Univariate linear regression of order m

A linear regression model of order m based on a single variable (univariate) can be expressed with the following equation:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \ldots\ldots\ldots + \beta_m X^m + \varepsilon = \sum_{i=0}^{m} \beta_i X^i + \varepsilon$$

If m = 1 this general equation can be reduced to that referred to simple linear regression, i.e., to first order univariate linear regression.

A more general expression for linear regression model of order m is:

$$Y = \underbrace{f(X, \beta_0 \beta_1 \ldots\ldots\ldots\ldots \beta_m)}_{\text{Deterministic component}} + \boxed{\varepsilon}$$

Deterministic component

Random component

For a single observation the equation becomes:

$$Y_i = f(X_i, \beta_0 \beta_1 \ldots\ldots\ldots\ldots \beta_m) + \varepsilon_i$$

Thus, the random error can be expressed as:

$$\varepsilon_i = Y_i - f(X_i, \beta_0 \beta_1 \ldots\ldots\ldots\ldots \beta_m)$$

The estimate of the m+1 (p) model parameters, β_0, β_1,…β_m, can be made using the Ordinary Least Squares (OLS) approach, that minimizes the sum of squared errors:

$$\min \sum_i \varepsilon_i^2$$

Supposing that n values of response Y are obtained by setting as many values of the independent variable (regressor) X, the following system of equations can be written:

$$\begin{cases} Y_1 = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \dots\dots + \beta_m X_1^m + \varepsilon_1 \\ Y_2 = \beta_0 + \beta_1 X_2 + \beta_2 X_2^2 + \dots\dots + \beta_m X_2^m + \varepsilon_2 \\ \dots\dots\dots\dots\dots\dots \\ Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \dots\dots + \beta_m X_i^m + \varepsilon_i \\ \dots\dots\dots\dots\dots\dots \\ Y_n = \beta_0 + \beta_1 X_n + \beta_2 X_n^2 + \dots\dots + \beta_m X_n^m + \varepsilon_n \end{cases}$$

The system can be solved more easily by adopting an approach based on matrices.

At this aim, the following column vectors are defined:

$$\mathbf{y} = \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ Y_i \\ . \\ . \\ Y_n \end{bmatrix} \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ . \\ \beta_i \\ . \\ . \\ \beta_m \end{bmatrix} \qquad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ \varepsilon_i \\ . \\ . \\ \varepsilon_n \end{bmatrix}$$

Responses vector (n×1)

Parameters vector (p×1)

Random errors vector (n×1)

In addition, matrix **X**, including values assumed by powers of the independent variable X (from X^0, i.e. 1, to X^m), is introduced:

$$\mathbf{X} = \begin{pmatrix} 1 & X_1 & X_1^2 & X_1^3 & . & . & X_1^m \\ 1 & X_2 & X_2^2 & X_2^3 & . & . & X_2^m \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 1 & X_n & X_n^2 & X_n^3 & . & . & X_n^m \end{pmatrix}$$

In this case the first column is made up only of «1» values, to account for the presence of the β_0 term in the model equation.

The system of n equations with p (i.e., m + 1) unknowns can thus be written in matricial notation:

$$\underset{(n\times1)}{\mathbf{y}} = \underset{(n\times p)}{\mathbf{X}}\underset{(p\times1)}{\boldsymbol{\beta}} + \underset{(n\times1)}{\boldsymbol{\varepsilon}}$$

The OLS criterion, i.e., finding $\min \sum_i \varepsilon_i^2$, can be written in matricial notation as: $\min(\boldsymbol{\varepsilon}^{\mathrm{T}}\boldsymbol{\varepsilon})$ i.e., finding the minimum of the scalar (or inner) product between column vector ε and its transpose (indicated as $\boldsymbol{\varepsilon}^{\mathrm{T}}$) .

Since:

$$\begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_n \end{bmatrix}^{\mathrm{T}} = [\, \boldsymbol{\varepsilon}_1 \; \boldsymbol{\varepsilon}_2 \; \ldots \; \boldsymbol{\varepsilon}_n \,]$$

$$\boldsymbol{\varepsilon}^{\mathrm{T}}\boldsymbol{\varepsilon} = \sum_i \varepsilon_i^2$$

Since, in matricial notation, $\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$, the quantity S to be minimized can be expressed as:

$$S = \sum_{i=1}^{n} \varepsilon_i^2 = \boldsymbol{\varepsilon}^{\mathrm{T}}\boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

It is worth noting that products $\mathbf{y}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta}$ and $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{y}$ are equal, thus S can be expressed as:

$$S = \mathbf{y}^{\mathrm{T}}\mathbf{y} - 2\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{y} + \boldsymbol{\beta}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})\boldsymbol{\beta}$$

By analogy with the OLS procedure referred to simple linear regression, the minimization of S is obtained by equalizing to 0 its first derivative with respect to $\boldsymbol{\beta}^{\mathrm{T}}$, i.e., to a row vector (the transpose of vector $\boldsymbol{\beta}$) including values of all regression parameters:

$$\frac{\partial S}{\partial \boldsymbol{\beta}^{\mathrm{T}}} = \mathbf{0}$$

Notably, the first term in the expression of S, i.e., the scalar product $\mathbf{y}^{\mathrm{T}}\mathbf{y} = y_1^2 + y_2^2 + \ldots. + y_k^2$, does not depend on $\boldsymbol{\beta}^{\mathrm{T}}$, thus its derivative with respect to $\boldsymbol{\beta}^{\mathrm{T}}$ is zero.

As for the remaining two terms, some general rules on derivatives involving vectors/matrices need to be introduced for their calculation.

First, given a scalar function y(**x**), where **x** is a row vector, the first derivative of y(**x**) with respect to **x** is a row vector expressed by the following formulation:

$$\frac{\partial y}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y}{\partial x_1} & \frac{\partial y}{\partial x_2} & \cdots & \frac{\partial y}{\partial x_n} \end{bmatrix}$$

If the scalar function is $y(\mathbf{b}^T) = \mathbf{b}^T\mathbf{a} = a_1b_1 + a_2b_2 + \ldots. + a_kb_k$, then:

$$\frac{\partial y}{\partial(\boldsymbol{b}^T)} = \frac{\partial\,(\boldsymbol{b}^T\boldsymbol{a})}{\partial(\boldsymbol{b}^T)} = \boldsymbol{a}$$

Consequently, the first derivative of the second term of S with respect to $\boldsymbol{\beta}^T$ is:

$$\frac{\partial\left(-2\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y}\right)}{\partial\boldsymbol{\beta}^T} = -2\mathbf{X}^T\mathbf{y}$$

Another general rule for derivatives of vectors/matrices has to be exploited to calculate the first derivative of the third term of S:

$$\frac{\partial \boldsymbol{\beta}^T (\mathbf{X}^T\mathbf{X})\boldsymbol{\beta}}{\partial \boldsymbol{\beta}^T}$$

In this case the $\mathbf{X}^T\mathbf{X}$ product corresponds to a symmetric matrix.

In fact, if we consider the following **X** matrix:

$$\mathbf{X} = \begin{pmatrix} 1 & X_1 & X_1^2 \\ 1 & X_2 & X_2^2 \\ . & . & . \\ . & . & . \\ 1 & X_n & X_n^2 \end{pmatrix}$$

its transpose is:

$$\mathbf{X}^T = \begin{pmatrix} 1 & 1 & . & . & 1 \\ X_1 & X_2 & . & . & X_n \\ X_1^2 & X_2^2 & . & . & X_n^2 \end{pmatrix}$$

The $\mathbf{X}^T\mathbf{X}$ product corresponds to the matrix:

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} n & \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 & \sum_{i=1}^{n} X_i^3 \\ \sum_{i=1}^{n} X_i^2 & \sum_{i=1}^{n} X_i^3 & \sum_{i=1}^{n} X_i^4 \end{pmatrix}$$

The following general case can thus be considered as a model of the **$\beta^T (X^T X) \beta$ term** :

$$\mathbf{y} = \mathbf{b}^T \mathbf{A}\, \mathbf{b} \qquad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \qquad \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix}$$

$$y(\mathbf{b}^T) = \begin{bmatrix} b_1 & b_2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} b_1 & b_2 \end{bmatrix} \begin{bmatrix} b_1 a_{11} + b_2 a_{12} \\ b_1 a_{12} + b_2 a_{22} \end{bmatrix} = b_1^2 a_{11} + 2 b_1 b_2 a_{12} + b_2^2 a_{22}$$

$$\frac{\partial y(\mathbf{b}^T)}{\partial b_1} = 2(b_1 a_{11} + b_2 a_{12}) \qquad \frac{\partial y(\mathbf{b}^T)}{\partial b_2} = 2(b_1 a_{12} + b_2 a_{22})$$

$$\frac{\partial y(\mathbf{b}^T)}{\partial \mathbf{b}^T} = \frac{\partial \mathbf{b}^T \mathbf{A} \mathbf{b}}{\partial \mathbf{b}^T} = 2\mathbf{A}\mathbf{b}$$

Starting from this general example, the first derivative of the third term of S can be easily calculated:

$$\frac{\partial\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta}\right)}{\partial\boldsymbol{\beta}^{\mathrm{T}}} = 2(\mathbf{X}^{\mathrm{T}}\mathbf{X})\boldsymbol{\beta}$$

Consequently:

$$\frac{\partial S}{\partial\boldsymbol{\beta}^{\mathrm{T}}} = -2\mathbf{X}^{T}\mathbf{y} + 2(\mathbf{X}^{\mathrm{T}}\mathbf{X})\boldsymbol{\beta}$$

If the OLS estimator of vector **β** is indicated as **b**, the following equation must be valid:

$$-2\mathbf{X}^{T}\mathbf{y} + 2(\mathbf{X}^{\mathrm{T}}\mathbf{X})\mathbf{b} = 0$$

$$(\mathbf{X}^{\mathrm{T}}\mathbf{X})\mathbf{b} = \mathbf{X}^{\mathrm{T}}\mathbf{y} \Rightarrow \underset{(p\times 1)}{\mathbf{b}} = \underset{(p\times p)}{(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}}\,\underset{(p\times 1)}{\mathbf{X}^{\mathrm{T}}\mathbf{y}}$$

As an example, the solution obtained in the case of second order univariate regression is described in the following:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon$$

$$\mathbf{X} = \begin{pmatrix} 1 & X_1 & X_1^2 \\ 1 & X_2 & X_2^2 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & X_n & X_n^2 \end{pmatrix} \qquad \mathbf{X}^T = \begin{pmatrix} 1 & 1 & \cdot & \cdot & 1 \\ X_1 & X_2 & \cdot & \cdot & X_n \\ X_1^2 & X_2^2 & \cdot & \cdot & X_n^2 \end{pmatrix} \qquad \mathbf{X}^T\mathbf{X} = \begin{pmatrix} n & \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 & \sum_{i=1}^{n} X_i^3 \\ \sum_{i=1}^{n} X_i^2 & \sum_{i=1}^{n} X_i^3 & \sum_{i=1}^{n} X_i^4 \end{pmatrix}$$

$$\begin{pmatrix} n & \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 & \sum_{i=1}^{n} X_i^3 \\ \sum_{i=1}^{n} X_i^2 & \sum_{i=1}^{n} X_i^3 & \sum_{i=1}^{n} X_i^4 \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & \cdot & \cdot & 1 \\ X_1 & X_2 & \cdot & \cdot & X_n \\ X_1^2 & X_2^2 & \cdot & \cdot & X_n^2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_n \end{pmatrix}$$

$$\mathbf{X}^T\mathbf{X} \quad \mathbf{b} \quad = \quad \mathbf{X}^T \quad \mathbf{y}$$

If scalar products between matrices and vectors are developed, the following equations are obtained:

$$b_0 n + b_1 \sum_{i=1}^{n} X_i + b_2 \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} Y_i$$

$$b_0 \sum_{i=1}^{n} X_i + b_1 \sum_{i=1}^{n} X_i^2 + b_2 \sum_{i=1}^{n} X_i^3 = \sum_{i=1}^{n} X_i Y_i$$

$$b_0 \sum_{i=1}^{n} X_i^2 + b_1 \sum_{i=1}^{n} X_i^3 + b_2 \sum_{i=1}^{n} X_i^4 = \sum_{i=1}^{n} X_i^2 Y_i$$

Equations for b_0 ad b_1 referred to simple linear regression can be easily obtained from the first two equations after removing the term including b_2.

Once estimates of all parameters are obtained, the fitted model is easily constructed.
If p = m+1 parameters are considered, the following equation is obtained:

$$\hat{Y}_i = b_0 + b_1 X_i + b_2 X_i^2 + \ldots\ldots. + b_m X_i^m$$

and residuals e_i are calculated: $e_i = Y_i - \hat{Y}_i$

Estimate for σ^2

In order to estimate $\boldsymbol{\sigma}^2$ the sum of squares referred to residuals (errors), SSE, is considered:

$$\mathrm{SSE} = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} e_i^{\,2} = \mathbf{e}^T\mathbf{e} = (\mathbf{y} - \mathbf{Xb})^T(\mathbf{y} - \mathbf{Xb}) =$$

$$= \mathbf{y}^T\mathbf{y} - \mathbf{b}^T\mathbf{X}^T\mathbf{y} - \mathbf{y}^T\mathbf{Xb} + \mathbf{b}^T\mathbf{X}^T\mathbf{Xb} =$$

$$= \mathbf{y}^T\mathbf{y} - 2\mathbf{b}^T\mathbf{X}^T\mathbf{y} + \mathbf{b}^T\mathbf{X}^T\mathbf{Xb}$$

Since $(\mathbf{X}^T\mathbf{X})\mathbf{b} = \mathbf{X}^T\mathbf{y}$ the equation can be written as:

$$\mathrm{SSE} = \mathbf{y}^T\mathbf{y} - \mathbf{b}^T\mathbf{X}^T\mathbf{y}$$

By analogy with the E(SSE) value found for simple linear regression, the expectation of SSE is:

$$\mathrm{E(SSE)} = \sigma^2(n-p)$$

thus an unbiased estimator of $\boldsymbol{\sigma}^2$ is: $s^2 = \hat{\sigma}^2 = \dfrac{\mathrm{SSE}}{n-p}$

Expectation and variance of vector b

As demonstrated below, vector **b** is an unbiased estimator of vector $\boldsymbol{\beta}$, i.e.: $\mathrm{E}(\mathbf{b}) = \boldsymbol{\beta}$

$$\mathbf{b} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\varepsilon}$$

$(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{X}$ corresponds to the identity matrix:

$$I_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

Consequently: $\mathbf{b} = \boldsymbol{\beta} + (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\varepsilon}$

and, considering that $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$, the following equation can be written:

$$\mathrm{E}(\mathbf{b}) = \mathrm{E}\left[\boldsymbol{\beta} + (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\varepsilon}\right] = \mathrm{E}(\boldsymbol{\beta}) + (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\,\mathrm{E}(\boldsymbol{\varepsilon}) \;= \boldsymbol{\beta}$$

Variance of **b** (variance-covariance matrix) can be obtained by exploiting the general rule:

$$\mathrm{Var}(\mathbf{A}\boldsymbol{\varepsilon}) = \mathbf{A}\,\mathrm{Var}(\boldsymbol{\varepsilon})\mathbf{A}^{\mathrm{T}}$$

and considering the equation $\mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$, i.e., the assumption that the error is the same for all responses (homoschedasticity).

Indeed:

$$\mathrm{Var}(\mathbf{b}) = \mathrm{Var}\left[(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}\right] = \left[(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\right]\mathrm{Var}(\mathbf{y})\left[(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\right]^{T} =$$

$$= \left[(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\right]\mathrm{Var}(\boldsymbol{\varepsilon})\left[\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\right] = \left[(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\right]\sigma^2\mathbf{I}\left[\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\right]$$

$$= \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} = \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$$

Note that:

1) Var(**y**) = Var(**ε**) since each of the components of vector **y**, *i.e.*, experimental values y_i, are the sum of a deterministic part, which has no variance, and of ε_i, the component of vector **ε**.

2) since $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ is a symmetrical matrix, also $(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$ is symmetrical, thus its transpose is equal to it.

Note that the variance of **b** can be obtained also using the following equations:

$$\mathrm{Var}(\mathbf{b}) = \mathrm{E}\left[(\mathbf{b}-\boldsymbol{\beta})(\mathbf{b}-\boldsymbol{\beta})^T\right] = \mathrm{E}\left[\begin{bmatrix} b_0-\beta_0 \\ b_1-\beta_1 \\ \cdots \\ b_m-\beta_m \end{bmatrix}\begin{bmatrix} b_0-\beta_0 & b_1-\beta_1 & \cdots & b_m-\beta_m \end{bmatrix}\right] =$$

external product of matrices

$$= \mathrm{E}\begin{bmatrix} (b_0-\beta_0)(b_0-\beta_0) & (b_0-\beta_0)(b_1-\beta_1) & \cdots & (b_0-\beta_0)(b_m-\beta_m) \\ (b_1-\beta_1)(b_0-\beta_0) & (b_1-\beta_1)(b_1-\beta_1) & \cdots & (b_1-\beta_1)(b_m-\beta_m) \\ \cdots & \cdots & \cdots & \cdots \\ (b_m-\beta_m)(b_0-\beta_0) & (b_m-\beta_m)(b_1-\beta_1) & \cdots & (b_m-\beta_m)(b_m-\beta_m) \end{bmatrix} =$$

$$= \begin{bmatrix} \mathrm{E}\left[(b_0-\beta_0)(b_0-\beta_0)\right] & \mathrm{E}\left[(b_0-\beta_0)(b_1-\beta_1)\right] & \cdots & \mathrm{E}\left[(b_0-\beta_0)(b_m-\beta_m)\right] \\ \mathrm{E}\left[(b_1-\beta_1)(b_0-\beta_0)\right] & \mathrm{E}\left[(b_1-\beta_1)(b_1-\beta_1)\right] & \cdots & \mathrm{E}\left[(b_1-\beta_1)(b_m-\beta_m)\right] \\ \cdots & \cdots & \cdots & \cdots \\ \mathrm{E}\left[(b_m-\beta_m)(b_0-\beta_0)\right] & \mathrm{E}\left[(b_m-\beta_m)(b_1-\beta_1)\right] & \cdots & \mathrm{E}\left[(b_m-\beta_m)(b_m-\beta_m)\right] \end{bmatrix} =$$

$$= \begin{bmatrix} \boxed{\mathrm{Var}(b_0)} & \mathrm{Cov}(b_0, b_1) & \cdots & \mathrm{Cov}(b_0, b_m) \\ \mathrm{Cov}(b_1, b_0) & \boxed{\mathrm{Var}(b_1)} & \cdots & \mathrm{Cov}(b_1, b_m) \\ \cdots & \cdots & \cdots & \cdots \\ \mathrm{Cov}(b_m, b_0) & \mathrm{Cov}(b_m, b_1) & \cdots & \boxed{\mathrm{Var}(b_m)} \end{bmatrix} = \begin{bmatrix} \boxed{\sigma^2 C_{00}} & \sigma^2 C_{01} & \cdots & \sigma^2 C_{0m} \\ \sigma^2 C_{10} & \boxed{\sigma^2 C_{11}} & \cdots & \sigma^2 C_{1m} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma^2 C_{m0} & \sigma^2 C_{m1} & \cdots & \boxed{\sigma^2 C_{mm}} \end{bmatrix}$$

Since:

$$\mathrm{Var}(\mathbf{b}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1} = \begin{bmatrix} \sigma^2 C_{00} & \sigma^2 C_{01} & \cdots & \sigma^2 C_{0m} \\ \sigma^2 C_{10} & \sigma^2 C_{11} & \cdots & \sigma^2 C_{1m} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma^2 C_{m0} & \sigma^2 C_{m1} & \cdots & \sigma^2 C_{mm} \end{bmatrix}$$

The variance of single regression parameters, b_i, can be expressed as:

$$\mathrm{Var}(b_i) = \sigma^2 C_{ii}$$

where C_{ii} is the diagonal element of matrix $(X^TX)^{-1}$ corresponding to b_i.

The covariances between regression parameters can be obtained by multiplying the non diagonal terms of matrix $(X^TX)^{-1}$ by σ^2.

It can be demonstrated that vector **b** represents the most efficient unbiased estimator of β, i.e., it has the minimum variance among unbiased linear estimators.

Confidence intervals for b_i and $\hat{Y}_0$ values at a α significance level

Since the variance of a specific parameter b_i can be calculated as follows:

$$\hat{V}(b_i) = \hat{\sigma}^2 C_{ii} \quad \text{with} \quad \hat{\sigma}^2 = \frac{SSE}{n-p}$$

the confidence interval for b_i at a α significance level can be calculated as follows:

$$b_i \pm t_{n-p,(1-\alpha/2)} \left[V(b_i)\right]^{1/2} = b_i \pm t_{n-p,(1-\alpha/2)} \left[\hat{\sigma}^2 C_{ii}\right]^{1/2}$$

By analogy with simple linear regression, a confidence interval can be calculated also for $\hat{Y}_0$, i.e., the value of response predicted by the model for a specific x_0:

$$\hat{Y}_0 = \mathbf{x}_0^T \mathbf{b} \quad \Rightarrow \quad \hat{Y}_0 = \begin{bmatrix} 1 & X_0 & X_0^2 \ldots\ldots & X_0^m \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ . \\ . \\ b_m \end{bmatrix}$$

Consequently:

$$E(\hat{Y}_0) = \mathbf{x}_0^T \boldsymbol{\beta}$$

$$\mathrm{Var}(\hat{Y}_0) = \mathrm{Var}(\mathbf{x}_0^{\mathbf{T}}\mathbf{b}) = E\left\{\left[\hat{Y}_0 - E(\hat{Y}_0)\right]\left[\hat{Y}_0 - E(\hat{Y}_0)\right]^T\right\} =$$

$$= \mathrm{E}\left[(\mathbf{x}_0^{\mathbf{T}}\mathbf{b} - \mathbf{x}_0^{\mathbf{T}}\boldsymbol{\beta})\ (\mathbf{x}_0^{\mathbf{T}}\mathbf{b} - \mathbf{x}_0^{\mathbf{T}}\boldsymbol{\beta})^T\right] = \mathrm{E}\left[\mathbf{x}_0^{\mathbf{T}}(\mathbf{b} - \boldsymbol{\beta})\ (\mathbf{b} - \boldsymbol{\beta})^T\mathbf{x}_0\right] =$$

$$= \mathbf{x}_0^{\mathbf{T}}\ \mathrm{E}\left[(\mathbf{b} - \boldsymbol{\beta})\ (\mathbf{b} - \boldsymbol{\beta})^T\right]\mathbf{x}_0 = \mathbf{x}_0^{\mathbf{T}}\ \mathrm{Var}(\mathbf{b})\mathbf{x}_0 = \mathbf{x}_0^{\mathbf{T}}\sigma^2(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{x}_0$$

According to the original assumption on response normality, the following relation can thus be written:

$$\hat{Y}_0 \sim N(\mathbf{x}_0^{\mathbf{T}}\boldsymbol{\beta}, \sigma^2\mathbf{x}_0^{\mathbf{T}}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{x}_0)$$

The confidence interval for $\hat{\mathbf{Y}}_0$ at a α significance level can thus be expressed as:

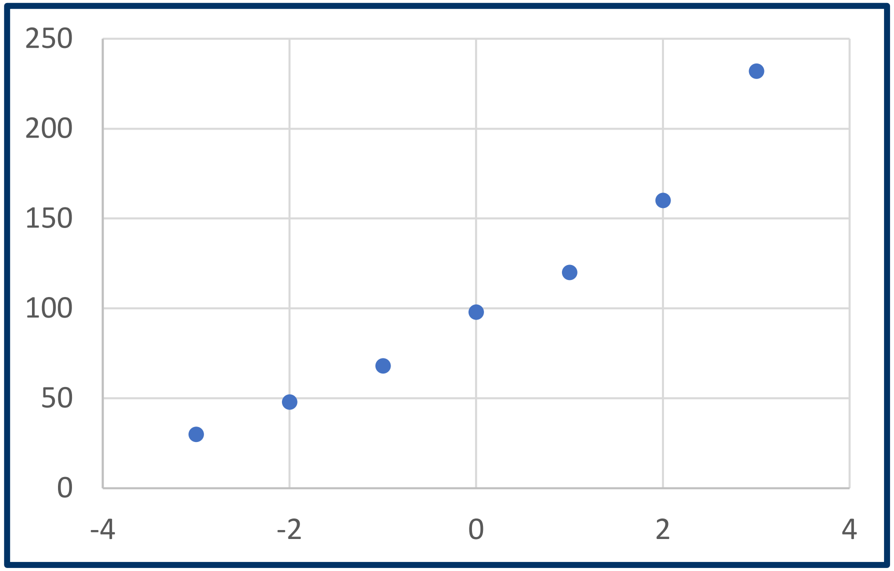
$$\mathbf{x}_0^{\mathbf{T}}\mathbf{b} \pm t_{n-p,(1-\alpha/2)}\hat{\sigma}\ \left[\mathbf{x}_0^{\mathbf{T}}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{x}_0\right]^{1/2}$$

where: $\hat{\sigma} = \left[\dfrac{\mathrm{SSE}}{n-p}\right]^{1/2}$

A numerical example for univariate regression of order 2

Let us suppose that the following seven couples of (x,y) values are available:

Controlled variable x	Measured y
-3	30
-2	48
-1	68
0	98
+1	120
+2	160
+3	232



A curvature is clearly observed, thus the following second order model can be adopted:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon$$

In this case: m = 2, p = 3, df = n-p = 7 – 3 = 4

Moreover:

$$\mathbf{X} = \begin{pmatrix} 1 & -3 & 9 \\ 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix} \qquad \mathbf{X}^T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -3 & -2 & -1 & 0 & 1 & 2 & 3 \\ 9 & 4 & 1 & 0 & 1 & 4 & 9 \end{pmatrix}$$

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} 7 & 0 & 28 \\ 0 & 28 & 0 \\ 28 & 0 & 196 \end{pmatrix} \qquad (\mathbf{X}^T\mathbf{X})^{-1} = \begin{pmatrix} 0.333 & 0.000 & -0.048 \\ 0.000 & 0.036 & 0.000 \\ -0.048 & 0.000 & 0.012 \end{pmatrix}$$

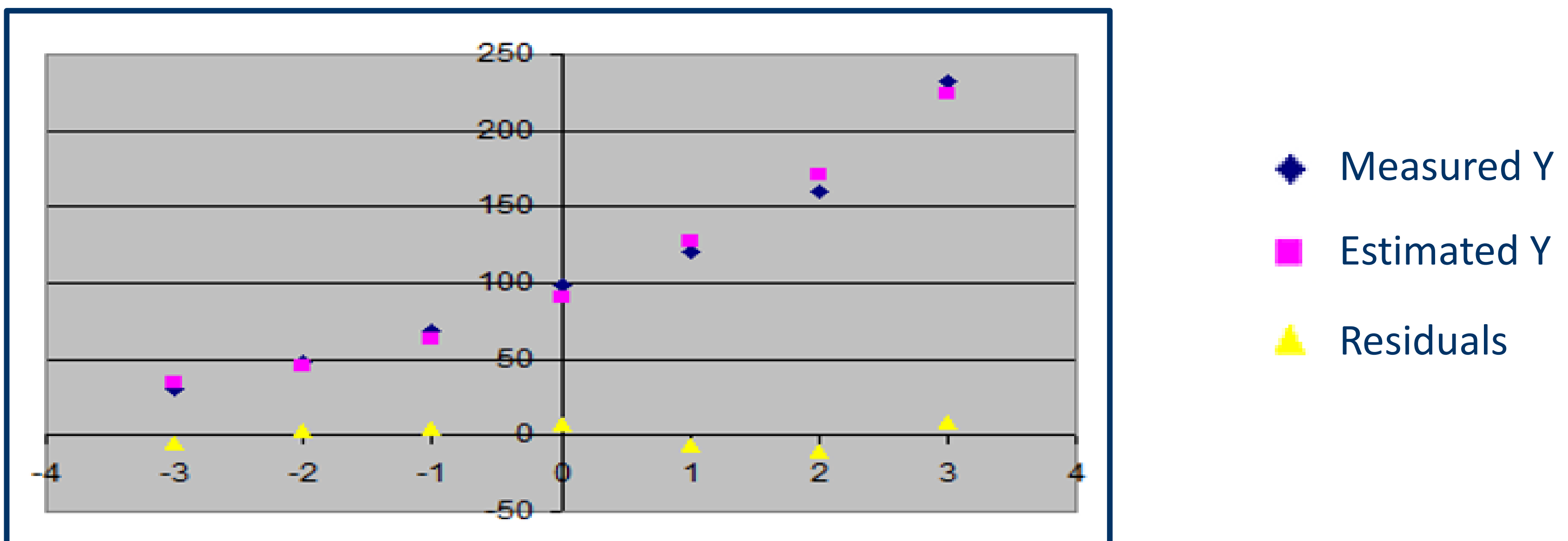
Note than both matrices are symmetric

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\,\mathbf{X}^T\mathbf{y} = \begin{pmatrix} 91.143 \\ 31.500 \\ 4.214 \end{pmatrix}$$

$$\Rightarrow \quad Y = 91.143 + 31.500\,X + 4.214\,X^2$$

Predicted response values and residuals can now be easily calculated and plotted:

Controlled variable x	Measured y	Estimated y	Residual	Squared residual
-3	30	34.5	-4.5	20.25
-2	48	45	3	9
-1	68	63.8	4.2	17.64
0	98	91.1	6.9	47.61
+1	120	126.8	-6.8	46.24
+2	160	171	-11	121
+3	232	223.6	8.4	70.56



Since SSE = 332.3, $\hat{\sigma}^2 = \frac{\text{SSE}}{n-p}$ =332.3/4= 83.075 and:

$$\text{Var}(\mathbf{b}) = \hat{\sigma}^2(\mathbf{X}^T\mathbf{X})^{-1} = 83.075 \begin{vmatrix} 0.333 & 0 & -0.048 \\ 0 & 0.036 & 0 \\ -0.048 & 0 & 0.012 \end{vmatrix} = \begin{vmatrix} \mathbf{27.5} & 0 & -4.0 \\ 0 & \mathbf{3.0} & 0 \\ -4.0 & 0 & \mathbf{1} \end{vmatrix}$$

95% confidence intervals for parameters b_i are given by the following equation:

$b_i \pm t_{4,\,0.975}\,[\hat{V}(b_i)]^{1/2}$ where: $\hat{V}(b_i) = \hat{\sigma}^2 C_{ii}$

correspond to red diagonal elements reported in the matrix shown before, thus:

b_0) $91.143 \pm 2.77\,(27.5)^{1/2} = 91 \pm 15$

b_1) $31.500 \pm 2.77\,(3)^{1/2} = 32 \pm 5$

b_2) $4.214 \pm 2.77\,(1)^{1/2} = 4 \pm 3$

The confidence interval for $\hat{Y}_0$ corresponding to $x_0 = 0.5$ is obtained as follows. First the calculation of $\hat{Y}_0$ is made:

$$\hat{Y}_0 = \mathbf{b}^T \mathbf{x}_0 = [91.143 \quad 31.5 \quad 4.214] \begin{pmatrix} 1 \\ 0.5 \\ 0.25 \end{pmatrix} = 107.94$$

then the general expression of the interval is considered:

$$\mathbf{x}_0^T \mathbf{b} \pm t_{n-p,(1-\alpha/2)} \left[\hat{\sigma}^2 \mathbf{x}_0^T (\mathbf{X}^T\mathbf{X})^{-1} \mathbf{x}_0\right]^{1/2}$$

Since:

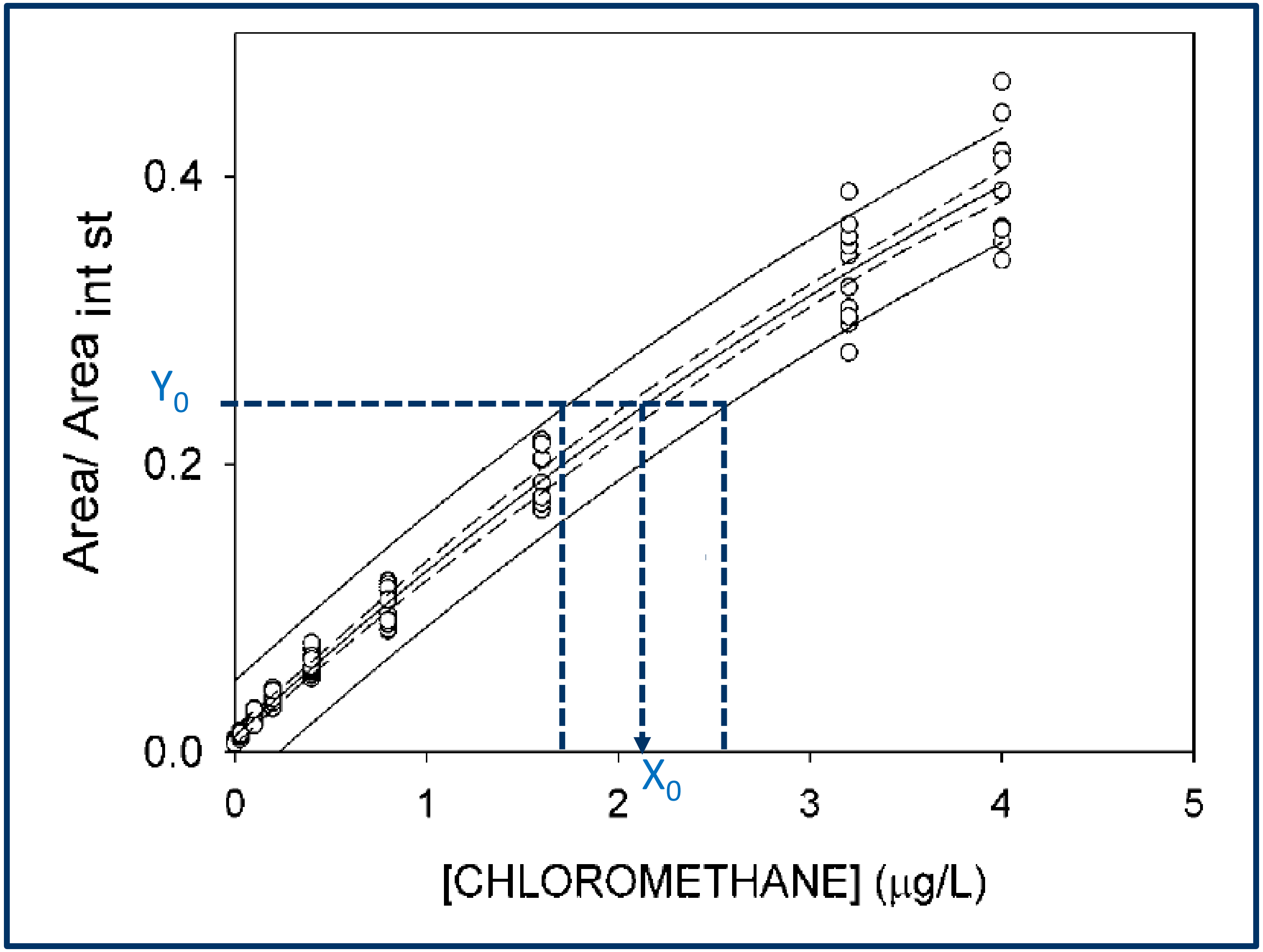
$$\mathbf{x}^T_0(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_0 = 0.319$$

the 95% confidence interval is expressed as:

$$107.94 \pm 2.77\ [83.075 \times 0.319]^{1/2} = 108 \pm 14$$

It is worth noting that inverse regression can be performed also in the case of second order univariate linear regression, i.e., by calculating prediction bands and then extrapolating the confidence interval for a specific x_0 value, once the corresponding y_0 value is fixed.

An example of the procedure is shown in the following figure:



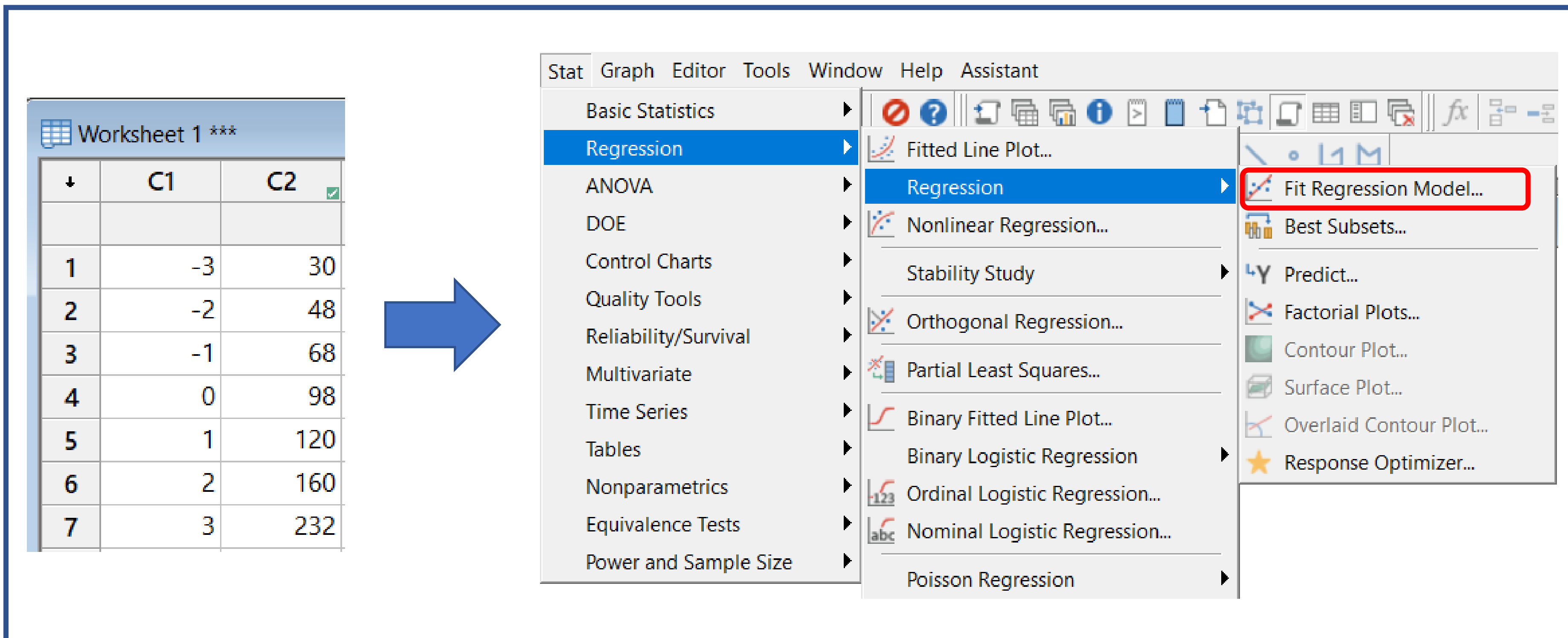
Notably, confidence bands are well distinct from prediction bands due to low precision.

Use of Minitab 18 to perform univariate regression of order 2

Minitab 18 is able to perform univariate regression of different orders.

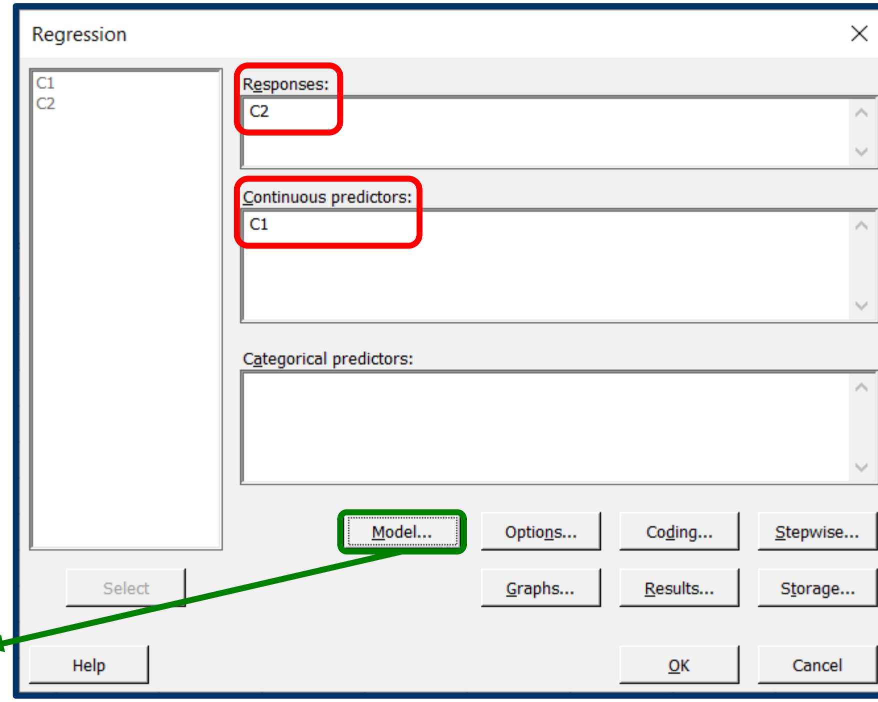
The procedure is started by introducing data referred to variable X and to response Y in columns C1 and C2, respectively.

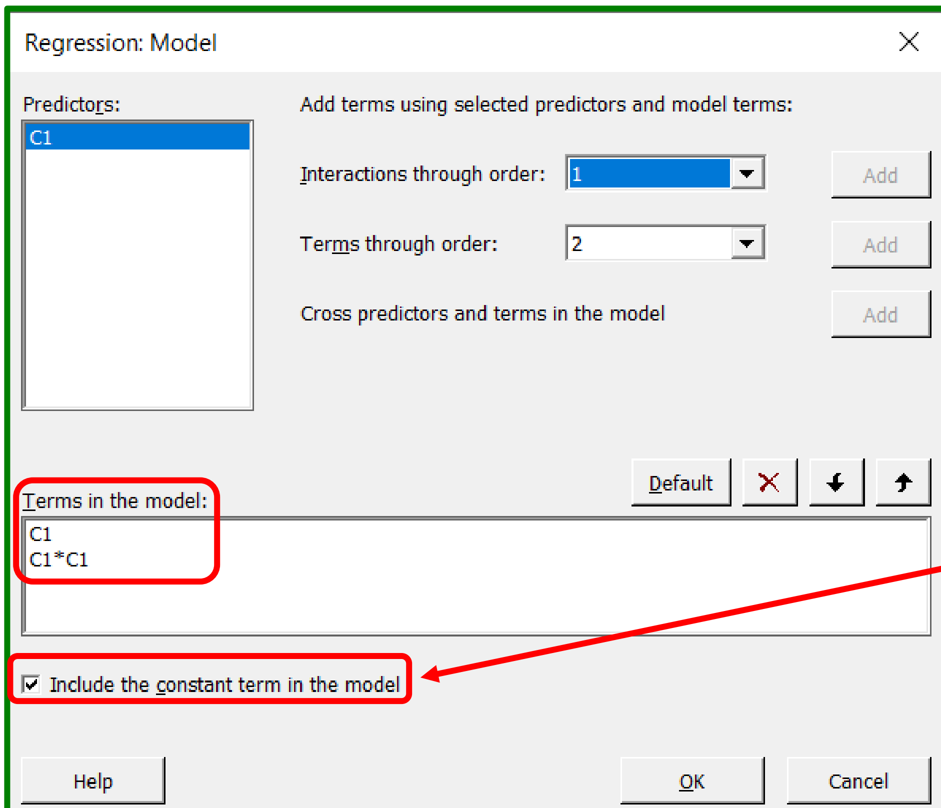
The choice of a regression model can be done by accessing the Stat > Regression menu, and then the Regression > Fit Regression Model… option.



Columns reporting values for response and independent variable (a continuous predictor, in the specific case) are selected in the main window for regression.

The Model... window can be opened afterwards, and the order of regression can be fixed in the *Terms through order* box.

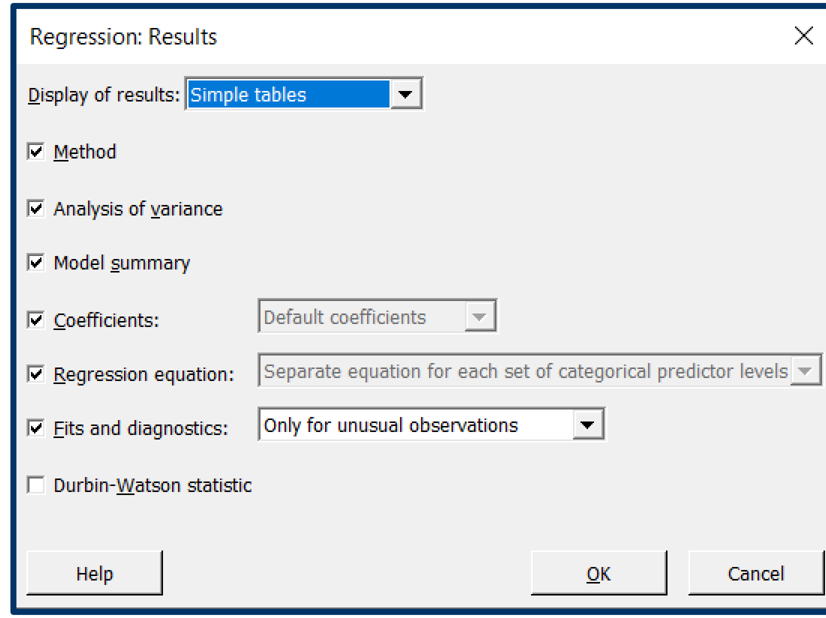




If value 2 is selected in the box, like in the present case, terms in the model are expressed as C1 and C1*C1 (i.e., $C1^2$).
Note that the constant term can be included in the model by selecting the appropriate box.

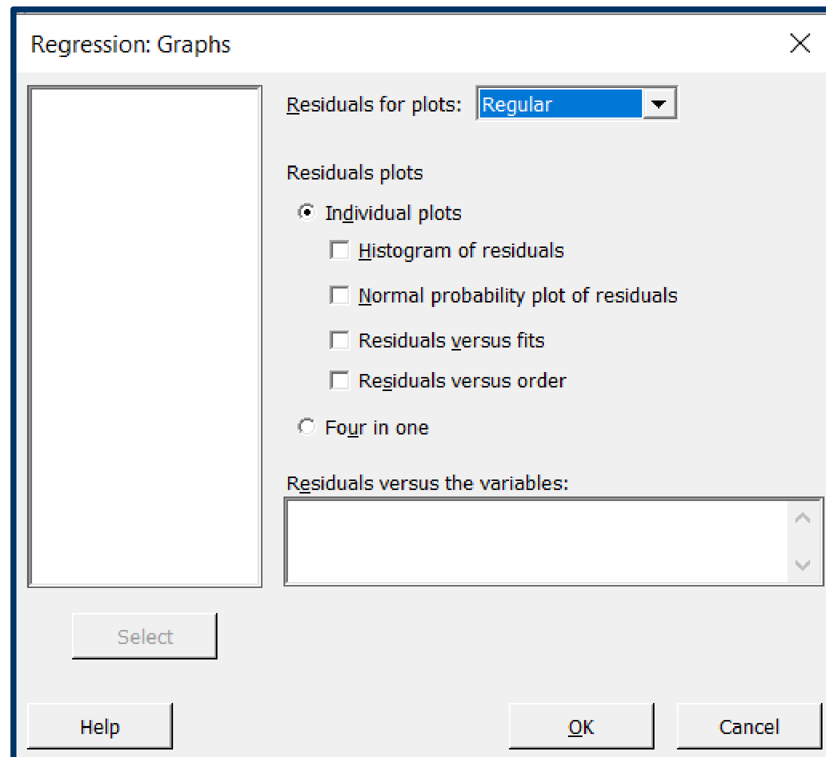
When only one predictor is present, interaction terms and cross predictors cannot be selected.

Several types of results can be selected to be displayed in the Results menu, including the Analysis of variance.

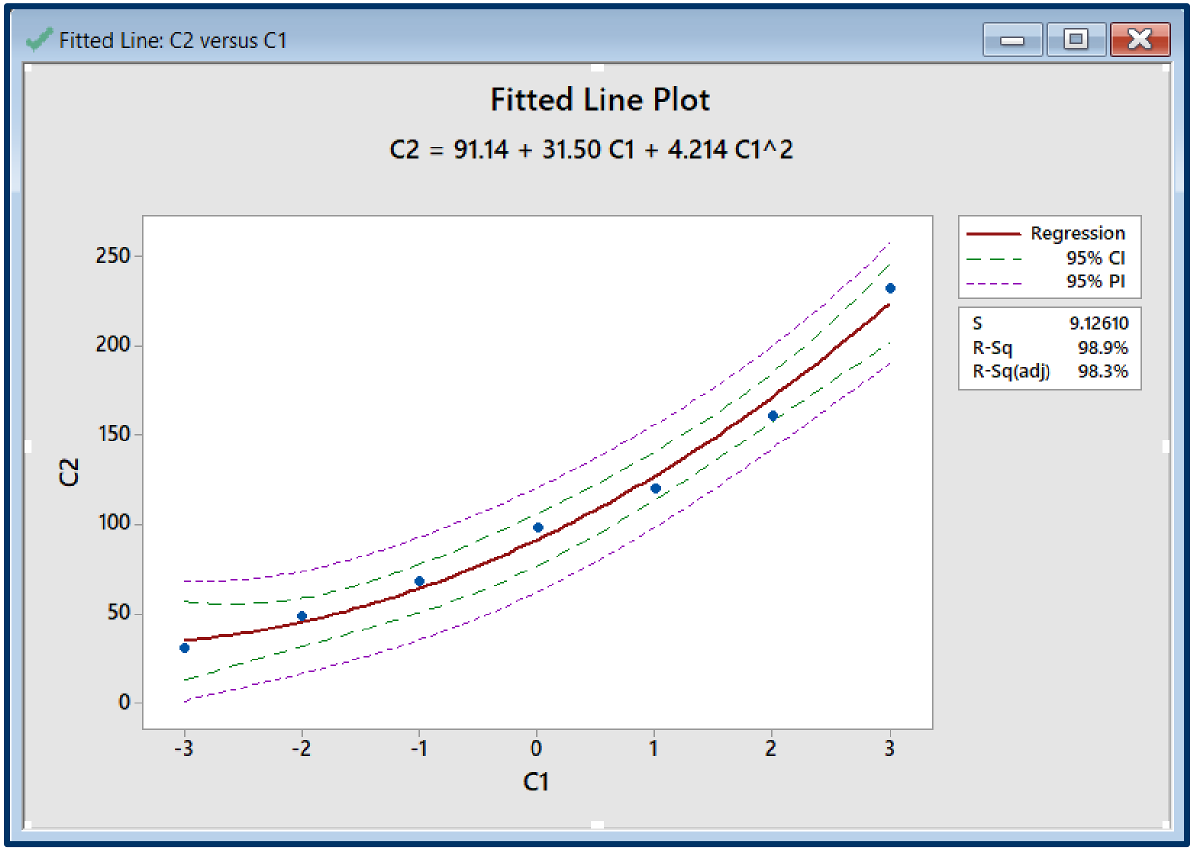


The Graphs menu enables the selection of plots related to residuals, that can be displayed together with the regression curve.

The latter is always accompanied by confidence and prediction bands, calculated for the probability level (e.g., 95%) selected in the Options menu of the main Regression window.



The regression curve obtained with data shown before, accompanied by confidence (CI) and prediction (PI) bands is reported in the following figure:



The value of correlation coefficient (R-Sq) is also reported, along with the S value, which corresponds to:

$$\hat{\sigma} = \left[\frac{\text{SSE}}{n-p} \right]^{1/2}$$

Information related to regression is reported, in tabular form, inside the Session window:

The regression equation and the values of coefficients, accompanied by their standard errors (SE) are indicated.

Moreover, the ANOVA table for regression is reported.
In this table F-values enable the evaluation of significance for the regression and, in particular, for terms referred to the first (C1) and the second (C1*C1) power of the variable value.

As inferred from the corresponding p-values, the model is correct, since both terms are significant (p values are lower than 0.05).

Note that T-values correspond to ratios between coefficient values and the corresponding standard errors.

Regression Analysis: C2 versus C1

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	29274.9	14637.4	175.75	0.000
C1	1	27783.0	27783.0	333.59	0.000
C1*C1	1	1491.9	1491.9	17.91	0.013
Error	4	333.1	83.3		
Total	6	29608.0			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
9.12610	98.87%	98.31%	92.88%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	91.14	5.27	17.30	0.000	
C1	31.50	1.72	18.26	0.000	1.00
C1*C1	4.214	0.996	4.23	0.013	1.00

Regression Equation

C2 = 91.14 + 31.50 C1 + 4.214 C1*C1