Stats
Friday, 1 October 2021 7:59 PM

Stats / Probability

- Descriptive — [insights of data]
- Inferential — C.I, Hypothesis Testing, A/B, CLT

Random Variable

Population Vs. Sample.

[Sample Bias]

30 states
10000
⇒ 3L

$$\mu = \frac{1}{1.4B}\sum_{i=1}^{1.4B} x_i$$

(mu)

μ → Population mean

$$\bar{x} = \frac{1}{3L}\sum_{i=1}^{3L} x_i$$

Sample mean

→ Stratified sampling

Data
- Structured
- Unstructured (Deep Learning)

Types of Data
- Numeric
  - Continous
  - Discrete — SAT, GRE, Step count
- Categorical
  - Ordinal — Grade [A, B, C, D, E], A ≥ A
  - Nominal — Country, Zipcode / Pincode

x = [8, 10, 5, 4, 2, 11], 2000

n = 6

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{8+10+5+4+2+11}{6}$$

summation

weighted mean

$$\bar{x} = \frac{\sum_{i=1}^{n}(x_i \cdot w_i)}{\sum_{i=1}^{n} w_i}$$

$$\frac{8\times 0.9 + 10\times 1 \ldots}{0.9 + 1 + \ldots}$$

w = [0.9, 1, 1, 1, 1, 1]

Median

x = [10, 8, 4, 1, 7, 3]

median(x) =
1. Sort the data
2. the middle value is the median

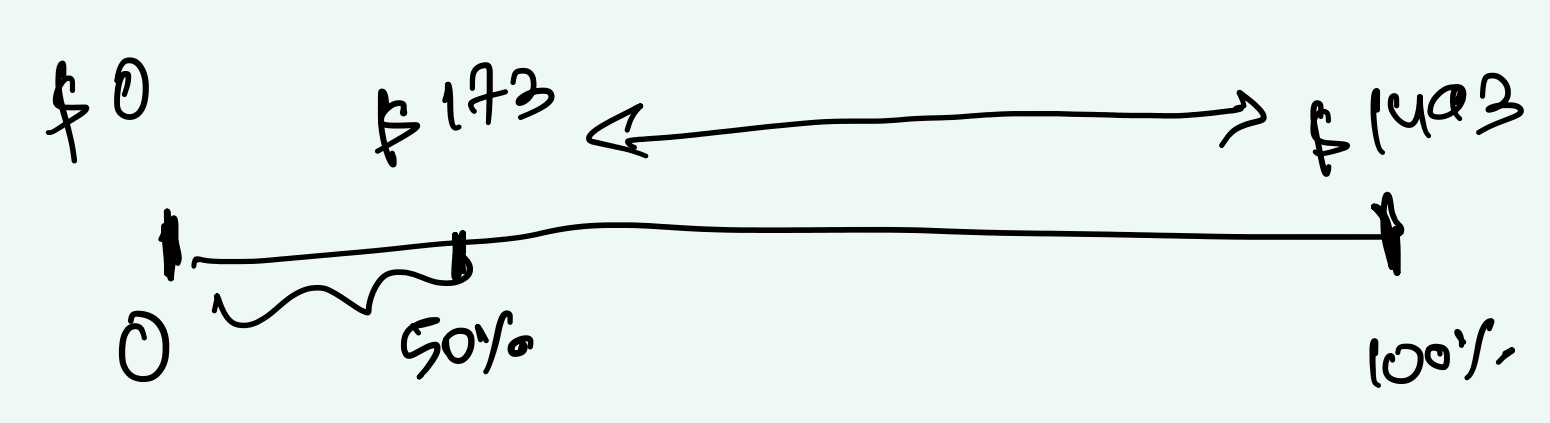
① [1, 3, 4, 7, 8, 10, 2000, 2500]

✓ n = Even ⇒ avg $\left[\frac{n}{2}, \frac{n+1}{2}\right]$ ≈ 5.5

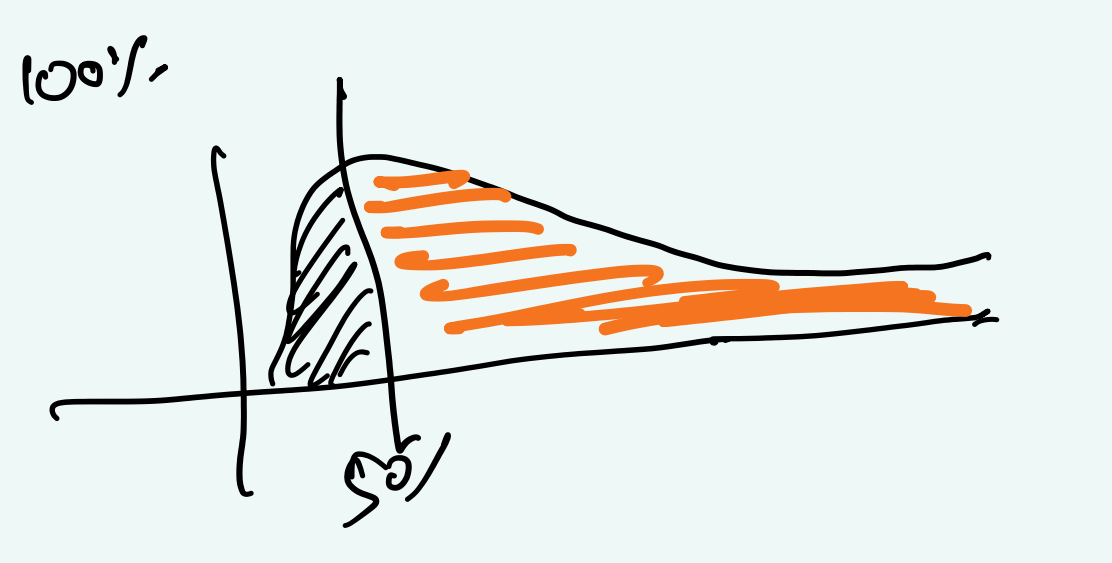
✓ n = odd $\frac{n+1}{2}$ ≈ 7

$\frac{7+8}{2}$ ⇒ 7.5

Mean Vs. Median



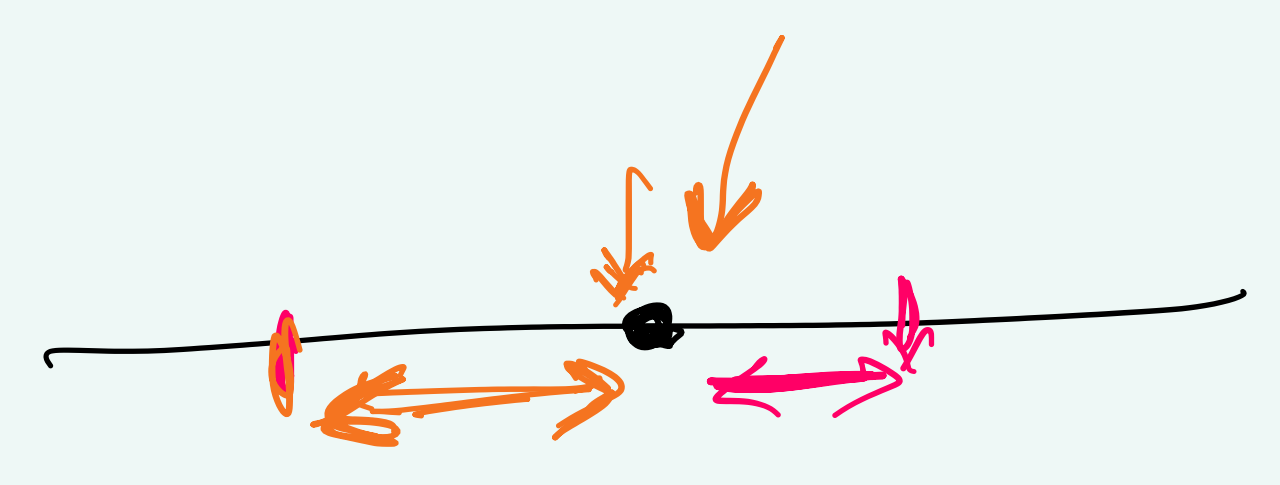
Skewness



Variability

$$\text{Mean absolute deviation} = \frac{1}{n}\sum_{i=1}^{n} |x_i - \bar{x}|$$

1. take mean
2. Subtract mean from all the points (abs)
3. Mean(take)



Standard deviation / Variance

$$\text{Variance} = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

n−1 : Bias Estimate

s = standard deviation.

Variance = s^2

s = $\sqrt{\text{Variance}}$

MAD
Median Absolute Deviation ⇒ $x = [x_1, x_2, x_3, \ldots, x_n]$

median(x) = m

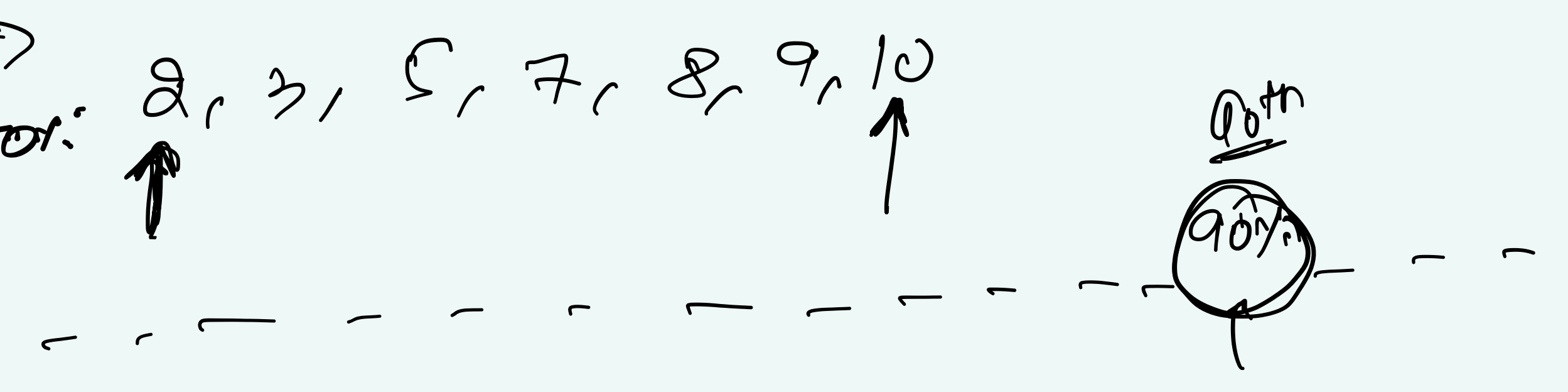
$$\text{median}(|x_1 - m|, |x_2 - m|, |x_3 - m|, \ldots, |x_n - m|)$$

p^{th} percentile ⇒ ← p values → x

Percentile

[9, 10, 7, 8, 5, 2, 3]

or: 2, 3, 5, 7, 8, 9, 10



Parameter Vs. Statistic.

μ