Numerical Methods

Numerical Methods of Describing Data

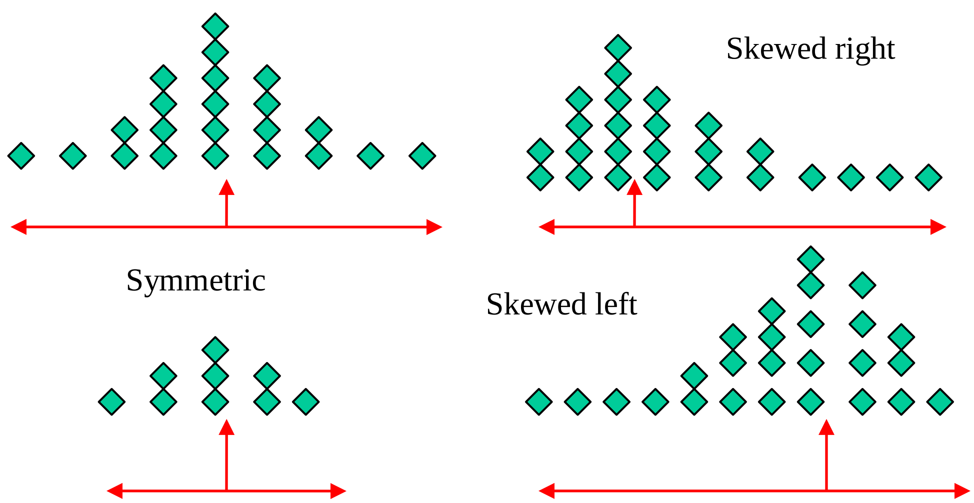
- **Location or central tendency:**
  - **Mean**: $\text{average} = \dfrac{sum}{number\ of\ observations}$ (data: 2, 9, 1, 5, 6)
  - **Median**:
    - Value such that 50% of the observations are smaller and 50% of the observations are larger.
    - The position of the median is $0.5(n+1)$ after the data has been sorted.
    - Odd sized sample = middle number (data: 2, 4, 9, 8, 6, 5, 3)
    - Even sized sample = 2 middle numbers / 2 (data: 2, 4, 9, 8, 6, 5)

When to use median?

  - **Mode**: the value in a data set that appears most frequently

SetA: 2, 4, 9, 8, 8, 5, 3 SetB: 2, 2, 9, 8, 8, 5, 3 SetC: 2, 4, 9, 8, 5, 3

Click for an example.



Where is the data centered on the horizontal axis, and how does it spread out from the center?

Can you locate the mean, median, and mode from the above graphs?

Note:

- Symmetric: Mean = Median
- Skewed right: Mean > Median
- Skewed left: Mean < Median

Scatter, spread, or dispersion

- **Range**: difference between the largest and the smallest observations in a set of data. Only two measurements are used. (data: 5, 12, 6, 8, 14)

$$Range = X_{largest} - X_{smallest}$$

- **Variance**: *average of the squares of the individual deviations*
  - **Deviation** is defined as *the difference between the individual observation and the arithmetic mean*

Use all measurements

Population Variance

$$\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}$$

Sample Variance

$$s^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$$

- **Standard Deviation**: the square root of variance.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}} \qquad s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}}$$

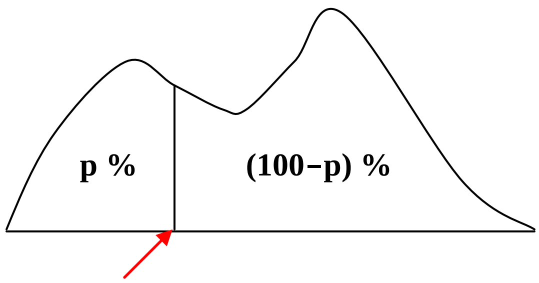
Use s to estimate σ : If we divide the total of deviations by $n-1$, we will get a better estimate of σ

Understanding Variation:

- **The more the data is spread out, the larger the range, variance, and standard deviation**
- **The more concentrated the data (homogenous), the smaller the range, variance, and standard deviation**
- **If all the observations are the same, the range, variance, and standard deviation = 0**
- **None of these measures can be negative**

Measures of Relative Standing:
Where does one particular measurement stand in relation to the other measurements in the data set?

- How many measurements lie below the measurement of interest? This is measured by the ***p*th percentile.**



- The **lower quartile (Q1)** is the value of *x* which is larger than 25% and less than 75% of the ordered measurements.

- The **upper quartile (Q3)** is the value of *x* which is larger than 75% and less than 25% of the ordered measurements.

- How many standard deviations away from the mean does the measurement lie? This is measured by the ***z*-score.**

$$z-\text{score}=\frac{x-\bar{x}}{s}$$ (***The number of standard deviations from mean***)

Example: If $s=2$ and $\bar{x}=5$, then $x=9$ lies $z=2$ standard deviations from mean

Example:

The prices ($) of **18** brands of walking shoes:

40 60 65 **65** **65** 68 68 70 70 70 70 70 70 **74** **75** 75 90 95

Position of Q1 = .25(18 + 1) = 4.75

Position of Q3 = .75(18 + 1) = 14.25

✓ Q1is **3/4** of the way between **the 4th and 5th** ordered measurements, or

Q1 = 65 + .75(65 - 65) = 65.

✓ Q3 is **1/4** of the way between **the 14th and 15th** ordered measurements, or

Q3 = 75 + .25(75 - 74) = 75.25