L12: Monday, Feb. 13

Housekeeping.

- Week 1 summary + discussion forum post due 11:59 p.m. today

- Optional homework to be posted on Canvas tonight — sol'ns will be posted on Wednesday

- Optional review session Tuesday 5-6 p.m. B205

- Exam 1 on Friday!

  Wednesday's class will be a review session

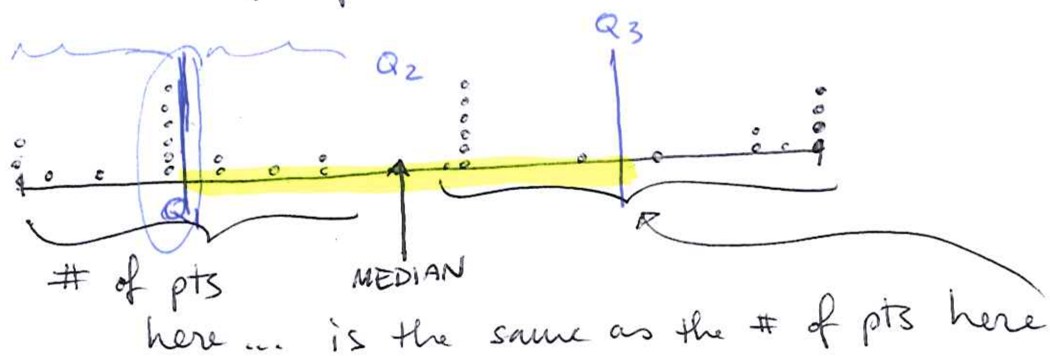
QUESTIONS?

Last time ...

- Mean, median, mode
- Quartiles, IQR, box & whisker plots
- Standard deviation
- z-scores

This time ... Review of the above topics

In the same way that finding the <u>median</u> splits the data into <u>two</u> parts, each with an equal # of points:



The <u>QUARTILES</u> split the data into <u>four</u> parts, each with an equal # of points:

<u>Reminder</u>: quartiles are <u>numbers</u> — numerical values — not ranges or intervals.

To find quartiles:

(1) Find the median of the data; that's Q_2.

(a) If # of pts (call it m) is odd, the median is the $\left(\frac{m+1}{2}\right)^{th}$ point in the sorted list.

(b) If m is even, then median = mean of $\left(\frac{m}{2}\right)^{th}$ & $\left(\frac{m}{2}+1\right)^{th}$ pts

(2) Take all points less than the median, and find the median of those; that's Q1.

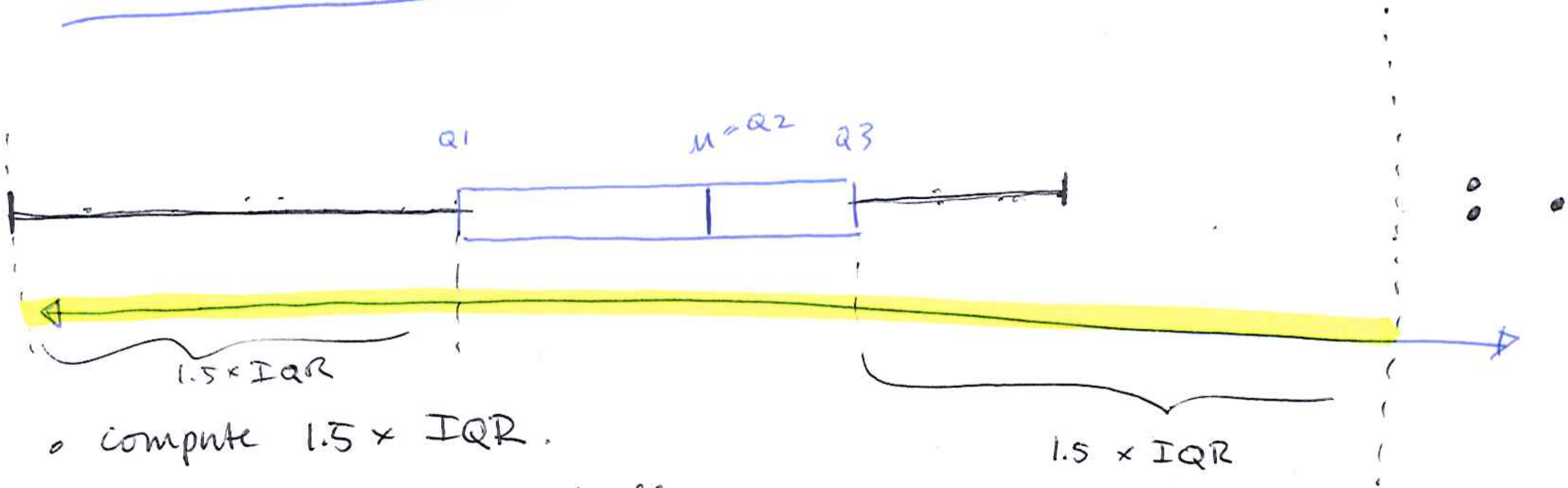
(3) Take all points greater than median, find the median of those; that's Q3.

The Interquartile Range (or IQR) is also a <u>number</u>: IQR = Q3 − Q1

<u>Percentiles</u> (similar to quartiles) divide data into 100 parts, each with (more or less) the same # of data points.

Q1 = P25, Q2 = P50 = MED, Q3 = P75

Box + whisker plots.



- compute 1.5 × IQR.
- plot outliers individually
- draw the whiskers extending to the furthest points that aren't outliers

Std. Dev.

Measures the "typical" spread of the data about the mean (IQR measures the spread abt. the median).

Naïve idea:
- ✓ Find the mean
- ✓ Find the dist. from. ea pt. to the mean
- ✓ Take the avg. of all distances.

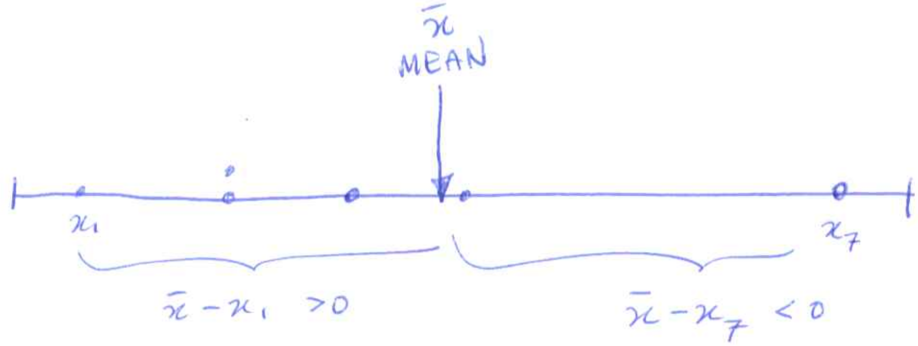
Data = $\{x_1, x_2, x_3, \dots, x_n\}$

✓ Mean is $\bar{x} = \dfrac{x_1 + x_2 + \cdots + x_n}{n} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$

$$\left.\begin{array}{llll} \text{Dist. from} & x_1 & \text{to } \bar{x} \text{ is} & \bar{x} - x_1 \\ \text{— " —} & x_2 & \text{— " —} & \bar{x} - x_2 \\ \text{— " —} & x_3 & \text{— " —} & \bar{x} - x_3 \\ & & & \vdots \end{array}\right\} \text{the set of distances from data pts. to mean}$$

Mean is $\dfrac{(\bar{x} - x_1) + (\bar{x} - x_2) + \cdots + (\bar{x} - x_n)}{n} = \dfrac{1}{n}\sum_{i=1}^{n} (\bar{x} - x_i$

Example:



Problem: Positive & negative distances cancel each other out in the sum....

A slightly modified algorithm: Instead of finding the mean of all distances, find the mean of the squares of the distances.

$$\text{VARIANCE} := s^2 := \frac{(\bar{x}-x_1)^2 + (\bar{x}-x_2)^2 + \cdots + (\bar{x}-x_m)^2}{\underbrace{m-1}_{\text{for technical reasons.}}}$$

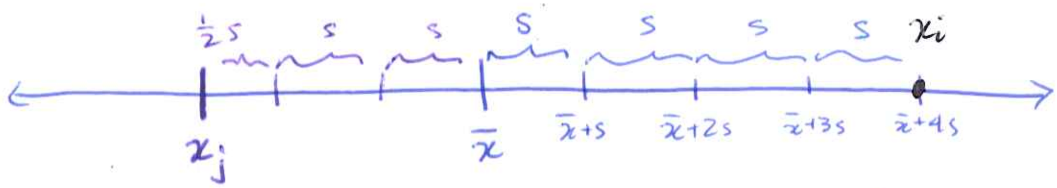
Standard deviation is $s := \sqrt{s^2}$.

$$s = \sqrt{\frac{1}{m-1} \sum_{i=1}^{m} (\bar{x}-x_i)^2}$$

while the S.D. is a characteristic of a data set, the z-score is a ——"—— point.

For a data pt. x_i from a set whose mean is $\bar{x}$ & s.d. is s,

$$z\text{-score of } x_i = \frac{x_i - \bar{x}}{s}$$



For this case, the z-score of x_i is 4.

The z-score of x_j is -2.5