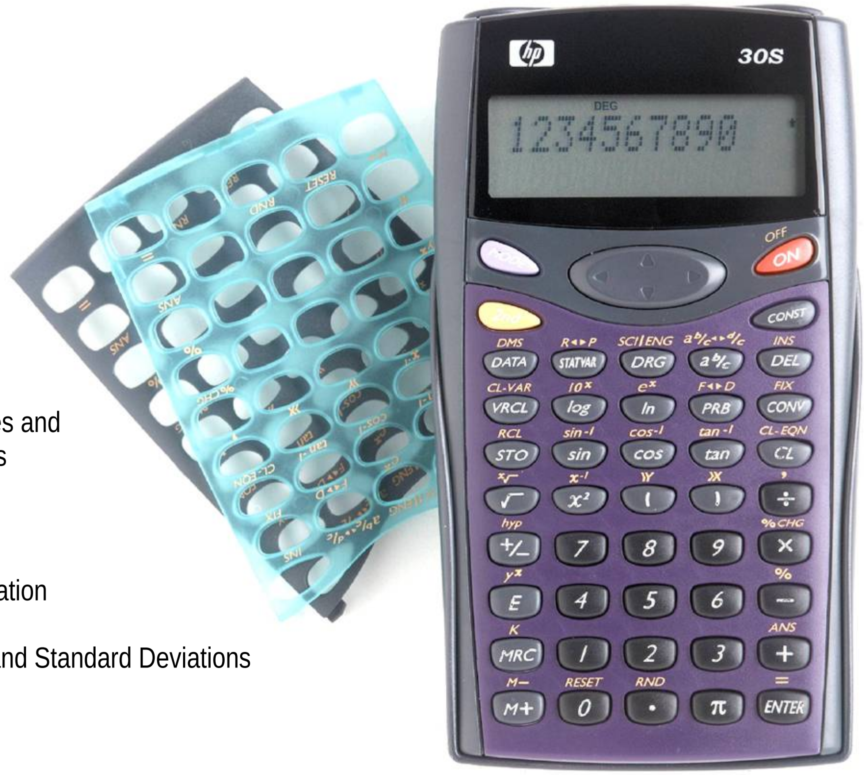
hp calculators



HP 30S Statistics – Averages and
Standard Deviations

Average and Standard Deviation

Practice Finding Averages and Standard Deviations

Average and standard deviation

The HP 30S 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, …, x_n is defined as the sum of the quantities divided by the number of quantities:

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

  These quantities can have frequencies f_1, f_2, …, f_n so that $\sum f_i = n$, in which 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,…, w_n is $\left(\sum w_i x_i\right)/\left(\sum w_i\right)$. On the HP 30S averages can be calculated by selecting the $\bar{x}$ and $\bar{y}$ options of the STATVAR menu in 1-VAR STAT mode (up to 40 data items plus frequencies) or 2-VAR STAT mode (up to 40 pairs of data items). 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. They can be calculated as follows:

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

  where n is the number of data points. The sample standard deviation is calculated using n – 1 as the divisor. The HP 30S 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.

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

Practice finding averages and standard deviations

Example 1: 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?

Solution: First of all press MODE / select CLR-DATA ENTER to be sure that no statistical data remains from previous calculations. Now, let's input the data in 1-VAR STAT mode:

MODE 1 select 1-VAR ENTER DATA

1 9 8 E 3 ▼ ▼ 1 8 5 E 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 E 5 ▼ ▼ 1 8 9 E 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 STATVAR ▶ to display the average and then ▶ to display the sample standard deviation.

Answer: 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] $100 \cdot (1 - \frac{1}{4^2}) = 93.75\%$ 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.

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'll input the data in 2-VAR mode, storing the high temperatures in x_i and the low ones in y_i:

MODE 1 select CLR-DATA ENTER MODE 1 select 2-VAR ENTER DATA

1 1 ▼ ▼ 1 4 ▼ 0 ▼ 1 0 ▼ +/– 1 ▼ 8 ▼ +/– 6 ▼ 9 ▼ +/– 5 ▼ 8 ▼ +/– 4 ▼ 7 ▼ +/– 3 ▼

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

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 30S STATVAR menu to calculate weighted averages; but, as noted on page 2, the weighted average calculation is

[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 $\mu \pm 2\sigma$.

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:

[MODE] [1] select CLR-DATA [ENTER] [MODE] [1] select *1-VAR* [2] [ENTER] [DATA]

[1] [.] [5] [6] ▼ [1] [5] ▼ [1] [.] [6] [4] ▼ [7] ▼ [1] [.] [7] ▼ [1] [0] ▼ [1] [.] [5] [8] ▼ [1] [7] ▼

Now, simply press [STATVAR] ► to display the answer.

Alternatively, since the weighted average is defined as $\bar{x}_w = \left(\sum w_i x_i\right)/\left(\sum w_i\right)$, we can calculate it as follows:

$$\bar{x}_w = \frac{\sum xy}{\sum y}$$

provided the number of gallons purchased is stored in y_i, as above but this time in 2-VAR mode because the $\sum xy$ and $\sum y$ calculations are present in the STATVAR menu only in 2-VAR mode:

[MODE] [1] select 2-VAR [ENTER] [STATVAR] ◄ ◄ ◄ [÷] [STATVAR] ◄ ◄ ◄ ◄ [ENTER] [ENTER]

Answer: The average price per gallon Emma has paid this week while showing houses 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: 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 in 1-VAR mode—remember to clear the previous data first, press [MODE] [1] select CLR-DATA [ENTER] , and then:

[MODE] [1] select 1-VAR [ENTER] [DATA]

[1] [0] [4] [5] ▼ ▼ [3] [2] [0] [0] ▼ ▼ [1] [3] ▼ ▼ [2] [5] ▼ ▼ [4] [5] ▼ ▼ [2] [9] [0] ▼ ▼ [9] [7] [0] ▼ ▼ [8] ▼

We can now find the second moment by pressing:

[STATVAR] ◄ [÷] [STATVAR] [ENTER] [ENTER]

[2] In 2-VAR mode the number of gallons purchased will be considered an independent variable, and therefore the resulting average will be wrongly calculated as (1.56 + 1.64 + 1.70 + 1.58) / 4 = 1.62.

To find the coefficient of variation, press: STATVAR ▶ ▶ ÷ STATVAR ▶ ENTER 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.)