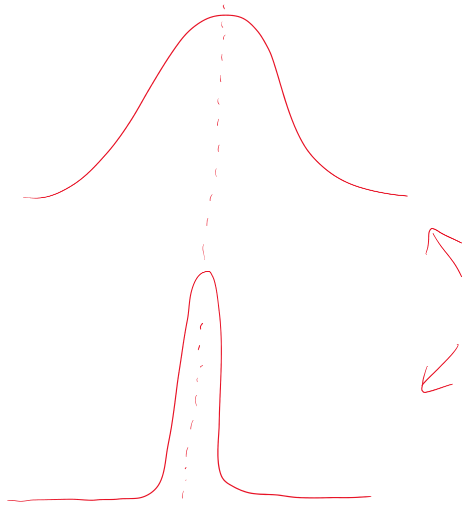
Section 6.2: Measures of Variability / spread

Thursday, March 14, 2024 1:22 PM



two distributions with the same mean

measure of spread - indication of how "wide" or "spread out" a data set is

when do you want a small spread?

- when trying for uniformity
  example: manufacturing identical objects

when do you want a large spread?

- when you are trying to make distinctions
  high quality versus low quality
  rankings

---

range - difference between the maximum and minimum values

difference is 54 - 3 = 51

example: 3, 7, 15, 42, 54

range is 51

↑
single number

good part : easy to calculate

bad part : almost completely useless

→ heavily influenced by outliers

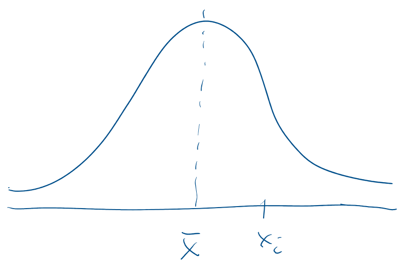
→ only depends on the values of two data points out of the entire set

---

the annoying measures to calculate

- variance
- standard deviation ⇐ most commonly used

a little bit of background (will not be tested)



consider some point in the data set : x_i

- how far from the mean is x_i?

$(x_i - \bar{x})$

if we add up all these values as is, the positive values will cancel the negative values and we'll end up with zero but if we square $(x_i - \bar{x})$ and take the sum

$$\sum (x_i - \bar{x})^2$$

then this result is a measure of how far away from the mean the data points are

population variance:

population mean

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

population size

Greek letter sigma (lower case)

population standard deviation

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

the sample population variance has a similar (but not identical) calculation and is written s^2

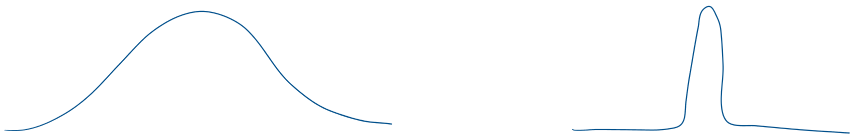
$$s = \sqrt{s^2}$$

---

things I would like you to know:

- the standard deviation (std dev) is a measure of how "wide" or spread out a data set is

- it measures, on average, how far away from the mean the data points are



larger std dev | smaller std dev

	mean	std dev
population	μ	σ
sample	$\bar{x}$	s

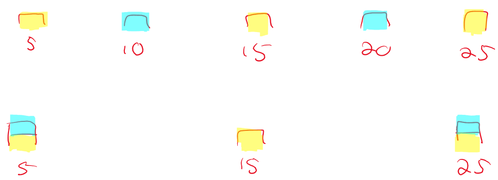
example: for the following pairs of data sets, which one has the higher std dev? or are they the same?

Set A: 5, 10, 15, 20, 25
Set B: 5, 5, 15, 25, 25

answer:

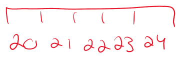
A:

B:



answer: B because the data points shaded in blue are further from the middle

Set A: 20, 21, 22, 23, 24
Set B: 30, 31, 32, 33, 34

set A : 

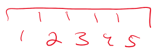
same shape but shifted

set B : 

A and B have <u>same</u> std dev

set A : 1, 2, 3, 4, 5

set B : 2, 4, 6, 8, 10

set A: 

set B: 

answer: B has higher std dev because points are further apart

(in fact, it's exactly twice as much)

Section 6.2: cont'd 2024/03/18

example: The Gizmo store has to raise its prices due to a platinum shortage. Every device in the store has a different price.

a) If every device has its price raised by \$5, what happens to the mean, median, range, and std dev of the prices of the devices? Be as specific as you can!

mean :
median :
} increase by \$5

range :
std dev : } same

b) Only one device needs to have its price increased. The most expensive device has its price increased by \$25. What happens to the mean, median, and range?

mean : increases

median : stay same* (* unless there are two or fewer devices)

range : increase by \$25

c) All devices increase in price by 10%. What happens to the mean, median, range, and std dev? Be as specific as you can!

all increase by 10%