hp calculators

HP 9s Statistics – Averages and Standard Deviations

Average and Standard Deviation

Practice Finding Averages and Standard Deviations

Average and standard deviation

The HP9s provides several functions to calculate statistics, i.e. quantities that describe some properties of a sample or of the whole population (for the latter 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 9s averages and the other statistics can be calculated in Statistics mode (2ndF STAT) provided the number of different values is not greater than 80 (although the frequency or the number of occurrences of each item can be up to 255).

- 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. They can be calculated as follows:

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

$$S = \sqrt{\frac{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}{n-1}} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

  where n is the number of data points. Note that the sample standard deviation is calculated using n – 1 as the divisor. The HP 9s 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. The standard deviation cannot be negative. Its square is known as the variance.

- $\sum x$ and $\sum x^2$ which are useful in calculating other statistics.

Practice finding averages and standard deviations

Example 1: The following table shows the number of votes obtained by a political party in all the local elections in Barcelona since the democracy was restored in Spain:

1979	1983	1987	1991	1995	1999	2003
272512	412991	400280	328282	347083	313623	254223

What is the average of these results? What is the standard deviation?

Solution: First of all, let's enter Statistics mode by pressing 2ndF STAT . Every time we select another mode, the statistics data is cleared so that we can be confident that no data remains from previous calculations. Now, let's input the number of votes:

DATA DEL [1] 2 7 2 5 1 2 DATA DEL 4 1 2 9 9 1
DATA DEL 4 0 0 2 8 0 DATA DEL 3 2 8 2 8 2
DATA DEL 3 4 7 0 8 3 DATA DEL 3 1 3 6 2 3
DATA DEL 2 5 4 2 2 3 ENTER

Note that the ENTER key enters the seventh value.

To calculate the average press $\bar{x}$ Σx^2 (i.e. X–M). The data can be considered as a population because it comprises all the local elections called since the restoration of the democracy in that country. Therefore, we are interested in the population standard deviation, which can be found by pressing 2ndF MR (if the 2ndF key is not pressed first, it is the sample standard deviation that will be displayed).

Answer: Rounding to the nearest vote, the average of the number of votes is 332713 and the population standard deviation is 55272. Note that *at least* [2] $100 \cdot (1 - \frac{1}{2^2}) = 75\%$ of the results of this party will fall within two standard deviations on either side of this average, i.e. between 222169 and 443257, unless the population (i.e. the society) changes!

Example 2: Below is a chart of daily high and low temperatures for a week of July 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

Solution: We must solve this problem in two steps since there's no "2-VAR" mode on the HP 9s. First, we'll calculate the average high temperature. Press 2ndF STAT 2ndF STAT to clear the data entered in the previous example. And then press:

DATA DEL 1 1 DATA DEL 1 4 DATA DEL 1 0 DATA DEL 8 DATA DEL 9 DATA DEL 8 DATA DEL 7 ENTER

The average is displayed by pressing $\bar{x}$ Σx^2 (i.e. X–M). Now, clear the data again and enter the low temperatures:

2ndF STAT 2ndF STAT
DATA DEL 1 DATA DEL 0 DATA DEL 1 +/– DATA DEL 6 +/– DATA DEL 5 +/– DATA DEL 4 +/– DATA DEL 3 +/– ENTER

Once again, press $\bar{x}$ Σx^2 (i.e. X–M) to find the average.

[1] That is, press M+ . The 2ndF key should not be pressed because it would select the DEL function instead. In STAT mode the memory functions are not available, but the M register is *not* lost.
[2] Remember that this is true regardless of the way the data is distributed. Depending on the distribution, this percentage can actually increase. For example, if the data is normally distributed (which seems unlikely in this example), 95.5% of the data points will fall within $\mu \pm 2\sigma$.

Answer: The average high and low temperatures were 9.6 and –2.6, respectively.

Example 3: Emma has bought gas this week while showing houses at four gasoline stations as follows:

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

What is the average price of the gasoline purchased?

Solution: In this case we have to calculate a weighted average. You won't find a function on the HP 9s keyboard to calculate weighted averages; but ,as noted on page 2, the weighted average calculation is mathematically equivalent to the calculation of the average of grouped data (i.e. data that occurs with given frequencies). Therefore, the average price can be calculated as follows:

2ndF STAT 2ndF STAT (to clear the statistics data):

DATA DEL 1 • 5 6 x 1 5 (data items 1 through 15 are equal to 1.56),
DATA DEL 1 • 6 4 x 7 (data items 16 through 22 are 1.64),
DATA DEL 1 • 7 x 1 0 (data items 23 through 32 are 1.70),
DATA DEL 1 • 5 8 x 1 7 ENTER (data items 33 through 49 are 1.58)

There are 49 items in total: $n = \sum w_i = 15 + 7 + 10 + 17 = 49$. Simply press $\bar{x}$ Σx^2 (i.e. X-M) to display the answer.

Answer: The average price per gallon Emma has paid this week while showing houses is slightly less than \$1.61.

Example 4: Find the second moment about the origin and the coefficient of variation of the following data. If the data comes from the same population, what can we say about it?

1045	3200	13	25	45	290	970	8

Solution: The *r*th moment about a value *a* is defined as: $m_r = \frac{\sum (x-a)^r}{n}$. If *a* = 0 and *r* = 2 then $m_2 = \frac{\sum x^2}{n}$.

The coefficient of variation is defined as: $CV = \frac{S}{\bar{x}}$. It is often given as a percentage, that is: $(S/\bar{x}) \times 100$.

Let's input the data—remember to clear the previous data first: 2ndF STAT 2ndF STAT :

DATA DEL 1 0 4 5 DATA DEL 3 2 0 0 DATA DEL 1 3 DATA DEL 2 5 DATA DEL 4 5 DATA DEL 2 9 0 DATA DEL 9 7 0 DATA DEL 8 ENTER

We can now find the second moment by pressing:

2ndF $\bar{x}$ Σx^2 ÷ n Σx (i.e.)) ENTER

To display the coefficient of variation press: S σ (i.e. MR) ÷ $\bar{x}$ Σx^2 (i.e. X-M) ENTER .

Answer: $m_2 = 1544988.5$. Rounding to two decimal digits, CV=1.57, or 157%. 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.)