Measures of Dispersion

Greg C Elvers, Ph.D.

1

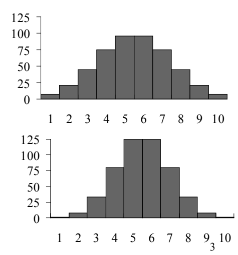
Definition

- *Measures of dispersion* are descriptive statistics that describe how similar a set of scores are to each other
  - The more similar the scores are to each other, the lower the measure of dispersion will be
  - The less similar the scores are to each other, the higher the measure of dispersion will be
  - In general, the more spread out a distribution is, the larger the measure of dispersion will be

2

Measures of Dispersion

- Which of the distributions of scores has the larger dispersion?
- The upper distribution has more dispersion because the scores are more spread out
  - That is, they are less similar to each other



3

Measures of Dispersion

- There are three main measures of dispersion:
  - The range
  - The semi-interquartile range (SIR)
  - Variance / standard deviation

4

The Range

- The *range* is defined as the difference between the largest score in the set of data and the smallest score in the set of data, $X_L - X_S$
- What is the range of the following data:
  4 8 1 6 6 2 9 3 6 9
- The largest score (X_L) is 9; the smallest score (X_S) is 1; the range is $X_L - X_S = 9 - 1 = 8$

When To Use the Range

- The range is used when
  - you have ordinal data or
  - you are presenting your results to people with little or no knowledge of statistics
- The range is rarely used in scientific work as it is fairly insensitive
  - It depends on only two scores in the set of data, X_L and X_S
  - Two very different sets of data can have the same range:
    1 1 1 1 9 vs 1 3 5 7 9

The Semi-Interquartile Range

- The *semi-interquartile range* (or *SIR*) is defined as the difference of the first and third quartiles divided by two
  - The first quartile is the 25^{th} percentile
  - The third quartile is the 75^{th} percentile
- $SIR = (Q_3 - Q_1) / 2$

SIR Example

- What is the SIR for the data to the right?
- 25 % of the scores are below 5
  - 5 is the first quartile
- 25 % of the scores are above 25
  - 25 is the third quartile
- $SIR = (Q_3 - Q_1) / 2 = (25 - 5) / 2 = 10$

2	
4	
6	← 5 = 25^{th} %tile
8	
10	
12	
14	
20	
30	← 25 = 75^{th} %tile
60	

When To Use the SIR

- The SIR is often used with skewed data as it is insensitive to the extreme scores

9

Variance

- *Variance* is defined as the average of the square deviations:

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

10

What Does the Variance Formula Mean?

- First, it says to subtract the mean from each of the scores
  - This difference is called a *deviate* or a *deviation score*
  - The deviate tells us how far a given score is from the typical, or average, score
  - Thus, the deviate is a measure of dispersion for a given score

11

What Does the Variance Formula Mean?

- Why can't we simply take the average of the deviates? That is, why isn't variance defined as:

$$\sigma^2 \neq \frac{\sum(X-\mu)}{N}$$

This is not the formula for variance!

12

What Does the Variance Formula Mean?

- One of the definitions of the *mean* was that it always made the sum of the scores minus the mean equal to 0
- Thus, the average of the deviates must be 0 since the sum of the deviates must equal 0
- To avoid this problem, statisticians square the deviate score prior to averaging them
  - Squaring the deviate score makes all the squared scores positive

What Does the Variance Formula Mean?

- Variance is the mean of the squared deviation scores
- The larger the variance is, the more the scores deviate, on average, away from the mean
- The smaller the variance is, the less the scores deviate, on average, from the mean

Standard Deviation

- When the deviate scores are squared in variance, their unit of measure is squared as well
  - E.g. If people's weights are measured in pounds, then the variance of the weights would be expressed in pounds^2 (or squared pounds)
- Since squared units of measure are often awkward to deal with, the square root of variance is often used instead
  - The standard deviation is the square root of variance

Standard Deviation

- Standard deviation = $\sqrt{\text{variance}}$
- Variance = standard deviation^2

Computational Formula

- When calculating variance, it is often easier to use a computational formula which is algebraically equivalent to the definitional formula:

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

- σ^2 is the population variance, X is a score, μ is the population mean, and N is the number of scores

Computational Formula Example

X	X^2	$X-\mu$	$(X-\mu)^2$
9	81	2	4
8	64	1	1
6	36	-1	1
5	25	-2	4
8	64	1	1
6	36	-1	1
$\Sigma = 42$	$\Sigma = 306$	$\Sigma = 0$	$\Sigma = 12$

Computational Formula Example

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

$$= \frac{306 - \frac{42^2}{6}}{6}$$

$$= \frac{306 - 294}{6}$$

$$= \frac{12}{6}$$

$$= 2$$

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

$$= \frac{12}{6}$$

$$= 2$$

Variance of a Sample

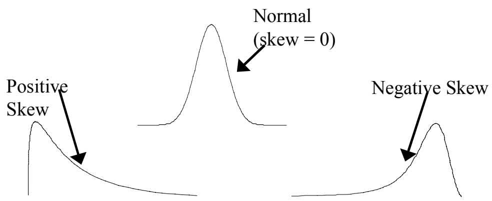
- Because the sample mean is not a perfect estimate of the population mean, the formula for the variance of a sample is slightly different from the formula for the variance of a population:

$$s^2 = \frac{\sum (X - \overline{X})^2}{N-1}$$

- s^2 is the sample variance, X is a score, $\overline{X}$ is the sample mean, and N is the number of scores

Measure of Skew

- *Skew* is a measure of symmetry in the distribution of scores



Measure of Skew

- The following formula can be used to determine skew:

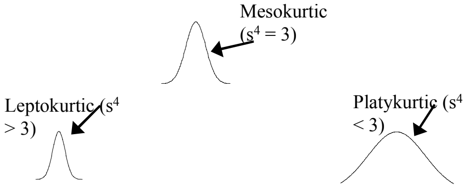
$$s^3 = \frac{\frac{\Sigma\left(X-\overline{X}\right)^3}{N}}{\sqrt{\frac{\Sigma\left(X-\overline{X}\right)^2}{N}}}$$

Measure of Skew

- If $s^3 < 0$, then the distribution has a negative skew
- If $s^3 > 0$ then the distribution has a positive skew
- If $s^3 = 0$ then the distribution is symmetrical
- The more different s^3 is from 0, the greater the skew in the distribution

Kurtosis (Not Related to Halitosis)

- Kurtosis measures whether the scores are spread out more or less than they would be in a normal (Gaussian) distribution



Kurtosis

- When the distribution is normally distributed, its kurtosis equals 3 and it is said to be *mesokurtic*
- When the distribution is less spread out than normal, its kurtosis is greater than 3 and it is said to be *leptokurtic*
- When the distribution is more spread out than normal, its kurtosis is less than 3 and it is said to be *platykurtic*

Measure of Kurtosis

- The measure of kurtosis is given by:

$$s^4 = \frac{\sum \left(\dfrac{X - \overline{X}}{\sqrt{\dfrac{\sum \left(X - \overline{X} \right)^2}{N}}} \right)^4}{N}$$

s^2, s^3, & s^4

- Collectively, the variance (s^2), skew (s^3), and kurtosis (s^4) describe the shape of the distribution