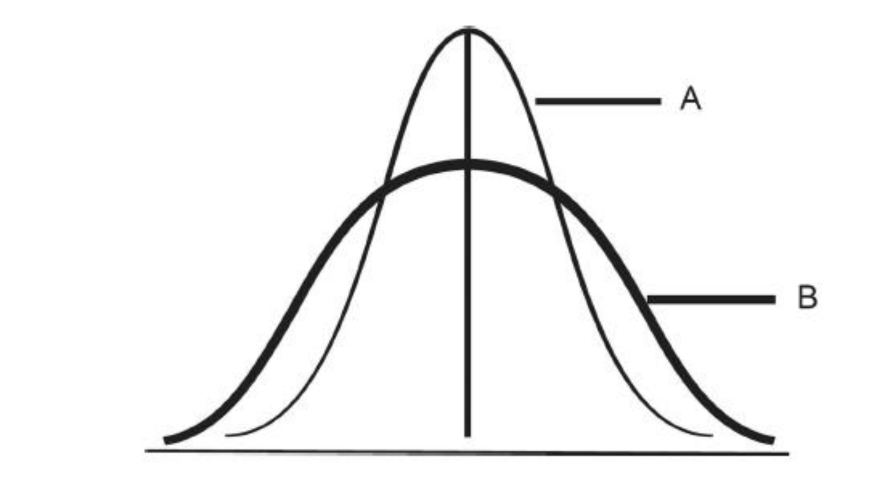
Measures of Variability

Class Notes by A.Sandanasamy, BHC.

Measures of variability provide information about how similar or different the scores are in relationship to other scores in the distribution.



Both of the distributions below have the same mean; however, they are quite diffe rent in the variability of their scores.

Three common measures of variability are the ***range,***
the interquartile range,
and the standard deviation.

THE RANGE

subtracting the lower limit of the lowest score from the upper limit of the highest score.

$$R = X_{\mathrm{UL-High}} - X_{\mathrm{LL-Low}}$$

For Example

For the following set of scores,

3, 3, 5, 7, 7, 7, 8, 9

$$R = 9.5 - 2.5 = 7$$

I. Range

A. Determine the range for the following set of scores:

3, 3, 3, 3, 4, 5, 5, 6, 6, 6, 6, 7, 10, 11, 11, 11

$R =$ ______________

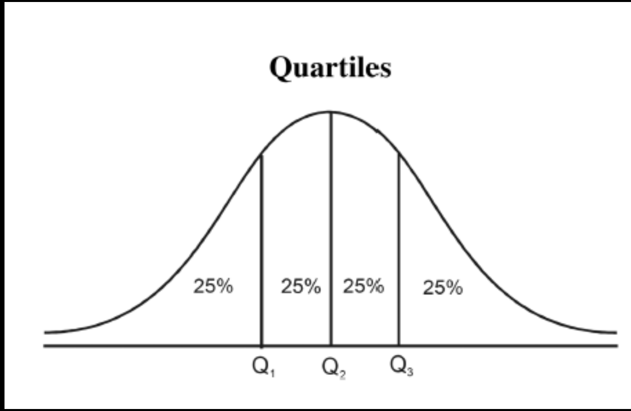
B. Replace the last score of 11 in the above distribution with a score of 54, and calculate the range.

$R =$ _____________

INTERQUARTILE RANGE

the range of scores from the middle 50% of a distribution.

To determine the interquartile range, we first divide the distribution into four equal parts, which produces three ***quartiles***



Q1 = the point at or below which 25% of the scores lie
Q2 = the point at or below which 50% of the scores lie
Q3 = the point at or below which 75% of the scores lie

Steps for Determining the Interquartile Range

1. Arrange scores in order from low to high.
2. Divide the distribution of scores into four equal parts.[1]
3. Find the points below which 25% of the scores and 75% of the scores lie.
4. Identify the two scores that surround these points.
5. Determine the means of each of these two pairs of scores to determine Q_1 and Q_3.
6. Subtract Q_1 from Q_3. The resulting value is the interquartile range, which, as you can see, is simply the distance between the third and first quartiles. Here is a formula for accomplishing these steps:

$$IQR = Q_3 - Q_1$$

Compute the interquartile range for the following scores:

85, 115, 90, 90, 105, 100, 110, 110, 95, 110, 95, 100

25% of the scores lie below this point.

75% of the scores lie below this point.

85 90 90 ⇓ 95 95 100 | 100 105 110 ⇓ 110 110 115

$$Q_1 = \frac{90+95}{2} = 92.50 \qquad Q_3 = \frac{110+110}{2} = 110$$

$$IQR = Q_3 - Q_1 = 110 - 92.50 = 17.50$$

Thus, the range of the middle 50% of the distribution is 17.50.

Determine the IQR for the following set of scores:

36, 42, 30, 7, 51, 29, 45, 35, 44, 53, 32, 50, 28, 43, 33, 29

$IQR =$ ___________

7 28 29 29 $\Downarrow$ 30 32 33 35 | 36 42 43 44 $\Downarrow$ 45 50 51 53

$$Q_1 = \frac{29+30}{2} = 29.5$$

$$Q_3 = \frac{44+45}{2} = 44.5$$

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 44.5 - 29.5 \\ &= 15 \end{aligned}$$

STANDARD DEVIATION

The **standard deviation** is a measure of the amount of variation or dispersion of a set of values.
A low standard deviation indicates that the values tend to be close to the mean of the set, while a high standard deviation indicates tha t the values are spread out over a wider range.

***Definitional formulas** are written the way that statistics are defined.*

These formulas usually involve more computations but are helpful when
initially learning statistics because they guide the learner th rough the
process of what is measured.

***Computational formulas**, on the other hand, are easier to us e with a*
handheld calculator and will result in the same mathematic

Definitional Formula

Formula	Explanation
$\sigma = \sqrt{\dfrac{\Sigma (X - \mu)^2}{N}}$	• σ = population standard deviation • $\sum$ = sum of… • X = each value • μ = population mean • N = number of values in the population

Deviation scores are obtained by subtracting the mean from the raw scores in a distribution.

sum of squares	variance	standard deviation
$SS = \sum (X - \mu)^2$	$\sigma^2 = \frac{SS}{N}$	$\sigma = \sqrt{\frac{SS}{N}}$

Calculate Standard deviation
17,24,22,26,18

X	μ	$(X-\mu)$	$(X-\mu)^2$
17	21.4	−4.4	19.36
24	21.4	2.6	6.76
22	21.4	0.6	.36
26	21.4	4.6	21.16
18	21.4	−3.4	11.56
$\Sigma X = 107$	$\mu = \frac{\Sigma X}{N} = \frac{107}{5} = 21.4$	$\Sigma(X-\mu) = 0$	$\Sigma(X-\mu)^2 = 59.20$

$$\sigma = \sqrt{\frac{\Sigma(X-\mu)^2}{N}} = \sqrt{\frac{59.20}{5}} = \sqrt{11.84} = 3.44$$

III. Population Standard Deviation (Definitional Formula)

Using the definitional formula, determine the sum of squares, the variance, and the standard deviation for the following numbers of cups of coffee consumed by each of the workers at a print shop on a given day:

3, 3, 0, 1, 1, 2, 2, 2, 6, 0

(*Continued*)

$$\sigma = \sqrt{\frac{\sum (X-\mu)^2}{N}} = \sqrt{\frac{28}{10}} = \sqrt{2.8} = 1.67$$

$SS = 28$

$\sigma^2 = 2.8$

$\sigma = 1.67$

Computational Formula

The computational formula does not require that means and devi ation scores be calculated, but simply makes use of the raw scores. For this reason, it is also referred to as the *raw score formula.*

$$\sigma = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N}}$$

where: ΣX^2 = sum of the squared raw scores

$(\Sigma X)^2$ = square of the sum of the raw scores

Guide for Computational Formula for Population Standard Deviation

1. Create two columns, X and X^2, and list the raw scores under X.
2. Square the individual raw scores and place these values in the X^2 column.
3. Sum the X column to obtain ΣX.
4. Sum the X^2 column to obtain ΣX^2.
5. Place these values into the formula along with the appropriate N.
6. Square the sum of the raw scores and divide the result by N to determine $\frac{(\Sigma X)^2}{N}$.
7. Subtract this result from ΣX^2. This value is SS.
8. Divide SS by N. This value is the variance (σ^2).
9. Find the square root of the variance to obtain the standard deviation (σ).

X	X^2
17	289
24	576
22	484
26	676
18	324
$\Sigma X = 107$	$\Sigma X^2 = 2349$

$$\sigma = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N}}$$

$$\sigma = \sqrt{\frac{2349 - \frac{(107)^2}{5}}{5}} = \sqrt{\frac{2349 - 2289.8}{5}} = \sqrt{\frac{\overset{\text{Sum of Squares}}{\downarrow}\atop 59.20}{5}} = \sqrt{\overset{\text{Variance}}{\downarrow}\atop 11.84} = \overset{\text{Standard Deviation}}{\downarrow}\atop 3.44$$

USING SAMPLES TO ESTIMATE POPULATION STANDARD DEVIATIONS

Samples are, by definition, smaller than populations, and you are more likely to find more deviant scores in a larger group than in a smaller one.

For instance, you are more likely to find a person with an IQ of 120 or higher in a group of 500 than in a group of 20.

Thus, samples do not reflect the true variability of their parent populations.

For this reason, if you are using sample data to estimate the population standard deviation, an adjustment in the standard deviation formula is n ecessary.

Specifically, you will need to use *$n - 1$, rather than N , in the denominat or.*

Using N in the denominator:

$$\sigma = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N}}$$

$$= \sqrt{\frac{6401 - \frac{(177)^2}{5}}{5}}$$

$$= 5.20$$

Using $n - 1$ in the denominator:

$$s = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1}}$$

$$= \sqrt{\frac{6401 - \frac{(177)^2}{5}}{5-1}}$$

$$= 5.81$$