Panel 1

Last time:

- Mean and Mode for Frequency Distr / Histograms

Data: 1, 5, 9, 3, 3, 6, 8, 3, 5, 10

| | count | frequ. |
|---|---|---|
| 1-4 | 4 | 4/10 = 40% |
| 4-7 | 3 | 30% |
| 7-10 | 3 | 30% |
| | 10 | |

- Shapes of Distributions

normal, skewed left/right

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

1

Panel 2

| | count | frequ. | cum |
|---|---|---|---|
| 1-4 | 4 | 40% | 40 |
| 4-7 | 3 | 30% | 70 |
| 7-10 | 3 | 30% | 100 |

Data: 1, 5, 9, 3, 3, 6, 8, 3, 5, 10

Median: 1st category with cumulative % ≥ 50%

use 2 for 1st cat. => 4 × 2 = 8
5 for 2nd => 3 × 5 = 15
8 for 3rd cat => 3 × 8 = 24
47 => $\bar{x} = \frac{47}{10} = 4.7$

Mode: 3, given 1st category

Median: 5, given 2nd cat.

1, 3, 3, 3, 5, 5, 6, 8, 9, 10

Mean: 4, given 4.7 ≈ 5.3

Homogeneous or heterogeneous distribution?

2

Panel 3

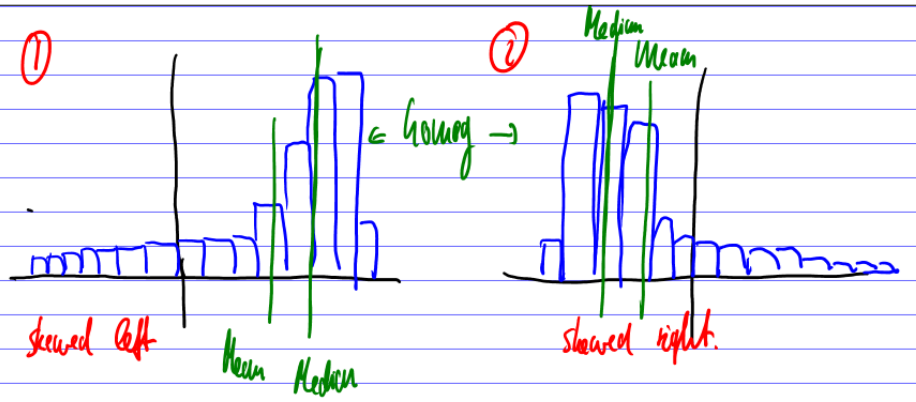
Homogeneous vs. Heterogeneous Distribution

homogeneous distribution is one where most of data is clustered around one value

heterogeneous dist. is one where all categories are equally likely, approx.

3

Panel 4




4

Panel 5

# Measures of Variability (Dispersion)

Buy a machine that makes 10 inch nails.

A: ~~8~~, 9, 10, 11, ~~12~~ (2, 1) $\bar{x} = 10$

B: 2, 3, 10, 17, 18 $\bar{x} = 10$ ← bad machine

I opt machine A, since those nails have less variation around the mean

① The Range: Range = Max - Min, e.g. A: 12 - 8 = 4 (16)

B: 18 - 2 = 16

5

Panel 6

Variance: "avg. distance of all data to the mean"

Variance: $\frac{1}{n-1} \sum (x - \bar{x})^2$

2 Below: $\sigma^2$ = pop. variance

$s^2$ = sample variance

Data: 8, 9, 10, 11, 12

$\bar{x} = 10$

Machine A

| $x$ | $x - \bar{x}$ | $(x - \bar{x})^2$ |
|---|---|---|
| 8 | -2 | 4 |
| 9 | -1 | 1 |
| 10 | 0 | 0 |
| 11 | 1 | 1 |
| 12 | 2 | 4 |
| | | 10 |

$\bar{x} = \frac{10}{4} = 2.5$

6

Panel 7

Data: 2, 3, 10, 17, 18

$\bar{x} = 10$

| $x$ | $x-\bar{x}$ | $(x-\bar{x})^2$ |
|---|---|---|
| 2 | −8 | 64 |
| 3 | −7 | 49 |
| 10 | 0 | 0 |
| 17 | 7 | 49 |
| 18 | 8 | 64 |
| | | 226 |

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

Machine B has larger variance (56.5) over machine A (2.5). Thus, machine A is more uniform.

7

Panel 8

Bad news about variance: Hard to compute Shortcut

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

A:

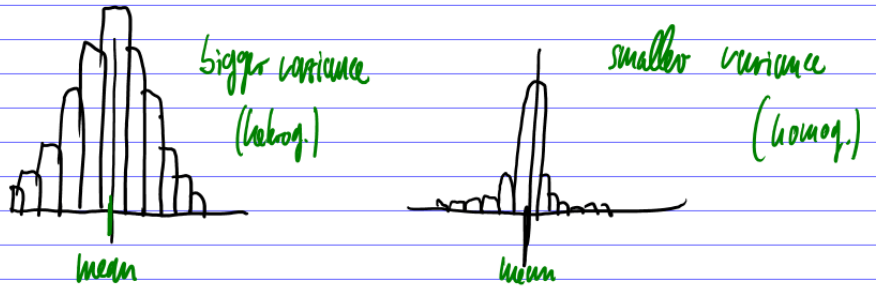
| $x$ | $x^2$ |
|---|---|
| 8 | 64 |
| 9 | 81 |
| 10 | 100 |
| 11 | 121 |
| 12 | 144 |
| 50 | 510 |

$s^2 = \frac{1}{4}\left(510 - \frac{(50)^2}{5}\right) = 2.5$

8

Panel 9

Variance shows how much the data is spread about mean.




Variance has the wrong unit! So:

Standard Deviation: $s = \sqrt{s^2}$

i.e. $\sqrt{\text{variance}} = \sqrt{2.5}$

$\sigma = \sqrt{\sigma^2}$

(hit square-root button on calculator)

9