Statistics – Variability around the Central Tendency / Dispersion

Range, Variance, Standard Deviation

How do values vary around the mean / median?

Dispersion – minimum, maximum & range

The range is the distance between the largest and the smallest value in the data set.

$$range = maximum\ value - minimum\ value$$

1, 2, 2, 2, 4, 4, 6

$$range = 6 - 1 = 5$$

Dispersion – mean deviation

The mean deviation is the average of deviations of individual observations from the arithmetic mean.

1, 2, 2, 2, 4, 4, 6

mean = 3

$$\frac{(1-3)+(2-3)+(2-3)+(2-3)+(4-3)+(4-3)+(6-3)}{7} = 0$$

Dispersion – variance σ^2 (population)

The variance is the average of the squared deviations from the arithmetic mean.

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

σ^2: variance (population)
μ: mean (population)
X_i: value of observation i
N: total number of observations

1, 2, 2, 2, 4, 4, 6

mean = 3

$$\sigma^2 = \frac{(1-3)^2+(2-3)^2+(2-3)^2+(2-3)^2+(4-3)^2+(4-3)^2+(6-3)^2}{7}$$

$$= 2.57$$

Dispersion – variance s^2 (sample)

The variance is the average of the squared deviations from the arithmetic mean.

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

s^2: variance (sample)
$\bar{X}$: mean (sample)
X_i: value of observation i
n : total number of observations

1, 2, 2, 2, 4, 4, 6

mean = 3

$$s^2 = \frac{(1-3)^2+(2-3)^2+(2-3)^2+(2-3)^2+(4-3)^2+(4-3)^2+(6-3)^2}{7-1}$$

$$= 3$$

Dispersion – standard deviation σ / s

Pitfall of variance: Hard to interpret (squared units)

Solution: use standard deviation, which is the square root of variance.

$$\sigma = \sqrt{\sigma^2}$$

$$s = \sqrt{s^2}$$

1, 2, 2, 2, 4, 4, 6

mean = 3

$$\sigma = \sqrt{2.57} = 1.60$$

$$s = \sqrt{3} = 1.73$$