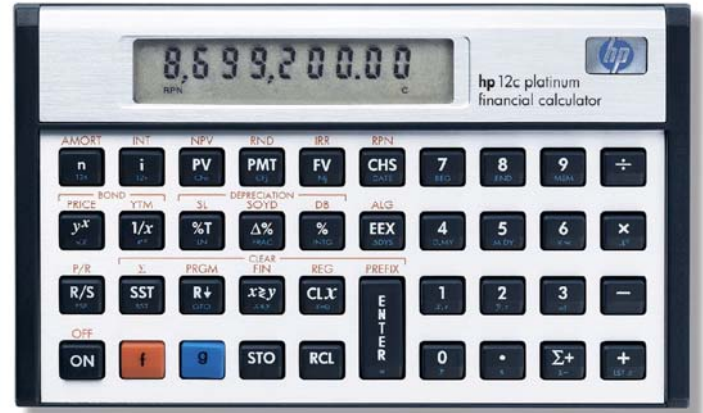**hp calculators**

HP 12C Platinum
Statistics - correlation coefficient

The correlation coefficient

HP12C Platinum correlation coefficient

Practice finding correlation coefficients and forecasting

The correlation coefficient

Statistics can be understood as a set of tools involving the study of methods and procedures for collecting, classifying, and analyzing data. Statistical tools also offer the means for making scientific inferences from summarized data, like linear regression and forecasting. Both linear regression and forecasting are computed with the parameters that define a curve (or line) that best represents the sample, given the plotted data of a two-variable sample is close enough to a straight line that it can be represented this way. The correlation coefficient r is in the range ($-1 \leq r \leq 1$) and measures how close the sample is to a straight line. The closer the sample is to a straight line, the closer $|r|$ will be to 1 and the closer the forecasted points will be to the sample behavior.

HP12C Platinum correlation coefficient

On the HP12C Platinum, statistics data are stored as a set of summations resulting from the originally collected data. The collected data set must be typed in prior to using any statistics features available in the HP12C Platinum. The HP12C Platinum memory organization allows the study of statistic data organized as one- or two-variable sample. As a general procedure, data is always collected as a pair of n numbers, or (x_n, y_n) values, and the HP12C Platinum computes the following summations:

$$\sum x_n \qquad \sum y_n \qquad \sum (x_n)^2 \qquad \sum (y_n)^2 \qquad \sum (x_n \times y_n) \qquad \text{Figure 1}$$

With these values updated and stored in memory, the HP12C Platinum computes the correlation coefficient with the following expression:

$$r = \frac{\left[\sum xy - \dfrac{\sum x \sum y}{n} \right]}{\sqrt{\left[\sum x^2 - \dfrac{\left(\sum x\right)^2}{n} \right] \times \left[\sum y^2 - \dfrac{\left(\sum y\right)^2}{n} \right]}} \qquad \text{Figure 2}$$

There are also functions to forecast values for each of the two variables. These functions are associated to $\boxed{\hat{x},r}$ (forecast x given y) and $\boxed{\hat{y},r}$ (forecast y given x) keys. Each time a new forecast is calculated, the correlation coefficient is computed too, and pressing $\boxed{x \gtrless y}$ right after $\boxed{g}\,\boxed{\hat{x},r}$ or $\boxed{g}\,\boxed{\hat{y},r}$ shows its value. The HP12C Platinum uses the following expressions to compute the forecasted values:

$$\hat{y} = A + Bx \qquad\qquad \hat{x} = \frac{y - A}{B}$$

given:

$$B = \frac{\sum xy - \dfrac{\sum x \sum y}{n}}{\sum x^2 - \dfrac{\left(\sum x\right)^2}{n}} \qquad \text{and} \qquad A = \bar{y} - B\bar{x}$$

Figure 3

Practice finding the correlation coefficient

Example 1: A land researcher wants to compute the relationship between the constructed area and the land area of a community in order to suggest the construction area for a new home with a land area of 12500 sq yards and the suggested land area for a construction with 3520 sq yards based on forecasted values. He also wants to know how good this relationship fits in a straight line to know if his suggestions are valid. The table below summarizes his measurements.

Land Area (sq yards)	Construction Area (sq yards)	Land Area (sq yards)	Construction Area (sq yards)
12000	3120	9000	2080
10000	2560	10000	2700
11000	2920	13000	3280
14000	3300	12000	3080

Figure 4

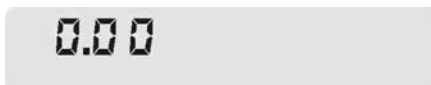Solution: Be sure to clear the statistics / summation memories before starting the problem.

f Σ

0.00

Figure 5

Then enter the first data point.

3 1 2 0 ENTER 1 2 0 0 0 Σ+

1.00

Figure 6

The first entered value (construction area) is used as the y variable and the second value (land area) is used as the x variable. The display shows the number of entries each time Σ+ is pressed. Make sure that all data is entered:
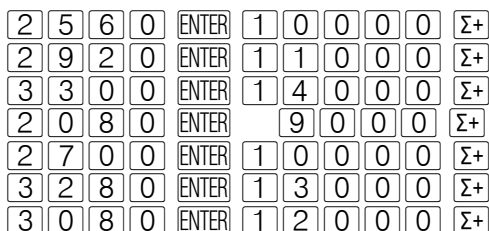
2 5 6 0 ENTER 1 0 0 0 0 Σ+
2 9 2 0 ENTER 1 1 0 0 0 Σ+
3 3 0 0 ENTER 1 4 0 0 0 Σ+
2 0 8 0 ENTER 9 0 0 0 Σ+
2 7 0 0 ENTER 1 0 0 0 0 Σ+
3 2 8 0 ENTER 1 3 0 0 0 Σ+
3 0 8 0 ENTER 1 2 0 0 0 Σ+

8.00

Figure 7

hp calculators

Since the y-values are the construction area, the forecast construction area (y) for a 12500 sq yards land area is computed by pressing:

⌐1⌐⌐2⌐⌐5⌐⌐0⌐⌐0⌐ ⌐g⌐⌐ŷ,r⌐

$$3,140.38$$

Figure 8

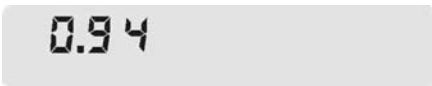The correlation coefficient is automatically computed each time a forecasting is performed. Press:

⌐x⋛y⌐

$$0.94$$

Figure 9

The land area is stored as x-values, so for a 3520 sq yards construction area, the forecasted land area is:

⌐3⌐⌐5⌐⌐2⌐⌐0⌐ ⌐g⌐⌐x̂,r⌐

$$14,140.22$$

Figure 10

Although the correlation coefficient has the same value for the same sample, it's easy to check for it:
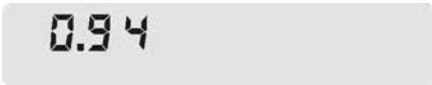
⌐x⋛y⌐

$$0.94$$

Figure 11

Answer: For a construction area of 3,520 sq yards, the estimated land area needed is approximately 14,140 sq yards. For a land area of 12,500 sq yards, a construction area of approximately 3,140 sq yards is recommended. The sample shows a correlation coefficient of 0.95, so the forecast is close to the actual data.

Example 2:   A stockholder observes the foreign stock market for some time in order to compose a curve that relates the amount of investment with amount of earnings in a particular brand. He decides to analyze the data and use the correlation coefficient to measure the margin of error when predicting the amount of earnings given the amount of investment. If the correlation coefficient is lower than 0.80, then he will not use the data to make future predictions. The collected data so far is as follows:

Investment amount	Amount of earnings	Investment amount	Amount of earnings
$1,200,000	$91,000	$1,450,000	$112,000
$1,000,000	$98,000	$1,300,000	$109,000
$900,000	$85,000	$1,150,000	$99,000

Solution: Be sure to clear the statistics / summation memories before starting the problem.
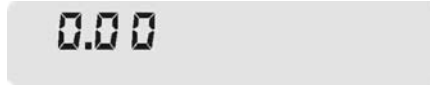
[f] [Σ]

```
0.00
```

Figure 12

Consider that each pair must be entered prior to add it to the statistics summations.
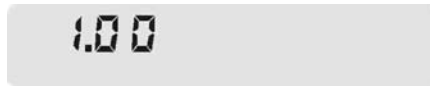
[1][2][0][0][0][0][0] [ENTER] [9][1][0][0][0] [Σ+]

```
1.00
```

Figure 13

Remember that the display shows the number of entries each time [Σ+] is pressed. Make sure that all data is entered:

[1][4][5][0][0][0][0] [ENTER] [1][1][2][0][0][0] [Σ+]
[1][0][0][0][0][0][0] [ENTER] [9][8][0][0][0] [Σ+]
[1][3][0][0][0][0][0] [ENTER] [1][0][9][0][0][0] [Σ+]
[9][0][0][0][0][0] [ENTER] [8][5][0][0][0] [Σ+]
[1][1][5][0][0][0][0] [ENTER] [9][9][5][0][0][0] [Σ+]

To compute the correlation coefficient, press:

[1] [g] [x̂,r] [x≷y]   (Note: Any value will do in place of the 1 shown here)

```
0.84
```

Figure 14

Answer: The correlation coefficient for the collected data is 0.84, and it is higher than the stockholder defined. Therefore, the available data will be used to predict future investments.