Proceedings of the 2014 International Conference on Industrial Engineering and Operations Management
Bali, Indonesia, January 7 – 9, 2014

Kronecker Models Versus Scheffé's Mixture Models

Javier Cruz-Salgado
Graduate School of Industrial Engineering and Manufacturing
Omega #201, Fracc. Delta
CIATEC, León, Guanajuato 37545, México

Sergio Alonso, Roberto Zitzumbo
Materials Research
Omega #201, Fracc. Delta
CIATEC, León, Guanajuato 37545, México

Abstract

In this article we compared the Kronecker model against the Scheffé model for a mixture experiment. The use of various diagnostic measures provides an effective tool in making such a comparison, particularly when the experimental region is highly constrained. We investigated conditioning using the variance inflation factor, the maximum and minimum eigenvalues of the information matrix, and the conditional number, to assess conditioning. The pseudocomponents transformation for lower bounds (L-pseudocomponents) is also discussed. A numerical example is presented.

Keywords
Kronecker model, Scheffé model, mixture constraints, L-pseudocomponents.

1. Introduction
Your paper should be formatted for 8 1/2" by 11" US standard paper format.

Mixtures experiments are those in which the response variables depend only on the relative proportion of the ingredients or components of the mixture. These proportions are connected by a linear restriction,

$$x_1 + x_2 + \cdots + x_q = 1 \tag{1}$$

Some examples of designs for mixtures include different bread flour blends, chemicals with different additives, different types of wines, and food consisting of different ingredients. Data from experiments using mixtures are usually modeled by using quadratic polynomial models Scheffe (S-model). For a review of these models, see Cornell J. A., (1990).

A general mixture model, in matrix terms, can be presented as $Y = X\beta + \epsilon$ or $E(Y) = X\beta$. To estimate the parameters in β via least squares, the following expression can be used:

$$\hat{\beta} = (XX')^{-1}X'y, \tag{2}$$

Where the covariance matrix is $V(\hat{\beta}) = (XX')^{-1}\sigma^2$. The vector of fitted values is given by $\hat{y} = X\beta$ and the residual vector is $e = y - \hat{y} = y - X\hat{\beta}$. Usually the vector ε is assumed to follow a normal distribution, that is $\epsilon \sim N(0, \sigma^2)$.

In addition to equation (1), some restrictions on the proportions of the mixture may exist. In the development of wood plastic composites, for example, a coupling agent is required to ensure that the wood flour is integrated into the polymeric matrix. The relative proportion of such an agent is usually at most 2.5%, which is extremely small range in terms of mixtures. This can cause difficulties in adjusting the model derived from ill-conditioning; that is,

the columns of the corresponding matrix $\mathbf{X}$ can be almost linearly dependent. As a consequence, $\mathbf{X'X}$ (called 'information matrix' to estimate $\boldsymbol{\beta}$) required to compute the estimated parameters $\hat{\boldsymbol{\beta}}$ in (2), could be almost singular. The consequences of an ill-conditioning matrix $\mathbf{X'X}$ are that the least squares estimates of the parameters in $\boldsymbol{\beta}$ have a large standard error, and are highly correlated. Furthermore, those estimates are highly dependent on the precise location of the points of design (Prescott, Dean, Draper, & Lewis, 2002).

A large number of settings of a mixture model can be obtained by an appropriate substitution of equation (1). All full model parameterizations lead to the same predictions and prediction intervals thereof. Surprisingly, however, different parameterizations lead to information matrices with different degrees of ill-conditioning (Prescott, Dean, Draper, & Lewis, 2002).
An alternative polynomial model form is the Slack Variable model (SV-model), which is obtained by designating one mixture component as a slack variable; that is, by expressing it in terms of the remaining $k-1$ components using equality (1), then substituting it into S-model. The purpose of this process is to produce mixture models that depend on $\mathbf{k-1}$ independent variables. The pros and cons of using the SV-model, as an alternative to the S-model, have generated many discussions between researchers and practitioners. This issue has been discussed recently in both Cornell J. A., (2000) and Khuri (2005).

The idea behind using a slack variable undermines the fundamental property of the mixture designs, which is that the relative proportions of the blend components are not independent (JA Cornell, 2000). Piepel (2009) considered four situations in which the SV-model is used and explained, for the four situations, that it is generally preferable to use a different approach to the mixing experiment. A second alternative model, introduced by Draper and Pukelsheim (1998), is the Kronecker quadratic model (K-model). This model is less susceptible to ill-conditioning and the associated problem of unstable models with highly correlated parameters and large standard error (Prescott, Dean, Draper, & Lewis, 2002). The K-model is discussed in more detail below.

The aim of this paper is to show that the K-model is a quadratic model specification that is less susceptible to ill-conditioning, compared to the S-model, when working with full quadratic models. A contrary case occurs when working with reduced quadratic models or sub-models either of the K-model or the S-model. This statement is also true if the original proportions of the ingredients are transformed into pseudocomponents (JA Cornell, 1990). Several diagnostic measurements are shown in this article. A Cornell (2000) example is used for comparison between the S and the K-model. We also investigate the use of pseudocomponents to improve conditioning in the S-model and the K-model.

1. Notation

In this article we consider quadratic mixture models. The most common of these is the canonical model of the second order Scheffé (S-model). This has the general form

$$E(Y) = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_q x_q + \beta_{12} x_1 x_2 + \cdots + \beta_{q-1,q} x_{q-1} x_q. \quad (3)$$

The alternative Kronecker model (K-model) contains only terms of second order and has the form

$$E(Y) = \alpha_{11} x_1^2 + \alpha_{22} x_2^2 + \cdots + \alpha_{qq} x_q^2 + \alpha_{12} x_1 x_2 + \cdots + \alpha_{q-1,q} x_{q-1} x_q. \quad (4)$$

A full quadratic mixture S-model has $(q+1)/2$ terms, just as the quadratic K-model. Any nonsingular mixture model with $(q+1)/2$ terms can be transformed in the quadratic K-model by replacing the linear term $\mathbf{x_i}$ by

$$x_i = x_i(X_1 + x_2 + \cdots + x_q) = x_i^2 + \sum_{j \neq i} x_i x_j. \quad (5)$$

For $\mathbf{1 \leq i \leq q}$, and replacing the constant term by

$$1 = \sum_{i=1}^{q} x_i = \sum_{i=1}^{q} x_i^2 + \sum_{i=1}^{q} \sum_{j=1(j \neq i)} x_i x_j. \quad (6)$$

For example, the quadratic S-model (3) with $\mathbf{q = 4}$ becomes

$$E(Y) = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{14} x_1 x_4 + \beta_{23} x_2 x_3 + \cdots + \beta_{34} x_3 x_4 \quad (7)$$

By applying the transformation (5) into each linear term, it then becomes

$$E(Y) = \alpha_1 x_1^2 + \alpha_2 x_2^2 + \alpha_3 x_3^2 + \alpha_4 x_4^2 + (\beta_{12} + \beta_1 + \beta_2)x_1 x_2 + (\beta_{13} + \beta_1 + \beta_3)x_1 x_3 + (\beta_{14} + \beta_1 + \beta_4)x_1 x_4 + (\beta_{23} + \beta_2 + \beta_3)x_2 x_3 + (\beta_{24} + \beta_2 + \beta_4)x_2 x_4 + (\beta_{34} + \beta_3 + \beta_4)x_3 x_4$$

,

which is identical to the quadratic K-model with four components

$$E(Y) = \alpha_{11} x_1^2 + \alpha_{22} x_2^2 + \alpha_{33} x_3^2 + \alpha_{44} x_4^2 + \alpha_{12} x_1 x_2 + \alpha_{13} x_1 x_3 + \alpha_{14} x_1 x_4 + \alpha_{23} x_2 x_3 + \cdots + \alpha_{24} x_2 x_4 + \alpha_{34} x_3 x_4,$$

(8)

2. Some Diagnostic Measures

As mentioned earlier, in mixtures experiments it is common that one or more components of a mixture have a very small interval, causing a very restricted experimental region for that component. This results in collinearity (or linear dependence) between the columns of the matrix X in Y = Xβ + ε. One way to reduce collinearity is considering eliminating, in the full model, one or more terms that contribute significantly to the collinearity. Therefore, selecting the submodel that best fits the data is of great interest in highly restricted regions (Khuri, 2005).

Below we briefly explain some diagnostic measures that can help to detect or identify collinearity (Montgomery, 2012) and (JA Cornell, 1990).

2.1 Multiple correlation coefficient

We defined x_j as the jth column of X and X_j as the matrix that results when the column x_j is deleted from X. Then R_j^2 is the multiple correlation coefficient obtained by regressing x_j on X_j. When the column of constants of the X matrix is not available, the unadjusted multiple correlation coefficient can be obtained by

$$R_j^2 = \frac{x_j' x_j (x_j' x_j)^{-1} x_j' x_j}{x_j' x_j} \quad (9)$$

For $j = 1, \ldots, p$.

2.2 Variance Inflation Factor

The variance inflation factor (VIF) associated with the estimated regression coefficient βj is given by

$$VIF(\beta_j) = (1 - R_j^2)^{-1}, \quad j = 1, \ldots, p, \quad (10)$$

2.3 Condition Number

Allow $\lambda_{max} \geq \lambda_2 \geq \cdots \geq \lambda_{p-1} \geq \lambda_{min}$ to be the p eigenvalue of $X'X$, which are p solutions to the determinant equation

$$|X'X - \lambda I| = 0,$$

which is a polynomial with p roots.

There are many definitions of the condition number of a matrix (Pearson, 1974). The general definition used in applied statistics is the square root of the ratio of the maximum to the minimum eigenvalues of X'X, which is denoted by

$$cond(X'X) = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}} \quad (8)$$

The above quantities are called collinearity diagnostic measures. Small values of λ_{min} and large values of λ_{max} indicate the presence of collinearity. Lower values of the condition number indicate more stability or more conditioning in the least squares estimate, and also show small values of FIV.

3. Il –Conditioning Example

This example was introduced by Cornell (2000) for explaining the comparison between both S and SV-models when the experimental space is highly restricted. Over a region, one or more components of the mixture have a very small range causing collinearity between the columns of X on the adjusted model. This example is based on a pharmaceutical study, which has three co-solvents denoted by x1, x2 and x3 with water as x4. The experimental region is defined by the following restrictions:

$$0.10 \leq x_1 \leq 0.40, \quad 0.10 \leq x_2 \leq 0.40$$
$$0.005 \leq x_3 \leq 0.30, \quad 0.30 \leq x_4 \leq 0.795 \tag{9}$$

A 10-point D-optimal design was selected to fit a quadratic model. The design consists in 6 of the 10 extreme vertices and midpoints of the edges of the region defined by the constraints in equation (9). The experiment was replicated once. We can note by the restrictions in (9) that x3 has a very small proportion range. Table 1 shows the experimental arrangement in natural units and the 20 values of the solubility response.

Table 1. Ten-point design in the in original components along with the solubility values of two replicates.

Blend	Polyethylene Glycol x_1	Glycerin x_2	Polysorbate 60 x_3	Water x_4	Solubility Y	
1	0.400	0.270	0.030	0.300	7.7	9.1
2	0.100	0.400	0.030	0.470	6.6	5.2
3	0.100	0.100	0.030	0.770	3.3	4.8
4	0.400	0.295	0.005	0.300	9.5	8.2
5	0.100	0.100	0.005	0.795	3.9	3.4
6	0.100	0.400	0.005	0.495	6.9	6
7	0.280	0.400	0.020	0.300	10.2	11.1
8	0.400	0.100	0.020	0.480	11.7	12.6
9	0.400	0.200	0.005	0.395	10.7	11.8
10	0.200	0.400	0.005	0.395	8.7	9.5

Table 2 shows the VIFs for the adjusted models (3) and (4) with the values of Table 1.

Table 2. VIF values and condition numbers for the cuadratic S-model and K-model for the exaple (original components)

S-model		K-model	
Coefficient	VIF	Coeficient	VIF
β_1	4,375	α_{11}	616.6
β_2	15,336	α_{22}	2,094.2
β_3	1,288,200	α_{33}	1115.9
β_4	662.18	α_{44}	284.2
β_{12}	4,876.5	α_{12}	481
β_{13}	102,380	α_{13}	159.5
β_{14}	1,601.4	α_{14}	174.79
β_{23}	115,120	α_{23}	177.42
β_{24}	9,068	α_{24}	1,398.78
β_{34}	371,850	α_{34}	546.53
Condition no.	76,034	Condition no.	22,118

Table 2 shows that all of the K-model VIFs are significantly lower for the 10 parameters (p), compared to the S-model. This is a clear indication that an overall improvement in stability and conditioning is provided by the full quadratic K-model. The maximum value of the eigenvalues of the information matrix is lower in the K-model than in the S-model (2.4 compared to 7.67). The minimum value of the eigenvalues for the K-model is greater in comparison with the S-model ($4.9 * 10^{-5}$ compared to $1.3 * 10^{-5}$). The condition number of the matrix information for the K-model (22.118) is much less than that of S-model (76.034) indicating a more stable analysis.

The $n = 20$ solubility values in Table 1 were adjusted with the S-quadratic model, as well as the K-quadratic model, by applying the transformation (5) into the S-model. The models become, respectively

S-model:

$$\hat{Y} = -97.78x_1 - 112.97x_2 - 21538.3x_3 - 21.65x_4 + 323.23x_1x_2 + 22382.6x_1x_3 + \cdots + 215.94x_1x_4 + 22445x_2x_3 + 226.38x_2x_4 + 22345.3x_3x_4 \quad (10)$$

K-model:

$$\hat{Y} = -97.78x_1^2 - 112.97x_2^2 - 21538.3x_3^2 - 21.65x_4^2 + 112.48x_1x_2 + 746.52x_1x_3 + \cdots + 96.5x_1x_4 + 793.7x_2x_3 + 91.76x_2x_4 + 785.3x_3x_4 \quad (11)$$

It is important to mention that the disproportionately large values of the coefficients in both (10) and (11) associated with the factor x3 are somewhat disturbing. The next point of interest is the reduction of the S-model and the quadratic K-model, in an effort to get rid of the terms that have a coefficient estimate disproportionately large. Table 4 lists the results of fitting models to the data of Table 1 using the techniques forward (F), backward (B) and stepwise regression (St.) at 0.10 significance level. It is noteworthy that for the three techniques (F, B and St.) the same 6-parameters submodel was obtained.

In Table 4 it can be seen that the values of the estimated coefficients, both of the Kronecker and Sheffer submodels, are lower than the complete model coefficients (10) and (11). Moreover, the VIFs values of the Sheffer submodel, compared to those of the K-model, do not differ significantly, except the linear term VIF x_1 (473.59) of the S-model and the quadratic term VIF x_1^2 (101.86) of the K-model. The maximum value of the eigenvalues of the information matrix of the Kronecker submodel is less than the Scheffé submodel (2.21 compared to 7.42) but the minimum value of the eigenvalues for the Kronecker submodel is much smaller than the Scheffé submodel ($2.0*10^{-6}$ compared with $4.0*10^{-4}$). The condition number of the information matrix of Kronecker submodel (882.11) is larger than the Scheffé submodel (134.92), which indicates that the reduced S-model provides a more stable analysis than the reduced K-model.

Tabla 4. Best submodel and corresponding VIF and condition number (original components)

Variable constant	Scheffé quadratic model		Kronecker quadratic model	
	Coefficient (β_i)	VIF	Coefficient (α)	VIF
x_1	-53.04	473.59	-	
x_2	8.39	16.56	-	
x_3	18.03	3.34	-	
x_4	-7.86	18.66	-	
x_1x_2	82.17	114.74	37.52	37.23
x_1x_3	-		-	
x_1x_4	170.32	288.60	109.42	81.85
x_2x_3	-		-	
x_2x_4	-		-	
x_3x_4	-		-	
x_1^2	-		-53.04	101.86
x_2^2	-		8.39	14.96
x_3^2	-		18.03	1.92
x_4^2	-		-7.86	11.54
Condition no.	134.92		882.11	

4. Use of Pseudo components

The coordinates of the experimental region in the previous example (section 4) is redefined using the following L-pseudo components (JA Cornell, 1990).

$$A = \frac{x_1 - 0.10}{1 - 0.505}$$

$$B = \frac{x_2 - 0.10}{0.495}$$

$$C = \frac{x_3 - 0.005}{0.495}$$

$$D = \frac{x_4 - 0.30}{0.495}$$

Table 5 shows the experimental arrangement redefined by the aforementioned L-pseudo components, along with the 20 values of the solubility response. Table 6 shows the VIFs models adjusted for (3) and (4) using the values in Table 5.

Table 5. Ten-point design in the L-pseudo components in along with the solubility values of two replicates.

	L-Pseudo components				Solubility	
Blend	A	B	C	D	Y	
1	0.606	0.343	0.051	0	7.7	9.1
2	0	0.606	0.051	0.343	6.6	5.2
3	0	0	0.051	0.949	3.3	4.8
4	0.606	0.394	0	0	9.5	8.2
5	0	0	0	1	3.9	3.4
6	0	0.606	0	0.394	6.9	6
7	0.364	0.606	0.03	0	10.2	11.1
8	0.606	0	0.03	0.364	11.7	12.6
9	0.606	0.202	0	0.192	10.7	11.8
10	0.202	0.606	0	0.192	8.7	9.5

Table 6. VIF values and condition numbers for the quadratic S-model and K-model for the example (L-pseudo components)

S-model		K-model	
Coefficient	VIF	Coefficient	VIF
β_1	356.94	α_{11}	121.42
β_2	965.39	α_{22}	313.59
β_3	186,230	α_{33}	425.01
β_4	2.37	α_{44}	1.86
β_{12}	933.16	α_{12}	217.98
β_{13}	30,452	α_{13}	84.77
β_{14}	50.62	α_{14}	14.16
β_{23}	34,750	α_{23}	101.16
β_{24}	419.37	α_{24}	157.9
β_{34}	59,717	α_{34}	162.9
Condition no.	16,819.68	Condition no.	6,086.15

Table 6 shows again that all of the K-model VIFs, are significantly lower for the 10 parameters (p) compared to the S-model. This remains a clear indication that an overall improvement in stability and conditioning is provided by the full quadratic K-model. From the experimental arrangement, redefined by the L-pseudo components, it is again found that the maximum value of the eigenvalues of the information matrix is lower in the K-model than in the S-model (3.78 compared to 6.69). The minimum value of the eigenvalues for the K-model is larger in comparison to that of the S-model ($1.02 * 10^{-7}$ compared with $2.3 * 10^{-8}$). The condition number of the matrix information

for the K-model (6086.15) remains much lower than that of the S-model (16819.68), indicating a more stable analysis.

The $^{11-20}$ solubility values in Table 5 were adjusted with the S-quadratic model, as well as the quadratic K-model, by applying the transformation (5) into the S-model. The models become respectively
S-model (L-pseudo components):

$$\hat{Y} = -7.75A - 13.56B - 5023.86C + 3.65D + 79.12AB + 5315.21AC + 52.88AD + 5330.63BC + 55.41BD + 5305.96CD \quad (12)$$

K-model (L-pseudo components):

$$\hat{Y} = -7.75A^2 - 13.56B^2 - 5023.86C^2 + 3.65D^2 + 57.81AB + 283.60AC + 48.78AD + 293.21BC + 45.50BD + 285.75CD \quad (13)$$

With the experimental arrangement redefined by the L-pseudo components, the values of the estimated coefficients in both (12) and (13), were significantly reduced. Although the coefficient of the C factor remains disproportionately long. The next step would be to reduce both the S-model and the K-quadratic model to reduce the estimated coefficient C. Table 7 shows the results of fitting the models to the data presented in Table 5 using the techniques of forward (F), backward (B) and stepwise regression (St.) at 0.10 significance level. It is noteworthy that for the three techniques (F, B and St.) the same 6 parameters sub-model was obtained.

Table 7. Best sub-model and corresponding VIF and condition number (L-pseudo components).

Variable constant	Scheffé quadratic model		Kronecker quadratic model	
	Coefficient (β)	VIF	Coefficient (α)	VIF
A	2.32	30.38	-	
B	7.42	3.03	-	
C	8.04	1.99	-	
D	3.74	1.53	-	
AB	20.11	22.04	29.85	9.94
AC	-		-	
AD	41.66	9.14	47.72	4.90
BC	-		-	
BD	-		-	
CD	-		-	
A^2	-		2.32	13.72
B^2	-		7.42	2.71
C^2	-		8.04	1.96
D^2	-		3.74	1.29
Condition no.	32.17		411.39	

Table 7 shows that the values of the estimated coefficients, both of the Sheffe sub-model as well as those of the K-model, are smaller than the complete models (12) and (13), and much lower in comparison to those in (10) and (11). It can also be stated that the VIFs of the Kronecker sub-model are slightly lower than those of the sub-model Sheffer. The maximum value of the eigenvalues of the information matrix of the Kronecker submodel is less than the Scheffé sub-model (3.77 compared to 6.61) but the minimum value of the eigenvalues for the Kronecker sub-model is much smaller than the Scheffé sub-model ($2.22*10^{-5}$ compared with $6.3*10^{-3}$). The condition number of the information matrix of Kronecker sub-model (411.39) is greater than the Scheffé sub-model (32.17), which indicates that the S-reduced model provides a more stable analysis than the K-model.

5. Conclusions

It is common, in mixture experiments with highly constrained experimental regions, to have problems adjusting models derived from ill-conditioning. The estimated coefficients obtained by adjusting standard models such as the S-model and the K-model, seem not to be representative of the data. In this article we compared the quadratic Kronecker model set against the Scheffe's quadratic model, commonly used in experiments with mixtures. Conditioning was assessed in both adjusted models using the variance inflation factor, both maximum and minimum eigenvalues of the information matrix, and condition number. As was shown by using the K-model, ill-conditioning can be reduced compared to using the full S-model. Moreover, to evaluate the conditioning in reduced models, it was found that the Scheffe sub-model indicates a more stable analysis compared to the Kronecker sub-model. Reducing both models produced more consistent coefficient estimates compared to observations.

Acknowledgments

The authors thank the National Council for Science and Technology (CONACYT) for the scholarship granted to M.Sc. Cruz-Salgado and Leslie Ann Knapp for the English language revision.

References

Cornell, J. A. (1990). Experiments with Mixtures. New York: Wiley.

Cornell, J. A. (2000). Fitting a slack-variable model to mixture data: some questions raised. Journal of Quality Technology, 32, pp. 133–147.

Khuri, A. I. (2005). Slack-variable Models Versus Scheffe's Mixture Models. Journal of Applied Statistics, 9, 887–908.

Montgomery, D. C. (2012). Introduction to Linear Regression Analysis. New Jersey: Wiley.

Pearson, C. E. (1974). Handbook of Applied Mathematics. New York: Van Nostrand Reinhold.

Peixoto, J. L. (1990). A Property of well-formulated polynomial regression models. The American Statistician, 44 26-30.

Piepel, G. F., & Landmesser, S. M. (2009). Mixture Experiment Alternatives to the Slack Variable Approach. Quality Engineering, 262-276.

Prescott, P., Dean, A. M., Draper, N. R., & Lewis, S. M. (2002). Mixture Experiments: