Measures of Spread + Distribution

Consider 2 wrestling Teams:

	Misfits	Bulk + Brute
average weight:	222.5 lbs	220lbs
average height:	6'1"	6'

Are these teams evenly matched?

To really examine a set of data, you must look at 3 things: centre, spread, distribution

Misfits		Bulk + Brute	
Mad Dog Max	5'2" 320lbs	Hulk Hogan	6' 220 lbs
String Bean Sam	7' 125lbs	Mr T	6' 220 lb

Measures of Spread (Dispersion)

- a way to state how the data is dispersed or spread out about the centre. The most common measure of spread is called the standard deviation.

more common $\Rightarrow$ S_x $\rightarrow$ standard deviation of a sample

σ_x $\rightarrow$ (sigma x) standard deviation of the population.
↑ little sigma

- standard deviation is the "average distance" from the mean of all data pieces

"big" sigma $\Rightarrow$ sum

$$S_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

x_i → a data piece; $\bar{x}$ → mean; n → sample size

$$S_x = \sqrt{\frac{(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + (x_3-\bar{x})^2 + \ldots + (x_n-\bar{x})^2}{n-1}}$$

Example: The amount of TV watched over the weekend (hrs).

5, 3, 10, 7, 9, 5, 8, 4, 12, 7

Find the mean ($\bar{x}$): $\bar{x} = \frac{5+3+10+7+9+5+8+4+12+7}{10}$

$n = 10$
$\bar{x} = 7$

$\bar{x} = \frac{70}{10}$

$\bar{x} = 7$

To calculate S_x:

x	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
5	-2	4
3	-4	16
10	3	9
7	0	0
9	2	4
5	-2	4
8	1	1
4	-3	9
12	5	25
7	0	0
TOTAL		72

$$S_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

$$S_x = \sqrt{\frac{72}{10-1}}$$

$$S_x = \sqrt{\frac{72}{9}} = 8 \text{ (variation)}$$

$S_x \doteq 2.83$

The standard deviation of a population is basically the same but we divide by n instead of $n-1$

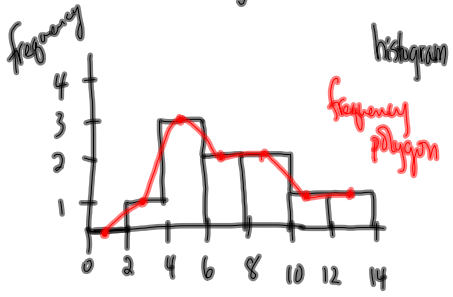
$$\sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

Data Distribution

- the shape of the data when graphed in a histogram

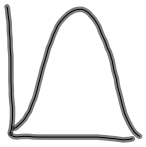
bin	hrs of TV	Frequency
→	0-2	0
	2-4	1
	4-6	3
	6-8	2
	8-10	2
	10-12	1
includes 12 →	12-14	1
		10

does not include

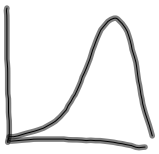


shows the distribution of the data

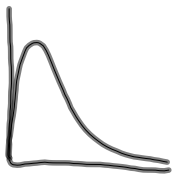
Frequency Polygons



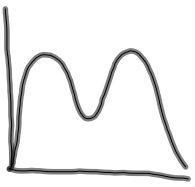
- normal distribution
- central peak.
- roughly symmetrical



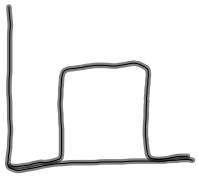
- left-skewed
- tail is to the left



- right-skewed
- tail is to the right



- bimodal distribution
- 2 distinct peaks



← uniform distribution.

Practice: p176/177 #1-4