

# Stat 101L: Lecture 5

**Measure of Center**

◆ Sample mean

$$\bar{y} = \frac{\text{Total}}{n} = \frac{(\sum y_i)}{n}$$

1

---

---

---

---

---

---

---

---

**Sample Mean**

◆  $Total = 8669$

◆  $n = 24$

$$\bar{y} = \frac{\text{Total}}{n} = \frac{8669}{24} = 361.2$$

2

---

---

---

---

---

---

---

---

**Mean or Median?**

- ◆ The sample mean is the balance point of the distribution.
- ◆ The sample median divides the distribution into a lower and an upper half.
- ◆ For skewed data, the mean is pulled in the direction of the skew.

3

---

---

---

---

---

---

---

---

# Stat 101L: Lecture 5

## Numerical Summaries

- ◆ How much variation is there in the data?
- ◆ Look for the spread of the distribution.
- ◆ What do we mean by “spread”?

4

---

---

---

---

---

---

---

---

## Measures of Spread

- ◆ Sample Range
  - The distance from the minimum and the maximum.  
Range =  $(378 - 349) = 29$  grams
  - The length of the interval that contains 100% of the data.
  - Greatly affected outliers.

5

---

---

---

---

---

---

---

---

## Quartiles

- ◆ Medians of the lower and upper halves of the data.
- ◆ Trying to split the data into fourths, quarters.

6

---

---

---

---

---

---

---

---



# Stat 101L: Lecture 5

## Box Plots

- ◆ Establish an axis with a scale.
- ◆ Draw a box that extends from the lower to the upper quartile.
- ◆ Draw a line from the lower quartile to the minimum and another line from the upper quartile to the maximum.

10

---

---

---

---

---

---

---

---

## Outlier Box Plots

- ◆ Establishes boundaries on what are “usual” values based on the width of the box.
- ◆ Values outside the boundaries are flagged as potential outliers.

11

---

---

---

---

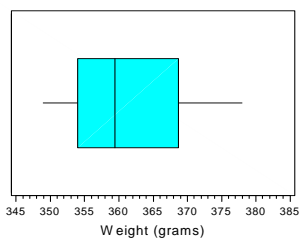
---

---

---

---

## Contents of Cans of Cola



12

---

---

---

---

---

---

---

---

# Stat 101L: Lecture 5

## Measures of Spread

- ◆ Based on the deviation from the sample mean.
- ◆ Deviation

$$(y - \bar{y})$$

13

---

---

---

---

---

---

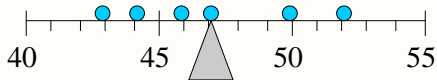
---

---

## 9-hole Golf Scores

46, 44, 50, 43, 47, 52

$$\bar{y} = \frac{282}{6} = 47 \text{ strokes}$$



14

---

---

---

---

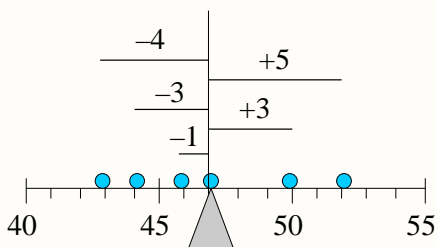
---

---

---

---

## Deviations



15

---

---

---

---

---

---

---

---

# Stat 101L: Lecture 5

## Sample Variance

Almost the average squared deviation

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

16

---

---

---

---

---

---

---

---

## Sample Variance

$$s^2 = \frac{(16 + 9 + 1 + 25 + 9)}{5} = \frac{60}{5}$$
$$= 12 \text{ strokes}^2$$

17

---

---

---

---

---

---

---

---

## Sample Standard Deviation

$$s = \sqrt{s^2} = \sqrt{\frac{(\sum (y - \bar{y})^2)}{n - 1}}$$
$$s = \sqrt{12} = 3.46 \text{ strokes}$$

18

---

---

---

---

---

---

---

---

# Stat 101L: Lecture 5

Which summary is better?

- ◆ For symmetric distributions use the sample mean,  $\bar{y}$ , and sample standard deviation,  $s$ .
- ◆ For skewed distributions use the five number summary.

19

---

---

---

---

---

---

---

---

Why?

- ◆ For symmetric distributions the sample mean and sample median should be approximately equal so either would work.
- ◆ We will see in Chapter 6 why the sample standard deviation is best for symmetric distributions.

20

---

---

---

---

---

---

---

---

Why?

- ◆ For skewed distributions, the sample mean and standard deviation will be affected by the skew and/or potential outliers. The five number summary displays the skew and is not affected by outliers.

21

---

---

---

---

---

---

---

---