Multiple Regression: The General Linear Model

- In the multiple linear regression case, the dependent variable $y$ is to be related to more than one independent variable, say, $x_1, x_2, \ldots, x_k$.
- The model we use is called a general linear model – its form is
  \[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon \]
- Our assumptions about the $\epsilon$’s are the same as for the simple linear regression case, namely the $\epsilon$’s are independent with mean $E(\epsilon)$ zero and variance $\sigma^2_\epsilon$, and they are normally distributed.
Data consist of \( n \) cases of \( k + 1 \) values denoted by

\[
(y_i, x_{i1}, x_{i2}, \ldots, x_{ik}), \quad i = 1, 2, \ldots, n
\]

Using this full notation we can write the model as

\[
y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \epsilon_i
\]

The form of the general linear model, as given, is understood to permit polynomial terms like \( x_2^3 \), cross-product terms like \( x_2 x_3 \) – etc. –
• Thus the polynomial model in one variable

\[ y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \epsilon \]

fits the form of the general linear model

\[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon \]

with \( x_1 \equiv x, \ x_2 \equiv x^2, \ x_3 \equiv x^3 \).

• The multiple regression model

\[
y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 \\
+ \beta_5 x_1^3 + \beta_6 x_2^3 + \beta_7 x_1 x_2 + \epsilon
\]
can also be written as

\[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 \]

\[ + \beta_6 x_6 + \beta_7, x_7 + \epsilon \]

with \( x_3 = x_1^2, \ x_4 = x_2^2, \ x_5 = x_3^2, \ x_6 = x_4^3, \ x_7 = x_1 x_2. \)

- Even terms like \( \log(x_3), \ e^{x_2}, \ \cos(x_4), \) etc., are permitted.
- The model is called a linear model because its expression is linear in the parameters i.e., the \( \beta \)'s. The fact that it may involve nonlinear functions of the \( x \)'s is irrelevant.
• The concept of interaction will be discussed later.

• Here let us just state that the presence (or absence) of interaction between two independent variables (say) $x_1$ and $x_2$ can be tested by including product terms such as $x_1 x_2$ in the model.

• Thus the model:

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

allows us to test whether interaction between $x_1$ and $x_2$ exists.

• The interpretation of this model is postponed till later.
The Multiple Regression Model

• When a general linear model relates $y$ to a set of quantitative independent variables ($x$’s) that may include squared terms, product terms, etc., we have a multiple regression model.

• When the independent variables are dummy variables coded to 0 or 1 representing values of qualitative independent variables or levels of treatment factors, the resulting models are called analysis of variance models.

• Just as we did in Chapter 11 for the simple linear regression model $y = \beta_0 + \beta_1 x + \epsilon$, we will consider least squares estimation of the parameters $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ in the general linear model.
• These least square estimates of the parameters will be denoted by $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$.

• Then confidence intervals and tests of hypotheses concerning the parameters will be discussed as well as inferences about $E(y)$ and a future $y$ observation.

• Computing quantities of interest like the $\hat{\beta}_i$, SSE, SSREG, MSE, MSREG, -etc- must be done using a computer because the amount of arithmetic required is very large.

• In order to demonstrate exactly what computations are required (so you will understand what the computer is doing) it is easiest to use matrix notation.
Multiple Regression Computations Using Matrix Notation

• Using matrix notation, we can define vectors \( y, \beta, \) and \( \epsilon \) and the matrix \( X \), as follows with the objective being to give a compact form of definition of quantities like least squares point and interval estimates, test statistics, etc.

• These are defined as follows:

\[
\begin{align*}
y \quad &\quad n \times 1 = \\
&= \begin{pmatrix}
y_1 \\
y_2 \\
\vdots \\
y_n
\end{pmatrix} \\
X \quad &\quad n \times (k+1) = \\
&= \begin{pmatrix}
1 & x_{11} & x_{12} & \ldots & x_{1k} \\
1 & x_{21} & x_{22} & \ldots & x_{2k} \\
1 & x_{31} & x_{32} & \ldots & x_{3k} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & x_{n1} & x_{n2} & \ldots & x_{nk}
\end{pmatrix}
\end{align*}
\]
\[ \beta \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}_{(k+1) \times 1} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}_{n \times 1} \]

- Now the model can be written in terms of matrices as

\[ y = X \beta + \epsilon \]

- The following quantities, defined in terms of the above, are useful in multiple regression computations that follow.
- \( X'X \) is the matrix of sums of cross products of the various columns of \( X \).
• The element in the \(i^{\text{th}}\) row and the \(j^{\text{th}}\) column of \(X'X\) is
\[
X'X = \left( \sum_{t=1}^{n} x_{ti} x_{tj} \right)
\]

• Denote the elements of the inverse of this matrix by \(v_{ij}\) where \(i = 1, 2, \ldots, k\), \(j = 1, 2, \ldots, k\). In this notation
\[
(X'X)^{-1} = \begin{bmatrix}
v_{00} & v_{01} & v_{02} & \cdots & v_{0k} \\
v_{10} & v_{11} & v_{12} & \cdots & v_{1k} \\
v_{20} & v_{21} & v_{22} & \cdots & v_{2k} \\
\vdots \\
v_{k0} & v_{k1} & v_{k2} & \cdots & v_{kk}
\end{bmatrix}
\]

where \(X'X\) and \((X'X)^{-1}\) are \((k + 1) \times (k + 1)\) matrices.
• Note also that $v_{ij} = v_{ji}$ for all $(i,j)$ i.e., the $(X'X)^{-1}$ matrix is symmetric.

• The least squares estimates $\hat{\beta}_i$ of the parameters $\beta_i$ in the model are elements of the vector $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k)$ where $\hat{\beta}$ is the solution to the normal equations

$$X'X \hat{\beta} = X'y$$

and is given by

$$\hat{\beta} = (X'X)^{-1}X'y$$

• Note that without using matrix notation the normal equations are the following
\[
\sum y = n \hat{\beta}_0 + \sum x_1 \hat{\beta}_1 + \cdots + \sum x_k \hat{\beta}_k \\
\sum x_1 y = \sum x_1 \hat{\beta}_0 + \sum x_1^2 \hat{\beta}_1 + \cdots + \sum x_1 x_k \hat{\beta}_k \\
\vdots \\
\sum x_k y = \sum x_k \hat{\beta}_0 + \sum x_k x_1 \hat{\beta}_1 + \cdots + \sum x_k^2 \hat{\beta}_k
\]

- Thus \( \hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k \) are the solutions to this set of simultaneous equations.

- The standard deviations of the \( \hat{\beta}_i \) are

\[
\sigma_{\hat{\beta}_0} = \sigma_\epsilon \sqrt{v_{00}} \\
\sigma_{\hat{\beta}_1} = \sigma_\epsilon \sqrt{v_{11}} \\
\vdots \\
\sigma_{\hat{\beta}_k} = \sigma_\epsilon \sqrt{v_{kk}}
\]
• The sum of squares of residuals (or SSE) is
\[ \text{SSE} = \mathbf{y}'\mathbf{y} - \hat{\mathbf{\beta}}' \mathbf{X}'\mathbf{y} \]

• The mean square error, i.e., the \( s_\varepsilon^2 \) which we use to estimate \( \sigma_\varepsilon^2 \) is
\[ s_\varepsilon^2 = \text{MSE} = \frac{\text{SSE}}{n - (k + 1)} . \]

• Thus the standard errors of \( \hat{\beta}_j \) are:
\[
\begin{align*}
  s_{\hat{\beta}_0} &= s_\varepsilon \sqrt{v_{00}} \\
  s_{\hat{\beta}_1} &= s_\varepsilon \sqrt{v_{11}} \\
  &\vdots \\
  s_{\hat{\beta}_k} &= s_\varepsilon \sqrt{v_{kk}}
\end{align*}
\]
• Confidence intervals for the $\beta_j$ therefore have the form

$$\hat{\beta}_j \pm t_{\alpha/2} \cdot s_\epsilon \sqrt{v_{jj}}, \quad j = 0, 1, 2, \ldots, k$$

where $t_{\alpha/2}$ is the upper $(1 - \alpha/2)$ percentile of the $t$-distribution with $df = n - (k + 1)$.

• To test $H_0 : \beta_j = 0$ vs. $H_a : \beta_j \neq 0$ use the t-statistic

$$t_j = \frac{\hat{\beta}_j}{s_\epsilon \sqrt{v_{jj}}}$$

with $df = n - (k + 1)$, for each $j = 1, \ldots, k$. 
• A 100(1-\(\alpha\))% confidence interval for \(E(y_{n+1})\), the population mean of \(Y_{n+1}\) at \((x_{n+1,1}, x_{n+1,2}, \ldots, x_{n+1,k})\)

\[
\hat{y}_{n+1} \pm t_{\alpha/2} \cdot s_{\epsilon} \sqrt{\ell'_{n+1}(X'X)^{-1}\ell_{n+1}}
\]

where \(\ell'_{n+1} = (1, x_{n+1,1}, x_{n+1,2}, \ldots, x_{n+1,k})\)

• A 100(1-\(\alpha\))% prediction interval for a new \(y_{n+1}\) simply adds one under the square root above.

\[
y_{n+1} \pm t_{\alpha/2} s_{\epsilon} \sqrt{1 + \ell'_{n+1}(X'X)^{-1}\ell_{n+1}}
\]
Inference On Multiple Regression Model

- In the general linear model the $\beta$’s are sometimes called partial slopes. $\beta_j$ is the expected change in $y$ for unit change in $x_j$ (when all other $x$’s are held constant) i.e., when the other $x$ variables are fixed at a set of values.

- Analysis of variance table for multiple regression

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>k</td>
<td>SSReg</td>
<td>MSReg=$SSReg/k$</td>
<td>$F=MSReg/MSE$</td>
</tr>
<tr>
<td>Error</td>
<td>$n-(k+1)$</td>
<td>SSE</td>
<td>MSE=$SSE/(n-(k+1))$</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>$n-1$</td>
<td>SSTot</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
• The computed F-statistic is used to test the hypothesis 
  \( H_0 : \beta_1 = \beta_2 = \cdots \beta_k = 0 \) vs. \( H_a : \) at least one \( \beta_j \neq 0 \).

• The null hypothesis is rejected if the computed F-value exceeds the \( F_{\alpha,k,n-(k+1)} \) percentile of the F-distribution for selected \( \alpha \) level.

• Quantities like \( s_{\hat{\beta}_j} \), and confidence interval (band) for \( E(y) \), or prediction interval (band) for a future \( y \) cannot reasonably be computed by hand.

• To obtain intervals at \( x \) values for which \( y \) was not observed, include an \( (n + 1)^{th} \) row of \( x \) values in the data set with a missing value (i.e. a period) for the value of \( y \). That is the
line of data is in the form

\[(\cdot, x_1, x_2, \ldots, x_k)\]

- Note that when we say “band” for \(E(y)\) and future \(y\) we are talking about a multivariate band, i.e., a “band” in more than two dimensions.
- Using the period in the \(y\)-value position tells JMP that the \(y\) observation is to be predicted. JMP will ignore the row of \((k + 1)\) \(x\) values when computing sums of squares, least squares estimates, etc., but will the compute predicted value \(\hat{y}\), confidence interval for \(E(y)\), and prediction interval for future \(y\) observation at the given set of \((k + 1)\) \(x\) values.
• The coefficient of determination \( R^2_{y \cdot x_1, x_2, \ldots, x_k} \) is computed as

\[
R^2_{y \cdot x_1, x_2, \ldots, x_k} = \frac{S_{yy} - SSE}{S_{yy}}
\]

• Generally, \( R^2_{y \cdot x_1, \ldots, x_k} \) is the proportion of the variability in \( y \) that is accounted for by the independent variables \( x_1, x_2, \ldots, x_k \) in the model.
More Inference On Multiple Regression Model

• Other than a full model (i.e., one involving all the $x$ variables being considered at the time) one often looks at reduced models (ones involving some but not all of the full set of independent variables.)

• Estimates computed for full and reduced models will usually all be different.

• If the $x_1$ variable is contained in both models, for example, the estimate of the $\beta_1$ coefficient ($\hat{\beta}_1$) will not be the same for the two least squares fits. The Example 12.18 (page 658) illustrates this.
• The term **multicollinearity** is used to describe the situation wherein some of the independent variables have strong correlations amongst themselves.

• If we are using both the high school GPA \((x_1)\), and the SAT score \((x_2)\) in a multiple regression model to predict college GPA \((y)\), it is very likely that the two variables \(x_1\) and \(x_2\) are correlated.

• If correlation is very high for a pair of \(x\) variables the effect is that the estimated coefficients \((\hat{\beta}'s)\) of one or both of these variables will have **very high standard errors**, resulting in the \(t\)-tests for these variables \(H_0 : \beta_j = 0\) vs. \(H_a : \beta_j \neq 0\) failing to reject the respective null hypotheses.
The investigator may make a decision on which of these variables will be retained in the model.

Such a decision is usually based on subject matter knowledge about the predictive value of these variables to the model.

Variables may be retained in the model even in the presence of some multicollinearity if they contribute types of information needed to explain the variability in $y$ other variables in the model may not be able to provide.

In the following section, we will discuss a statistic called vif that may be calculated to determine the variables that are involved in multicollinearity.
Statistics useful for explaining multicollinearity

- The estimated standard errors of the $\hat{\beta}_j$’s can be expressed as functions of $R^2$ values obtained by regressing $x_j$ on the rest of the $x$ variables.

$$s_{\hat{\beta}_j} = s_\epsilon \sqrt{\frac{1}{S_{x_jx_j}(1 - R^2_{x_j \cdot x_1, \ldots, x_j-1, x_j+1, \ldots, x_k})}}$$

- where $S_{x_jx_j} = \sum x_j^2 - \frac{(\sum x_j)^2}{n}$

- and $R^2_{x_j \cdot x_1, x_2, \ldots, x_j-1, x_j+1, \ldots, x_k}$ is the coefficient of determination for regressing $x_j$ on $x_1, x_2, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k$
• That is, $R^2_{x_j \cdot x_1, x_2, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k}$ is the coefficient of determination computed using the variable $x_j$ as the dependent variable and all of the other $x$’s in the original model as the independent variables.

• This definition is useful for illustrating the effect of multicollinearity on the estimation of each coefficient $\beta_j$.

• If the $x_j$ variable is highly collinear with one or more of the other $x$ variables, then $R^2_{x_j \cdot x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k}$ will be large.

• Note that the denominator of the formula above for $s_{\hat{\beta}_j}$ contains the term $(1 - R^2_{x_j \cdot x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k})$.
Thus the denominator of the formula for $s_{\hat{\beta}_j}$ will be small if $R^2_{x_j \cdot x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k}$ is large. This will possibly result in a very large standard error for $\hat{\beta}_j$.

A large standard error for $\hat{\beta}_j$ shows that the coefficient $\beta_j$ will be estimated inaccurately.

Thus the formula for $s_{\hat{\beta}_j}$ shows directly the effect of severe multicollinearity on the estimation.

The quantity variance inflation factor or vif for a coefficient $\beta_j$ is defined as

$$vif_j = \frac{1}{1 - R^2_{x_j \cdot x_1, x_2, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k}}$$
• The vif measures the factor by which the variance (square of the standard error) of $\hat{\beta}_j$ increases because of multicollinearity.

• If the vif is 1 for a particular variable, there is no multicollinearity problem at all with that variable.

• As a rule of thumb a vif of greater than 10 indicates severe collinearity of the $x_j$ with other $x$’s and the standard error of the corresponding estimated coefficient will be inflated.
One noticeable effect of high standard error of $\hat{\beta}_j$ is that it may result in an incorrect sign for $\hat{\beta}_j$.

For example, one would expect the sign of $\hat{\beta}_1$ to be positive if subject matter knowledge suggests that the mean $E(Y)$ must increase when the value of $x_1$ increases (when the other $x$ variable values are unchanged).

However, if the parameter is estimated with high variability, the point estimate of the parameter may show up as a negative value.
Inferences Concerning a Set of $\beta$’s

- We know that we can test hypothesis about any one of the $\beta$’s in the model using a $t$-test.
- That is, a hypothesis like $H_0 : \beta_2 = 0$ employs the test statistic $t = \hat{\beta}_2 / s_{\hat{\beta}_2}$.
- Sometimes we wish to test a hypothesis whether two or more of the $\beta$’s in the model are zero against the alternative hypothesis that at least one of these $\beta$’s is not zero.
- Here we shall assume that $k$ is the total number of predictors ($x$’s) and that $g$ is the number of coefficients considered to be nonzero where, obviously, $g < k$. 
• Thus \( k - g \) represents the number of variables with coefficients hypothesized to be zero.

• To test this hypothesis we use an \( F \) test. To formulate it we first fit the two models:

**Full Model:** The model involving all \( x \)'s.

\[
y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon
\]

**Reduced Model:** The model involving only those \( x \)'s from the full model whose \( \beta \) coefficients are not hypothesized to be zero.

\[
y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_g x_g + \epsilon
\]
• Notice that the $k - g$ variables corresponding to the coefficients to be tested equal to zero are assumed to be the last $k - g$ variables in the full model.

• This can be done without any loss of generality since the order that the variables appear in the full model can always be re-arranged.

• The reason for doing this is that it is easier to state the hypotheses needed to be tested if the two models are formulated as above.

• In this situation, the hypotheses to be tested are

$$H_0 : \beta_{g+1} = \beta_{g+2} = \cdots = \beta_k = 0$$

$$H_a : \text{at least one of these is not zero}.$$
Now the $F$-test statistic for testing $H_0$ is
\[
F = \frac{[\text{SSReg(Full)} - \text{SSReg(Reduced)}]/(k - g)}{\text{MSE(Full)}}
\]

- The numerator and denominator degrees of freedom here for the $F$ statistic are $k - g$ and $n - (k + 1)$, respectively.
- When this computed $F$ exceeds the $F$-value obtained from the $F$-tables for a chosen $\alpha$, $H_0$ is rejected.
Example 12.18 illustrates this kind of a test. The two models considered are:

- \[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \epsilon \]
- \[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon \]

The null hypothesis \( H_0 : \beta_3 = \beta_4 = 0 \) tests whether \( x_3 \) and \( x_4 \) taken together contribute significant predictive value to be included in the Full Model.

Fitting the two models, using the same 20 observations in each case, yields the following:
Full Model:

Prediction Equation:

\[ \hat{y} = -2.784 + 0.02679x_1 + 0.50351x_2 + 0.74293x_3 + 0.05113x_4 \]

ANOVA Table:

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS of</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>4</td>
<td>24.0624</td>
<td>6.0156</td>
<td>39.65</td>
</tr>
<tr>
<td>Error</td>
<td>15</td>
<td>2.27564</td>
<td>0.1517</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td>26.338</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Reduced Model:

Prediction Equation:

\[
\hat{y} = -0.8709 + 0.0394x_1 + 0.828x_2
\]

ANOVA Table:

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS of</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>2</td>
<td>2.9125</td>
<td>1.456</td>
<td>1.06</td>
</tr>
<tr>
<td>Error</td>
<td>17</td>
<td>23.425</td>
<td>1.378</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td>26.338</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
• Thus the test for $H_0 : \beta_3 = \beta_4 = 0$ is:

$$T.S. \quad F = \frac{[SSReg(\text{Full}) - SSReg(\text{Reduced})]/(k - g)}{\text{MSE(\text{Full})}}$$

$$= \frac{(24.0624 - 2.9125)/(4 - 2)}{.1517} = 69.7$$

• For $\alpha = 0.01$, the percentile from the F table is $F_{0.01,2,15} = 6.36$. Thus we reject $H_0$ at $\alpha = 0.01$.

• In this case we have evidence that Access and Structure variables do actually contribute predictive value to the model.
Including Interaction Terms in the Model

- The concept of interaction between two independent variables (say) $x_1$ and $x_2$ is discussed in Section 12.1.
- By definition $x_1$ and $x_2$ do not interact if the expected change in $y$ for unit change in $x_1$, does not depend on $x_2$.
- To illustrate, consider the first-order model in only $x_1$ and $x_2$,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

- That is, consider $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$
● When the value of $x_2$ is kept fixed, $E(y)$ is a straight line function of $x_1$.

● Different values of $x_2$ give parallel lines.

● Thus the expected change in $y$ for unit change in $x_1$ will remain $\beta_1$ regardless of what the value of $x_2$ is.
Now suppose we have the model
\[ E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 \]
\[ = \beta_0 + (\beta_1 + \beta_3 x_2) x_1 + \beta_2 x_2 \]

When the value of \( x_2 \) is kept fixed, \( E(y) \) is still a straight line function of \( x_1 \).
• However, the slope is $\beta_1 + \beta_3 x_2$, which obviously is a function of $x_2$.
• Therefore the slope now depends on the value of $x_2$.
• So, as you see in the picture, the lines are not parallel because the slopes are different for different values of $x_2$.
• This model can be used to represent interaction between $x_1$ and $x_2$, if it exists.
• To detect interaction in regression data, we may construct a scatterplot of the data with the numbers identifying values of the variable $x_2$: 
Here the scatterplot suggests parallel traces, hence no interaction is indicated.

But if the plot looks like the one below, interaction is indicated.
• However, this type of a plot may not always be useful. Subject matter knowledge of the experimenter may suggest that certain interactions terms are needed in the model.

• A term in the model like $\beta x_2 x_3$ is used to account for interaction (a 2-factor interaction term, we say) between $x_2$ and $x_3$. 
Higher order interactions (3-factor, 4-factor, etc.) are conceivable, like $\beta x_1 x_3 x_4$ or $\beta x_1 x_2 x_3 x_4$, but we seldom consider other than possibly 2-factor interactions, mainly because such models become difficult to interpret and therefore not useful.