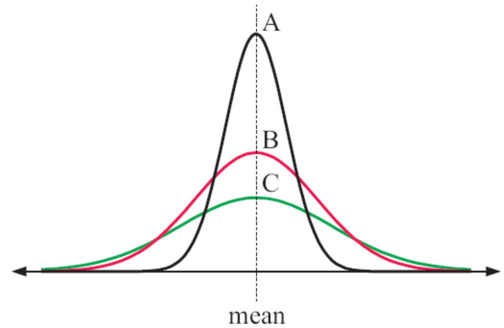
11-3 Measures of Variability

Dec 12-11:01 AM



Nov 6-8:01 AM

Variance

The sample variance, s^2, is the sum of the squared differences between each observation and the sample mean divided by the sample size minus 1.

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

The population variance, σ^2, is the sum of the squared differences between each observation and the population mean divided by the population size, N.

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

Nov 6-8:07 AM

Standard deviation

The sample standard deviation, s_n, is the (positive) square root of the variance, and is defined as:

$$s_n = \sqrt{s_n^2} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}} \qquad \text{(IB)}$$

$$s_{n-1} = \sqrt{s_{n-1}^2} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{n}{n-1}} \cdot s_n$$

$$\text{or, } s_n = \sqrt{\frac{n-1}{n}} \cdot s_{n-1}$$

The population standard deviation is:

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

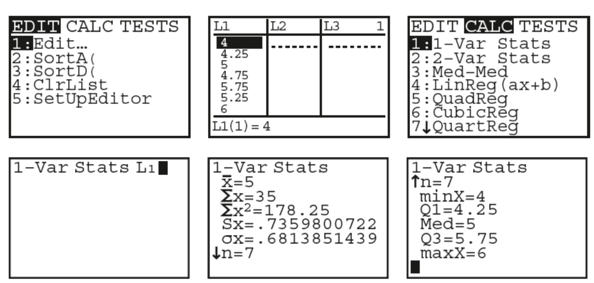
Nov 6-8:14 AM

In business, investors invest their money in stocks whose prices fluctuate with market conditions. Stocks are considered risky if they have high fluctuations. Here are the closing prices of two stocks traded on Vienna's stock market for the first seven business days in September 2007:

Stock A	Stock B
4	1
4.25	3
5	2.5
4.75	5
5.75	7
5.25	6.5
6	10
$\bar{x}_A = 5$ Median (A) = 5	$\bar{x}_B = 5$ Median (B) = 5

Nov 6-8:32 AM

Here is how you can use your TI GDC:



Nov 6-8:38 AM

The following sets of data are test results obtained by a group of students on two tests in which the maximum score was 20.

Test A

4	12	11	10	5	10	12	12	6	8	19	13	3
7	11	13	4	9	12	10	6	13	19	11	3	12
14	11	6	13	16	11	5	10	12	13	7	8	13
14	6	10	12	10	7	10	12	10				

Test B

9	8	10	10	8	9	10	11	8	8	11	10	9
8	11	10	9	8	10	11	8	9	11	10	9	8
11	11	9	9	11	10	8	9	11	19	8	9	11
11	8	8	11	10	8	9	10	10				

Nov 6-8:46 AM

a.) Find the mean and standard deviation of group A by hand.

b.) Find the mean and standard deviation of group B on you G.C.

c.) If the purpose of tests are to rank students in order of their performance, which test is more effective at doing that? Why?

Nov 7-10:16 AM

Percentiles and quartiles

Data must first be in ascending order.

Percentiles separate large ordered data sets into hundredths. The pth percentile is a number such that p per cent of the observations are at or below that number.

To score in the 90th percentile indicates 90% of the test scores were less than or equal to your score. An excellent performance! You scored in the upper 10% of all persons taking the test.

Quartiles are descriptive measures that separate large ordered data sets into four quarters.

Nov 7-10:25 AM

- **First quartile, Q_1**
  The first quartile, Q_1, is another name for the 25th percentile. The first quartile divides the ordered data such that 25% of the observations are at or below this value. Q_1 is located in the $0.25(n + 1)$st position when the data is in ascending order. That is,

  $$Q_1 = \frac{n+1}{4} \text{ ordered observation}$$

- **Third quartile, Q_3**
  The third quartile, Q_3, is another name for the 75th percentile. The third quartile divides the ordered data such that 75% of the observations are at or below this value. Q_3 is located in the $0.75(n + 1)$st position when the data is in ascending order. That is,

  $$Q_3 = \frac{3(n+1)}{4} \text{ ordered observation}$$

- **The median**
  The median is the 50th percentile, or the second quartile, Q_2.

Nov 7-11:13 AM

Stock A	Stock B
4	1
4.25	3
5	2.5
4.75	5
5.75	7
5.25	6.5
6	10
$\bar{x}_A = 5$ Median (A) = 5	$\bar{x}_B = 5$ Median (B) = 5

Nov 7-11:15 AM

Interquartile range

The interquartile range (IQR) measures the spread in the middle 50% of the data; it is the difference between the observations at the 25th and the 75th percentiles:

$$\text{IQR} = Q_3 - Q_1$$

Five-number summary

Five-number summary refers to the five descriptive measures: minimum, first quartile, median, third quartile, maximum.
Clearly, $X_{minimum} < Q_1 < \text{Median} < Q_3 < X_{maximum}$.

Nov 7-11:21 AM

Box-and-whisker plot

Whenever we have a five-number summary, we can put the information together in one graphical display called a **box plot**, also known as a **box-and-whisker** plot. In the student expenditure data, the IQR is €11. This is evident in the box plot below, where the IQR is the difference between 50 and 61.

Let us make a box plot with the student expense data.

- Draw an axis spanning the range of the data. Mark the numbers corresponding to the median, minimum, maximum, and the lower and upper quartiles.
- Draw a rectangle with lower end at Q1 and upper end at Q3, as shown below.
- To help us consider outliers, mark the points corresponding to lower and upper fences. Mark them with a dotted line since they are not part of the box. The fences are constructed at the following positions:
  - Lower fence: $Q_1 - 1.5 \times \text{IQR}$ (Here it is $50 - 1.5(11) = 33.5$.)
  - Upper fence: $Q_3 + 1.5 \times \text{IQR}$ (Here it is $61 + 1.5(11) = 77.5$.)

  Any point beyond the lower or upper fence is considered an **outlier**.
- Mark any outlier with an asterisk (*) on the graph (shown below).
- Extend horizontal lines called 'whiskers' from the ends of the box to the smallest and largest observations that are not outliers. In the first case these are 38 and 68, while in the second they are 38 and 67.

Nov 7-11:23 AM

Test A

4	12	11	10	5	10	12	12	6	8	19	13	3
7	11	13	4	9	12	10	6	13	19	11	3	12
14	11	6	13	16	11	5	10	12	13	7	8	13
14	6	10	12	10	7	10	12	10				

Test B

9	8	10	10	8	9	10	11	8	8	11	10	9
8	11	10	9	8	10	11	8	9	11	10	9	8
11	11	9	9	11	10	8	9	11	19	8	9	11
11	8	8	11	10	8	9	10	10				

a.) Find the median, Q1, and Q3 for each test.

b.) Find the interquartile range.

c.) Find the 5 number summary and create a box and whisker plot that describes the data.

d.) Using the box and whisker plot for the two tests, compare and contrast the spread of the data.

Nov 7-11:17 AM