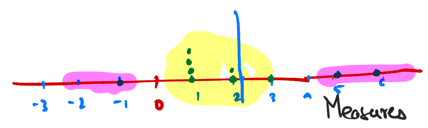
$1, 1, 2, 2, 3, 1, 1, -1, 5, 6$



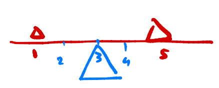
Measures of Location

$\bar{x} = \frac{1+1+2+2+3+1+1-1+5+6}{10} = \frac{21}{10} = 2.1$

Given sample data $\{x_1, x_2, \ldots, x_n\}$

⇝

Want: a "number" that best locates points in the dataset.



$\frac{1+5}{2} = 3$

① Sample Mean

$$\boxed{\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n}}$$

= "arithmetic mean / average".

Note: The "true mean" for the population is denoted by $\boxed{\mu}$.

⇝

μ is fixed and usually unknown

⇝

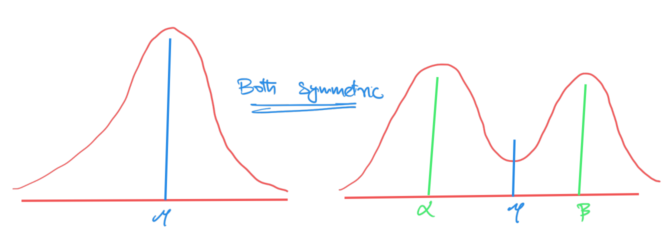
Want to use $\bar{x}$ to estimate μ.

⇝

might not always be a good idea!!

Applicability of the Mean

① mean (sample/population) is a good measure of location if data is symmetric and unimodal.



mean is a good measure of location

mean is not a good measure of location.

③ Sample mean is sensitive to individual data points and outliers.

$\{1, 2, 3\}$ $\bar{x} = 2$

$\{1, 2, 30\}$ $\bar{x} = 11$

⇝

mean is "pulled" in the direction of outlying points

⇝

not a good measure of location when data has "outliers".

② Percentiles

$p = 0.9 \rightsquigarrow 100 \cdot p = 90^{th}$ percentile

⇝

90% of the data is less than the 90th percentile.

Let $p \in (0,1) \rightsquigarrow$ the $(100 \cdot p)^{th}$ percentile for the sample data $\{x_1, x_2, \ldots, x_n\}$ is a number satisfying

⇝

$(100 \cdot p)\%$ of data less than this number.

Important Percentiles:

① Median $\rightsquigarrow 50^{th}$ percentile.

Given a sample $\{x_1, x_2, x_3, \ldots, x_n\}$

⇝

to calculate sample median $\tilde{x}$

⇝

$1, 1, 2, 2, 3, 1, 1, -1, 5, 6$

⇝

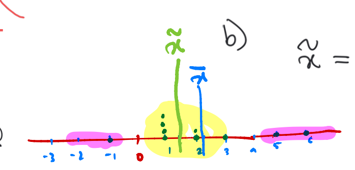
$-1, 1, 1, 1, 1, 2, 2, 3, 5, 6$

$1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ 10$

$\tilde{x} = 1.5$

a) order the dataset in increasing order

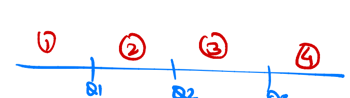
b)



$$\tilde{x} = \begin{cases} \left(\frac{n+1}{2}\right)^{th} \text{ ordered value if } n \text{ is odd.} \\ \text{avg of } \left(\frac{n}{2}\right)^{th} \text{ and } \left(\frac{n}{2}+1\right)^{th} \text{ ordered values if } n \text{ is even.} \end{cases}$$

The true population median is denoted as $\tilde{\mu}$

② Quartiles:



$Q1 \rightsquigarrow 1^{st}$ Quartile $\rightsquigarrow 25^{th}$ percentile

$Q2 \rightsquigarrow 2^{nd}$ Quartile $\rightsquigarrow 50^{th}$ percentile

$Q3 \rightsquigarrow 3^{rd}$ Quartile $\rightsquigarrow 75^{th}$ percentile.

Note: a) $\tilde{x} = \cancel{Q3}$ Q2 $\rightsquigarrow$ quantile in R.

b) the quartiles are robust to minor changes in data values.

c) robust to outliers

d) have poor arithmetic properties.

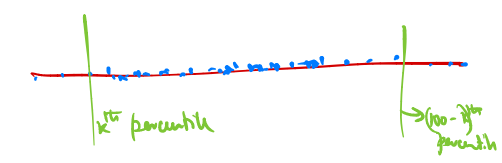
③ Trimmed Means

⇝

Combines arithmetic properties of the mean and the robustness of percentiles

Let $k \in (0, 50)$.

⇝



$\boxed{k\% \text{ trimmed mean} := \bar{x}_k := \text{avg of data values betwn } k^{th} \text{ and } (100-k)^{th} \text{ percentiles.}}$

In practice $k = 10$.

Note: $\bar{x}_0 = \bar{x}$ = sample mean

$\bar{x}_{50} = \tilde{x}$ = sample median.