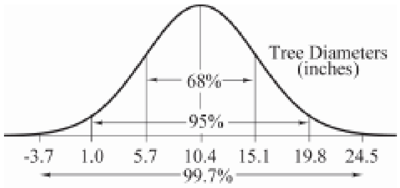
# Day 4 Homework Answers:

30. Trees.

a) The Normal model for the distribution of tree diameters is at the right.




b) Approximately 95% of the trees are expected to have diameters between 1.0 inch and 19.8 inches.

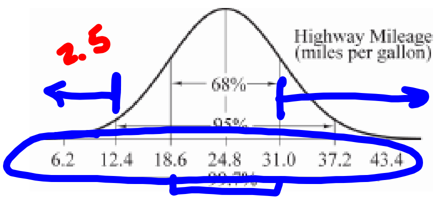
c) Approximately 2.5% of the trees are expected to have diameters less than an inch.

d) Approximately 34% of the trees are expected to have diameters between 5.7 inches and 10.4 inches.

e) Approximately 16% of the trees are expected to have diameters over 15 inches.

31. Guzzlers?

a) The Normal model for auto fuel economy is at the right.




b) Approximately 68% of the cars are expected to have highway fuel economy between 18.6 mpg and 31.0 mpg.

c) Approximately 16% of the cars are expected to have highway fuel economy above 31 mpg.

d) Approximately 13.5% of the cars are expected to have highway fuel economy between 31 mpg and 37 mpg.

e) The worst 2.5% of cars are expected to have fuel economy below approximately 12.4 mpg.

32. Trees, part II.

The use of the Normal model requires a distribution that is unimodal and symmetric. The distribution of tree diameters is neither unimodal nor symmetric, so use of the Normal model is not appropriate.

33. Car speeds, the picture.

The distribution of cars speeds shown in the histogram is unimodal and roughly symmetric, and the normal probability plot looks quite straight., so a normal model is appropriate.

## The Standard Deviation as a Ruler

- The trick in comparing very different-looking values is to use standard deviations.
- 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.

 Chapter 5, Slide 7

## 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.
  - A *z*-score of 2 says that a data value is 2 standard deviations above the mean.
  - A *z*-score of –1.6 means a data value is 1.6 standard deviations below the mean.

## 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.

## *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.

## 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.
- A data value that sits right at the mean, has a *z*-score equal to 0.
- A *z*-score of 1 means the data value is 1 standard deviation above the mean.
- A *z*-score of –1 means the data value is 1 standard deviation below the mean.

 Chapter 5, Slide 14

## When Is a *z*-score **BIG**?

- How far from 0 does a *z*-score have to be to be interesting or unusual?
- There is no universal standard, but the larger a *z*-score is (negative or positive), the more unusual it is.
- Often we consider a z-score greater than 2 or less than -2 to be *roughly* an indication of unusualness. But every data set is different, so be careful!

 Chapter 5, Slide 15

## Keeping your signs straight

- 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.
- Pro-tip: if you forget which order to subtract, use common sense! Above or below? Positive or negative?

 Chapter 5, Slide 16

## z-scores

- It is hard to underestimate the power and usefulness of standardized scores.
- You will calculate them in this course a countless number of times. Especially in the final unit.

 Chapter 5, Slide 11

Examples from textbook pg. 116

pg. 33

1. Standardizing Skiing Times

The men's super combined skiing event debuted in the 2010 Winter Olympics in Vancouver. It consists of two races: a downhill and a slalom. Times for the two events are added together, and the skier with the lowest total time wins. At Vancouver, the mean slalom time was 52.67 seconds with a standard deviation of 1.614 seconds. The mean downhill time was 116.26 seconds with a standard deviation of 1.914 seconds. Bode Miller of the U.S., who won the gold medal with a combined time of 164.92 seconds, skied the slalom in 51.01 seconds and the downhill in 113.91 seconds. **On which race did he do better compared to the competition?**

$$z_{slalom} = \frac{51.01 - 52.67}{1.614} = -1.03$$

$$z_{downhill} = \frac{113.91 - 116.26}{1.914} = -1.23$$

His downhill time if even more below the mean than his slalom time. Therefore, his downhill time was more remarkable.

2. Your statistics teacher has announced that the lower of your two test scores will be dropped. You got a 90 on test 1 and an 80 on test 2. You're all set to drop the 80 until she announces that she grades "on a curve." She standardized the scores in order to decide which is the lower one. If the mean on the first test was 88 with a standard deviation of 4 and the mean on the second test was 75 with a standard deviation of 5. Which one will be dropped?

$$z_1 = \frac{90-88}{4} = \frac{2}{3} = .5$$

$$z_2 = \frac{80-75}{5} = \frac{5}{5} = 1$$

Test 1 will be dropped because it's standardized score (z-score) is closer to the mean.

3. The calorie content for 23 variations of Kellogg's cereals averages 109 calories per serving with a standard deviation of 22.2 calories. The mean fiber content for these cereals is 2.7 grams per serving with a standard deviation of 3.2 grams. A service of Kellogg's All-Bran with Extra Fiber has a very low 50 calories and a very high 14 grams of fiber. Which is more remarkable - the calorie content or the fiber content? Explain.

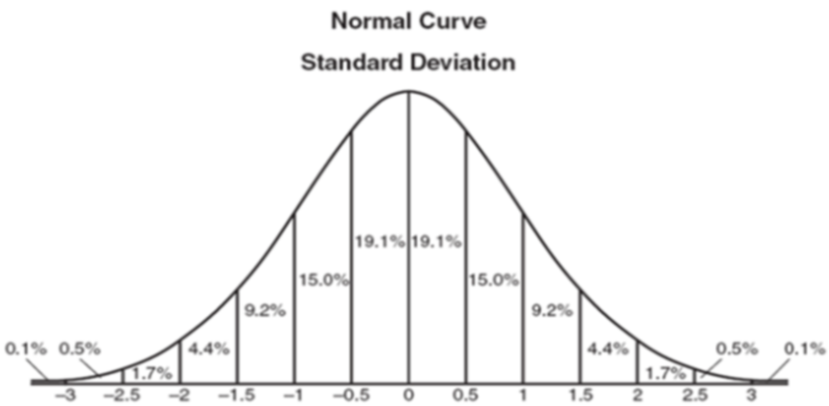
$$z_{Fiber} = \frac{14 - 2.7}{3.2} = 3.53$$

$$z_{Cal} = \frac{50 - 109}{22.2} = -2.66$$

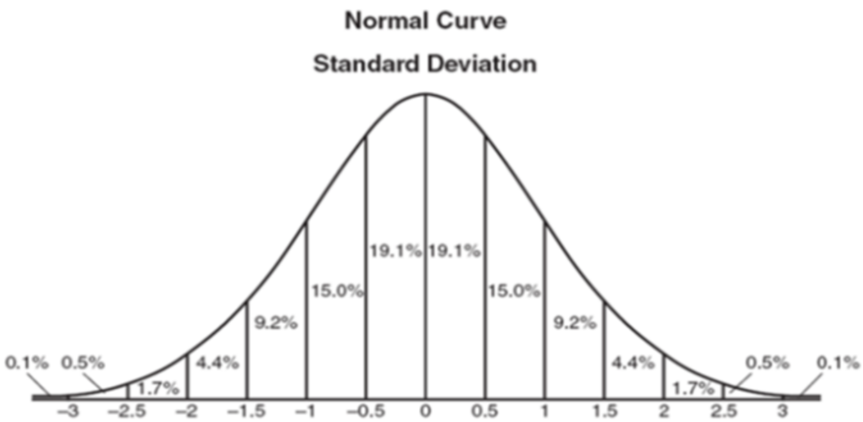
The fiber content is more remarkable because it is furthest from the mean.

Homework: Pg. 131 #14, 15, 16, 17

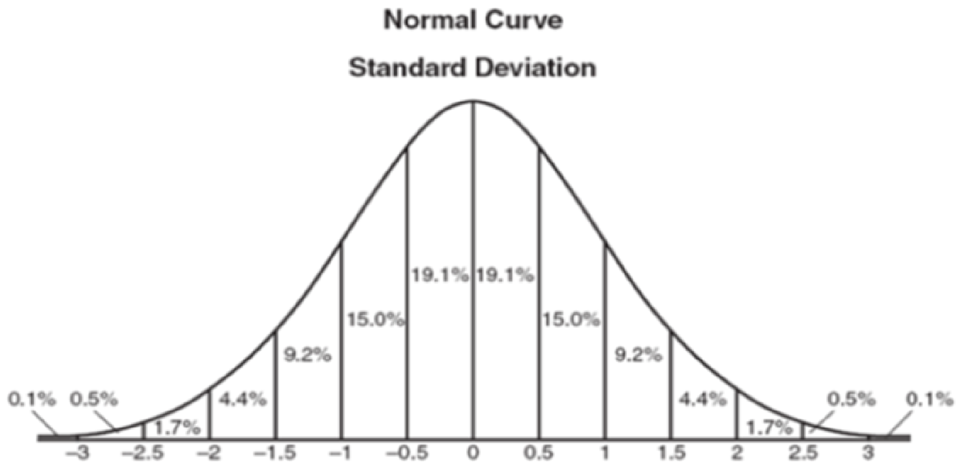
14)



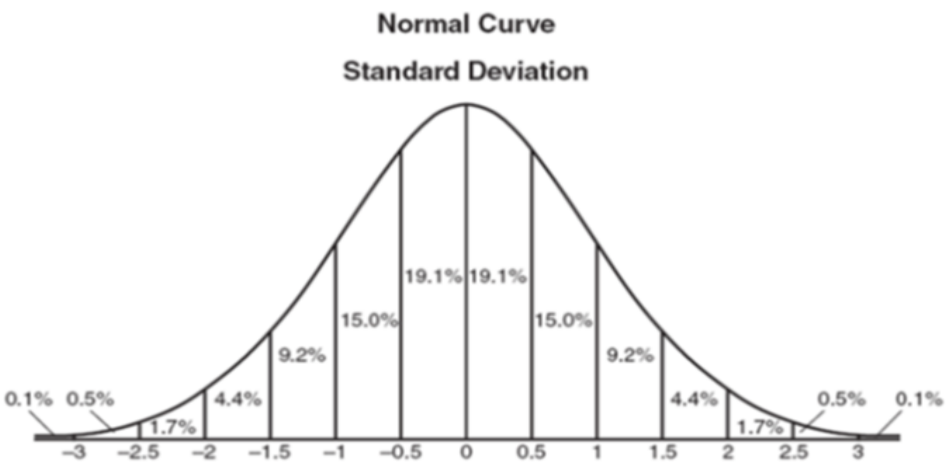
15)


Normal Curve
Standard Deviation
0.1%
0.5%
1.7%
4.4%
9.2%
15.0%
19.1%
19.1%
15.0%
9.2%
4.4%
1.7%
0.5%
0.1%
−3
−2.5
−2
−1.5
−1
−0.5
0
0.5
1
1.5
2
2.5
3

16)


Normal Curve
Standard Deviation
0.1%
0.5%
1.7%
4.4%
9.2%
15.0%
19.1%
19.1%
15.0%
9.2%
4.4%
1.7%
0.5%
0.1%
−3
−2.5
−2
−1.5
−1
−0.5
0
0.5
1
1.5
2
2.5
3

17)


Normal Curve
Standard Deviation
0.1%
0.5%
1.7%
4.4%
9.2%
15.0%
19.1%
19.1%
15.0%
9.2%
4.4%
1.7%
0.5%
0.1%
–3
–2.5
–2
–1.5
–1
–0.5
0
0.5
1
1.5
2
2.5
3