11.0 MODEL BUILDING
(Updated Spring 2005)

Model Building Steps:

1. Postulation of a tentative model form:
   \[ y = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2 + \varepsilon \]
   This is a linear model with interaction term. If we add second order terms, then a full second order model is
   \[ y = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2 + b_{11}x_1^2 + b_{22}x_2^2 + \varepsilon \]

2. Design of an appropriate experiment - Supplies data to
   - Estimate model parameters
   - Place model in jeopardy

   Linear model \( \rightarrow 2^k \) factorial design

   2nd order model \( \rightarrow \) Central Composite Design (\( 3^k \) factorial designs)

3. Fitting of the model to the data
   - Are all the terms necessary?

4. Diagnostic checking of the model
   - Are the model terms sufficient?
   - Lack of fit, model adequacy.

Diagnostic Checking of the fitted Model

- Plots of the model residuals, \( e = y - \hat{y} \)
- Should be centered about mean of zero
- Should be normally distributed
- Scatter in residuals should not be dependent on any of the x’s, time, y or \( \hat{y} \), or any of the variables related to the experiment.

- Normal Prob Plot of the Resids
- Resids vs. time order of the tests
- Resids vs. \( \hat{y} \)
- Resids vs. \( x_1, x_2, \ldots \)
- For a replicated expt. plot (\( y_{ij} - \bar{y}_i \)) vs. \( \bar{y}_i \)

In calculating a pooled sample variance (ANOVA, replicated factorial designs, independent t-test, etc.) we assumed that the variance was constant for all the responses. We use residual plots to check this, but there is an
analytical test for replicated experiments.

**Bartlett’s test for Homogeneity of the Variance (M.S. Bartlett)**

Previously, to calculate \( s_{eff}^2 = \frac{4}{N} s_P^2 \)

\[ s_p^2 = \frac{\sum s_i^2}{m} \] uses H0: \( \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_m^2 \)

To test this hypothesis,

\[ \chi^2_{calc} = 2.3026 \frac{q}{c} \]

where

\[ q = (N - m) \log_{10} s_P^2 - \sum_{i=1}^{m} (n_i - 1) \log_{10} s_i^2 \]

where \( N = n_1 + n_2 + \ldots + n_m \)

\[ c = 1 + \frac{1}{3(m - 1)} \left( \frac{1}{\sum_{i=1}^{m} \frac{1}{n_i - 1}} - \frac{1}{n - m} \right) \]

Compare \( \chi^2_{calc} \) to \( \chi^2_{m - 1, 1 - \alpha} \). Sensitive to normality of y’s. n should be greater than 4.