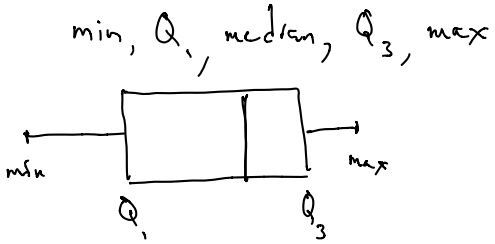
Stat 145, Fri 12-Feb-2021 -- Fri 12-Feb-2021
Biostatistics
Spring 2021

Friday, February 12th 2021

Wk 2, Fr
Topic:: Measures of spread
Topic:: Correlation
Read:: Lock5 2.3, 2.5

Start/join an Etherpad at link
https://pad.disroot.org/p/s145-12feb2021-gXX (Feb. 12, not Feb. 11)
with same groups as yesterday (I'll change them around, but not daily)
XX = 01 for Latvaitis, Morren, Aardema
= 02 for Toldy, Bultje, Katje
= 03 for Triezenberg, Pastoor, Lemon
= 04 for Steen, Ching, Tanis
= 05 for Nedd, Rai, Brink
= 06 for Ochiagha, Anderson, Cheek
= 07 for Arthur, Stob, Sytsema
= 08 for Johnson, Opalewski, Haveman
= 09 for Rudy, Krikke
= 10 for Wolf, Schneider, Wakeman

Range, interquartile range (IQR), standard deviation
- said to be measures of "spread" (or "variation")
- valid only for a quantitative variable
- what they are



IQR = $Q_3 - Q_1$ fairly resistant

seeing in a box plot

range = max − min (using 2 nos. in 5-number summary)

sensitive to outliers

seeing in a box plot
1.5 x IQR criterion for outliers

Quantitative Data

Spread: range, IQR, s

have mean $\bar{x}$ (computed from $x_1, x_2, \ldots, x_n$)

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

$x_i - \bar{x}$ amt. x_i deviates from $\bar{x}$

$(x_i - \bar{x})^2$ is squared deviation

s standard deviation = $\sqrt{\text{variance}}$

sensitive to outliers

- commands for computing in R

Q1: Using as few commands as possible, find the mean, sd, range, and IQR for the eruptions of Old Faithful (data frame called "faithful").

- resistance to outliers?

- standard deviation
  has units like the data
  often used, itself, as a unit of measurement:
  How many std devs from the mean are you?
  Z-score

IQ mean = 100, sd = 15

you: 120

$Z = \frac{120 - 100}{15} = \frac{20}{15} = 1.33$

$Z = \frac{(\text{observed number}) - \text{mean}}{\text{std dev.}}$

Q2:

: Who performed better?
Millie with score of 1410 on the SAT (mean = 1026, sd = 209), or
Michal with score of 27 on the ACT (mean = 20.8, sd = 4.8), or

example calculating it: use small data set (9, 21, 13, 29) $n = 4$

Must get $\bar{x} = \frac{1}{4}(9+21+13+29) = 18$

obs.	deviations from $\bar{x}$	squared deviations
9	$9-18 = -9$	$(-9)^2 = 81$
21	$21-18 = 3$	9
13	$13-18 = -5$	25
29	$29-18 = 11$	121
		236 ← SSD

$\sum (x_i - \bar{x})^2 = 236$

$\text{var} = \frac{1}{n-1}\sum (x_i - \bar{x})^2 = \frac{1}{3}(236) = 78.\overline{6}$

$s = \sqrt{\text{var.}} = \sqrt{78.\overline{6}} = 8.869$