Section 1: cont'd

Monday, March 13, 2017 12:56 PM

For large data sets, repeated data values can be quite common. In this situation, it can be more convenient to list the frequency or relative frequency of each measurement.

(5)

temperature (°C)	frequency
22	11
23	6
25	3
	20 ← number of points

sample

mean: $\bar{x} = \dfrac{22 \cdot 11 + 23 \cdot 6 + 25 \cdot 3}{20}$

$= \dfrac{455}{20} = 22.75\ ^\circ C$

median: $n = 20$ position $= \dfrac{n+1}{2} = 10.5$

average of 10th and 11th points

if you were to write at the data set, the 10th and 11th points would both be 22, so

median = 22 °C

(6)

mass (g)	relative frequency

84	0.1
85	0.85
86	0.05

] sum is 0.95 > 0.5, so 85 will contain median

mean: $\bar{x} = \dfrac{84(0.1) + 85(0.85) + 86(0.05)}{1}$

$= 84.95$ g

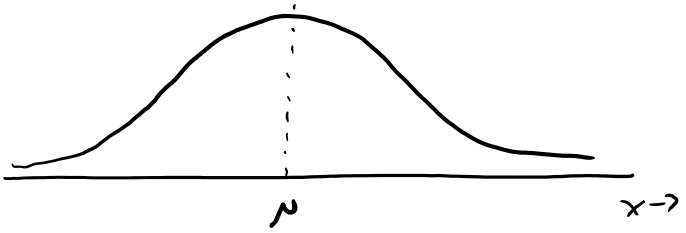
median: with the measurements ordered, to find the middle you add the relative frequencies until 0.5 is first reached or exceeded

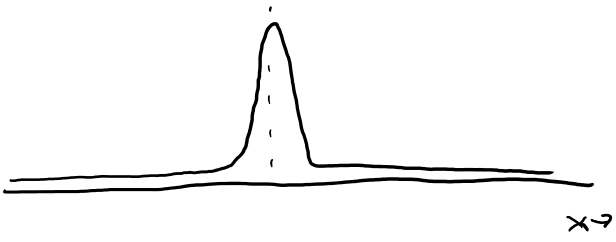
median = 85 g

---

measures of spread:

two different data sets can have the same mean but still be very different





x-7

note: When do you want a small spread?

When trying to manufacture identical objects

when do you want a large spread?

when you are trying to make distinctions

variance / standard deviation

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

definition:

population variance

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

σ

Greek letter sigma

standard deviation

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

example: for a population, calculate σ

x	$x-\mu$	$(x-\mu)^2$
2	-4	16
5	-1	1
8	2	4
9	3	9

$$\mu = \frac{2+5+8+9}{4} = 6$$

$$\sigma^2 = \frac{\Sigma (x_i - \mu)^2}{n} = \frac{16+1+4+9}{4} = 7.5$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{7.5} \approx 2.7$$

definition: for a sample of measurements

sample variance:

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

sample std dev

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

---

note: the bigger the variance/SD, the more spread out the data is

the unit for SD is the same as for the measurements

treat the data set as a population unless it is specified as a sample

(7) Which sample is more spread out?

a) 1, 4, 10

b) 31, 36, 38

$\bar{x} = \frac{1+4+10}{3} = 5$

$\bar{x} = \frac{31+36+38}{3} = 35$

a)

x	$x - \bar{x}$	$(x-\bar{x})^2$
1	-4	16
4	-1	1
10	5	25

$s^2 = \frac{16+1+25}{3-1} = 21$

b)

x	$x - \bar{x}$	$(x-\bar{x})^2$
31	-4	16
36	1	1
38	3	9

so (b) is less spread out

$s^2 = \frac{16+1+9}{3-1} = 13$

(8)

	Machine 1	Machine 2
$\bar{x}$	355.8	355.2
s	0.3	1.4

$\Rightarrow$ 355 mL cans of pop

a) Machine 2 is more accurate since 355.2 is closer to the target of 355

b) Machine 1 is more precise since less

variability