

Diseño de Experimentos para Mezclas con R

¹Javier Cruz-Salgado y ²Sandra Paola Garcidueñas-González

¹Jefe del Departamento de Investigación y Desarrollo Tecnológico, Universidad Politécnica del Bicentenario

jcruzs@upbicentenario.edu.mx.

²Inteligencia Mercados, Investigadora Independiente.

sandragarciduenas@hotmail.com

Recibido: 30 de Octubre

Aceptado: 12 Noviembre

RESUMEN

Los diseños de experimentos para mezclas (DEPM) son una estrategia ordenada y eficiente para el desarrollo y mejoramiento de productos del tipo de mezcla, como gasolinas, pegamentos, compuestos plásticos, etc. En el DEPM normalmente se utiliza el modelo de Scheffe para el análisis de los datos experimentales, una característica particular de este modelo es que no tiene el término intercepto. Esta característica del modelo de Scheffe puede originar problemas en el análisis si no se ajusta el modelo adecuadamente. Existen softwares estadísticos comerciales que utilizan el modelo de Scheffé de manera correcta, sin embargo, adquirir alguno de estos software puede ser una limitante debido al costo. R es un software estadístico de libre acceso, el cual puede ser utilizado para aplicar el DEPM, siempre que se aplique el procedimiento correcto. En este artículo se describe un procedimiento para ajustar el modelo de Scheffé utilizando R, para de esta forma poder aplicar el DEPM, sin tener que adquirir un software comercial.

Palabras claves: Diseño de mezclas, Modelo de Scheffe, Análisis estadístico con R.

1. INTRODUCCIÓN

Los Diseños de Experimentos para Mezclas (DEPM) son empleados comúnmente en diversas industrias, como: la química, la farmacéutica, la alimenticia y recientemente en la industria del plástico (Chatree, Thanate, & Wiriya, 2013). En dichos experimentos, los productos son desarrollados mediante la mezcla de diferentes ingredientes o componentes. A diferencia del diseño de experimentos convencional, en el DEPM los factores del diseño son las proporciones de los componentes de la mezcla (Cornell, 2002). Existen DEPM especiales en donde se consideran factores adicionales como la cantidad total de la mezcla (Piepel & Cornell, 1985), ciertas variables del proceso (Kowalski, Cornell, & Vining, 2002) e incluso existen DEPM en donde un componente de una mezcla principal, es a su vez, una mezcla de otros componentes (Kang, Joseph, & Brenneman, 2011). En este artículo nos enfocamos solo en los DEPM donde los únicos factores del diseño son las proporciones de los componentes de la mezcla; las diferentes variantes del DEPM.

Para un DEPM con q componentes, sus proporciones x_1, x_2, \dots, x_q deben satisfacer la siguiente restricción:

$$x_1 + x_2 + \dots + x_q = 1 \quad (1)$$

Esta restricción diferencia al DEPM de otros tipos de diseños experimentales (Cornell, 2002). Comúnmente la región experimental está sujeta a restricciones adicionales de la forma:

$$a_i \leq x_i \leq b_i$$

(2)

Las restricciones del tipo Ec. (2) se presentan cuando las proporciones de uno o más componentes están limitadas a un intervalo. Para los DEPM no solo la región experimental queda restringida, también el modelo utilizado debe satisfacer esta restricción.

Existen DEPM con restricciones múltiples de las proporciones de los componentes de la forma:

$$C_k \leq \sum_{i=1}^q A_{ki}x_i \leq D_k \quad k = 1, 2, \dots, K \quad (3)$$

tales como el presentado por (Piepel G. F., 2005).

El modelo más comúnmente empleado en los DEPM es el modelo de Scheffé desarrollado por Scheffé H. (Scheffe, 1958). Existen modelos de mezclas alternativos al modelo de Scheffé, como el modelo de variable de holgura (Slack-Variable model) (Cruz-Salgado, 2015) y el modelo de Kronecker. Cada modelo tiene sus pros y sus contras, el modelo de variable de holgura ofrece una interpretación más clara y directa para interpretar los efectos de los componentes principales de manera individual, sin embargo, al aplicar el modelo de variable de holgura, uno de los componentes se utiliza como variable de holgura y queda fuera del análisis, lo cual puede tener ciertas complicaciones si el componente utilizado como variable de holgura tiene un fuerte efecto sobre la variable de respuesta (Kang, Cruz-Salgado, & Brenneman, 2015). El modelo de Kronecker en ciertos casos presenta una matriz de información mejor condicionada que la matriz de información del modelo de Scheffé (Prescott, Dean, Draper, & Lewis, 2002), pero este es un modelo poco común ya que todos sus términos son de orden superior, lo cual puede tener ciertas complicaciones en su interpretación.

En este artículo solo se hace mención del modelo de mezclas de Scheffé ya que es el modelo más comúnmente empleado en el DEPM. El modelo de Scheffé se deriva del modelo general polinomial utilizado en los modelos de regresión lineal mediante la sustitución de la restricción de la Ec. (1) en una forma específica. Para un polinomio de grado uno con dos componentes,

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2, \quad (4)$$

donde y representa una variable de respuesta y las β 's son parámetros desconocidos a ser estimados. Aplicando la restricción de la Ec. (1) es posible transformar el polinomio general en el modelo-S mediante la sustitución de la Ec. (1) en la Ec. (4), de esta forma:

$$x_1 + x_2 = 1,$$

$$y = \beta_0(1) + \beta_1x_1 + \beta_2x_2,$$

sustituyendo (1) en (4) se tiene:

$$y = \beta_0(x_1 + x_2) + \beta_1x_1 + \beta_2x_2$$

resolviendo se obtiene

$$y = (\beta_0 + \beta_1)x_1 + (\beta_0 + \beta_2)x_2.$$

Así, el modelo de Sheffé lineal tiene la forma:

$$y = \sum_{i=1}^q \beta_i x_i,$$

(5)

El modelo-S cuadrático con q componentes tiene la forma:

$$y = \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^{q-1} \sum_{j=i+1}^q \beta_{ij} x_i x_j,$$

(6)

El lector notara que el modelo de Scheffé no contiene el termino intercepto u ordenada al origen β_0 , esto es debido a la incorporación de la restricción de la Ec. (1), esto es un punto importante a tomar en cuenta al realizar un análisis en un DEPM, ya que si el análisis sin tomar en cuenta la forma del modelo de Scheffé la interpretación y conclusiones pueden ser erróneas. La gran mayoría de los investigadores que aplican el DEPM comúnmente emplean el modelo-S, incluso los softwares estadísticos más populares como: Minitab®, Statgraphics® y Design Expert®, utilizan el modelo de Scheffé para el ajuste de los datos experimentales. Si se utiliza alguno de estos softwares comerciales se podría tener la confianza de que el análisis será llevado acabo con la forma del modelo de Scheffé, sin embargo, la adquisición de un software comercial no siempre está al alcance de todos los investigadores o estudiantes que deseen aplicar el DEPM.

Existe un software estadístico de libre acceso llamado R, R es un software estadístico gratuito, el cual puede ser instalado en una gran variedad de plataformas como: UNIX, Windows y MacOS. Con R es posible realizar cualquier análisis que ofrezca cualquier de los tres software comerciales antes mencionados. Sin embargo, al ser un software de libre acceso no cuenta con una interface tan intuitiva como la de los softwares comerciales.

El objetivo de este artículo es brindar a los investigadores interesados en aplicar el DEPM una guía sobre como ajustar y analizar el modelo de Scheffé, utilizando el software de libre acceso R, para de esta forma superar la limitante económica de los softwares comerciales.

2. DESCARGA DE R Y DOCUMENTOS DE AYUDA

La mejor manera de aprender un nuevo lenguaje de programación es probando los comandos. R puede ser descargado gratuitamente de:

<https://www.r-project.org/>

En la página principal de R se puede obtener una gran cantidad de material para aprender a utilizar R. El manual de usuario puede ser descargado en diversos formatos de:

<https://cran.r-project.org/manuals.html>

Así mismo, existe una inmensa cantidad de libros relacionados a R, en la siguiente dirección se pueden encontrar 152 libros relacionados a R:

<https://www.r-project.org/doc/bib/R-books.html>

este artículo no pretende ser un tutorial sobre el uso general de R, específicamente se limita al ajuste y análisis del modelo de Scheffé en el DEPM.

3. EJEMPLO DE DEPM EN INDUSTRIA DEL PLÁSTICO

En (Cruz-Salgado & Alonso-Romero, 2015) los autores describen un experimento para optimizar las propiedades mecánicas de un compuesto de madera plástica (WPC) con tereftalato de etileno (PET) como matriz polimérica. En este artículo se presenta el mismo experimento pero utilizando el enfoque de DEPM. Los puntos experimentales del DEPM se determinaron mediante un diseño de experimentos para mezclas D-óptimo (Cornell, 2002). Los componentes analizados, fueron: PET (x_1), madera en polvo (x_2) y un agente de acoplamiento E-GMA (x_3). Estos componentes están sujetos a la restricción de la Ec. (1), ya que la suma de los porcentajes de los componentes debe representar el 100%. Además, los componentes están sujetos a restricciones del tipo de la Ec. (2). Las restricciones inferiores y superiores de los porcentajes de contribución de los componentes de la mezcla, del tipo de la Ec. (2), se muestran en la Tabla 1.

Tabla1: Componentes de la mezcla y sus restricciones.

Componente	Restricción de la fracción (wt%)
PET (x_1)	$60 \leq x_1 \leq 90$
Madera (x_2)	$10 \leq x_2 \leq 40$
E-GMA (x_3)	$0 \leq x_3 \leq 3$

El diseño consta de tres corridas de puntos del modelo, 5 para estimar la falta de ajuste y 5 réplicas, dando un total de 13 corridas experimentales. El diseño experimental se muestra en la Tabla 2.

Tabla 2: Diseño de experimentos y variables de respuesta.

Corridas	Fracciones de los componentes de la mezcla			Variables de respuesta		
	(% en peso)			(kgf/mm^2)		
	PET (x_1)	Madera (x_2)	E-GMA (x_3)	Flexión (Y_1)	Compresión (Y_2)	Tensión (Y_3)
1	87	10	3	3	4.8	9.71
2	87	10	3	3.02	4.91	9.72
3	60	40	0	1.44	5.06	10.05
4	82.125	17.125	0.75	2.83	5.07	10
5	60	37	3	1.52	5.22	10.37
6	90	10	0	3.19	4.81	9.35
7	87	10	3	3.02	4.94	9.82
8	67.125	32.125	0.75	1.83	5.33	10.28
9	60	40	0	1.36	5.04	10.29
10	60	37	3	1.49	5.23	10.44
11	90	10	0	3.32	4.89	9.53
12	74.25	24.25	1.5	2.15	5.25	10.31
13	67.125	30.625	2.25	1.74	5.39	10.6

3.1 AJUSTE DEL MODELO DE SCHEFFÉ EN R

Para realizar el análisis de los datos experimentales mostrados en la Tabla 1, primeramente se ajusta el modelo cuadrático de Scheffé mostrado en la Ec. (6) a los datos. Para esto, se requiere cargar los datos en R. La forma más sencilla para cargar los datos es utilizando una hoja de notas, es posible copiar y pegar los datos de la Tabla 1 en un bloc de notas, de forma que se obtenga algo similar a la Figura 1.

x1	x2	x3	y1	y2	y3
87	10	3	3	4.8	9.71
87	10	3	3.02	4.91	9.72
60	40	0	1.44	5.06	10.05
82.125	17.125	0.75	2.83	5.07	10
60	37	3	1.52	5.22	10.37
90	10	0	3.19	4.81	9.35
87	10	3	3.02	4.94	9.82
67.125	32.125	0.75	1.83	5.33	10.28
60	40	0	1.36	5.04	10.29
60	37	3	1.49	5.23	10.44
90	10	0	3.32	4.89	9.53
74.25	24.25	1.5	2.15	5.25	10.31
67.125	30.625	2.25	1.74	5.39	10.6

Figura 1: Datos del diseño experimental en bloc de notas.

El bloc de notas se guarda en el disco duro de la pc en una dirección conveniente. Para cargar los datos a R, se escribe la siguiente instrucción en la pantalla de comandos:

```
dat<-read.table(file.choose(), header=T)
```

“dat” sera el nombre que se asignará a la matriz de datos, “read.table” es la función de R que permite cargar los datos, seleccionando el archivo del bloc de notas donde capturamos la información. Al escribir la instrucción en la pantalla de comandos de R se abrirá una ventana emergente como se muestra en la Figura 2.

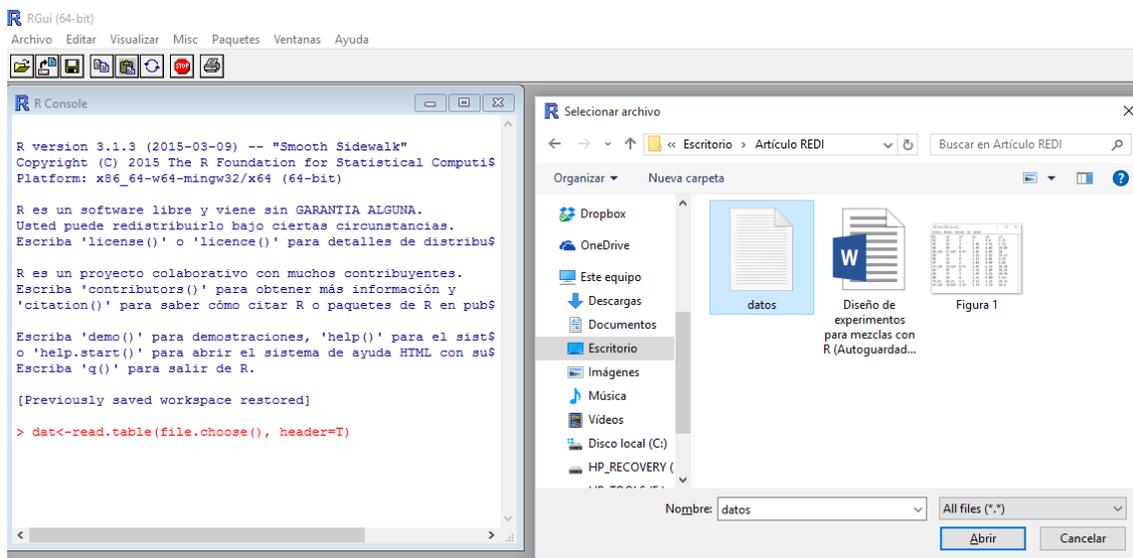


Figura 2: Ventana emergente R.

Al seleccionar el archivo del bloc de notas, la matriz del diseño se guardara en la memoria de R. Una vez cargados los datos, se puede proceder a indicar el modelo cuadrático de Scheffé que será ajustado a los datos, para esto se debe escribir la siguiente instrucción en la pantalla de comandos:

```
Scheffel<-y1~-1+x1+x2+x3+I(x1*x2)+I(x1*x3)+I(x2*x3)
```

```
Scheffe2<-y2~-1+x1+x2+x3+I(x1*x2)+I(x1*x3)+I(x2*x3)
```

```
Scheffe3<-y3~-1+x1+x2+x3+I(x1*x2)+I(x1*x3)+I(x2*x3)
```

Con la instrucción anterior se cargaran tres modelos cuadráticos en R, uno para cada una de las tres variables de respuesta. Nótese que en la instrucción de los tres modelos se incluye el término “-1” después del símbolo “~”, esto le indica a R que el modelo no tiene el termino intercepto (β_0). Es indispensable incluir en la instrucción del modelo el término “-1”, de otra forma R asumirá que el modelo tiene el termino intercepto, lo que puede dar como resultado que los estadísticos calculados sean incorrectos y por consecuencia las conclusiones serán incorrectas también.

Cuando se han indicado los modelos que se ajustaran a los datos experimentales, se procede a escribir la instrucción para realizar el ajuste en la pantalla de comandos, la instrucción es la siguiente:

```
scheffe1_fit<-lm(scheffe1, data=dat)
```

```
scheffe2_fit<-lm(scheffe2, data=dat)
```

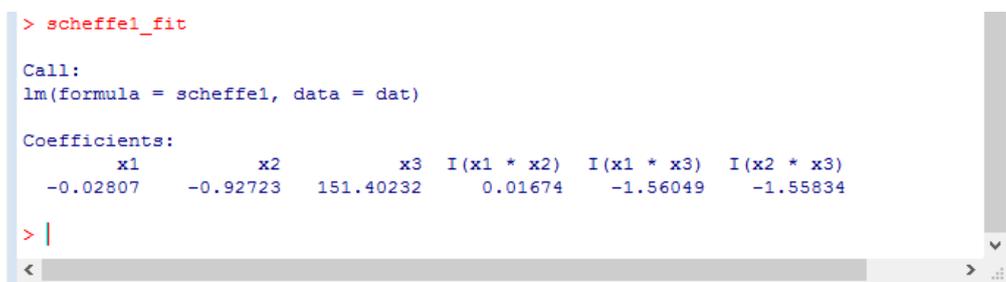
```
scheffe3_fit<-lm(scheffe3, data=dat)
```

La función “lm()” es la función de R para ajustar un modelo lineal, la sintaxis es muy sencilla el termino “scheffe1”, en la primer línea y dentro de la función “lm()”, es el modelo que se ajustara a los datos, el término “data=dat” indica cuales son los datos que se ajustaran.

En este punto los tres modelos ya fueron ajustados a los datos experimentales, para ver el modelo ajustado, con los parámetros estimados mediante mínimos cuadrados solo tenemos que escribir el nombre del modelo que dimos de alta en R, por ejemplo, para el modelo de la variable de respuesta “y1”, escribimos en la pantalla de comandos la instrucción:

```
scheffe1_fit
```

En la consola se mostrara el modelo mediante los coeficientes, como se muestra en la Figura 3.



```
> scheffe1_fit
Call:
lm(formula = scheffe1, data = dat)

Coefficients:
      x1          x2          x3  I(x1 * x2)  I(x1 * x3)  I(x2 * x3)
-0.02807  -0.92723  151.40232   0.01674  -1.56049  -1.55834

> |
```

Figura 3: Coeficientes del modelo ajustado.

Para ver los coeficientes de los otros dos modelos se sigue el mismo procedimiento.

Para conocer el resumen estadístico de los modelos ajustados se escribe la siguiente instrucción en la consola de comandos:

```
summary(scheffel_fit)
```

En la consola de comandos se mostrara la información de los residuales, los coeficientes estimados junto con sus errores estándar estimados, el valor-t y el valor-p. Así mismo, se mostrara El error estándar residual, el coeficiente de correlación R^2 y su versión ajustada R_{aju}^2 , el estadístico F y el valor P de la regresión. El la Figura 4 se muestra un ejemplo:

```
R Console
> summary(scheffel_fit)

Call:
lm(formula = scheffel, data = dat)

Residuals:
    Min       1Q   Median       3Q      Max
-0.074128 -0.033022  0.005085  0.022918  0.055940

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
x1          -0.028069   0.024333  -1.154   0.2866
x2          -0.927226   0.327522  -2.831   0.0254 *
x3          151.402315  53.512164   2.829   0.0254 *
I(x1 * x2)    0.016736   0.006065   2.760   0.0281 *
I(x1 * x3)   -1.560490   0.551280  -2.831   0.0254 *
I(x2 * x3)   -1.558343   0.551211  -2.827   0.0255 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05147 on 7 degrees of freedom
Multiple R-squared:  0.9998,    Adjusted R-squared:  0.9995
F-statistic: 4773 on 6 and 7 DF,  p-value: 2.821e-12

> |
<
```

Figura 4: Resumen estadístico del modelo ajustado.

Para obtener un intervalo de confianza para los coeficientes estimado, se puede utilizar la función “confint()”. En la Figura 5 se muestra un ejemplo de la función “confint()”.

```
> confint(scheffel_fit)
                2.5 %          97.5 %
x1          -0.085607821   0.02946997
x2          -1.701692091  -0.15275948
x3          24.866155300 277.93847493
I(x1 * x2)   0.002394909   0.03107751
I(x1 * x3)  -2.864061018  -0.25691971
I(x2 * x3)  -2.861749497  -0.25493670

> |
<
```

Figura 5: Intervalo de confianza de los coeficientes estimados.

Es posible obtener información de los residuales mediante su representación gráfica, para esto se utiliza la siguiente instrucción:

```
plot(scheffel_fit)
```

Al escribir la instrucción en la consola de comandos se debe teclear enter para que se generen las diferentes graficas de información de los residuales. La primera gráfica que se genera es la gráfica de los residuales contra los valores ajustados. La gráfica de los residuales contra los valores ajustados de la variable de respuesta y_1 se muestra en la Figura 6.

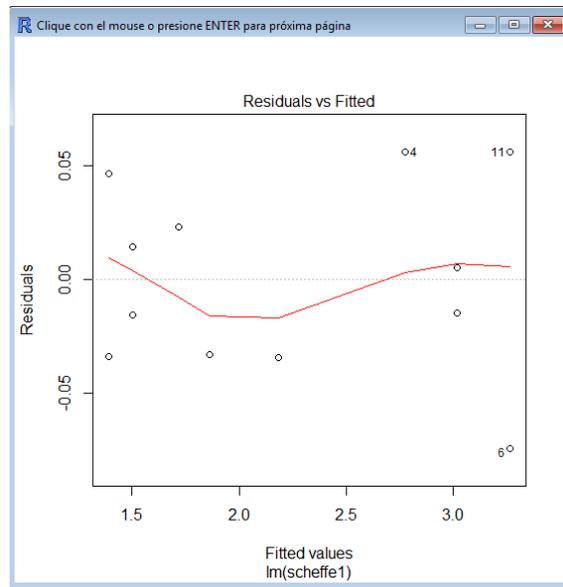


Figura 6: Gráfica de residuales contra valores ajustados.

Es posible obtener también la tabla del análisis de la varianza del modelo ajustado mediante la siguiente instrucción:

```
anova(scheffe1_fit)
```

Con la función “anova()” R muestra la información de la tabla del análisis de la varianza (ANOVA por sus siglas en ingles). Un ejemplo del análisis de la varianza se muestra en la Figura 7.

```
> anova(scheffe1_fit)
Analysis of Variance Table

Response: y1
      Df Sum Sq Mean Sq  F value    Pr(>F)
x1      1  74.220   74.220 28012.9755 7.173e-14 ***
x2      1   1.602    1.602  604.7768 4.684e-08 ***
x3      1   0.012    0.012   4.6190 0.06870 .
I(x1 * x2) 1   0.011    0.011   4.1230 0.08185 .
I(x1 * x3) 1   0.016    0.016   5.8848 0.04570 *
I(x2 * x3) 1   0.021    0.021   7.9926 0.02551 *
Residuals 7   0.019    0.003
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figura 7: Tabla de análisis de la varianza.

4. CONCLUSIONES

En el presente artículo se da una descripción de como ajustar el modelo cuadrático de Scheffé, para poder realizar el análisis de un DEPM. Dado que el modelo de Scheffé no tiene el término intercepto (β_0), se debe poner mucha atención a la hora de ajustar dicho modelo a un conjunto de datos

experimentales. Con lo presentado en este artículo, los investigadores interesados en aplicar un DEPM, cuentan con una alternativa de análisis de libre acceso, superando así la limitante económica del costo de adquirir un software estadístico comercial.

Referencias

- Chatree, H., Thanate, R., & Wiriya, T. (2013). Minimizing the creep of recycled polypropylene/rubberwood flour composites with mixture design experiments. *Journal of Composite Materials*, 1-10.
- Cornell, J. A. (2002). *Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data*. Wiley.
- Cruz-Salgado, J. (2015). Selecting the Slack Variable in Mixture Experiment. *Ingeniería Investigación y Tecnología*, 16(4), 613-623.
- Cruz-Salgado, J., Alonso-Romero, S., & Domínguez-Domínguez, J. (2015). Optimization of the Tensile and Flexural Strength of a Wood-PET Composite. *Ingeniería Investigación y Tecnología*, 105-112.
- Kang, L., Joseph, R. V., & Brennenman, W. (2011). Design and Modeling Strategies for Mixture-of-Mixtures Experiments. *Technometrics*, 125-136.
- Kowalski, S. M., Cornell, J. A., & Vining, G. G. (2002). Split-Plot Designs and Estimation Methods for Mixture Experiments With Process Variables. *Technometrics*, 72-79.
- L., K., J., C.-S., & W., B. (2015). Comparing the Slack-Variable Mixture Model with Other Alternatives. *Technometrics*, In press.
- P., P., A., M. D., Draper, N. R., & S., M. L. (2002). P. Prescott, A. M. Dean, Draper, NMixture Experiments: ILL-Conditioning and Quadratic Model Specification. *Technometrics*, 260-268.
- Piepel, G. F. (2005). Construction of a 21-component Layered Mixture Experiment Design Using a New Mixture Coordinate-Exchange Algorithm. *Quality Engineering*, 17, 579-594.
- Piepel, G., & Cornell, J. A. (1985). Models for Mixture Experiments When the Response Depends on the Total Amount. *Technometrics*, 219-227.
- Scheffe, H. (1958). Experiments with mixtures. *Journal of the Royal Statistical Society*, 20, 344-360.