APPENDIX-A

D.1. Estimation of model coefficients, \( \beta \) in a regression model.

Estimation of model coefficient, \( \beta_s \), in regression model is illustrated for the second order polynomial model used for fitting the data obtained in the optimization of temperature and pH for alcohol production employing the central composite experimental design, viz.,

\[
Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_1^2 x_1^2 + \beta_2^2 x_2^2 + \beta_{12} x_1 x_2 + \epsilon \quad (D.1)
\]

where \( Y \) denotes the observed response, \( x_1 \) and \( x_2 \) represent the levels of the real value of the variables, and \( \epsilon \) represents the random error in \( Y \). The random errors are assumed to be independently distributed variable with a zero mean and a common variance, \( \sigma^2 \).

In matrix notation the model in equation D.1, over \( N \) observations, is

\[
Y = X \beta + \epsilon \quad (D.2)
\]

where

\[
Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{pmatrix}, \quad X = \begin{pmatrix} 1 & X_1 & X_2 & X_1^2 & X_2^2 & X_1 X_2 \\ 1 & X_{11} & X_{12} & X_{13} & X_{14} & X_{15} \\ 1 & X_{21} & X_{22} & X_{23} & X_{24} & X_{25} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{N_1} & X_{N_2} & X_{N_3} & X_{N_4} & X_{N_5} \end{pmatrix}
\]

\( N \times 1 \) \hspace{0.5cm} \( N \times (5+1) \)
The normal equations are
\[ X'Xb = X'Y \] (D.3)

where \( X' \) is the transpose of matrix \( X \) and \( b \) is the matrix of coefficient estimates.

The solution to these normal equations are
\[ b = (X'X)^{-1} X'Y \] (D.4)

where \((X'X)^{-1}\) is the inverse of \(X'X\). Both \(X'X\) and \((X'X)^{-1}\) are symmetric matrices.

The fitted second-order model in the coded variable is
\[ Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_{11}x_1^2 + \beta_{22}x_2^2 + \beta_{12}x_1x_2 + \varepsilon \] (D.5)

The predicted values of the response under study (\(Y\)) are obtained by regressing the equation D.5 using MATLAB software.
D.2. Construction of the Analysis of Variance (ANOVA) table

After experimentation, the data are analyzed and the results of the analysis are displayed in a tabular form. The table is called Analysis of Variance (ANOVA) table. The entire data in the table represent measures of information concerning the separate sources of variation in the data.

The total variation in a set of data is called the total sum of squares (SST). The quantity SST is compared by summing the squares of the deviations of the observed response \( (Y_u) \) about their average value for \( N \) observations,

\[
\bar{Y} = \frac{(Y_1 + Y_2 + Y_3 + \ldots + Y_N)}{N}
\]

\[
SST = \sum_{u=1}^{N} (Y_u - \bar{Y})^2
\]  
(D.6)

The quantity SST associate with \( N-1 \) degrees of freedom since the sum of deviations, \( Y_u - \bar{Y} \), is equal to zero.

The total sum of square can be partitioned into two parts, the sum of squares due to regression (or sum of squares explained by the fitted model) and the sum of squares unaccounted for by the fitted model. The formulation for calculating the sum of squares due to regression (SSR) is

\[
SSR = \sum_{u=1}^{N} (Y(x_u) - \bar{Y})^2
\]  
(D.7)
The deviation $Y(x_u) - Y$ is the difference between the value predicted by the fitted model for the $u^{th}$ observation and the overall average of the $Y_u$'s. If the fitted model contains 'p' parameters, the number of degree of freedom associated with SSR is $p-1$.

The sum of squares unaccounted for by the fitted model (SSE) is

$$SSE = \sum_{u=1}^{N} [(Y_u - Y(X_u))^2]$$  \hspace{1cm} (D.8)

The quantity SSE is also called the sum of squares of the residual or the sum of squares of the errors. The number of degrees of freedom for SSE is the difference $(N-1)-(p-1) = N-p$.

The usual test of the significance of the regression equation is test of the null hypothesis $H_0$: all values of $\beta_1$ (excluding $\beta_0$) is not zero.

Assuming normality of the errors, the test of $H_0$ involves first calculating the value of the F-statistic by the following equation.

$$F = \frac{\text{Mean square regression}}{\text{Mean square residual}} = \frac{\text{SSR/(p - 1)}}{\text{SSE/(N - p)}}$$  \hspace{1cm} (D.9)

If the null hypothesis is true, the F-static in equation D.9 follows an F-distribution with $p-1$ and $N-p$ degrees of freedom in the numerator and in the denominator respectively. The second step of the test of $H_0$ is to compare the value of $F$ in equation D.9 to the table value, $F_{\alpha, p-1, N-p}$ which is the upper 100 $\alpha$ percent point of the F distribution with $p-1$ and $N-p$ degrees of freedom, respectively. If the value of $F$ in equation D.9 exceeds $F_{\alpha, p-1, N-p}$ then the null hypothesis is rejected at the $\alpha$ level of significance and it is inferred that the variation accounted for by the model is significantly greater than the unexplained variation.
Another accompanying statistic is the multiple coefficient of determination (R^2)

\[ R^2 = \frac{SSR}{SST} \]  
(D.10)

The value of R^2 is a measure of the proportion of total variation of the values of Y_u about the mean Y explained by the fitted model. It is often expressed in a percent. The analysis of variance table is shown in as table.

Analysis of Variance (ANOVA) table:

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Degree of freedom (df)</th>
<th>Sum of squares (SS)</th>
<th>Mean squares (MS)</th>
<th>F value</th>
<th>Probability&gt;F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Due to P-1 SSR</td>
<td>p-1</td>
<td>SSR</td>
<td>SSR/(p-1)</td>
<td>MSR/MSE</td>
<td></td>
</tr>
<tr>
<td>Regression model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual (error)</td>
<td>N-p</td>
<td>SSE</td>
<td>SSE/(N-p)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>N-1</td>
<td>SST</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ F = \frac{\text{Mean square regression}}{\text{Mean square residual}} \]