Section 1.4: cont'd

Thursday, January 11, 2018 12:25 PM

easiest measure of variability to calculate

range - the different between the max and minimum values

good part - easy to calculate

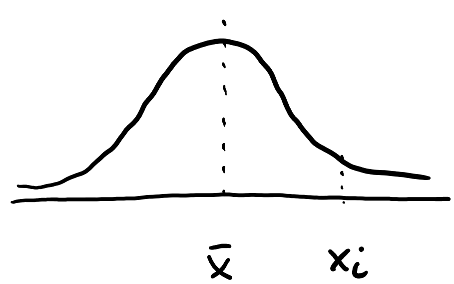
bad part - almost completely useless

→ heavily influenced by outliers

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

the annoying measures to calculate:

Variance:



sample data

consider some data point x_i on the above distribution

how far is x_i away from the mean? $(x_i - \bar{x})$

note: if you sum $\sum (x_i - \bar{x})$, you get zero

but if you sum $\sum (x_i - \bar{x})^2$, so all terms are non-negative,

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

population variance:

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

Greek letter "sigma" (lowercase)

N = size of population
μ = population mean

population standard deviation

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

---

sample variance

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

n = sample size

sample standard deviation:

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

---

note: the units of σ/s are the same as for $\mu/\bar{x}$

so if μ is the average of some lengths measured in metres, then σ is also in metres

a common convention (at least in physics), is to round σ/s to one sig fig, then round $\mu/\bar{x}$ to the same precision

calculator says $\mu = 58.593287\ldots$
$\sigma = 0.71285\ldots$

acceptable to say $\mu = 58.6$
$\sigma = 0.7$