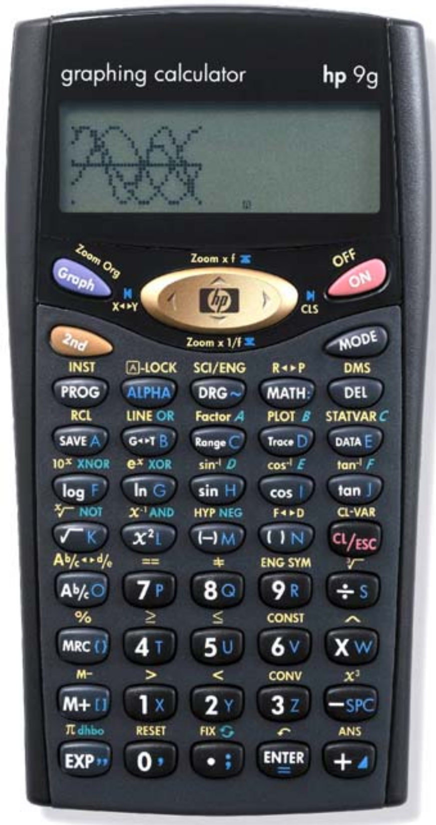
hp calculators



HP 9g Statistics – Averages and Standard Deviations

Average, Standard Deviation and other Statistics

Practice Finding Averages and Standard Deviations

Average, standard deviation and other statistics

The HP9g provides several functions to calculate statistics, i.e. quantities that describe some properties of a sample or of the whole population (in this last case, some authors prefer the term parameter), namely:

- Average or arithmetic mean (symbols: $\bar{x}, \mu$).The average of n quantities $x_1, x_2, \ldots, x_n$ is defined as the sum of the quantities divided by the number of quantities:

$$\bar{x} = \frac{\sum x_i}{n}$$

  These quantities can have frequencies $f_1, f_2, \ldots, f_n$ so that $\sum f_i = n$. In such case the average is $\left(\sum f_i x_i\right)/n$. A similar concept is that of the weighted average. The weighted average of n quantities each having weights $w_1, w_2, \ldots, w_n$ is $\left(\sum w_i x_i\right)/\left(\sum w_i\right)$. On the HP 9g averages can be calculated using the MATH function AVG (MATH MATH 3 Z) provided the number of values is not greater than ten; and also selecting the $\bar{x}$ and $\bar{y}$ options of the STATVAR menu in 1-VAR STAT mode (up to 30 data plus frequencies) or 2-VAR STAT mode (up to 30 pairs of data). Even though no specific weighted average function is provided, it can be easily calculated as shown in one of the examples below.

- Sample and population standard deviations (symbols: S and σ, respectively). The standard deviation is a measure of how dispersed the data values are about the average. The difference between the sample and the population standard deviation is that the former assumes the data is a sampling of a larger, complete set of data, whereas the latter assumes the data constitutes the complete set of data, and can be calculated as follows:

$$\sigma = \sqrt{\frac{\sum x_i^{\,2} - \frac{\left(\sum x_i\right)^2}{n}}{n}} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

  where n is the number of data points. The sample standard deviation is calculated using n – 1 as the divisor. The HP 9g can also calculate grouped standard deviation (when data points occur at given frequencies). It can be proved (Tchebycheff's inequality) that between the mean and $\pm k\cdot\sigma$ are at least $100 \cdot (1 - k^{-2})\%$ of the data points, regardless of the distribution of the data. (This is also true for the sample standard deviation, because $\sigma = S\sqrt{(n-1)/n} \Rightarrow S > \sigma$). The standard deviation cannot be negative. Its square is known as the variance.

- Minimum, maximum and range. The HP 9g functions MAX (MATH MATH 0 ,) and MIN (MATH MATH 1 X) return the largest (maximum) number and the smallest (minimum) number in the data set (up to ten numbers separated by commas). These values are also provided by the STATVAR menu in STAT mode. In this menu is also the range function, which is the difference between the maximum and the minimum.

- Coefficient of variation. (CV). This coefficient is a measure of the variability of the data. It is useful in comparing the *dispersions* of two data sets with different means. The square of CV is known as relative variance or relvariance. On the HP 9g, this statistic is included in the STATVAR menu, and is given as a percentage. It is defined as:

$$CV_x = \frac{S_x}{\bar{x}} \cdot 100$$

Practice finding averages and standard deviations

<u>Example 1:</u> The sales price of the last 10 homes sold in the Parkdale community were: \$198,000; \$185,000; \$205,200;\$225,300; \$206,700; \$201,850; \$200,000; \$189,000; \$192,100; \$200,400. What is the average of these sales prices and what is the sample standard deviation? Would a sales price of \$246,000 be considered unusual in the same community?

<u>Solution:</u> First of all press MODE 1 D-CL ENTER to ensure that no data remains from previous calculations. Now, let's input the data in 1-VAR STAT mode:

MODE 1 1-VAR ENTER DATA ENTER
1 9 8 EXP 3 ▼ ▼ 1 8 5 EXP 3 ▼ ▼ 2 0 5 2 0 0 ▼ 2 2 5 3 0
0 ▼ ▼ 2 0 6 7 0 0 ▼ ▼ 2 0 1 8 5 0 ▼ ▼ 2 EXP 5 ▼ ▼ 1 8 9
EXP 3 ▼ ▼ 1 9 2 1 0 0 ▼ ▼ 2 0 0 4 0 0 ▼

The average and standard deviation are both shown in the STATVAR menu. Press 2nd STATVAR ▶ to display the average and then ▶ to display the sample standard deviation.

<u>Answer:</u> The average of the sales prices is \$200,355 and the sample standard deviation is \$11,189.04. Within four standard deviations on either side of this average, i.e. between \$155,598.83 and \$245,111.18, *at least* [1] $100 \cdot \left(1 - \frac{1}{4^2}\right) = 93.75\%$ of all home sales prices will fall. If a home were to sell for \$246,000 in this area, it would be an unusual event.

<u>Example 2:</u> Below is a chart of daily high and low temperatures for a July week in Buenos Aires, Argentina. What were the average high and low temperatures for that week?

	Sunday	Monday	Tuesday	Wed.	Thurs.	Friday	Sat.
High	11	14	10	8	9	8	7
Low	1	0	–1	–6	–5	–4	–3

<u>Solution:</u> We'll input the data in 2-VAR mode, storing the high temperatures in x_i and the low ones in y_i:

MODE 1 D-CL ENTER MODE 1 2-VAR ENTER DATA ENTER
1 1 ▼ ▼ 1 4 ▼ 0 ▼ 1 0 ▼ (−) 1 ▼ 8 ▼ (−) 6 ▼ 9 ▼ (−) 5 ▼ 8 ▼
(−) 4 ▼ 7 ▼ (−) 3 ▼

Press 2nd STATVAR ▶ to display the average high temperature ($\bar{x}$) and then ▼ ▼ ▼ to display the average low temperature ($\bar{y}$).

<u>Answer:</u> The average high and low temperatures were 9.6 and –2.6, respectively.

[1] Remember that this is true regardless of how the data is distributed. Depending on the distribution, this percentage can actually increase. For example, if the data is normally distributed, 95.5% of the data points will fall within $\mu \pm 2\sigma$.

Example 3: What was the average price per gallon of gasoline Emma purchased from four locations shown below?

Gallons	15	7	10	17
Cost per gallon	$1.56	$1.64	$1.70	$1.58

Solution: In this case we have to calculate a weighted average. The HP 9g does not have a weighted average mean calculation built-in, but does provides some calculations that will solve this problem easily. The weighted average is defined as: $\overline{X}_w = \left(\sum w_i x_i\right) / \left(\sum w_i\right)$. If we store the number of gallons purchased in y_i, then it can be calculated as follows:

$$\overline{X}_w = \frac{\sum xy}{\sum y}$$

where the $\sum xy$ and $\sum y$ calculations are both included in the STATVAR menu (only in 2-VAR mode, which was already set in the previous example). Let's now input the data:

[MODE] [1] D-CL [ENTER] [DATA] [ENTER]
[1] [•] [5] [6] [▼] [1] [5] [▼] [1] [•] [6] [4] [▼] [7] [▼] [1] [•] [7] [▼] [1] [0] [▼] [1] [•] [5] [8] [▼]
[1] [7] [▼]

The weighted average is calculated as follows:

[2nd] STATVAR [▲] [▲] [▲] [▲] [◀] [÷] [2nd] STATVAR [▲] [▲] [▲] [▲] [ENTER] [ENTER]

While scrolling through the STATVAR menu, an error message can be displayed for a particular variable (e.g. OUT OF SPEC or DIVIDE BY 0) because it is not defined for our data, in such cases just press the cursor key again to select the next variable, there's no need to press [CL/ESC] (which exits the menu) to clear these warning messages!

Answer: The average price per gallon Emma has paid is slightly less than $1.61.

Example 4: Judging by the coefficient of variation, what can we say for the following data if it comes from the same population?

1045	3200	13	25	45	290	970	8

Solution: Let's input the data in 1-VAR mode. Press:

[MODE] [1] D-CL [ENTER] [MODE] [1] 1-VAR [ENTER] [DATA] [ENTER]
[1] [0] [4] [5] [▼] [▼] [3] [2] [0] [0] [▼] [▼] [1] [3] [▼] [▼] [2] [5] [▼] [▼] [4] [5] [▼] [▼] [2] [9] [0] [▼] [▼]
[8] [▼]

To display the coefficient of variation press: [2nd] STATVAR [▼] [▼]

Answer: Rounding to two decimal digits, CV=157.03%. The coefficient of variation of positive data coming from a homogeneous population is normally less than 100%. If it is greater than 150%, the data probably comes from *heterogeneous sources* (e.g. from people of different sex, age, etc.)