

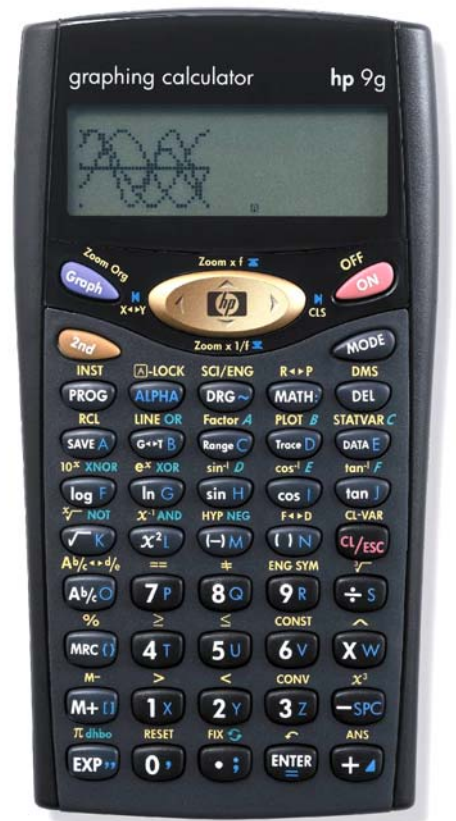


hp calculators

HP 9g Statistics – Non-Linear Regression

Non-Linear Regression

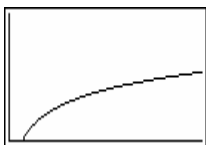
Practice Solving Non-Linear Regression Problems



Non-linear regression

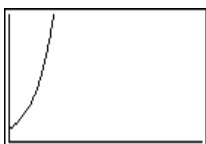
In addition to the linear regression (described in the HP 9g learning module *Statistics – Linear Regression*) the HP 9g provides five more types of regression of a variable y on a variable x , namely:

- ◆ **Logarithmic model:** $y = a + b \ln x$.



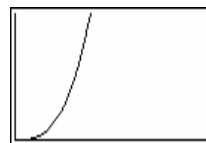
The logarithmic curve-fitting model requires positive x -values. An example of this model is the psychophysics law of Weber-Fechner.

- ◆ **Exponential model:** $y = a \cdot e^{bx}$.



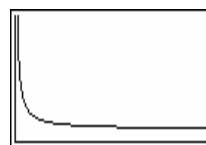
This model requires positive y -values. Examples of this model are the number of undecayed nuclei of radioactive materials after a period x of time, the discharging process in Capacitor-Resistance circuits, etc.

- ◆ **Power model:** $y = a \cdot x^b$.



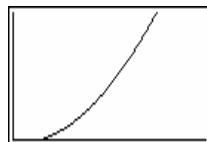
The power model requires positive x -values *and* positive y -values, otherwise the DOMAIN error occurs when attempting to display the STATVAR menu. Newton's law of gravitation is a classic example.

- ◆ **Inverse model:** $y = a + \frac{b}{x}$.



Zero cannot be one of the x -values, since it will trigger the DIVIDE BY 0. Examples are Curie-Weiss' law, the reactance of a capacitor, relation between potential energy and distance, etc.

- ◆ **Quadratic model:** $y = a + bx + cx^2$.



Bear in mind that this model needs at least *three* points. An example is the relation between kinetic energy and speed.

Regression curves can be found by following the same procedure used with linear regressions. That is, press MODE $(1X)$, select REG, and press ENTER . The REG menu is now displayed showing all the types of regression available. Select the desired one then press ENTER . Once the data has been entered, regression results are shown in the STATVAR menu (2nd STATVAR C), which displays the values of a and b (and c) and also the correlation coefficient r (except in the case of quadratic regression). Also in this menu are the functions \hat{y} and \hat{x} , which are used for predicting values. Let's illustrate all this with some examples.

Practice solving non-linear regression problems

Example 1: Enter the following points into your HP 9g as statistical data and plot them. Find the equation that best fits the points.

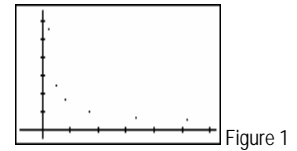
x	10	25	40	80	160	250
y	106	46	31	18.5	12.25	10

Solution: Let's first clear any previous data by pressing MODE $(1X)$, then select D-CL and press ENTER .

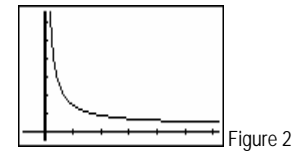
To plot the scatter graph (or scatterplot) we must set the 2-VAR mode, so press: MODE $(1X)$, select 2-VAR and press ENTER . Let's now enter the data: DATA E , select DATA-INPUT and ENTER , then press:

$(1X)$ (0) v $(1X)$ (0) $(6V)$ v
 $(2Y)$ $(5U)$ v $(4T)$ $(6V)$ v
 $(4T)$ (0) v $(3Z)$ $(1X)$ v
 $(8O)$ (0) v $(1X)$ $(8O)$ $(\cdot;)$ $(5U)$ v
 $(1X)$ $(6V)$ (0) v $(1X)$ $(2Y)$ $(\cdot;)$ $(2Y)$ $(5U)$ v
 $(2Y)$ $(5U)$ (0) v $(1X)$ (0) v

To plot the graph press: 2nd STATVAR C Graph . Figure 1 shows the resulting graph. These points seem to belong to a rectangular hyperbola. Let's try to verify it by finding the inverse regression equation. Press first: MODE $(1X)$, select REG, press ENTER , select INV and press ENTER again.



Regression variables are displayed in the 2nd STATVAR C menu, where we can read $a = 6$, $b = 1000$ and $r = 1$, which means that the fit is perfect. Without leaving the currently displayed menu, the curve can be plotted by pressing a single key: Graph . Figure 2 shows the displayed graph (notice the asymptote in the y axis).



Answer: The equation that best fits these points is $y = 6 + \frac{1000}{x}$.

Example 2: Using the data of the previous example, what are the predicted values of y for $x = 1000$ and for $x = 95$? Which prediction is more reliable? And what is the predicted value of x for $y = 35$?

Solution: As to the first question, we have to evaluate the regression line at $x = 1000$, and 95, that is to say: $y(1000)$ and $y(95)$. Actually, y and x are often written as \hat{y} and \hat{x} , respectively, when we are referring to predicted values. On the HP 9g their symbols are y' and x' and are displayed in the STATVAR menu. To calculate $\hat{y}(1000)$ and $\hat{y}(95)$, press:

2nd STATVAR C , select y' , ENTER $(1X)$ (0) (0) (0) v ENTER and v $(5U)$ $(5U)$ Cl/esc ENTER respectively.

Likewise, to find $\hat{x}(35)$, press:

2nd STATVAR C , select x' , ENTER $(3Z)$ $(5U)$ ENTER

Answer: $\hat{y}(95) = 16.52$, $\hat{y}(1000) = 7$ and $\hat{x}(35) = 34.48$ (to two decimal digits). $\hat{y}(95)$ is more reliable than $\hat{y}(1000)$ because the former is the result of an interpolation whereas the latter is an extrapolation.¹ Also, 95 is much closer to $\bar{x}=94.17$ than 1000 (\bar{x} being the arithmetic mean of the x values, and can be obtained by this sequence of keystrokes: $\text{2nd} \text{ MODE } \text{2-VAR}$, $\text{2nd} \text{ STATVAR } \text{C} \text{ } \text{>}$), which means that any variability in the former value will be far less significant than in the latter.

Example 3: The following data shows the number of AIDS cases diagnosed in the United States from the years 1981 to 1994. (From *The Spread of AIDS in the United States*, a data analysis activity for the HP 38G by Michael Grasse)

Year	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994
Cases	315	1184	3119	6276	11894	19195	28833	35662	42700	48092	58549	75908	73774	58225

Verify that from 1981 to 1987 the data fits an exponential model quite well. Considering all the years, what model seems to provide the best fit?

Solution: First, clear the previous data by pressing $\text{MODE} \text{ 1X}$, then selecting D-CL and finally pressing ENTER . Let's now enter the data, first press: DATA E , select DATA-INPUT and ENTER . (The 2-VAR mode is assumed to be set from the previous example). We'll enter the years 198x as x and 199x as 1x, but we'll enter up to 1987 for now:

$\text{1X} \text{ 3Z} \text{ 1X} \text{ 5U} \text{ 2Y} \text{ 1X} \text{ 1X} \text{ 8O} \text{ 4T} \text{ 3Z} \text{ 3Z} \text{ 1X} \text{ 1X} \text{ 9R} \text{ 4T} \text{ 6V} \text{ 2Y} \text{ 7P} \text{ 6V} \text{ 5U} \text{ 1X} \text{ 1X} \text{ 8O} \text{ 9R} \text{ 4T} \text{ 6V} \text{ 1X} \text{ 9R} \text{ 1X} \text{ 9R} \text{ 5U} \text{ 7P} \text{ 2Y} \text{ 8O} \text{ 8O} \text{ 3Z} \text{ 3Z}$

Let's find the correlation coefficient corresponding to an exponential relation:

$\text{MODE} \text{ 1X}$ select REG ENTER select "e^" $\text{ENTER} \text{ 2nd} \text{ STATVAR } \text{C} \text{ } \text{>}$

Rounding to four decimal digits $r = 0.9793$, which indicates that the exponential equation $y = 250.8619 e^{0.7307x}$ fits the data well unfortunately.

We'll enter now the rest of the data. Note that you don't need to be in the 2_VAR mode now, just press: DATA E , select DATA-INPUT ENTER , and then:

$\text{8O} \text{ 3Z} \text{ 5U} \text{ 6V} \text{ 6V} \text{ 2Y} \text{ 9R} \text{ 4T} \text{ 2Y} \text{ 7P} \text{ 0,} \text{ 0,} \text{ 1X} \text{ 0,} \text{ 4T} \text{ 8O} \text{ 0,} \text{ 9R} \text{ 2Y} \text{ 1X} \text{ 1X} \text{ 5U} \text{ 8O} \text{ 5U} \text{ 4T} \text{ 9R} \text{ 1X} \text{ 2Y} \text{ 7P} \text{ 5U} \text{ 9R} \text{ 0,} \text{ 8O} \text{ 1X} \text{ 3Z} \text{ 7P} \text{ 3Z} \text{ 7P} \text{ 7P} \text{ 4T} \text{ 1X} \text{ 4T} \text{ 5U} \text{ 8O} \text{ 2Y} \text{ 5U} \text{ 5U}$

¹ To understand the lack of validity in extrapolation, consider the number of AIDS diagnoses that the regression models in Example 3 predict for the year 1974, a year in which AIDS as a disease was unknown!

The correlation coefficient is now: $(\text{2nd STATVAR } C \gg) 0.9097$, which means that data (from 1981 to 1994) begins to diverge from an exponential model. In order to know what model best fits the data, we'll find the correlation coefficient associated with each of the models available on the HP 9g (except for the quadratic model because no correlation coefficient is provided. You can plot the resulting quadratic, however, and see how well it fits the data):

Linear fitting:	$(\text{MODE } 1 X \text{ REG } \text{ENTER} \text{ LIN } \text{ENTER} \text{ 2nd STATVAR } C \gg)$	$r = 0.9647$
Logarithmic fitting:	$(\text{MODE } 1 X \text{ REG } \text{ENTER} \text{ LOG } \text{ENTER} \text{ 2nd STATVAR } C \gg)$	$r = 0.8795$
Power fitting:	$(\text{MODE } 1 X \text{ REG } \text{ENTER} \text{ PWR } \text{ENTER} \text{ 2nd STATVAR } C \gg)$	$r = 0.9927$
Inverse fitting:	$(\text{MODE } 1 X \text{ REG } \text{ENTER} \text{ INV } \text{ENTER} \text{ 2nd STATVAR } C \gg)$	$r = -0.6694$

Solution: Of the considered models, the power curve $y = 3219.9124 \cdot x^{2.168}$ is the line that best fits the data. Notice that when taking all the years into account, the linear model is *better* than the exponential model.