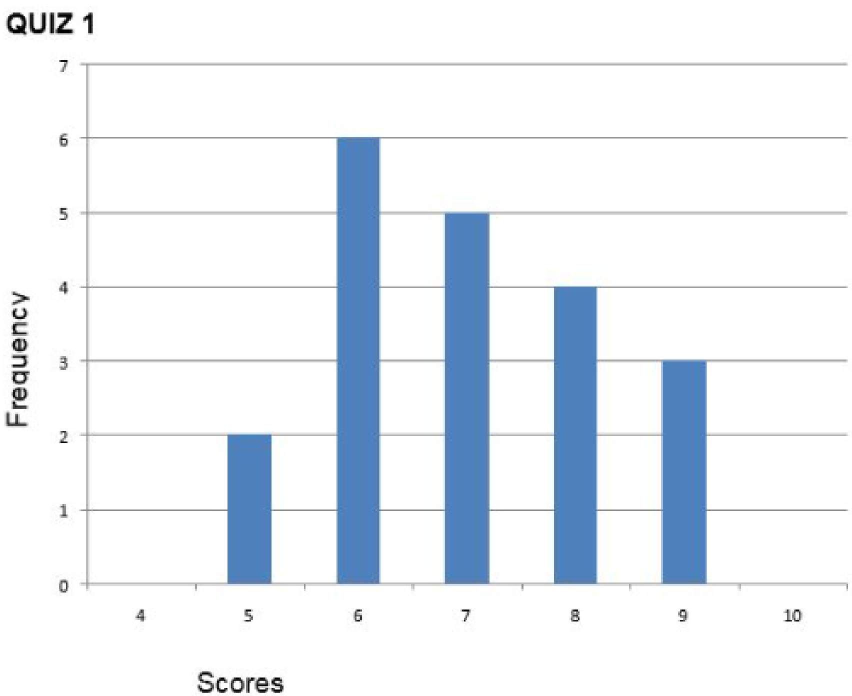
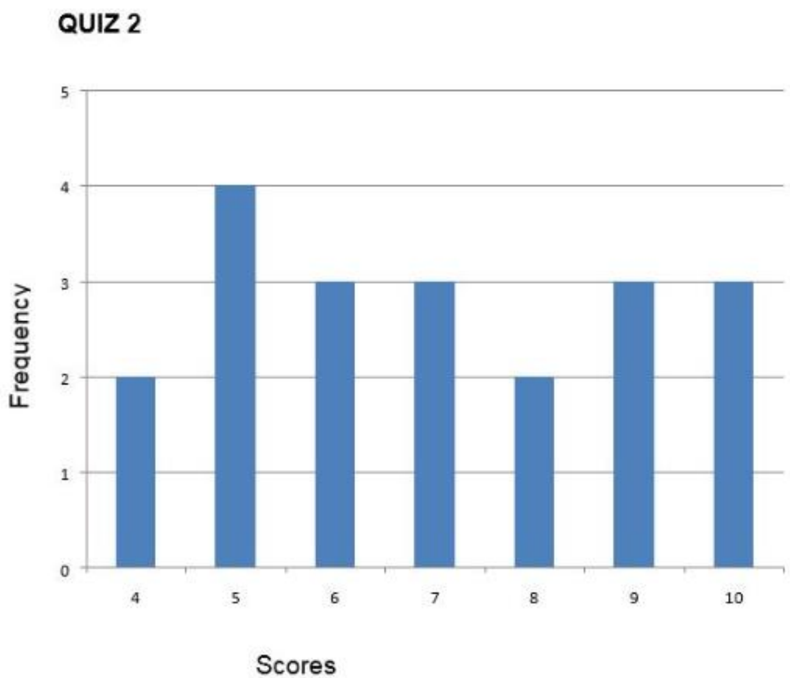
Measure of variability

Measure of variability

Variability refers to how spread scores are in a distribution out; that is, it refers to the amount of spread of the scores around the mean. For example, distributions with the same mean can have different amounts of variability or dispersion.

In the following two histograms, the distribution of scores for Quiz 1 and Quiz 2 are presented. Despite the equal means (the mean score for both quizzes is 7), the scores on Quiz 1 are more packed or clustered around the mean, whilst the scores on Quiz 2 are more spread out. Thus, the differences within the student group were greater on Quiz 2 than on Quiz 1.





Measure of variability

There are four frequently used measures of the variability of a distribution including ***range, interquartile range, variance and standard deviation***

Range ungroup data

The most basic measure of variation is the range, which is the distance from the smallest to the largest value in a distribution.

Formula:

R= HV – SV

Solution:

For the distribution of scores of Quiz 1 and Quiz 2, the range is:
Quiz 1. R = 9-5= 4
Quiz 1. R = 10-4= 6

Explanation:

Which shows (like the histograms above) that Quiz 2 scores have greater spread than Quiz 1 scores.
However, the range uses only two values in the data set, and one of these values may be an unusually large or small value

Measure of variability

Example 2:

Find the range in the following sets of data.

A. 15, 17, 18, 19, 20, 22, 22, 23, 24

B. 3, 7, 10, 12, 20, 25, 28, 35, 40

Solution:

A. R = 24 -15 = 9

B. R = 40 – 3 = 37

Measure of variability

Range of group data

The range of grouped data is estimated by subtracting the lower boundary of the lowest class interval from the upper boundary of the highest interval. $R = UB_h - LB_l$

Solving for the measure of variability for grouped data is the same as that of the ungrouped data, except that instead of working with individual scores, we are going to use the class mark of every interval

Let's try to solve for the range of the test scores found on table 1.1.

TABLE 1.1

FREQUENCY DISTRIBUTION OF THE TEST SCORES OF FIFTY (50) STUDENTS IN STATISTICS

Class Interval (X)	Frequency (f)
20-24	2
25-29	6
30-34	9
35-39	10
40-44	12
45-49	7
50-54	4
$i = 5$	$n = 50$

Solution:

$R = UB_h - LB_l$

$R = 54.5 – 19.5$

$R = 35$

Therefore, the range of the scores is 35

Measure of variability

Standard deviation

The standard deviation is the square root of the variance, and it is a useful measure of variability when the distribution is normal or approximately normal (see below on the normality of distributions). The proportion of the distribution within a given number of standard deviations (or distance) from the mean can be calculated.

A small standard deviation coefficient indicates a small degree of variability (that is, scores are close together); larger standard deviation coefficients indicate large variability (that is, scores are far apart).

Standard deviation for ungroup data:

$$\sigma = \sqrt{\frac{\sum (x_m - \mu)^2}{N}} \qquad \sigma = \sqrt{\frac{\sum (x_m - \bar{X})^2}{N-1}}$$

Standard deviation for group data:

$$\sigma = \sqrt{\frac{\sum f(x_m - \mu)^2}{N}} \qquad \sigma = \sqrt{\frac{\sum f(x_m - \bar{X})^2}{N-1}}$$

Measure of variability

Variance ungroup data

Is the average squared deviation from the mean. It uses the symbol σ^2or s^2 (the former is used for population data while the latter is used for sample data. The formula for the variance is expressed as

For population data $\sigma^2 = \frac{\sum (x-\mu)^2}{N}$ For sample data $s^2 = \frac{\sum (X-\bar{x})^2}{N-1}$

Where:
σ^2 is the variance of the population
s^2 is the variance of the sample mean
x is the values score
N is the number of values score
μ is the mean population
$\bar{x}$ is the sample mean

Steps in solving the variance:
1. Calculate the mean
2. Subtract the mean from each score to compute the deviation from mean score
3. Square each deviation score (multiply each score by itself)
4. Add up the squared deviation score to give the sum
5. Divide the sum by the number of scores

Measure of variability

Example 1: Calculate the variance of the following data:

8, 9, 10, 12, 17, 18, 18, 19, 20, 21

Solution:

Step 1: Solve for the mean.

$$\bar{X} = \frac{8+9+10+12+17+18+18+19+20+21}{10} = 15.2 \approx 15$$

Step 2 – 4:

x	$(x-\mu)$	$(x-\mu)^2$
8	-7	49
9	-6	36
10	-5	25
12	-3	9
17	2	4
18	3	9
18	3	9
19	4	16
20	5	25
21	6	36
		$\sum(x-\mu)^2 = 218$

Substituting the value in our formula:

$$\sigma^2 = \frac{\sum(X-\bar{X})^2}{10} = \frac{218}{10} = \mathbf{21.8}$$

Standard deviation

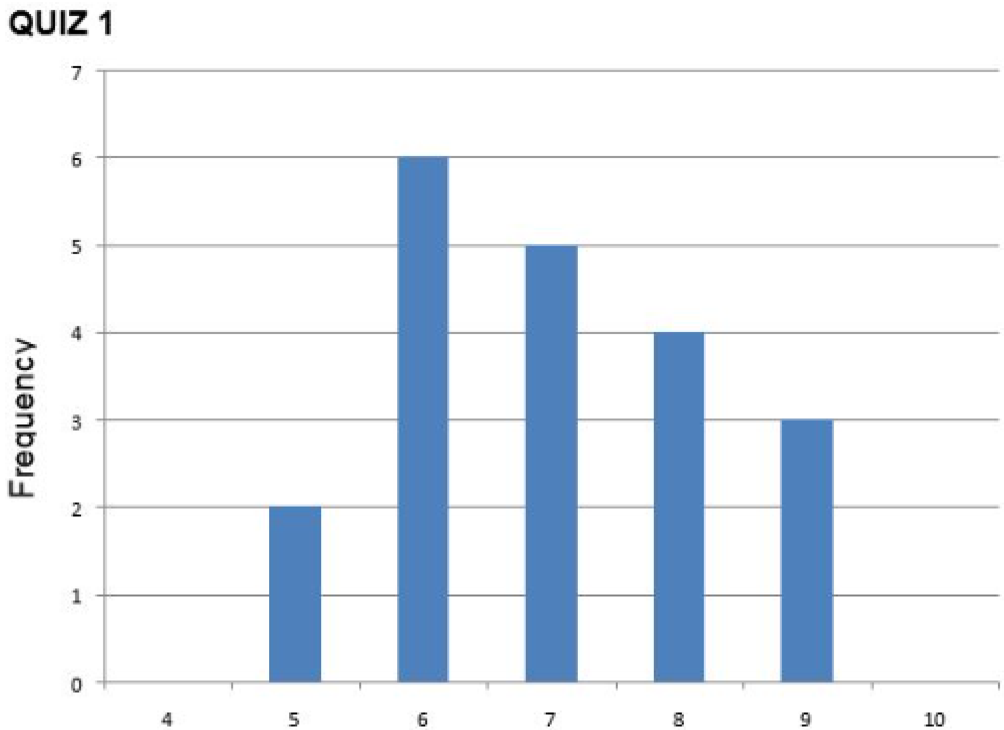
$$\sigma = \sqrt{\frac{\sum(x_m-\mu)^2}{N}}$$

$$\sigma = \sqrt{21.8}$$

$$\sigma = 4.6$$

Measure of variability

Example 2: Find the variance



Measure of variability

Solution:

Table 1.1

Scores (x)	(x - μ)	(x - μ)ˆ2
9	2	4
9	2	4
9	2	4
8	1	1
8	1	1
8	1	1
8	1	1
7	0	0
7	0	0
7	0	0
7	0	0
7	0	0
6	-1	1
6	-1	1
6	-1	1
6	-1	1
6	-1	1
6	-1	1
5	-2	4
5	-2	4
Means		
7	**0**	**1.5**

Population mean:

$$\mu = \frac{(9\,x\,3)+(8\,x\,4)+(7\,x\,5)+(6\,x\,6)+(5\,x\,2)}{20}$$

$$\mu = 7$$

Variance:

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

$$\sigma^2 = \frac{30}{20}$$

$$\sigma^2 = 1.5$$

Standard deviation

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

$$\sigma = \sqrt{1.5}$$

$$\sigma = 1.22$$

Measure of variability

Variance group data

The variance and standard deviation of group data are obtained in much the same way as those ungroup data. The procedure in solving the variance is:

1. Calculate the mean (round off to whole number)
2. Subtract the mean from each class mark
3. Square each deviation obtained in step 2
4. Multiply each squared deviation by the corresponding frequency
5. Take the sum of the results in step 4
6. Divide the sum in step by by n-1 if the values constitute a sample, or by N if they constitute a population

Symbolically, we get:

$\sigma^2 = \frac{\sum f(x_m - \mu)^2}{N}$ and $s^2 = \frac{\sum x_m f\left(X - \bar{X}\right)^2}{n-1}$

Measure of variability

Example: Find the variance of the test scores found in Table 1.1

Solution:

In here, it is best to use a table to have an organized solution.

X	f	X_m	fX_m	$\overline{X}$	$(x-\mu)$	$(x-\mu)^2$	$f(x-\mu)^2$
20-24	2	22	44	38	-16	256	512
25-29	6	27	162		-11	121	726
30-34	9	32	288		-6	36	324
35-39	10	37	370		-1	1	10
40-44	12	42	504		4	16	192
45-49	7	47	329		9	81	567
50-54	4	52	208		14	196	784
$i=5$	$n=50$		Σ=1905				Σ=3115

$$\sigma^2 = \frac{\sum f(X_m-\mu)^2}{N} = \frac{3115}{50} = \mathbf{62.3}$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{62.3} = \mathbf{7.89}$$

Therefore, the variance is 62.3, with a standard deviation of 7.89.

$$\mu = \frac{\sum f x_m}{n}$$

$$\mu = \frac{1905}{50}$$

$$\mu = 38.1$$

Measure of variability

TRY THIS!

Compute for the range, variance and standard deviation of the table on the daily allowances of 70 students enrolled in LORMA Colleges.

**Daily Allowance of
BS Nursing Students**

Daily Allowance (in Pesos)	Frequency
500 – 549	2
450 – 499	3
400 – 449	1
350 – 399	3
300 – 349	4
250 – 299	14
200 – 249	12
150 – 199	21
100 – 149	10

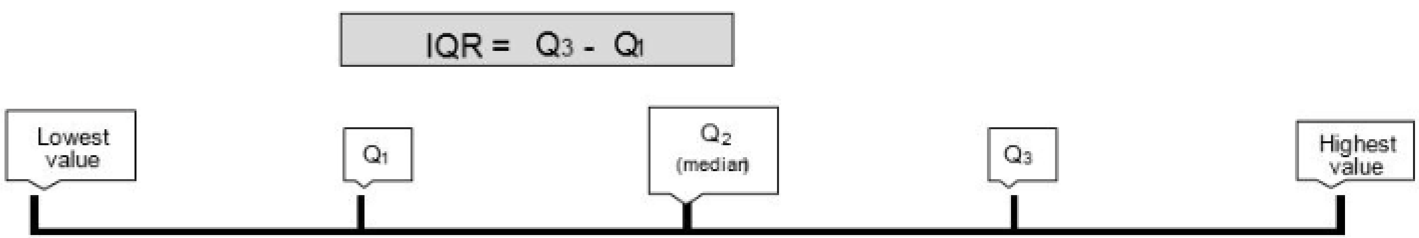
Measure of variability

Interquartile range

The interquartile range (IQR) is the range of the middle 50% scores in a distribution:

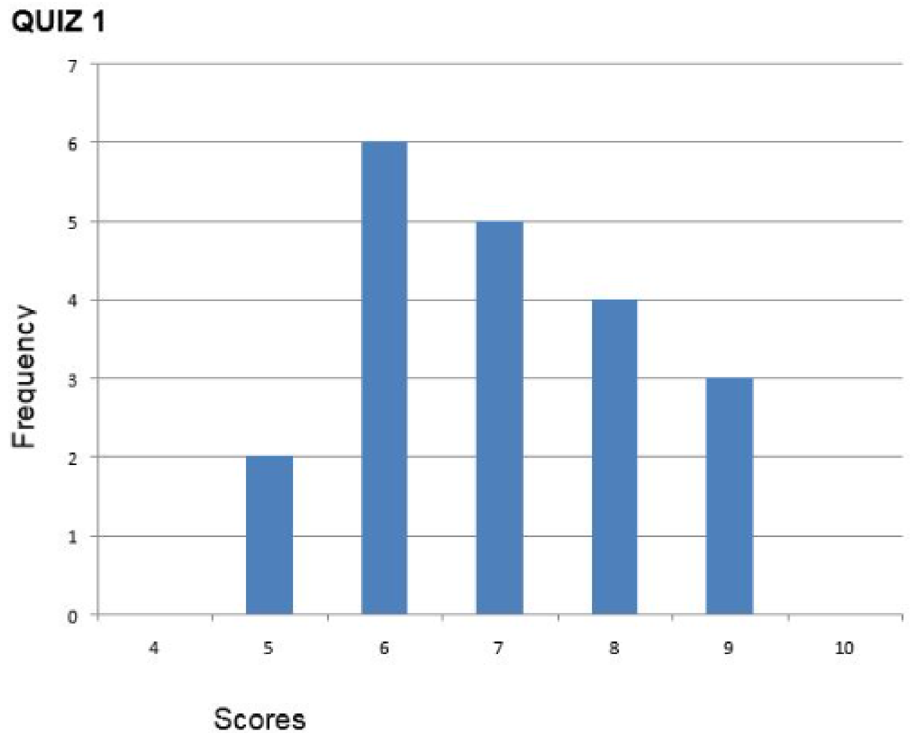
IQR= 75th percentile – 25th percentile

It is based on dividing a data set into quartiles. Quartiles are the values that divide scores into quarters. Q1 is the lower quartile and is the middle number between the smallest number and the median of a data set. Q2 is the middle quartile-or median. Q3 is the upper quartile and is the middle value between the median set and the highest value of a data set. The interquartile range formula is the first quartile subtracted from the third quartile



Measure of variability

Example:

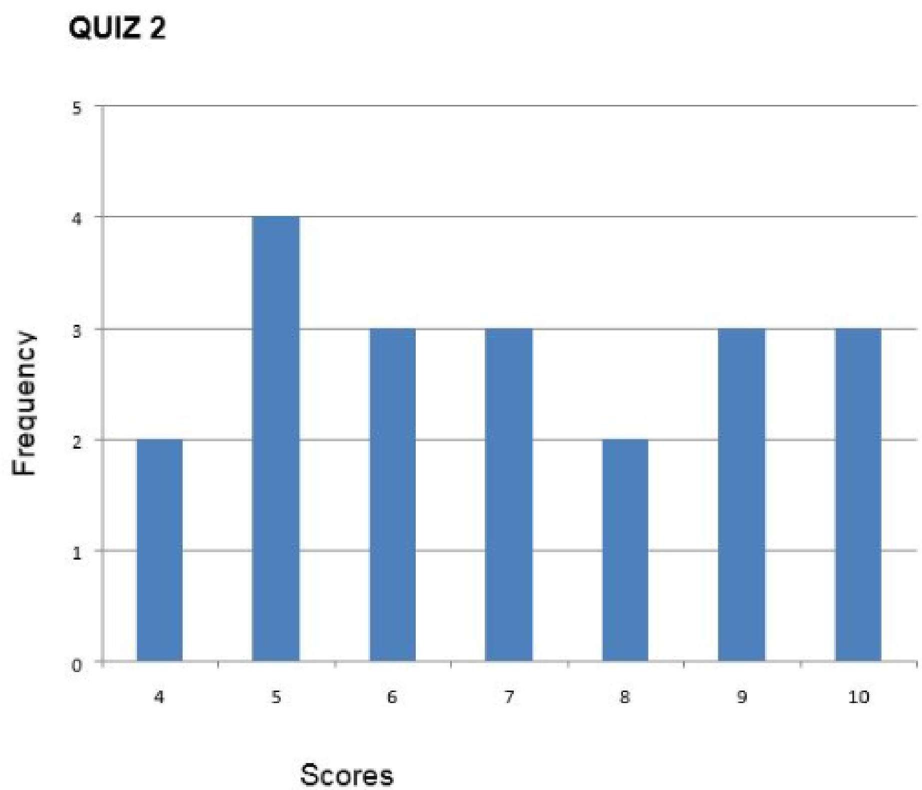


For *Quiz 1*, Q3 is 8 and Q1 is 6 . These are the scores:
5, 6, 7, 8, 9

If the median is 7, then Q1 is 6 (middle value between median and lowest value) and Q3 is 8 (middle value between median and highest value).
To calculate the IQR:
IQR= 8-6= 2

Measure of variability

Example:



For *Quiz 2*, Q3 is 9 and Q1 is 5. These are the scores:

4, 5, 6, 7, 8, 9, 10

The median is 7. To find Q1, we'll look at the lower half section of the distribution of scores: 4,5,6.

Q 1 is the median of this section of the distribution: 5

To find Q2, we'll look at the upper half section of the distribution of scores: 8, 9,10. Q3 is the median of this section of the distribution: 9.

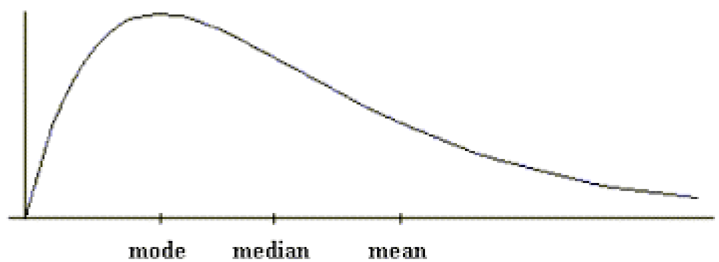
To calculate the IQR, knowing Q1 and Q3:

IQR= 9-5= 4

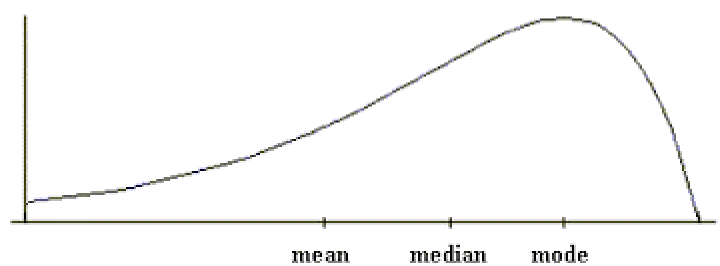
Measure of variability

Shapes of distributions: skewness and kurtosis
Distributions can be asymmetrical or skewed; that is, the tail of the distribution in the positive direction extends further than the tail in the negative direction, or vice versa.

A distribution with the longer tail extending in the positive direction is said to have a positive skew; it is skewed to the right.

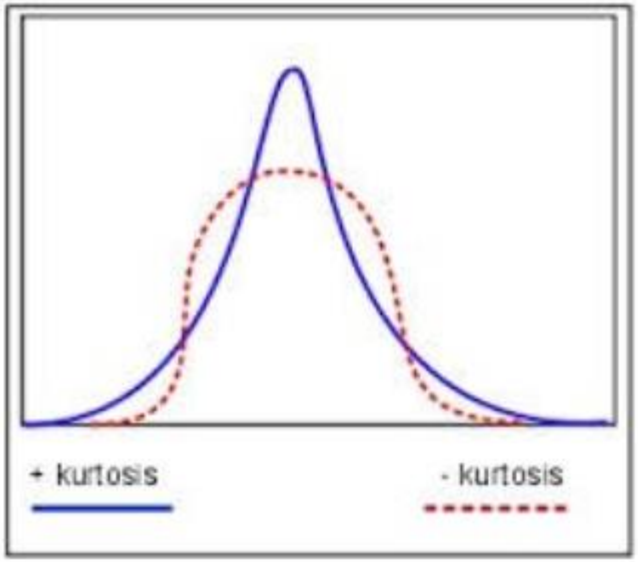


A distribution with the longer tail extending to the left is negatively skewed, or skewed to the left:



Measure of variability

Distributions also differ in terms of whether the data are peaked or flat. Distributions with positive kurtosis have a distinct peak near the mean and decline rapidly, whilst distributions with negative kurtosis tend to be more flat:



Measure of variability

Normal distributions

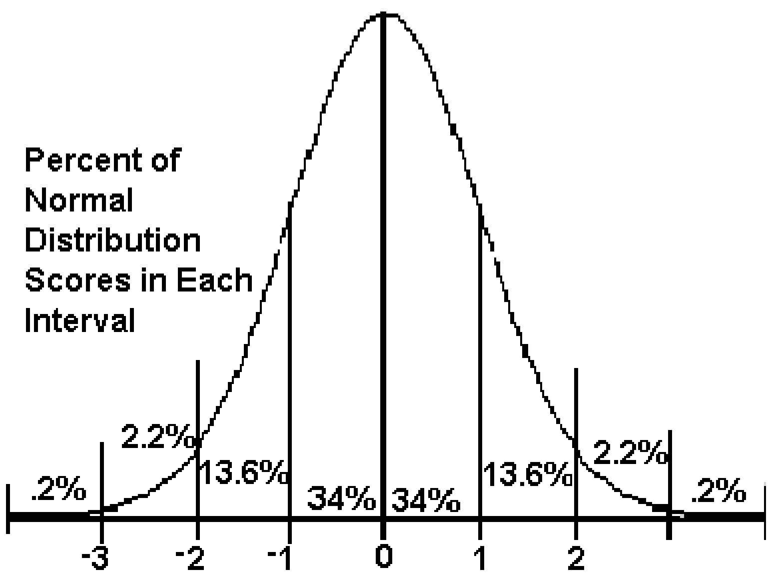
The normal distribution is the most important and commonly used distribution in statistics. It is also known as the bell curve or Gaussian curve. Even though normal distributions can differ in their means and standard deviation, they share some characteristics related to the distribution of scores:

- they are symmetric around their mean
- the mean, median, and mode are equal
- are denser in the center and less dense in the tails
- are defined by two parameters: the mean and standard deviation
- 68% of the area is within one standard deviation of the mean
- approximately 95% of the area is within two standard deviations of the mean
- 99.7 % of the area is within 3 standard deviations of the mean.

Measure of variability

The 68-95-99.7% rule

Standard normal deviations follow the 68-95-99.7% rule:



- 68% of scores are within one standard deviation of the mean.
- 95% of scores are within 2 standard deviations of the mean.
- 99.7% of scores are within 3 standard deviations of the mean.