The general linear model (also known as the multiple regression model):

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i \quad i = 1, 2, \cdots N \ (population\ model)$$

$$Y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \cdots + b_k x_{ki} + e_i \quad i = 1, 2, \cdots n \ (sample\ model)$$

In matrix notation:

$$\underset{(n\times 1)}{\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}} = \underset{\text{Design matrix } (n\times k)}{\begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix}} \underset{(k\times 1)}{\begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix}} + \underset{(n\times 1)}{\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}}$$

$$Y = Xb + e \ (estimated\ sample\ model)$$

$$Y = X\beta + \varepsilon \ (assumed\ population\ model)$$

In this multiple regression model $Y = X\beta + \varepsilon$:

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \quad \text{and} \quad X = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix}$$

β is a vector of parameters of the model, with length $p = k + 1$, and X is the $n \times p$ design matrix

β_0 is the intercept and $\beta_1, \beta_2, \ldots, \beta_k$ are the slope coefficients associated with $X_1, X_2, \ldots, X_k$ respectively – these are called *partial* regression coefficients.

The interpretation of a partial regression coefficient is the expected change in Y as the corresponding X changes by 1, *with all the other X variables held constant.*

However, the X variables are often inter-related and it is often not possible to change one X variable without changing some of the other X variables at the same time.

The model is “linear in the parameters”, i.e. a “linear” combination of the β coefficients, not necessarily a “linear” combination of the original X variables.

For example, the models:

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2$$

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$$

are both “linear”, and their associated design matrices are:

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix} \quad \text{and} \quad X = \begin{bmatrix} 1 & x_{11} & x_{21} & x_{11} \cdot x_{21} \\ 1 & x_{12} & x_{22} & x_{12} \cdot x_{22} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & x_{1n} \cdot x_{2n} \end{bmatrix}$$

These are both examples of what is called polynomial regression, since the models have the structure of being polynomials in the predictor X variables.

Even models that do not at first appear linear, can be “linearised”. For example:

$Y = \beta_0 x_1^{\beta_1} x_2^{\beta_2}$ (see Example 4: Black Cherry Trees at the back of the “brick”)

Here, we could take logs of all variables as a “linearising” transformation:

$$ln(Y) = ln(\beta_0) + \beta_1 ln(x_1) + \beta_2 ln(x_2) = \beta_0' + \beta_1 ln(x_1) + \beta_2 ln(x_2)$$

Similarly, $Y = \frac{\beta_0}{1+\beta_1 x_1 + \beta_2 x_2}$ can be “linearised” by taking reciprocals:

$$\frac{1}{Y} = \frac{1}{\beta_0} + \frac{\beta_1}{\beta_0} x_1 + \frac{\beta_2}{\beta_0} x_2 = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2$$

Then once we have estimated γ_0, γ_1, γ_2 we can also estimate β_0, β_1, β_2 using:

$$\beta_0 = \frac{1}{\gamma_0}, \quad \beta_1 = \frac{\gamma_1}{\gamma_0}, \quad \beta_1 = \frac{\gamma_2}{\gamma_0}$$

In the multiple regression model, $Y = X\beta + \varepsilon,$ the underlying model assumptions are now that the random error vector, ε, satisfies:

$$E[\varepsilon] = 0 \quad \text{and} \quad Var(\varepsilon) = \sigma^2 I = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} (n \times n)$$

So, as for simple linear regression (SLR), we assume uncorrelated (independent) and homoscedastic (constant variance) errors. As with SLR, we also assume that the errors are normally distributed.

Note that as for SLR regression, we can still draw simple plots of the estimated errors to assess these assumptions, using the residuals (the estimated errors). The main residual plot will still be a 2-dimensional plot of the (standardised) residuals against the fitted values, which is a known or “fixed” function of all the X variables.

Estimating the multiple regression model, $Y = X\beta + \varepsilon,$ is much the same as it was for SLR, as least squares estimation and the same matrix equations still hold (see pages 4 and 5 of chapter 2 of the notes for further details):

$$\hat{\beta} = b = (X^T X)^{-1}(X^T Y) \quad \text{and} \quad Var(b) = \sigma^2 (X^T X)^{-1}$$

We still need to estimate the error variance, σ^2, which is still based on the sum of squares of the estimated errors (the residuals, e), however, now that we have a model with p parameters rather than just 2, the residuals have to satisfy p linear constraints:

$$\sum_{i=1}^{n} e_i = 0, \quad \sum_{i=1}^{n} e_i x_{1i} = 0, \quad \sum_{i=1}^{n} e_i x_{2i} = 0, \quad \cdots \quad \sum_{i=1}^{n} e_i x_{ki} = 0$$

And the error degrees of freedom are now $n - p$:

$$\hat{\sigma}^2 = s^2 = \frac{SS_{Errors}}{(n-p)} = \frac{e^T e}{(n-p)}$$