Section 6.2: Measures of Spread / Variability

Thursday, November 02, 2023 5:15 PM

two distributions with the same mean

measures of spread - indication of how "wide" or how "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 vs. low quality
  - rankings

---

range - difference between the maximum and minimum values

example: 3, 7, 13, 42, 54

range: 51
single number

good part: easy to calculate

good part: easy to calculate

bad part: almost completely useless

- heavily influenced by outliers

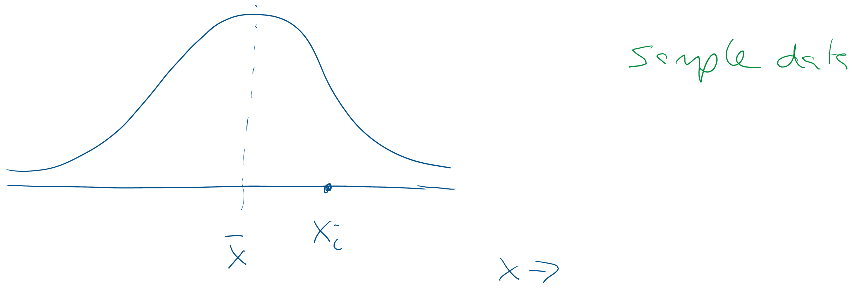
- only depends on the values of two data points out of the entire set

---

the annoying measures to calculate:

- variance
- standard deviation ← commonly used

a little bit of background on how you calculate these:



consider some point x_i in this data set

- how far from the mean is x_i?

$(x_i - \bar{x})$

if we add up all of these 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

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

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

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

population variance:

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

Greek letter sigma (lower case)

where μ = population mean
N = population size

population standard deviation:

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

sample variance:

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

$\bar{x}$ = sample mean
n = sample size

sample standard deviation:

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

---

things I would like you to know:

- the standard deviation (std dev) is

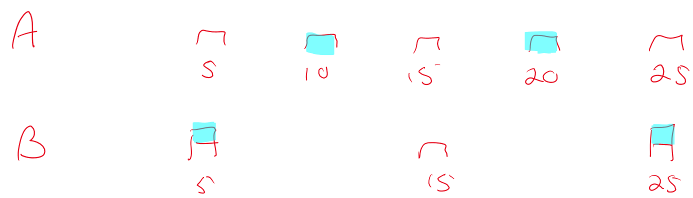
- the standard deviation (std dev) is a measure of how "wide" or "scattered" or "spread out" a distribution / data set is

- it measures how far, on average, each data point is from the mean

examples: for the following pairs of data sets which one has the higher standard deviation? or are they the same?

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

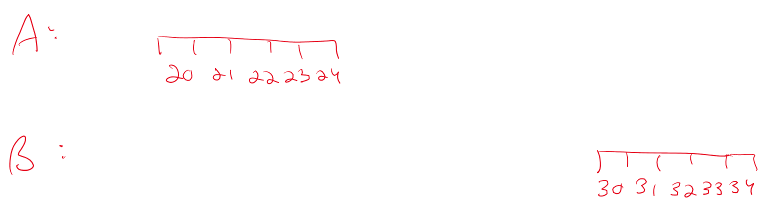
answer:



B because the data points shaded in blue are farther from the middle

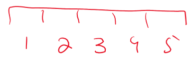
b) Set A: 20, 21, 22, 23, 24
   Set B: 30, 31, 32, 33, 34

answer:



A and B have same standard deviation

c) Set A: 1, 2, 3, 4, 5
   Set B: 2, 4, 6, 8, 10

answer: A : 

B : 

B has a higher std dev

– in fact, it's exactly twice as high

Section 6.2: cont'd 2023/11/07

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

a) If every device has its price increased by \$5, what happens to the mean, median, range, and standard deviation? Be as specific as you can!

answer: mean: increases by \$5
median: " "
range: stays same
std dev: " "

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

mean: increases (by how much will depend on how many devices you have – if you want to

devices you have
- if you want to be specific, it increases by $25/n$ where n is the number of devices)

median: stays the same *

(* unless there are two or fewer devices)

range: increases (by $25)

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

all increase by 10%

example: Consider the following data set: 7, 7, 7, 7, 7, and 7. Calculate the mean, median, range, and std dev.

mean and median: 7
range and std dev: 0