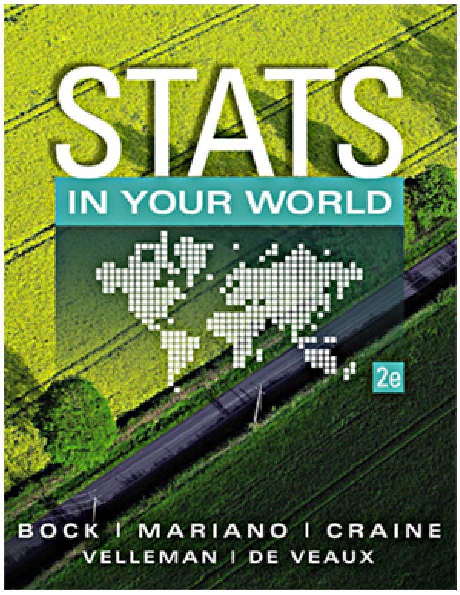
Statistics

Section 1

Chapter 5:

What's Normal?



Nov 15-9:18 AM

The Standard Deviation

Recall: The standard deviation, s, is just the square root of the variance and is measured in the same units as the original data.

$$s = \sqrt{\frac{\sum(y-\bar{y})^2}{n-1}}$$

Calculate the standard deviation of the following set:

10 13 14 19 11 16 20 22 19

$\bar{y} = 16$

$s = 3\sqrt{2} \approx 4.24$

$z_{14} \approx -.47$

$z_{20} \approx .94$

$z_{16} = 0$

$z_{22} \approx 1.41$

Oct 11-10:13 AM

The Standard Deviation as a Ruler

- COMPARING DIFFERENT VALUES:

  The trick in comparing very different-looking values (often in different variables) is to use standard deviations.

- COMPARING A VALUE TO THE GROUP:

  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 (particularly for symmetric data), the standard deviation plays a crucial role in how we look at data.

Nov 15-9:18 AM

Standardizing with *z*-scores

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

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

** s - Standard deviation*
y - individual value
$\bar{y}$ - mean of the data set

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

Recall: Measures of center are shifted and scaled. Measures of spread are scaled ONLY.

Nov 15-9:18 AM

Standardizing with z-scores

- Standardized values have no units. *Why*?

dividing by s cancels out unit

- z-scores measure the distance of each data value from the mean in standard deviations.
  - A negative z-score: value is below the mean
  - A positive z-score: value is above the mean.
- EX: a z-score of 2 says that a data value is 2 standard deviations above the mean.
- Ex: a z-score of –1.6 says that a data value is 1.6 standard deviations below the mean.

Nov 15-9:18 AM

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.

Long Jump Stem	Leaf	200-m Race Stem	Leaf
66	3	23	233
65	3	23	699
64	0578	24	22333334
63	368	24	5566677999
62	1	25	000234444
61	11235668	25	599
60	24689	26	1
59	267778		
58	8		
66	3=6.63 meters	23	2=23.2 seconds

6.21 26.1

Find the z-score of each red and green data value.

	LJ	200
Red	2.25	+.52
Green	1	+2.07

Nov 15-9:18 AM

Benefits of Standardizing (partner activity)

- How might college admissions departments use this method of comparison? What do you think they could compare this way?

1. Brainstorm answers to this question. – paper
2. Discuss your brainstorm. – with partner verbally
3. Write your thoughts. – on paper

Nov 18-9:26 AM

Shifting Data (RECALL...)

- Shifting data:
- Adding (or subtracting) a constant amount to each value has the same effect on the mean, median and all other measures of position.

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.

Nov 15-9:18 AM

z-scores (shift, scale, shape, center, spread)

Standardizing data into z-scores...

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

ALWAYS LEARNING PEARSON Chapter 5, Slide 9

Nov 15-9:18 AM

When is a z-score "BIG"?

- A z-score gives us an indication of how unusual a value is because it tells us relatively how far it is from the mean.
- A data value that sits right at the mean, has a z-score equal to 0.
- A z-score of 1 or –1 tells us that the data value is 1 standard deviation away from the mean.

ALWAYS LEARNING PEARSON Chapter 5, Slide 10

Nov 15-9:18 AM

When Is a *z*-score **BIG**? (*JUDGEMENT CALL*)

- How far from 0 does a z-score have to be to be interesting or unusual?
- If |z| is larger then the more "unusual" it is.
- No universal standard
- Often we consider |z| > 2 to be roughly an indication of "unusualness".
- But every data set is different, so be careful!

Nov 15-9:18 AM

Finding z-scores

For a data set with a mean of 45 and a standard deviation of 3.5 ...

a) Find the z-score of the value 34.5 and explain the meaning of this value.

b) Find the original value of a data point with a z-score of 3.2

c) Which is more "unusual" for the data set, 11 or 76?

d) Which is the more "unusual" data value:

a value of 13 in a set with $\bar{x}$ = 34 and s = 6

OR

a value of 22 in a set with $\bar{x}$ = 92 and s = 24

a) -3 b) 56.2 c) 11

d) 13

Nov 15-9:18 AM

z-scores & the Normal model

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 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 (or data value) is.

Nov 15-9:18 AM

Models and parameters

A *statistical model* is a way to simplify and/or represent reality. It can be displayed as an equation or a curve.

Numbers used to create or describe the model that don't come directly from the data are called *parameters*. They are numbers that we choose to help specify the model.

Nov 15-9:18 AM

Normal models

There is a Normal model for every possible combination of mean and standard deviation. The normal model looks like a "bell curve".

We write N(μ,σ) to represent a Normal model with a mean of μ and a standard deviation of σ.

We use Greek letters because this mean and standard deviation are not numerical summaries of the data. They are part of the model (parameters).

N(44,7)

 Chapter 5, Slide 14

Nov 15-9:18 AM

Calculating z-scores

- 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

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

Standard Normal

N(0,1)

 Chapter 5, Slide 15

Nov 15-9:18 AM

The Normal Model

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.

 Chapter 5, Slide 16

Nov 15-9:18 AM

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.

 Chapter 5, Slide 17

Nov 15-9:18 AM

Create histograms of each set of data.

Would applying a normal model be appropriate?

Could you apply the same normal model to each set? Support your answer.

*Find the z-score of 11 in each data set.

set 1	set 2
12	7
11	10
17	11
19	15
21	15
22	16
23	20
25	21
27	21
28	22
31	20
31	28
35	29
39	28
9	29
44	28
20	34
27	33
30	34
23	38
28	39
17	44

 Chapter 5, Slide 17

Nov 15-9:18 AM

The 68-95-99.7 Rule

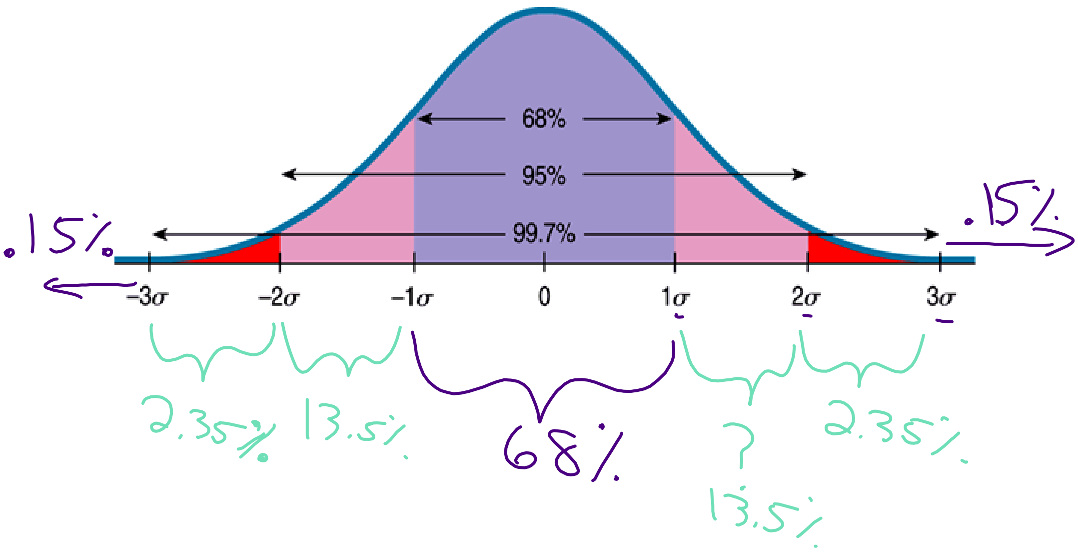
- Normal models give us an idea of how extreme a value is by telling us how likely it is to find one that far from the mean.
- We can find these numbers precisely, but it is useful to begin with a simple rule that tells us a lot about the Normal model…
- It turns out that in a Normal model:
  - about 68.3% (or 68%) of values fall within one standard deviation of mean
  - about 95% of values fall within two standard deviations of mean
  - about 99.7% of values (almost all) fall within three standard deviations of mean.

 Chapter 5, Slide 19

Nov 15-9:18 AM

The 68-95-99.7 Rule (cont.)

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

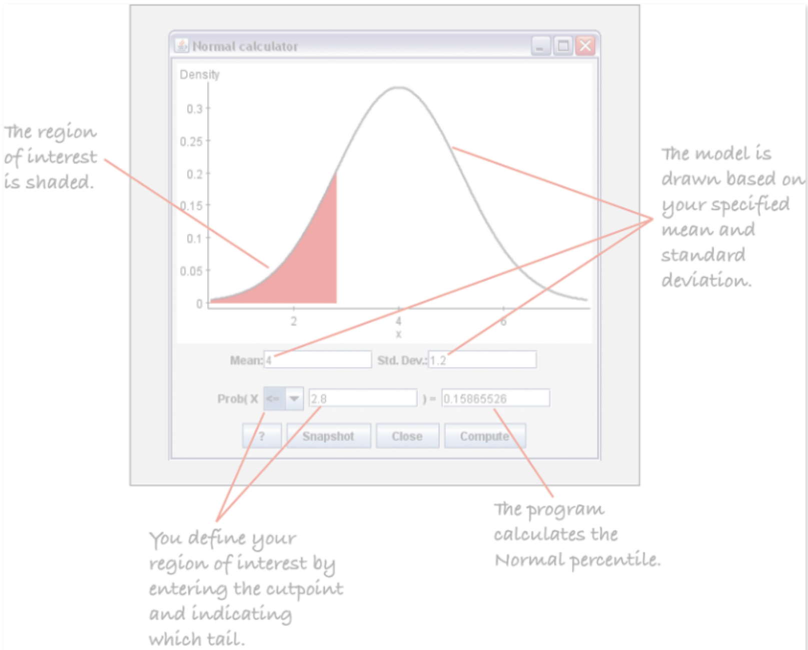


 Chapter 5, Slide 21

Nov 15-9:18 AM

Finding Normal Percentiles

- When a data value doesn't fall exactly 1, 2, or 3 standard deviations from the mean, we can look it up on StatCrunch. StatCrunch gives us a "calculator" to find these values.



 Chapter 5, Slide 22

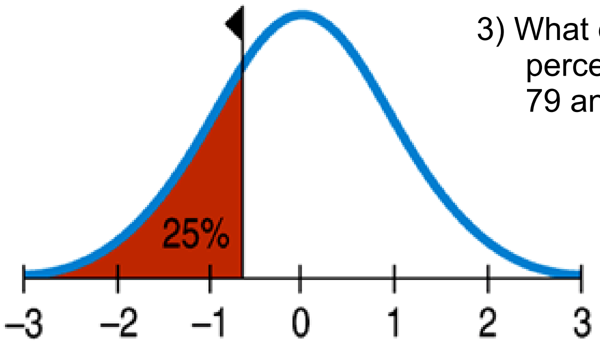
Nov 15-9:18 AM

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.

Examples: 1) What z-score represents the first quartile in a Normal model?

2) What is the data value of the third quartile in a set with a mean of 112 and a standard deviation of 1.54?

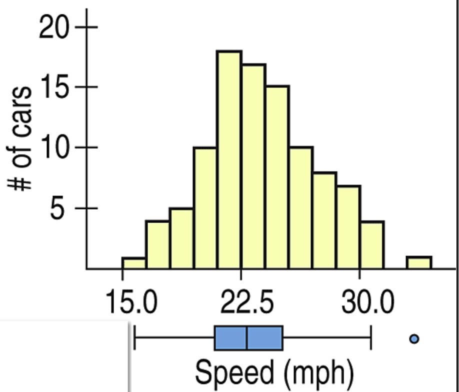


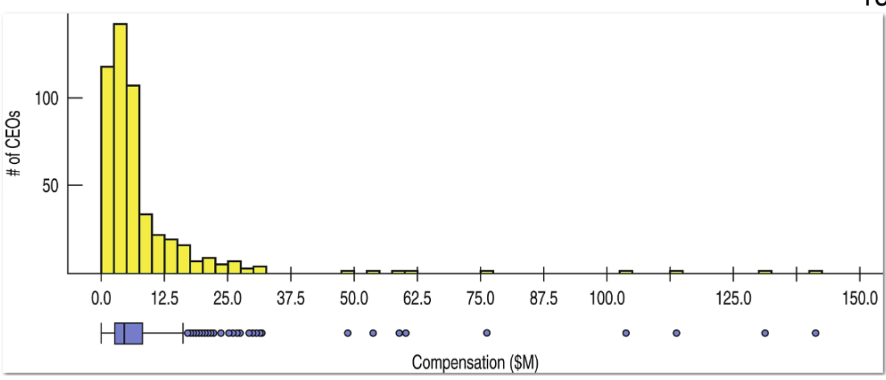
3) What data value represents the 95th percentile of a data set with a mean of 79 and a standard deviation of 6.79

Nov 15-9:18 AM

What Can Go Wrong?

- Don’t use a Normal model when the distribution is not unimodal and symmetric.
- Remember: symmetry is not necessarily influenced by a couple outliers.





Nov 15-9:18 AM

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 your results in the middle of a calculation.
- Don't worry about minor differences in results.

Nov 15-9:18 AM

What have we learned?

- *We've learned the power of standardizing data using standard deviation 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, even if the data does not follow a normal model.*
- *z-scores can identify unusual or surprising values among data.*
- *We've learned that the 68-95-99.7 Rule can be a useful rule of thumb for understanding distributions:*
  - *For unimodal and symmetric data, about 68% fall within 1 SD of mean, 95% fall within 2 SDs of mean, and 99.7% fall within 3 SDs of mean.*

Nov 15-9:18 AM