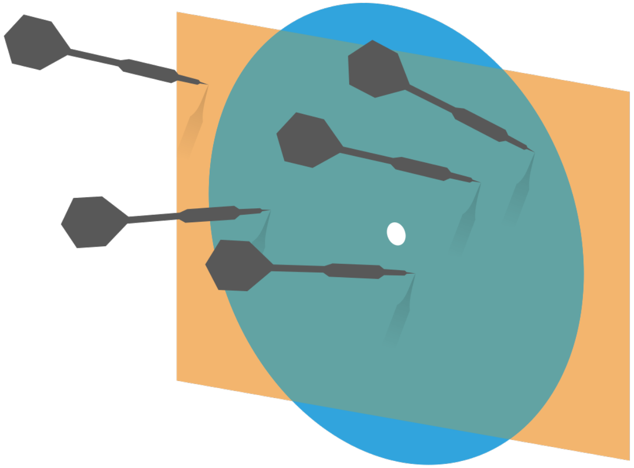
Probability and Statistics for Computer Science



Credit: wikipedia

"The statement that "The average US family has 2.6 children" invites mockery" – Prof. Forsyth reminds us about critical thinking

Hongye Liu, Teaching Assistant Prof, CS361, UIUC, 8.27.2020

Last lecture

- Welcome/Orientation
- Big picture of the contents
- Lecture 1 - Data Visualization & Summary (I)
- **Some feedbacks**

Warm up question:

- What kind of data is a letter grade?
- What do you ask for usually about the stats of an exam with numerical scores?

Objectives

- **Grasp** Summary Statistics
- Learn more Data Visualization for **Relationships**

Summarizing 1D continuous data

For a data set {x} or annotated as $\{x_i\}$, we summarize with: N items

- Location Parameters

  Mean (μ), Median, Mode

- Scale parameters

  Standard deviation (σ), Interquartile range (iqr)

  variance (σ^2)

Summarizing 1D continuous data

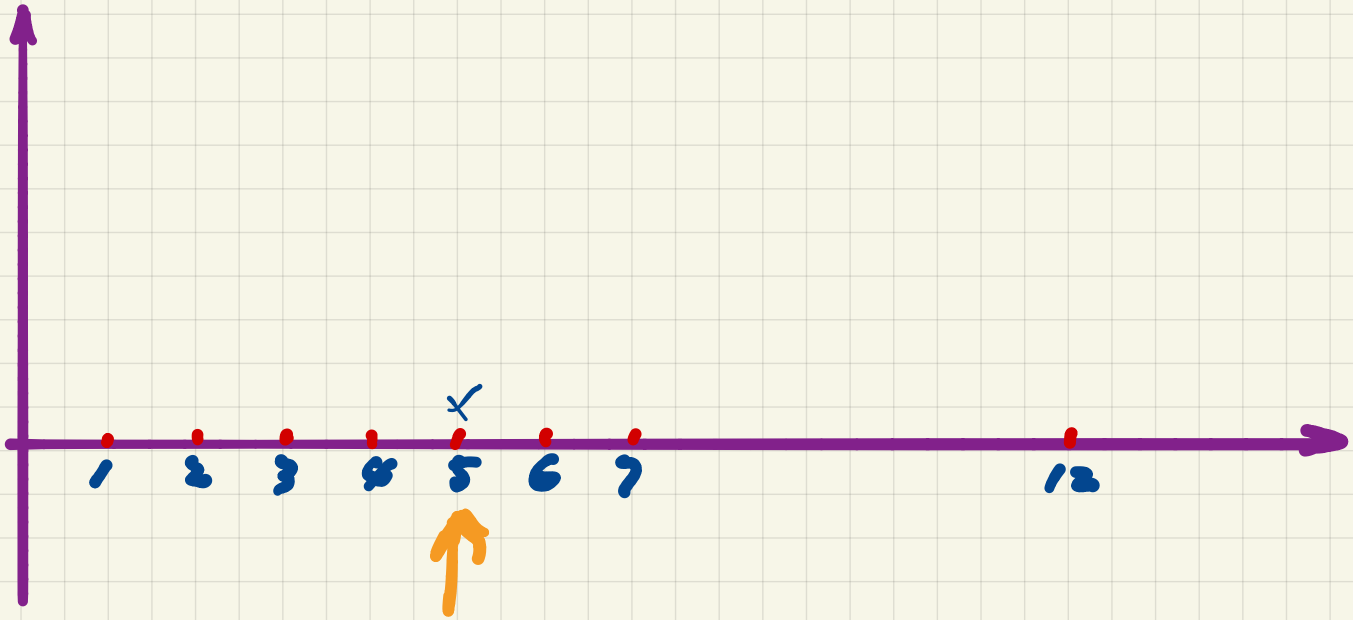
- Mean

$$mean(x_i) = \frac{1}{N}\sum_{i=1}^{N} x_i$$

It's the centroid of the data geometrically,
by identifying the data set at that point, you find
the center of balance.

$\{x_i\} \quad i \in [1, 8]$

$\{x_i\} = 1, \ 2, \ 3, \ 4, \ 5, \ 6, \ 7, \ 12$



$\sum (x_i - \text{mean}(\{x_i\})) = 0$

Properties of the mean

- Scaling data scales the mean

mean(a{x_i} + c)

= a mean({x_i}) + c

$$mean(\{k \cdot x_i\}) = k \cdot mean(\{x_i\})$$

- Translating the data translates the mean

$$mean(\{x_i + c\}) = mean(\{x_i\}) + c$$

Less obvious properties of the mean

- The signed distances from the mean sum to 0

$$\boxed{\sum_{i=1}^{N}(x_i - mean(\{x_i\})) = 0}$$

- The mean minimizes the sum of the squared distance from any real value

argument

$$\hat{\mu} = \underset{\mu}{argmin} \boxed{\sum_{i=1}^{N}(x_i - \mu)^2} = mean(\{x_i\})$$

Prove $\sum_{i=1}^{N} (x_i - \text{mean}(\{x_i\})) = 0$

LHS : $\sum_{i=1}^{N} (x_i) - \sum_{i=1}^{N} \text{mean}(\{x_i\})$

number

$\text{mean}(\{x_i\}) = \dfrac{\sum_{i=1}^{N} x_i}{N}$

LHS : $\sum_{i=1}^{N} (x_i) - \cancel{N} \cdot \dfrac{\sum^{N} x_i}{\cancel{N}} = 0$

Prove $\operatorname{argmin}_{\mu} \sum_i (x_i - \mu)^2 = \operatorname{mean}(\{x_i\})$

$$\frac{d(\Sigma f)}{d\mu} = \sum_{i=1}^{N} \frac{df}{d\mu}$$

$$= \sum_{i=1}^{N} \frac{df}{dg} \frac{dg}{d\mu}$$

$$= \sum_{i=1}^{N} 2g \cdot (-1)$$

$$= -\sum_{i=1}^{N} 2g = 0$$

$\Sigma g = 0$

$g = x_i - \mu$

$f = (x_i - \mu)^2$

$= g^2$

$$\frac{df}{d\mu} = \frac{df}{dg} \frac{dg}{d\mu}$$

$$\frac{df}{dg} = 2g$$

$$\frac{dg}{d\mu} = \frac{d(x_i - \mu)}{d\mu}$$

$= -1$

$$\sum_{i=1}^{N} g = 0 \qquad g_i = x_i - \mu$$

$$\sum_{i=1}^{N} (x_i - \mu) = 0$$

$$\sum_{i=1}^{N} (x_i) - \sum_{i=1}^{N} \mu = 0$$

$$\sum_{i=1}^{N} x_i - N \cdot \mu = 0$$

$$\hat{\mu} = \frac{\sum_{i=1}^{N} x_i}{N} = \text{mean}$$

$$\text{argmin}(\ldots) = \text{mean}$$

Q1:

- What is the answer for

  $mean(mean(\{x_i\}))$?

  A. $mean(\{x_i\})$ B. unsure C. 0

Standard Deviation (σ)

- The standard deviation

$f = (x-\mu)^2$

$\underset{\mu}{\operatorname{argmin}} \sum f = \text{mean}$

$$std(\{x_i\}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - mean(\{x_i\}))^2}$$

$$= \sqrt{mean(\{(x_i - mean(\{x_i\}))^2\})}$$

Q2. Can a standard deviation of a dataset be -1?

A. YES
B. NO

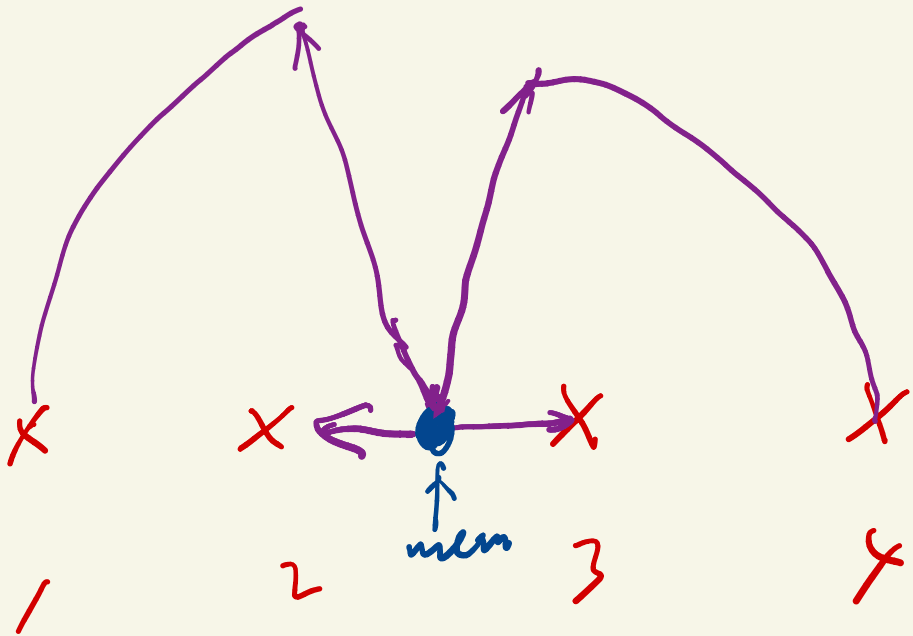
Properties of the standard deviation

- Scaling data scales the standard deviation

$$std(\{k \cdot x_i\}) = |k| \cdot std(\{x_i\})$$

- Translating the data does **NOT** change the standard deviation

$$std(\{x_i + c\}) = std(\{x_i\})$$

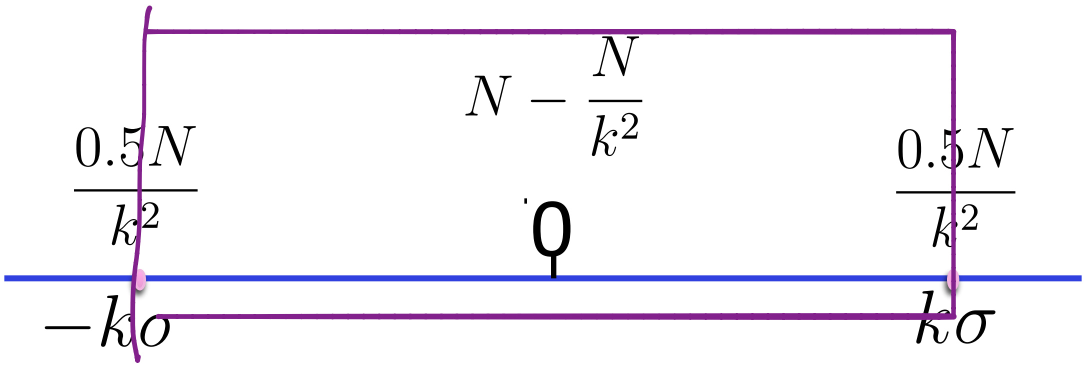


mem
1
2
3
4

Standard deviation: Chebyshev's inequality (1st look)

- At most $\boxed{\frac{N}{k^2}}$ items are k standard deviations (σ) away from the mean

- Rough justification: Assume mean =0



$$std = \sqrt{\frac{1}{N}[(N - \frac{N}{k})\underset{(0-0)}{0^2} + \frac{N}{k^2}(k\underset{-0}{\sigma})^2]} = \sigma$$

Variance (σ^2)

- Variance = (standard deviation)^2 σ^2

$$var(\{x_i\}) = \frac{1}{N}\sum_{i=1}^{N}(x_i - mean(\{x_i\}))^2$$

- Scaling and translating similar to standard deviation

$$var(\{k \cdot x_i\}) = \boxed{k^2} \cdot var(\{x_i\})$$

$$var(\{x_i + c\}) = var(\{x_i\})$$

Q3: Standard deviation

- What is the value of

  $std(mean(\{x_i\}))$?

A. 0 B. 1 C. unsure

Standard Coordinates/normalized data

- The *mean* tells where the data set is and the *standard deviation* tells how spread out it is. If we are interested only in comparing the $\boxed{\text{shape,}}$ we could

  define:

$$\widehat{x}_i = \frac{x_i - mean(\{x_i\})}{std(\{x_i\})}$$

for every i

- We say $\{\widehat{x}_i\}$ is in standard coordinates

Q4: Mean of standard coordinates

- μ of $\{\widehat{x_i}\}$ is:

  A. 1 B. 0 C. unsure

mean

$$\widehat{x_i} = \frac{x_i - mean(\{x_i\})}{std(\{x_i\})}$$

Q5: Standard deviation (σ) of standard coordinates

- σ of $\{\widehat{x_i}\}$ is:

  A. 1 B. 0 C. unsure

std

$$\widehat{x_i} = \frac{x_i - mean(\{x_i\})}{std(\{x_i\})}$$

Q6: Variance of standard coordinates

- Variance of $\{\widehat{x_i}\}$ is:

  A. 1 B. 0 C. unsure

$$\widehat{x_i} = \frac{x_i - mean(\{x_i\})}{std(\{x_i\})}$$

Q7: Estimate the range of data in standard coordinates

- Estimate as close as possible, 90% data is within:

  A. [-10, 10]

  B. [-100, 100]

  C. [-1, 1]

  $\boxed{\text{D.}}$ [-4, 4]

  E. others

$$\widehat{x_i} = \frac{x_i - mean(\{x_i\})}{std(\{x_i\})}$$

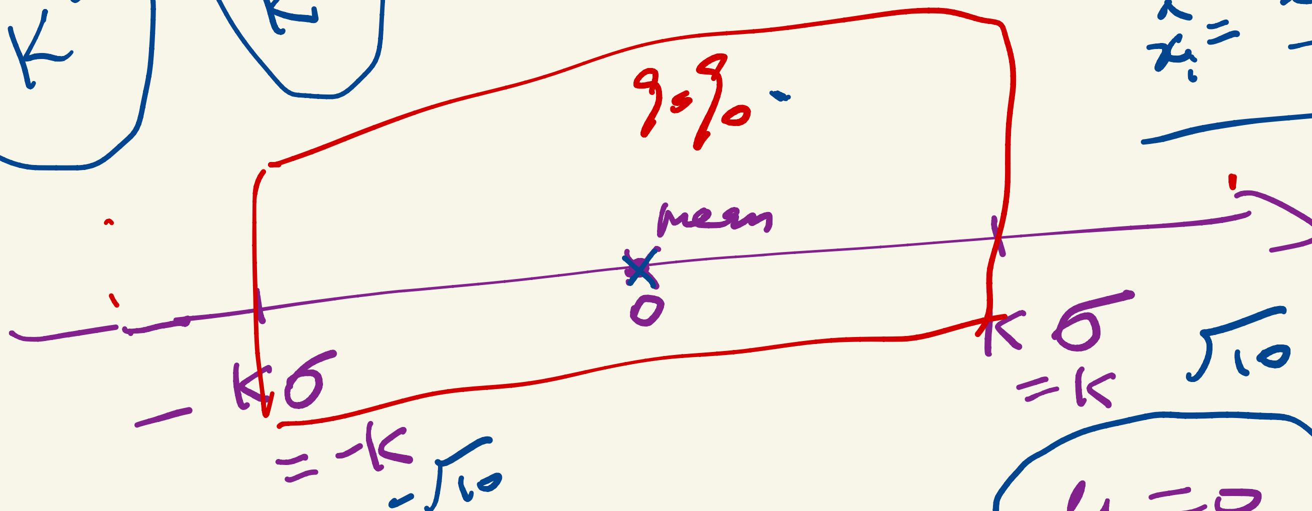
$\frac{N}{K^2}$

$\frac{1}{K^2} \le 10\%$

$K \ge \sqrt{10}$

standized data set

$\hat{x}_i = \frac{x_i - \mu}{\sigma}$



$\mu = 0$

$\sigma = 1$

Summary stats of standard Coordinates/normalized data

Standard Coordinates/normalized data to $\mu=0$, $\sigma=1$, $\sigma^2=1$

- Data in standard coordinates always has

  mean = 0; standard deviation =1;

  variance = 1.

- Such data is unit-less, plots based on this sometimes are more comparable

- We see such normalization very often in statistics

Additional References

- Charles M. Grinstead and J. Laurie Snell "Introduction to Probability”
- Morris H. Degroot and Mark J. Schervish "Probability and Statistics”

See you next time

See You!