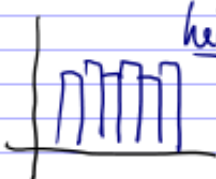


Panel 1

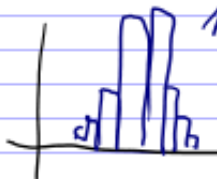
Last time:

- ~~Mean~~ ^{Median} and Mode for Frequency Distr / Histograms
 ↑
 numeric ord, numeric all

• Shapes of Distributions



heterog.



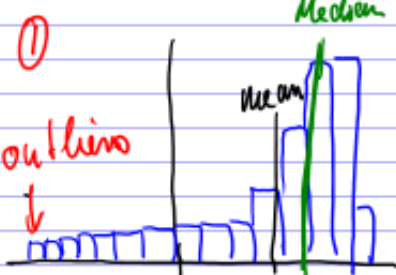
homog.

1, 2, 3

- Who is better, mean, median, or mode?

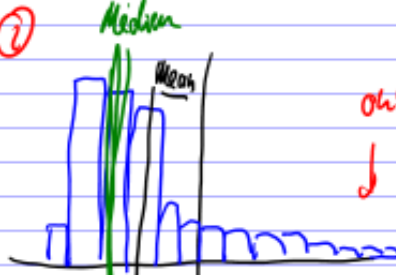
Panel 2

①



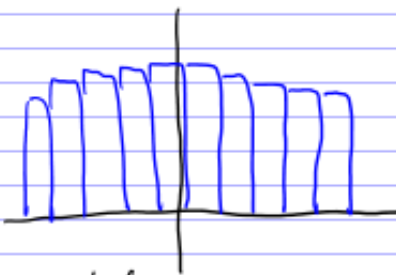
outliers ↓
skewed to left

②



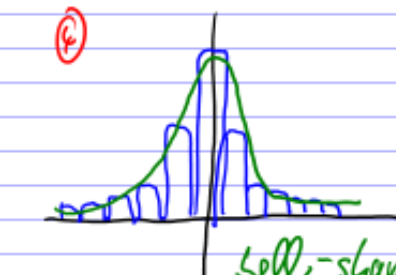
outliers ↓
skewed to right

③



heterog.

④



bell-shaped,
normal

Panel 3

Measures of Variability (Dispersion)

Want to buy a machine that makes nails, 10 inch ones.

Machine A: 8, 9, 10, 11, 12 $\bar{x} = 10$

Machine B: 7, 8, 10, 12, 13 $\bar{x} = 10$

Want to know how close these nails are together ✓
how close to the avg. they are.

Range: max - min

A: range = 4

B: range = 6

3

Panel 4

Variance: is the avg. distance of data to the mean.

$$\rightarrow \left(\frac{1}{n-1} \right) \sum (x - \bar{x})^2 \text{ is variance}$$

Two symbols $s^2 = \text{sample variance}$

$\sigma^2 = \text{pop. variance}$

(A)	x	x - \bar{x}	(x - \bar{x}) ²
	8	-2	4
	9	-1	1
	10	0	0
	11	1	1
	12	2	4
			<u>10</u>

$$\bar{x} = 10$$

$$s^2 = \frac{10}{4} = \underline{\underline{2.5}}$$

4

Panel 5

Data: 2, 3, 10, 17, 18

x	x - \bar{x}	(x - \bar{x}) ²
2	-8	64
3	-7	49
10	0	0
17	+7	49
18	8	64
		<u>226</u>

$\bar{x} = 10$

$\bar{x} = 10$

$s^2 = \frac{226}{4} = \underline{\underline{56.5}}$

Machine B is better, because its variance is smaller, so units are more uniform.

5

Panel 6

Variance is hard to compute: need short cut

$$s^2 = \frac{1}{n-1} \left[\sum x^2 - \frac{(\sum x)^2}{n} \right] = \frac{1}{n} \sum (x - \bar{x})^2$$

x	x ²
8	64
9	81
10	100
11	121
12	144
<u>50</u>	<u>510</u>

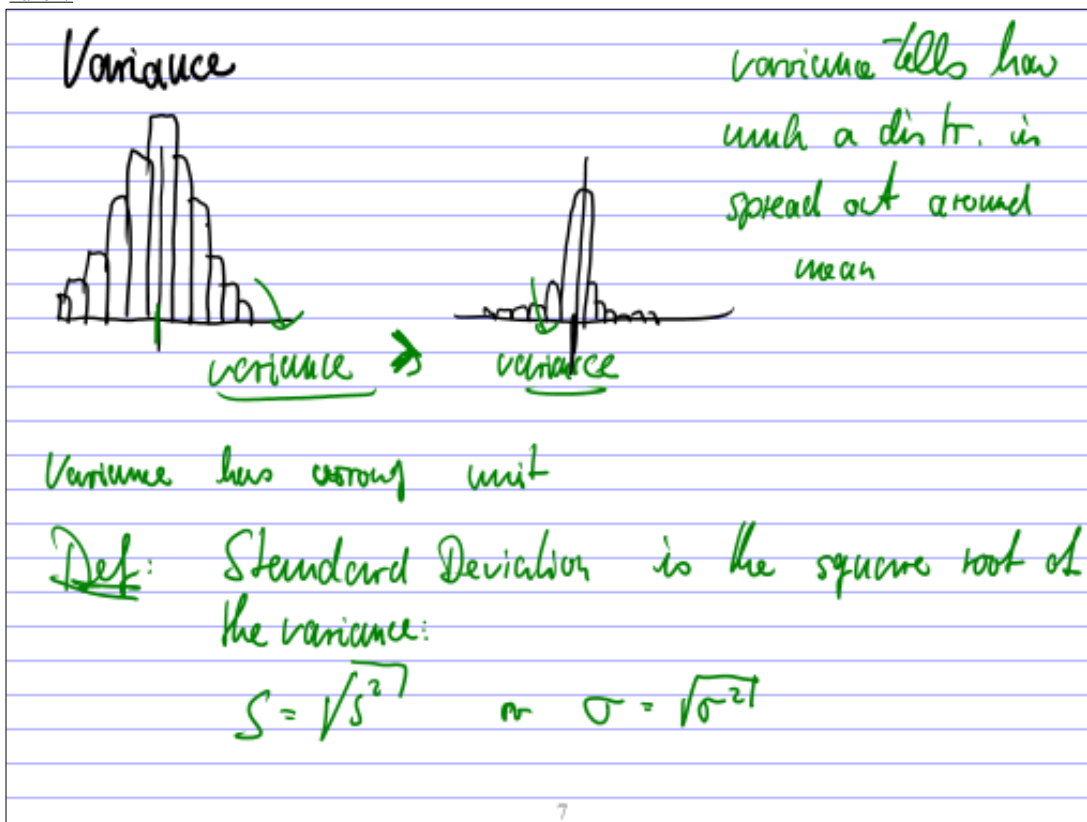
$s^2 = \frac{1}{4} \left(510 - \frac{(10)^2}{5} \right) =$

$\left(\bar{x} = \frac{50}{5} = 10 \right)$

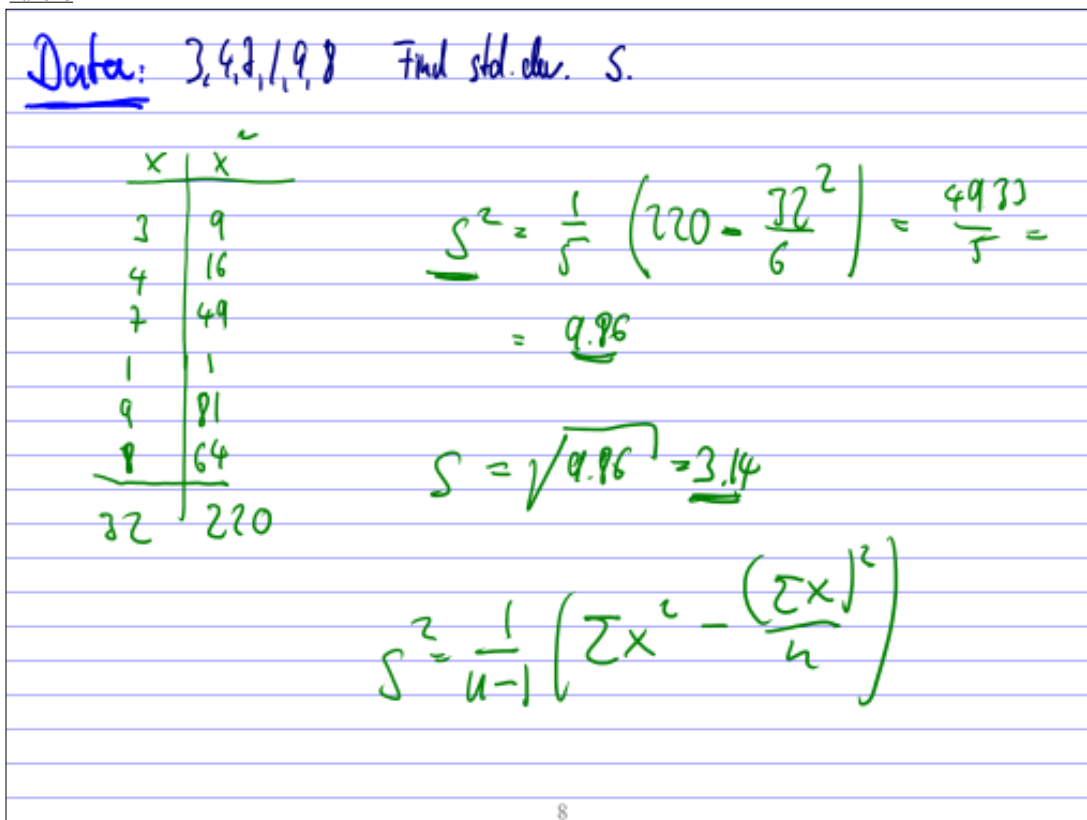
$\frac{1}{4} (510 - 100) = \frac{410}{4} = \underline{\underline{102.5}}$

6

Panel 7



Panel 8



Panel 9

More numeric descriptors: Quartiles

Lower Quartile Q_1 : that number s.t. 25% are less than it

Median Q_2 : # s.t. 50% are less

Upper Quartile Q_3 : number s.t. 75% are less than it

Ex: Data 1, 2, 3, 4, 5, 6, 7. Find quartiles

① Sort data

↓ median

② Split in half [1, 2, 3] & [5, 6, 7]

③ Split each half: [1] 2 [3] 4 [5] 6 [7]

↑
 Q_1

↑
 Q_3

9

Panel 10

How to find Quartiles

① Sort data, smallest to largest

② Find N (sample size)

③ Compute position $L_1 = 0.25 \cdot N$

→ if L_1 is whole #, pick between L_1 and next pos

→ else pick number after L_1 → Q_1

④ Compute position $L_3 = 0.75 \cdot N$

→ as above

→ Q_3

⑤ (the $L_2 = 0.5 \cdot N$ for median)

10

Panel 11

Ex: Data 1, 2, 3, 4, 5, 6 sorted, $N=6$

$$L_1 = 0.25 \cdot 6 = 1.5 \Rightarrow \text{pick \# at position } 2 \Rightarrow \underline{Q_1 = 2}$$

$$L_3 = 0.75 \cdot 6 = 4.5 \Rightarrow \text{pick \# at position } 5 \Rightarrow \underline{Q_3 = 5}$$

11

Panel 12

Data: 0, 87, 123, 253, 1, 103, 123, 265, 1, 112, 198

Sort: 0, 1, 87, 103, 112, 123, 123, 198, 253, 265

$$Q_1: L_1 = 0.25 \cdot 11 = 2.75$$

\Rightarrow look at position 3: $Q_1 = 1$

$$Q_3: L_3 = 0.75 \cdot 11 = 8.25$$

\rightarrow look at position 9: $Q_3 = 198$

(what if $L_1 = 3$ pick between 3rd and 4th: 44)

12