Sec. 13.3 - Measures of Dispersion

HW #3, 7, 11, 15, 21, 23,
29-33 all, 46*, 47*, 52, 53

* Remind me to talk about these two on Monday!

We can start with a couple of homework questions.

Oct 2-8:12 AM

First, wrap up discussion of weighted means.

The U.S. National Center for Health Statistics compiles data on the length of stay by patients in short-term hospitals and publishes its findings in Vital and Health Statistics. A random sample of 21 patients yielded the following data on length of stays, in days. Find the average length of stay.

# Days (x)	Frequency	$x \cdot f$
1	3	3
2	4	8
3	6	18
4	3	12
5	2	10
6	2	12
9	1	9
	21	72 days

$$\bar{x} = \frac{\sum_{i=1}^{7}(x_i \cdot f_i)}{\sum_{i=1}^{7}(f_i)} = \frac{72}{21} = 3.4$$

$\underbrace{1,1,1}_{3}, \underbrace{2,2,2,2}_{8}, 3 \ldots, 9$

Oct 1-9:00 AM

A small high school held a fundraiser to raise money to buy mosquito nets in Africa, which can help prevent the spread of malaria.

(www.nothingbutnets.net has sold 2,181,693 nets to date @ $10 / net. About 3,000 people die *each day* from malaria. The nets reduce infection rates by as much as 90% in high transmission areas.)

The high school has:

The freshmen raised an average of $80 / person. ↔ 20% Freshmen
The sophomores raised an average of $84 / person. ↔ 23% Sophomores
The juniors raised an average of $113 / person. ↔ 30% Juniors
The seniors raised an average of $98 / person. ↔ 27% Seniors

The teachers raised an average of $95 / person.

Who raised more money on average? The students, or the teachers?

$$\bar{x} = \frac{80+84+113+98}{4} = \$93.75$$ → assumes all classes are the same size.

$$\bar{x}_w = \frac{80(.2)+84(.23)+113(.30)+98(.27)}{100\%}$$

$$\boxed{\frac{\sum_{i=1}^{4}(x_i \cdot f_i)}{\sum_{i=1}^{4}(f_i)}}$$

$= \$95.68$ → students win!

Oct 1-12:25 PM

Sec. 13.3 - Measures of Dispersion (spread)
HW #3, 7, 11, 15, 21, 23,
29-33 all, 46*, 47*, 52, 53

A testing lab wishes to test two experimental hybrid cards to determine the gas mileage. The results in (mpg) are given below for each type of car.

Car A: 10, 20, 30, 40, 50, 60 → $\bar{x}_A = \frac{210}{6} = 35$ mpg
Car B: 25, 30, 35, 35, 40, 45 → $\bar{x}_B = \frac{210}{6} = 35$ mpg

Which car would be a better choice? Why?

range: $R_A = 60-10 = 50$
$R_B = 45-25 = 20$

Oct 2-8:40 AM

Measures of Spread

- The **range, R,** is the maximum number minus the minimum number.

  R = max - min

- The **standard deviation, s,** measures the spread of the data about the mean of the data set. It is often defined as the "average" amount by which scores in a distribution differ from the mean, ignoring the sign of the difference.

$$s = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1}}$$

Find the standard deviation of the data sets for Car A and Car B.

Car A: 10, 20, 30, 40, 50, 60
Car B: 25, 30, 35, 35, 40, 45

$$s = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1}}$$

Data Point, x	Deviation: $x-\bar{x}$	Square of deviation: $\left(x-\bar{x}\right)^2$
10	$10-35=-25$	$(-25)^2=625$
20	$20-35=-15$	$(-15)^2=225$
30	-5	$(-5)^2=25$
40	$+5$	25
50	$+15$	225
60	25	625
		1750

$\bar{x}_A = 35$

$$s = \sqrt{\frac{1750}{5}} = \sqrt{350} \approx 18.7 \text{ mpg}$$

Car A: 10, 20, 30, 40, 50, 60
Car B: 25, 30, 35, 35, 40, 45

Car B: $s_B = \sqrt{\frac{\sum_{i=1}^{6}(x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{(-10)^2+(-5)^2+0+0+5^2+10^2}{5}}$

$$= \sqrt{\frac{250}{5}} = \sqrt{50} \approx 7.1 \text{ mpg}$$

The **variance** is the square of the standard deviation, s^2. It is often calculated using the *computing formula:*

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right)$$

do not memorize

Find the mean and standard deviation of the data set, given that:

$$\sum_{i=1}^{32} x_i = 2{,}578 \qquad \sum_{i=1}^{32} x_i^2 = 212{,}152$$

$\bar{x} = \frac{2578}{32} \approx 80.56$

$s^2 = \frac{1}{31}\left(212{,}152 - \frac{1}{32}(2578)^2\right)$

$= \frac{1}{31}(212{,}152 - 207690.1)$

$\approx \frac{1}{31}(4461.9) \approx 143.9$ (variance)

So $s \approx \sqrt{143.9} \approx 12.0$

Oct 2-8:40 AM

Interpreting the Standard Deviation

Chebyshev's Theorem

For any set of numbers, regardless of the shape of the distribution,

The fraction of any data set lying within k (k > 1) standard deviations of the mean is at least

$$1 - \frac{1}{k^2}$$

If k = 1:

$1 - \frac{1}{1^2} = 0$

If k = 2:

$1 - \frac{1}{2^2} = 75\%$

So at least 75% of the data is within 2 st.dev of $\bar{x}$.

If k = 3:

$1 - \frac{1}{3^2} = .\bar{8} \approx 88.9\%$

Oct 2-8:40 AM

Data collection experiment:

A common way to measure reaction time is to have someone drop a ruler or yard stick and for you to catch it as quickly as possible. The distance the yard stick dropped gives a rough measure of your reaction time.

Carry out this experiment in pairs. We'll record the data in a spreadsheet for future analysis.

1. Decide who will drop and who will catch.
2. Try it a couple times for practice.
3. When ready, agree that the next trial will be the "real deal" and report your drop distance to me.
4. Reverse roles and repeat.

Oct 2-8:40 AM

Reaction Time Data:

	A	B	C	D
1	Drop distance		Sum =	183.8
2	8	✓	n =	24
3	3	✓	Mean =	7.658333
4	7	✓	Std Dev =	3.515422
5	5	✓		
6	11.5	✓		
7	4.5	✓		
8	5.5	✓		
9	14	✓		
10	16.5			
11	6	✓		
12	3	✓		
13	10	✓		
14	5	✓		
15	5	✓		
16	7	✓		
17	5.5	✓		
18	6	✓		
19	9.5	✓		
20	13	✓		
21	7	✓		
22	10	✓		
23	4.5	✓		
24	7	✓		
25	10.3	✓		

Chebyshev's Theorem says that at least 75% of data must be within 2 st.dev. of the mean.

What fraction of *our* data is within 2 st.dev. of the mean?

$\bar{x} \approx 7.7$

$s \approx 3.5$

$7.7 - 2(3.5) = 0.7$

$7.7 + 2(3.5) = 14.7$

$\frac{23}{24} \approx 96\%$

Oct 2-9:11 AM

Example: Chebyshev's Theorem

The mean price of houses in a certain neighborhood is \$150,000, and the standard deviation is \$15,000.

Approximate the price range within which at least 89% of the houses are priced.

$$89\% \approx 1-\frac{1}{K^2}$$

$$\boxed{K=3}$$

lower: $\bar{x}-3(s)$

$= 150{,}000 - 3(15{,}000) = \$105{,}000$

upper: $\bar{x}+3(s)$

$= 150{,}000 + 3(15{,}000) = \$195{,}000$

Oct 2-8:40 AM

Summary:

The range is a rough measure of spread. (R = max - min)

(?) Is it a robust statistic?

The standard deviation (and variance) are two other ways to measure spread.

(?) Is st.dev. a robust statistic?

$$s=\sqrt{\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}{n-1}}$$

Standard deviation can be interpreted as a sort of "average" deviation from the mean.

(a) Chebyshev's Theorem and (b) the Emperical Rule are two different tools for interpreting the st.dev.

(?) Which is more accurate?

(?) When can you use them?

Sec. 13.3 - Measures of Dispersion

HW #3, 7, 11, 15, 21, 23,

29-33 all, 46*, 47*, 52, 53

Oct 2-9:24 AM