score Big? (cont.)

- When we use the Normal model, we are assuming the distribution is Normal.
- We cannot check this assumption in practice, so we check the following condition:
  - **Nearly Normal Condition:** The shape of the data's distribution is unimodal and symmetric.
  - This condition can be checked with a histogram or a Normal probability plot (to be explained later).

Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley

Slide 6- 17

### The 68-95-99.7 Rule

- Normal models give us an idea of how extreme a value is by telling us how likely it is to find one that far from the mean.
- We can find these numbers precisely, but until then we will use a simple rule that tells us a lot about the Normal model...

Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley

Slide 6- 18

### The 68-95-99.7 Rule (cont.)

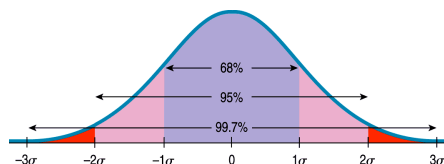
- It turns out that in a Normal model:
  - about 68% of the values fall within one standard deviation of the mean;
  - about 95% of the values fall within two standard deviations of the mean; and,
  - about 99.7% (almost all!) of the values fall within three standard deviations of the mean.

Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley

Slide 6- 19

### The 68-95-99.7 Rule (cont.)

- The following shows what the 68-95-99.7 Rule tells us:



Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley

Slide 6- 20

### The First Three Rules for Working with Normal Models

- Make a picture.
- Make a picture.
- Make a picture.
- And, when we have data, make a histogram to check the Nearly Normal Condition to make sure we can use the Normal model to model the distribution.

Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley

Slide 6- 21

### Finding Normal Percentiles by Hand

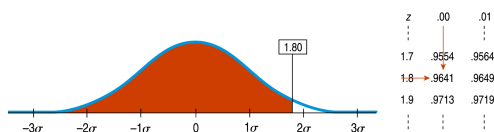
- When a data value doesn't fall exactly 1, 2, or 3 standard deviations from the mean, we can look it up in a table of **Normal percentiles**.
- Table Z in Appendix E provides us with normal percentiles, but many calculators and statistics computer packages provide these as well.

Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley

Slide 6- 22

### Finding Normal Percentiles by Hand (cont.)

- Table Z is the *standard Normal* table. We have to convert our data to z-scores before using the table.
- Figure 6.5 shows us how to find the area to the left when we have a z-score of 1.80:



Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley

Slide 6- 23

### Finding Normal Percentiles Using Technology

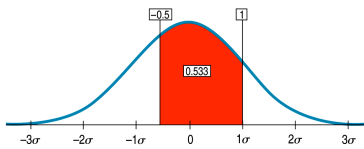
- Many calculators and statistics programs have the ability to find normal percentiles for us.
- The *ActivStats* Multimedia Assistant offers two methods for finding normal percentiles:
  - The "Normal Model Tool" makes it easy to see how areas under parts of the Normal model correspond to particular cut points.
  - There is also a Normal table in which the picture of the normal model is interactive.

Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley

Slide 6- 24

### Finding Normal Percentiles Using Technology (cont.)

The following was produced with the "Normal Model Tool" in *ActivStats*:

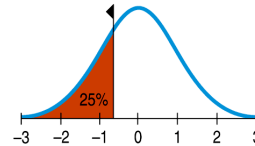


Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley

Slide 6-25

### From Percentiles to Scores: z in Reverse

- Sometimes we start with areas and need to find the corresponding z-score or even the original data value.
- Example: What z-score represents the first quartile in a Normal model?



Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley

Slide 6-26

### From Percentiles to Scores: z in Reverse (cont.)

- Look in Table Z for an area of 0.2500.
- The exact area is not there, but 0.2514 is pretty close.



- This figure is associated with  $z = -0.67$ , so the first quartile is 0.67 standard deviations below the mean.

Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley

Slide 6-27

### Are You Normal? How Can You Tell?

- When you actually have your own data, you must *check* to see whether a Normal model is reasonable.
- Looking at a histogram of the data is a good way to check that the underlying distribution is roughly unimodal and symmetric.

Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley

Slide 6-28

### Are You Normal? How Can You Tell? (cont.)

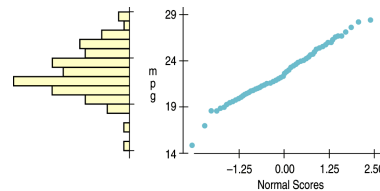
- A more specialized graphical display that can help you decide whether a Normal model is appropriate is the **Normal probability plot**.
- If the distribution of the data is roughly Normal, the Normal probability plot approximates a diagonal straight line. Deviations from a straight line indicate that the distribution is not Normal.

Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley

Slide 6-29

### Are You Normal? How Can You Tell? (cont.)

- Nearly Normal data have a histogram and a Normal probability plot that look somewhat like this example:

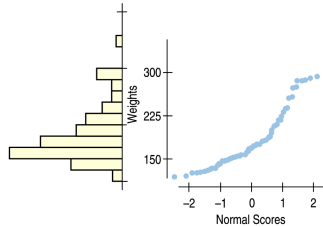


Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley

Slide 6-30

### Are You Normal? How Can You Tell? (cont.)

- A skewed distribution might have a histogram and Normal probability plot like this:

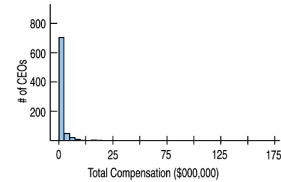


Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley

Slide 6- 31

### What Can Go Wrong?

- Don't use a Normal model when the distribution is not unimodal and symmetric.



Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley

Slide 6- 32

### What Can Go Wrong? (cont.)

- Don't use the mean and standard deviation when outliers are present—the mean and standard deviation can both be distorted by outliers.
- Don't round off too soon.
- Don't round your results in the middle of a calculation.
- Don't worry about minor differences in results.

Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley

Slide 6- 33

### What have we learned?

- The story data can tell may be easier to understand after shifting or rescaling the data.
  - Shifting data by adding or subtracting the same amount from each value affects measures of center and position but not measures of spread.
  - Rescaling data by multiplying or dividing every value by a constant changes all the summary statistics—center, position, and spread.

Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley

Slide 6- 34

### What have we learned? (cont.)

- We've learned the power of standardizing data.
  - Standardizing uses the SD as a ruler to measure distance from the mean (z-scores).
  - With z-scores, we can compare values from different distributions or values based on different units.
  - z-scores can identify unusual or surprising values among data.

Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley

Slide 6- 35

### What have we learned? (cont.)

- A Normal model sometimes provides a useful way to understand data.
  - We need to *Think* about whether a method will work—check the Nearly Normal Condition with a histogram or Normal probability plot.
  - Normal models follow the 68-95-99.7 Rule, and we can use technology or tables for a more detailed analysis.

Copyright © 2007 Pearson Education, Inc. Publishing as Pearson Addison-Wesley

Slide 6- 36