

UNIVERSIDAD DE COSTA RICA  
SISTEMA DE ESTUDIOS DE POSGRADO

ALGORITMO PARA LA REDUCCIÓN DE TRATAMIENTOS EN DISEÑOS  
EXPERIMENTALES CON MEZCLAS DE VÉRTICES EXTREMOS, BASADO EN LA  
RECUPERACIÓN DE LA FORMA REGULAR DE UN SIMPLEX

Tesis sometida a la consideración de la Comisión del Programa de Estudios de  
Posgrado en Estadística para optar al grado y título de Maestría Académica en  
Estadística

STEVEN GARCÍA GOÑI

Ciudad Universitaria Rodrigo Facio, Costa Rica

2025

## DEDICATORIA

A todas las personas que estuvieron durante este camino, que resultó más largo de lo planteado. A mi familia, amigos, amigas, compañeros y compañeras que me prestaron su oído en los momentos en los que lo necesité. A Carlos y Allan, que llegaron a nuestra vida de forma sorpresiva a poner todo de cabeza y que ahora forman parte de ella y sirven de motor.

*“Espero que la posteridad me juzgue con benevolencia, no solo por las cosas que he explicado, sino también por aquellas que he omitido voluntariamente, con el fin de permitirles disfrutar del placer que produce su descubrimiento”.*

*La geometría – René Descartes*

## AGRADECIMIENTO

A Patricia Ramírez Barrantes, a quien considero mi mentora y consejera, por mostrarme en el año 2016 el maravilloso mundo del diseño de experimentos, sus enseñanzas hicieron que me enamorara de este tema. A Ronny Pacheco Segura, quien desde el año 2014, ha sido un ejemplo a seguir, gracias por todas sus enseñanzas.

De igual manera agradezco a Ricardo Alvarado Barrantes y Shirley Rojas Salazar, me llevo la mejor de las impresiones de ustedes. Muchas gracias por su infinita paciencia para conmigo durante este trabajo final de graduación; la vida no siempre es color de rosa y su comprensión y buenos deseos, durante los múltiples retos personales que enfrenté, los llevo dentro. Muchas gracias además por su invaluable guía y dedicación en la revisión de este documento. También, a Gilbert Brenes Camacho, por tomarse todo el tiempo del mundo en atender mis consultas en el curso de Modelos Lineales Generalizados, mi primer curso del posgrado. Su vocación es admirable y me hicieron amar esta carrera. Finalmente, al resto del cuerpo docente del posgrado, por las enseñanzas que dejaron en mí.

“Esta tesis fue aceptada por la Comisión del Programa de Estudios de Posgrado en Estadística de la Universidad de Costa Rica, como requisito parcial para optar al grado y título de Maestría Académica en Estadística”

---

MSc. Patricia Ramírez Barrantes  
**Representante de la Decana  
Sistema de Estudios de Posgrado**

---

PhD. Ricardo Alvarado Barrantes  
**Director de Tesis**

---

PhD. Shirley Rojas Salazar  
**Asesora**

---

PhD. Ronny Pacheco Segura  
**Asesor**

---

PhD. Gilbert Brenes Camacho  
**Director  
Programa de Posgrado en Estadística**

---

Steven Ignacio García Goñi  
**Candidato**

## TABLA DE CONTENIDO

DEDICATORIA.....	ii
AGRADECIMIENTO .....	ii
TABLA DE CONTENIDO.....	iv
RESUMEN.....	vii
ABSTRACT .....	viii
LISTA DE CUADROS .....	ix
LISTA DE TABLAS .....	xi
LISTA DE FIGURAS .....	xii
1. INTRODUCCIÓN .....	1
1.1 Contexto.....	2
1.2 Justificación.....	5
1.3 Problema de investigación .....	7
1.4 Objetivos.....	8
1.4.1 Objetivo general .....	8
1.4.2 Objetivos específicos.....	8
2. MARCO TEÓRICO.....	10
2.1 Diseño de experimentos.....	10
2.1.1 Conceptos básicos.....	10
2.1.2 Principios básicos de experimentación .....	12
2.1.3 Historia breve del diseño de experimentos .....	13
2.2 Modelos de superficies de respuesta .....	15
2.3 Experimentos con mezclas.....	20
2.3.1 Diseños símples .....	24

2.3.2	Diseños símplex ampliados con puntos axiales.....	27
2.3.3	Diseños con restricciones de frontera.....	28
2.3.4	Algoritmos de generación de coordenadas.....	40
2.4	Polinomios canónicos .....	46
2.4.1	Colinealidad.....	50
2.4.2	Reducción del modelo.....	51
2.4.3	Falta de ajuste de los modelos.....	51
2.5	Otras formas de polinomios .....	52
2.6	Criterios para la selección entre diseños competentes .....	53
2.7	Función de deseabilidad .....	58
2.8	Experimentos con mezclas y variables de proceso.....	60
2.9	Experimentos mezcla – cantidad .....	62
3.	ABORDAJE METODOLÓGICO.....	64
3.1	Diseño del algoritmo.....	66
3.1.1	CONVEXSIM – I.....	67
3.1.2	CONVEXSIM – S.....	69
3.1.3	Ejemplo de aplicación de los algoritmos para $p = 3$ .....	71
3.1.4	Ejemplo de aplicación de los algoritmos para $p = 4$ .....	76
3.2	Diseño de la simulación .....	78
3.3	Caso práctico .....	85
3.3.1	Contexto del caso práctico .....	85
3.3.2	Componentes de la mezcla .....	88
3.3.3	Variables controladas .....	89
3.3.4	Variables no controladas.....	90

3.3.5	Variables de respuesta .....	91
3.3.6	Selección del diseño experimental.....	92
3.4	Herramientas de generación y análisis de los datos .....	92
4.	RESULTADOS DEL ESTUDIO DE SIMULACIÓN.....	94
4.1	Detalle de la simulación .....	94
4.2	Cantidad de corridas experimentales .....	101
4.3	Eficacia .....	106
4.4	Eficiencia.....	114
5.	CASO PRÁCTICO.....	119
5.1	Eficacia y eficiencia del diseño.....	119
5.2	Análisis de resultados .....	120
5.3	Optimización simultánea de las variables de respuesta .....	126
6.	CONCLUSIONES Y LIMITACIONES .....	128
6.1	Conclusiones.....	128
6.2	Limitaciones e investigaciones futuras.....	130
	BIBLIOGRAFÍA.....	132

## RESUMEN

La investigación que se desarrolla a continuación persigue como su objetivo el diseñar un algoritmo para la reducción de tratamientos en diseños experimentales con mezclas de vértices extremos generados por restricciones en la frontera superior, basado en la recuperación de la forma símplex, reduciendo a su vez el tamaño de la región de experimentación. El algoritmo es nombrado CONVEXSIM (CONvertir Vértices Extremos a SIMplex) y cuenta con dos variantes: I y S. Con estos se logra convertir en seis pasos a las restricciones originales en el límite superior en restricciones de frontera inferiores y superiores que decantan en una forma L – Símplex, reduciendo así el tamaño de la región de experimentación y la cantidad de tratamientos experimentales.

Mediante un estudio de simulación determinística, que tiene como insumo la variación de la cantidad de componentes ( $p$ ), la cantidad de restricciones sobre la frontera superior y el tamaño de estas, el grado del polinomio ( $m$ ) máximo que se puede estimar, así como la presencia o ausencia de puntos axiales y la cantidad total de réplicas; se concluye que el algoritmo conduce a la reducción en la cantidad de tratamientos, consiguiendo generar diseños que pueden ser eficaces y que la mayoría de las ocasiones son eficientes. El estudio del caso práctico permite demostrar que la principal desventaja de CONVEXSIM no es un obstáculo para la consecución de los objetivos de un programa de experimentación, obteniendo resultados razonables en función del contexto del experimento.

Finalmente, se subraya que el algoritmo diseñado no es la panacea, y que su principal desventaja recae en que al reducir la región de experimentación el óptimo global puede quedar fuera y no ser estudiado. Se recomienda que su uso esté sujeto al contexto del experimento y a la perspicacia de las personas experimentadoras para moverse dentro de la región de operabilidad mediante el uso de la experimentación secuencial.

## ABSTRACT

The research developed below aims to design an algorithm for reducing treatments in experimental designs with mixtures of extreme vertices generated by upper boundary constraints, based on the recovery of the simplex form, while simultaneously reducing the size of the experimental region. The algorithm, named CONVEXSIM (CONVert EXtreme vertices to SIMplex), has two variants: I and S. Through a six-step process, these variants transform the original upper boundary constraints into both lower and upper boundary constraints, resulting in an L-Simplex form, which reduces the size of the experimental region and the number of experimental treatments.

A deterministic simulation study was conducted using variables such as the number of components ( $p$ ), the number and magnitude of upper boundary constraints, the degree ( $m$ ) of the highest polynomial estimable, the presence or absence of axial points, and the total number of replicates. The findings demonstrate that the algorithm effectively reduces the number of treatments, generating designs that are not only practical but also, in most cases, efficient. A practical case study further demonstrates that the primary limitation of CONVEXSIM does not hinder achieving the objectives of an experimental program, yielding reasonable results within the context of the experiment.

Finally, it is emphasized that the designed algorithm is not a panacea, as its main limitation lies in the possibility that reducing the experimental region might exclude the global optimum, leaving it unexplored. It is recommended that the algorithm's use be contextualized to the experiment and supported by the experimenters' expertise, leveraging sequential experimentation to navigate the operability region effectively.

## LISTA DE CUADROS

Cuadro 2.1. Diferencia en el cálculo del criterio de A – Optimalidad y D – Optimalidad con base en las diferencias en el orden de las corridas experimentales.....	57
Cuadro 3.1. Resultados de la aplicación del paso iii. de CONVEXSIM – I sobre un diseño de vértices extremos con $p = 3$ .....	72
Cuadro 3.2. Rangos ( $R_i$ ) obtenidos en el paso iv. de CONVEXSIM – I sobre un diseño de vértices extremos con $p = 3$ .....	72
Cuadro 3.3. Resultados de la aplicación del paso iv. de CONVEXSIM – I sobre un diseño de vértices extremos con $p = 3$ .....	73
Cuadro 3.4. Resultados de la aplicación del paso v. de CONVEXSIM – I sobre un diseño de vértices extremos con $p = 3$ .....	73
Cuadro 3.5. Transformación en pseudocomponentes del diseño generado por el algoritmo CONVEXSIM – I, sobre un diseño de vértices extremos con $p = 3$ .....	74
Cuadro 3.6. Resultados de la aplicación de CONVEXSIM – S sobre un diseño de vértices extremos con $p = 3$ .....	75
Cuadro 3.7. Resultados de la aplicación de CONVEXSIM – I sobre un diseño de vértices extremos con $p = 4$ .....	77
Cuadro 3.8. Transformación en pseudocomponentes del diseño generado por el algoritmo CONVEXSIM – I y CONVEXSIM – S, sobre un diseño de vértices extremos con $p = 4$ .....	77
Cuadro 4.1. Cantidad de restricciones generadas para cada número de componentes variados en la simulación .....	95
Cuadro 4.2. Cantidad de restricciones inferiores y superiores válidas que fueron generadas por CONVEXSIM – I y CONVEXSIM – S .....	95
Cuadro 4.3. Cantidad de restricciones inferiores y superiores inválidas que fueron generadas por CONVEXSIM – I y CONVEXSIM – S .....	99
Cuadro 4.4. Cantidad de casos desechados por error en la función $X_{vert}$ de $\{mixexp\}$ o interrupción de R y RStudio .....	100

Cuadro 4.5. Ejemplo de matriz de diseño errónea para el conjunto de restricciones empleadas cuando $p = 3$ y $m = 3$ .....	107
Cuadro 5.1. Valores de optimalidad y eficiencia para el diseño simplex de la región de operabilidad completa, el diseño de vértices extremos original y el diseño creado con CONVEXSIM en proporciones y pseudocomponentes .....	120
Cuadro 5.2. Coeficientes de regresión, error estándar y FIV para la variable de respuesta densidad .....	121
Cuadro 5.3. Coeficientes de regresión, error estándar y FIV para la variable de respuesta oleosidad.....	125
Cuadro 5.4. Deseabilidad individual y valor esperado para cada variable de respuesta	127
Cuadro 5.5. Proporción en la que debe participar cada componente en la mezcla del producto cosmético .....	127

## LISTA DE TABLAS

Tabla 3.1. Resumen, rango de variación y descripción de las variables de entrada utilizadas en el diseño del estudio de simulación.....	79
Tabla 3.2. Variables de salida que son empleadas para medir la eficacia y la eficiencia de los diseños generados por CONVEXSIM.....	82
Tabla 3.3. Variables de operación determinadas a partir del experimento de Taguchi..	89

## LISTA DE FIGURAS

Figura 1.1. Modelo general de un proceso o sistema .....	1
Figura 2.1. La naturaleza secuencial de los RSM .....	16
Figura 2.2. Superficie de respuesta de primera orden y ruta del ascenso más empinado (perpendicular a la estimación) .....	18
Figura 2.3. Diseño central compuesto para $k = 2$ .....	19
Figura 2.4. Diseño de Box – Behnken para tres factores .....	19
Figura 2.5. Geometría del símplex .....	21
Figura 2.6. Espacio de mezclas para dos componentes .....	22
Figura 2.7. Sistema coordenado trilineal .....	22
Figura 2.8. Diseño símplex reticular $\{3, 2\}$ .....	25
Figura 2.9. Diseño símplex centroide para $p = 3$ y $p = 4$ .....	27
Figura 2.10. Diseño símplex reticular $\{3,2\}$ aumentado con axiales y centroide general	28
Figura 2.11. Subregión del espacio símplex original redefinida con pseudocomponentes inferiores .....	31
Figura 2.12. Región símplex invertida dentro de la región símplex original .....	33
Figura 2.13. Intersección entre el diseño símplex original y el diseño símplex invertido por pseudocomponentes de frontera superior .....	35
Figura 2.14. Región doblemente restringida con restricciones de frontera y restricciones multicomponente .....	45
Figura 2.15. Funciones de deseabilidad individuales para optimización simultánea .....	59
Figura 2.16. Experimento símplex – centroide de tres componentes con un arreglo factorial completo $2^3$ .....	61
Figura 2.17. Diseño símplex – centroide de tres componentes para tres niveles ( $A_i$ ) del total de una mezcla .....	63
Figura 3.1. Resumen de la secuencia de pasos para ambos algoritmos .....	70
Figura 3.2. Resultados geométricos de la aplicación de CONVEXSIM – I sobre un diseño de vértices extremos con $p = 3$ .....	74

Figura 3.3. Resultados geométricos de la aplicación de CONVEXSIM – S sobre un diseño de vértices extremos con $p = 3$ .....	75
Figura 3.4. Resultados geométricos de la aplicación de CONVEXSIM – I o CONVEXSIM – S sobre un diseño de vértices extremos con $p = 4$ .....	78
Figura 3.5. Esquematización del programa de experimentación para la producción de un exfoliante con subproductos de trigo.....	87
Figura 3.6. Diagrama de flujo de una corrida experimental.....	90
Figura 4.1. Representación en grupos de triadas, cuando el componente restante es $x_i = 0.05$ , de una región de experimentación reducida en la que no aplican los algoritmos CONVEXSIM .....	97
Figura 4.2. Representación en grupos de triadas de una región de experimentación reducida en la que no aplican los algoritmos CONVEXSIM .....	98
Figura 4.3. Reducción de corridas experimentales entre el diseño de vértices extremos y la alternativa con CONVEXSIM.....	101
Figura 4.4. Reducción de corridas experimentales entre el diseño de símplex reticular y la alternativa con CONVEXSIM, en función del grado del polinomio ( $m$ ) .....	102
Figura 4.5. Diferencia entre los promedios de la cantidad de corridas experimentales por cada tipo de experimento analizado en función de la cantidad de réplicas.....	103
Figura 4.6. Reducción de corridas experimentales entre el diseño de vértices extremos y la alternativa con CONVEXSIM, en función de la presencia o no de puntos axiales ....	104
Figura 4.7. Reducción de corridas experimentales entre el diseño de vértices extremos y la alternativa con CONVEXSIM en función de la cantidad de restricciones impuestas sobre la frontera superior en el diseño de vértices extremos.....	105
Figura 4.8. Cantidad de diseños en proporciones que cumplen con cada uno de los criterios de optimalidad según la cantidad de componentes.....	108
Figura 4.9. Cantidad de diseños en proporciones que cumplen con cada uno de los criterios de optimalidad según el grado del polinomio ( $m$ ).....	109
Figura 4.10. Cantidad de diseños en pseudocomponentes que cumplen con cada uno de los criterios de optimalidad según la cantidad de componentes .....	110

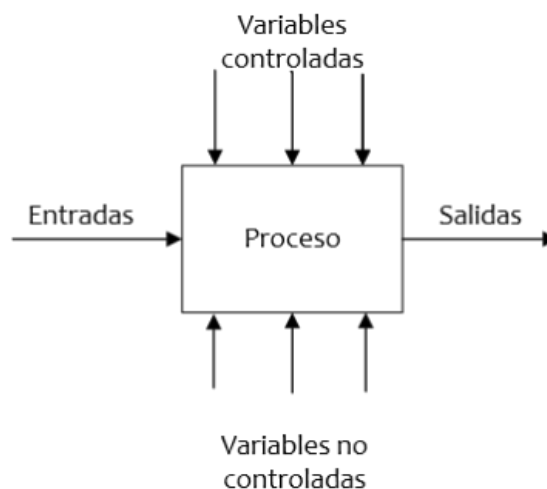
Figura 4.11. Cantidad de diseños en pseudocomponentes que cumplen con cada uno de los criterios de optimalidad según el grado del polinomio (m).....	111
Figura 4.12. Correlaciones entre las hipótesis de trabajo y los criterios de optimalidad para proporciones .....	112
Figura 4.13. Gráfico de correlaciones entre las hipótesis de trabajo y los criterios de optimalidad para pseudocomponentes.....	113
Figura 4.14. Cantidad de diseños en proporciones que cumplen con cada uno de los criterios de eficiencia según la cantidad de componentes .....	115
Figura 4.15. Cantidad de diseños en proporciones que cumplen con cada uno de los criterios de eficiencia según el grado del polinomio (m).....	116
Figura 4.16. Cantidad de diseños en pseudocomponentes que cumplen con cada uno de los criterios de eficiencia según la cantidad de componentes .....	117
Figura 4.17. Cantidad de diseños en pseudocomponentes que cumplen con cada uno de los criterios de eficiencia según el grado del polinomio (m) .....	118
Figura 5.1. Primer grupo de gráficos de contorno lineales para la variable de respuesta densidad.....	123
Figura 5.2. Segundo grupo de gráficos de contorno lineales para la variable de respuesta densidad.....	124
Figura 5.3 Tercer grupo de gráficos de contorno lineales para la variable de respuesta oleosidad.....	126

## 1. INTRODUCCIÓN

La estadística está definida como la ciencia de recolectar, analizar y extraer conclusiones de los datos para la toma de decisiones. En este sentido, los datos son usualmente recogidos a través de encuestas por muestreo, estudios observacionales o experimentos. En un estudio observacional, por ejemplo, las correlaciones pueden encontrarse y estimarse entre dos o más variables, pero no se puede probar que las relaciones detectadas sean de causa y efecto. Empero, en un experimento sí es posible, puesto que algunas variables son cambiadas intencionalmente mientras que otras se mantienen constantes, de tal forma que el efecto observado es causado por el cambio en la variable deliberadamente variada (Lawson, 2014).

Para Montgomery (2017) un experimento se define como una prueba o una serie de pruebas en las que se hacen cambios deliberados en las variables de entrada de un proceso o sistema para observar e identificar las razones de los cambios en la variable de salida, tal y como se detalla en la Figura 1.1.

Figura 1.1. Modelo general de un proceso o sistema



Fuente: Adaptado de Montgomery (2017)

En el ámbito de las ciencias y la ingeniería el diseño de experimentos (DdE) es una herramienta fundamental y se considera como parte del proceso científico y uno de los

medios para conocer el funcionamiento de sistemas y procesos. Esto naturalmente implica, la solución de problemas, mejoras u optimización de procesos o bien el diseño de nuevos productos y procesos, teniendo como fin generar conocimiento y aprendizaje de la manera más eficiente posible (Gutiérrez Pulido & de la Vara Salazar, 2012; Montgomery, 2017).

Técnicamente el diseño estadístico de experimentos se refiere al proceso para planear el experimento de tal forma que se recaben datos adecuados que puedan analizarse con métodos estadísticos que llevarán a conclusiones válidas y objetivas (Lawson, 2014; Montgomery, 2017).

### 1.1 Contexto

En los DdE clásicos existen tres principios fundamentales: aleatorización, replicación y bloqueo, lo que ayuda a evitar efectos de factores confusores<sup>1</sup>. Es el primero de ellos el que hace, en función del contexto, a los DdE costosos y difíciles de llevar a cabo. Por ejemplo, factores como el tiempo de exposición o temperatura de secado/cocción/entre otros suelen ser difíciles de cambiar y por motivos de costo, tiempo y utilización de recursos en general no es posible aleatorizar completamente el orden de las corridas experimentales (García-Goñi, 2024), incumpléndose el principio de aleatorización, pues prima la eficiencia de la ejecución del diseño. Para estas situaciones existen y se realizan DdE que permiten violar algunos de estos principios fundamentales, como los diseños anidados, la introducción del bloqueo, entre otros (Montgomery, 2017). En los DdE con mezclas los experimentos pueden volverse difíciles de ejecutar, debido a los principios previos o bien por las restricciones de frontera que se le impongan y a la cantidad y posición de sus puntos de diseño, donde acciones como las previas pueden ser insuficientes o no aplicables a todos los tipos de contextos.

---

<sup>1</sup> Un confusor o variable confusora es un factor distinto al predictor, que se asocia de forma independiente tanto con la variable predictora como con la variable respuesta (Quispe et al., 2020; Shapiro, 2008)

Como tal, hay una amplia variedad de tipos de diseños experimentales y la selección del apropiado depende de los objetivos de experimentación. Lawson (2014) propone que estos diseños se pueden agrupar de manera sencilla en dos categorías principales: i) Aquellos que se emplean para determinar las fuentes de variabilidad y ii) los que se utilizan para establecer las relaciones causa y efecto.

Dentro del marco del DdE clásico y los experimentos utilizados para la optimización de las relaciones causa – efecto se encuentran los modelos de superficie de respuesta (RSM, por sus siglas en inglés), éstos a su vez engloban a los experimentos con mezclas, los cuales pueden considerarse como un caso particular de los RSM, donde los niveles de cada “factor” (en el contexto de experimentos con mezclas se acostumbra llamar componentes a los factores) son dependientes y por tanto colineales (Myers et al., 2009). La multicolinealidad, que por lo general busca evitarse, se debe a que los componentes forman parte de una mezcla, lo que necesariamente implica que cada nivel de cada uno de ellos depende de los otros que componen el diseño. Por ejemplo, en un caso hipotético sobre un producto con peso final de 90 g compuesto de tres ingredientes (A, B y C) sucedería que, al aumentar la cantidad de A en 10 g, se debe disminuir a B, C o ambos en la misma cantidad.

Los experimentos con mezclas son importantes pues en la práctica muchos productos están conformados por la mezcla de dos o más ingredientes; y estos tienen una o más características que son de interés para una persona experimentadora. Algunos ejemplos de productos y sus propiedades son:

- i. La esponjosidad de una mezcla de queque conformada por polvo de hornear, harina, azúcar, agua y mantequilla.
- ii. La dureza o resistencia a la compresión (psi) de una mezcla de concreto formada por arena, agua y cemento.

- iii. La iluminación<sup>2</sup> (cd) y la duración de la iluminación (s) de bengalas formadas por la mezcla de magnesio, nitrato de sodio, nitrato de estroncio y aglutinante.

Los experimentos con mezclas pueden utilizarse para determinar si existen combinaciones de dos o más ingredientes que produzcan propiedades de un producto que sean más deseables que las que se pueden obtener con los compuestos individuales. En los casos enlistados anteriormente se asume que las propiedades de interés están funcionalmente relacionadas con la composición del producto y que, por tanto, al cambiar la composición de éste mediante la variación de la proporción de los ingredientes, cambiarían dichas propiedades (Cornell, 2002).

En general, el objetivo del DdE con mezclas es cuantificar la influencia que tienen los diferentes componentes sobre la respuesta, tanto en forma individual como en su acción conjunta con otros componentes (Cornell, 2002; Gutiérrez Pulido & de la Vara Salazar, 2012). Estos diseños trabajan bajo el supuesto de que las características de calidad de la mezcla dependen de las proporciones con las que participan los ingredientes y no de la cantidad absoluta de ellos (Cornell, 2002) y por lo general para cumplir este supuesto las personas experimentadoras fijan la cantidad total de los ingredientes en un solo valor para todas las mezclas de cada corrida experimental. Cuando este supuesto no es plausible o deseable de cumplir, pues se sabe que la cantidad total de la mezcla puede afectar la variable de respuesta<sup>3</sup>, se emplea el concepto de experimentos mezcla – cantidad (Cornell, 2002).

Los componentes, que son las proporciones en las que participa cada ingrediente, son representados por  $x_i$ , con  $i = 1, 2, \dots, p$ ; siendo  $p$  la cantidad de componentes. Estas proporciones deben ser no negativas ( $x_i \geq 0$ ) (Cornell, 2002; Jones

---

<sup>2</sup> Símbolo de candela, unidad básica del Sistema Internacional (SI) que mide la intensidad luminosa.

<sup>3</sup> Por ejemplo, al producir un helado, la cantidad total de la mezcla es una variable que afecta el tiempo de congelación y por lo tanto también las características de calidad.

& Aitchison, 1987) y no pueden ser más de 100 % ( $x_i \leq 1$ ). Además, con estos experimentos también se trata de modelar la respuesta para predecirla en cualquier formulación posible, y utilizar estos modelos con el propósito de encontrar la composición de la mezcla que proporcione mejores resultados.

Para estimar estos modelos es preciso tomar en cuenta las particularidades de los experimentos con mezclas, ya que los diseños factoriales tradicionales no son aplicables, y los polinomios estándar no son adecuados, pues algunos de sus parámetros carecen de sentido (Cornell, 2002; Gutiérrez Pulido & de la Vara Salazar, 2012).

## 1.2 Justificación

En algunos problemas de mezclas surgen restricciones sobre los componentes individuales, dentro de las cuales son comunes las restricciones sobre la frontera inferior, superior o ambas. En los experimentos con mezclas es común que estas restricciones existan, ya sea por motivos de seguridad, porque no se puede exceder la cantidad de un componente químico, o por utilidad práctica cuando hay tratamientos que no tienen sentido lógico<sup>4</sup>.

A los experimentos con mezclas también se le conocen como *símplex* cuando estos tienen una forma regular de  $p - 1$  dimensiones, por ejemplo, para tres componentes, de triángulo equilátero (Cornell, 2002). Cuando están presentes las restricciones sobre la frontera inferior ( $a_i$ ), la región factible del diseño sigue siendo un *símplex* ( $L - \text{símplex}$ ), pero se inscribe dentro de la región del *símplex* original conservando una forma regular de  $p - 1$  dimensiones. Esta situación puede simplificarse

---

<sup>4</sup> Por ejemplo, en un experimento para el diseño de un refresco natural e innovador, no sería práctico estudiar el tratamiento de los vértices de 100 % agua o 100 % fruta.

y analizarse mediante la introducción de pseudocomponentes<sup>5</sup>  $x'_i$  (Cornell, 2002; Crosier, 1984; McLean & Anderson, 1966).

Si las restricciones están presentes sobre el límite superior ( $b_i$ ), el diseño es un símplex (U – Símplex) si y solo si la suma de las  $p - 1$  fronteras superiores son menores o iguales a 1, situación que no suele ser común. El uso de pseudocomponentes es adecuado y la orientación de estos con forma símplex es opuesta a la dirección del símplex original; por lo tanto, los pseudocomponentes tienen efectos opuestos a los de los componentes originales. Es decir, si la respuesta aumenta con  $x_i$ , disminuye con  $x'_i$  (Crosier, 1984).

Cuando las restricciones de la frontera superior no cumplen con lo anterior o los componentes tienen restricciones tanto sobre la frontera superior como inferior, la región factible deja de ser un diseño símplex; y se convierte en un politopo<sup>6</sup> por lo general irregular. Los diseños que no son símplex se conocen como de vértices extremos e indefectiblemente, debido a la geometría irregular, implican un aumento en la cantidad de puntos de diseño, lo que aumenta las corridas experimentales y, por ende, un aumento de los recursos necesarios para su correcta consecución. Por otro lado, la cantidad de tratamientos de un experimento con mezclas también depende del grado del polinomio que se desea estimar y de la cantidad de componentes ( $p$ ).

Es por esto por lo que, aun con la eficiencia de recursos a la que conduce el uso del DdE (Montgomery, 2017); que los experimentos con mezclas de vértices extremos suelen ser ineficientes, pues la cantidad de tratamientos experimentales aumenta vertiginosamente, sobre todo cuando  $p$  es alto. Desde luego, a esto también se le agrega lo ya mencionado, que la experimentación clásica suele ser una tarea onerosa e intensiva

---

<sup>5</sup> Los pseudocomponentes son variables codificadas que se utilizan para simplificar la construcción de diseños y el ajuste de modelos. La transformación en pseudocomponentes depende de las restricciones impuestas sobre la frontera.

<sup>6</sup> Politopo es una generalización del concepto de polígono (2D) o poliedro (3D) a cualquier otra dimensión.

en el uso de recursos, que puede crecer aún más en función de la complejidad de la medición de la(s) variable(s) de respuesta y el contexto de la ejecución del experimento.

Ésta es la motivación por la que con esta investigación se busca diseñar y validar mediante criterios de optimalidad y eficiencia, un algoritmo que reduzca la cantidad de tratamientos experimentales de los diseños de vértices extremos, generados únicamente por restricciones en la frontera superior, al mínimo plausible. Ésta se basa en la idea de transformar las restricciones del límite superior de uno o varios componentes en restricciones en la frontera inferior; de tal forma que la región factible resultante sea una parte de la región de experimentación original con forma de símpex o L – Símpex, con el uso de pseudocomponentes, eliminando así la forma de politopo irregular.

Desde un punto de vista práctico, la utilidad de esta investigación radica en que la implementación de este algoritmo permitiría a las personas experimentadoras y a las organizaciones que utilizan este tipo de DdE reducir el uso de los recursos y posibilitar la ejecución y análisis de estos para los fines que lo conciban, en una o varias etapas de un programa de experimentación.

Sobre todo, en condiciones en las que la cantidad de corridas aumenten rápidamente, como cuando la cantidad de componentes es alta, se requiere estimar modelos complejos (de segundo orden y superiores) o se incluyen variables de operación (diseño multinivel); también cuando el costo de las corridas experimentales sea elevado o intensivo en recursos. Desde luego, se sopesan las posibles desventajas que le acaecen a la reducción de la región de experimentación.

### **1.3 Problema de investigación**

La reducción en la cantidad de tratamientos experimentales, devenida de la recuperación de politopo regular, implica una reducción de la región de experimentación a un espacio menor y por lo general inscrito dentro de la región de operabilidad.

Esta reducción puede resultar beneficiosa, ya que autores como Cornell (2002) indican que la estrategia de confinar una región dentro de la región símplex original permite aislar los esfuerzos de diseño y modelado a la subregión, aumentando también la precisión de las estimaciones del modelo, lo cual es un paso natural en el proceso de experimentación secuencial. A su vez puede tener desventajas, que incluyen la posibilidad de que el óptimo global se encuentre fuera de la región de experimentación confinada.

Con base en esto, se desea conocer si los diseños generados con este algoritmo son eficientes y eficaces para estimar modelos lineales, cuadráticos y cúbicos especiales, bajo criterios de optimalidad y eficiencia. Para ello se pretende variar las cantidades de componentes ( $p$ ), la cantidad y el tamaño de las restricciones en la frontera superior y la cantidad de réplicas experimentales, para estudiar el comportamiento de dichos criterios.

#### **1.4 Objetivos**

A continuación, se plantea el objetivo general y los objetivos específicos que se persiguen con esta investigación.

##### **1.4.1 Objetivo general**

Diseñar un algoritmo para la reducción de tratamientos en diseños experimentales con mezclas de vértices extremos generados por restricciones en la frontera superior, basado en la recuperación de la forma símplex, reduciendo a su vez el tamaño de la región de experimentación.

##### **1.4.2 Objetivos específicos**

1. Proponer un algoritmo que recupere la forma de politopo regular en los diseños experimentales con mezclas de vértices extremos generados por

restricciones de frontera superiores, haciendo posible aplicar un diseño símplex o L – Símplex.

2. Comparar la eficiencia y eficacia de los diseños, tanto en proporciones como en pseudocomponentes, en la región símplex original, la región generada por vértices extremos y la región generada con el algoritmo propuesto, mediante simulaciones.
3. Aplicar, a modo de ejemplo, el algoritmo propuesto en un DdE con mezclas, relacionado con una situación común al ejercicio de la ciencia e ingeniería, en el que la reducción de la cantidad de tratamientos experimentales sea deseable para reducir la utilización de recursos.

## 2. MARCO TEÓRICO

En el presente capítulo se presenta el desarrollo teórico de los DdE partiendo de lo general hasta lo específico con el DdE con mezclas. Se incluyen algoritmos de generación de coordenadas, criterios de selección entre experimentos competentes y funciones de optimización simultánea.

### 2.1 Diseño de experimentos

Observar un sistema o un proceso mientras este se encuentra en operación es una parte importante del proceso de aprendizaje, y es una parte integral del entendimiento de como los mismos funcionan; de esta forma se pueden plantear hipótesis o teorías sobre este funcionamiento. Con el uso del DdE se puede demostrar si esas teorías son correctas o no (Montgomery, 2017).

El DdE es una herramienta que se utiliza para entender las relaciones de causa – efecto en un sistema en el que deliberadamente se cambian las variables de entrada y se observan las variaciones en la salida que son provocadas por dichos cambios (Montgomery, 2017). Los experimentos usualmente involucran muchos factores, y por lo general el objetivo de una persona experimentadora es determinar la influencia que esos factores tienen sobre la variable de respuesta de un sistema.

#### 2.1.1 Conceptos básicos

Algunos conceptos básicos utilizados en el DdE, según Gutiérrez Pulido & de la Vara Salazar (2012) son:

i. Experimento:

Cambio en las condiciones de operación de un sistema o proceso que se hace con el objetivo de medir el efecto del cambio sobre una o varias propiedades del producto o resultado, permitiendo aumentar el conocimiento acerca del sistema.

- ii. **Unidad experimental:**

Es la entidad que recibe el tratamiento experimental y sobre la que se desean hacer inferencias a partir del experimento. Para cada estudio experimental esta debe definirse de manera cuidadosa.
- iii. **Variables de respuesta:**

Son las variables a través de las cuales se conoce el efecto de variar cada factor, en sus respectivos niveles, en el experimento. Éstas por lo general son características de calidad de un producto, o variables que miden el desempeño de un proceso.
- iv. **Variables controlables:**

Son variables que se pueden fijar en un nivel dado. Para cada una de estas existe la manera o el mecanismo para cambiar o manipular el nivel de operación; esta característica es la que hace posible que se pueda experimentar con ellos.
- v. **Variables no controlables o de ruido:**

Contrario al anterior, son variables que no se pueden controlar durante el experimento. Un ejemplo clásico son las variables ambientales.
- vi. **Factores estudiados:**

Son las variables de interés o que se investigan durante el experimento sobre cómo afectan a la variable de respuesta. Forman parte del conjunto de las variables controlables.
- vii. **Niveles:**

Las diversas condiciones que se asignan a cada factor estudiado en un diseño experimental.
- viii. **Tratamiento:**

Es cada combinación de niveles de todos los factores estudiados. También se conoce como punto de diseño.

ix. Error aleatorio:

Variabilidad que no se puede explicar por los factores estudiados y que es el resultado del pequeño efecto de los factores no estudiados y del error experimental.

x. Error experimental:

Es una parte del error aleatorio y refleja los errores de la persona experimentadora en la planeación y ejecución del experimento.

### 2.1.2 Principios básicos de experimentación

Los tres principios básicos de experimentación son: aleatorización, replicación y bloqueo (Fisher, 1971; Montgomery, 2017); y para ambos autores el principio de aleatorización es la piedra angular que sostiene el uso de métodos estadísticos en el diseño experimental. Por aleatorización se entiende que tanto el material experimental como el orden en que se realizan las corridas individuales del experimento se determinan al azar. Uno de los requisitos de los métodos estadísticos es que las observaciones (o los errores) sean variables aleatorias con distribuciones independientes. La aplicación del principio de aleatorización generalmente hace válida esta suposición (Montgomery, 2017).

El segundo principio corresponde a la replicación, que se define como una repetición independiente de los tratamientos. La aplicación de este principio tiene dos propiedades importantes:

- i. Permite al experimentador obtener una estimación del error experimental; esta se convierte en una unidad de medida básica para determinar si las diferencias entre niveles, observadas en los datos son realmente estadísticamente diferentes.
- ii. Si se usa la media muestral para estimar el efecto de un factor en el experimento, la realización de réplicas permite a la persona experimentadora obtener una mejor estimación de este efecto.

Es necesario hacer la distinción entre replicación y medidas repetidas, pues son conceptos distintos, las medidas repetidas en este contexto corresponderían a la toma de varias medidas sobre el mismo tratamiento.

El bloqueo es el tercer principio, y es una técnica de diseño utilizada para mejorar la precisión con la que se hacen las comparaciones entre los factores de interés. A menudo, el bloqueo se utiliza para reducir o eliminar la variabilidad transmitida por factores de ruido, es decir, factores que pueden influir en la respuesta experimental, pero en los que no se está directamente interesado. Al bloquear, se supone que el subconjunto de datos que se obtengan dentro de cada bloque (nivel particular del factor bloqueado), debe resultar más homogéneo que el conjunto total de datos (Gutiérrez Pulido & de la Vara Salazar, 2012; Montgomery, 2017).

### 2.1.3 Historia breve del diseño de experimentos

Hay cuatro eras en el desarrollo moderno del DdE:

- i. La era de la agricultura liderada por Sir Ronald Fisher, quien fue responsable por la estadística y el análisis de datos en la *Rothamsted Agricultural Experimental Station*, en Londres, Inglaterra. A través de su libro *The Design of Experiments*, publicado por primera vez en 1935, ha influido en la investigación agrícola, aportando métodos ahora usados en todo el mundo para evaluar los resultados de experimentos con muestras pequeñas. Fisher notó que las fallas en la forma de analizar experimentos obstaculizaban el análisis de los resultados experimentales y llevó a cabo el desarrollo de los conocimientos que decantaron en los principios básicos de experimentación ya mencionados. Además, introdujo el concepto de diseño factorial y análisis de varianza (Fisher, 1971; Gutiérrez Pulido & de la Vara Salazar, 2012; Montgomery, 2017).
- ii. Segunda era o era industrial, la cual se caracterizó por el desarrollo de los RSM por Box & Wilson (1951) quienes reconocieron y aprovecharon el hecho de que muchos experimentos en la industria son diferentes de sus homólogos agrícolas

en al menos dos maneras: 1) la variable de respuesta generalmente se puede observar casi inmediatamente y 2) el experimentador puede aprender rápidamente información crucial de un grupo pequeño de corridas experimentales que se pueden utilizar para planificar el próximo experimento. Éstas dos características reciben el nombre de inmediatez y secuencialidad (Box, 1999). Se puede afirmar que entre 1950 y 1980 el DdE se convirtió en una herramienta de aplicación frecuente en las áreas de investigación y desarrollo. Hasta la década de 1970, la aplicación a nivel de planta o manufactura no estaba generalizada, debido a la falta de recursos computacionales y la carencia de formación estadística por parte de ingenieros y especialistas de manufactura (Gutiérrez Pulido & de la Vara Salazar, 2012). Por su parte la investigación estadística sobre los experimentos con mezclas tuvo sus inicios en la década de 1950 con Quenouille (1953, 1959). Claringbold (1955) introdujo el diseño simplex con tres componentes y Scheffé (1958) introdujo el diseño simplex lattice (también llamado reticular). No obstante, casi toda la teoría y metodología que ha emanado de la comunidad estadística surgió entre el año 1960 y el año 2000 (Cornell, 2002).

- iii. La tercera era del DdE llega en 1970-1980 con el ingeniero japonés Genichi Taguchi (1987), principalmente, cuyos conceptos sobre diseño robusto de parámetros (RPD, por sus siglas en inglés) tuvieron un impacto significativo en la academia en el mundo occidental. Taguchi sugirió diseños factoriales altamente fraccionados y otros arreglos ortogonales junto con algunos métodos estadísticos novedosos. La metodología resultante generó mucha discusión y controversia. Parte de la controversia surgió porque la metodología de Taguchi fue defendida en occidente inicialmente (y principalmente) por los empresarios, mientras que la ciencia estadística subyacente no había sido adecuadamente revisada por pares.
- iv. La cuarta era ha incluido un renovado interés general en el diseño estadístico por parte de investigadores y profesionales y el desarrollo de muchos enfoques nuevos y útiles para los problemas experimentales en el mundo industrial,

incluyendo alternativas a los métodos técnicos de Taguchi que permiten que sus conceptos de ingeniería se lleven a la práctica de manera eficiente y efectiva (Montgomery, 2017). Esto continúa en la actualidad; incluso, en los últimos años, las universidades han incorporado el diseño de experimentos como materia obligatoria u operativa en la mayoría de las ingenierías (Gutiérrez Pulido & de la Vara Salazar, 2012).

## 2.2 Modelos de superficies de respuesta

Son una colección de técnicas matemáticas y estadísticas utilizadas para el modelado y análisis de problemas en los que la variable de respuesta de interés es influenciada por varias variables independientes y el objetivo es optimizar dicha respuesta. Aunque esta técnica tiene orígenes en el trabajo de Box & Wilson (1951), es con el auge de las computadoras que ha tenido un mayor desarrollo tanto en aspectos teóricos como prácticos.

Surgen o se utilizan porque en algunas ocasiones hay experimentos con los que no se obtienen los resultados buscados o las mejoras introducidas son insuficientes, por lo que es necesario experimentar de manera secuencial hasta alcanzar los resultados esperados, por ejemplo, después de una primera etapa experimental puede ser imperativo desplazar la región experimental en una dirección adecuada al objetivo del DdE o bien, explorar con detalle la región experimental inicial (Montgomery, 2017; Myers et al., 2009).

Por lo general en los problemas de RSM no se conoce la forma de la relación entre la variable de respuesta y las variables independientes; por tanto, el primer paso corresponde a encontrar una aproximación adecuada de la verdadera relación funcional entre variables dependientes e independientes. Para ello, frecuentemente, se utiliza un polinomio de orden inferior, por ejemplo, el de primer orden de la Ecuación (1):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon \quad (1)$$

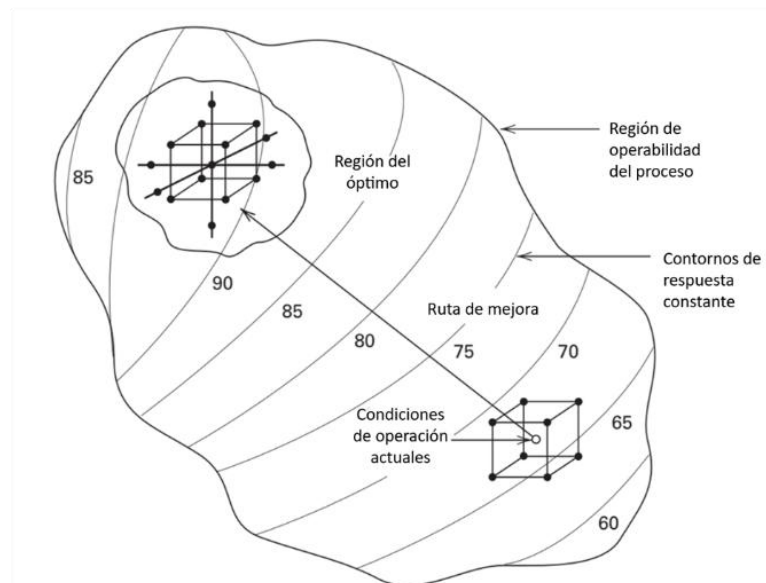
Si se encuentra una curvatura en el sistema, entonces se utiliza un polinomio de orden superior, como el modelo de segundo orden de la Ecuación (2):

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \beta_i x_i^2 + \sum_{i < j} \beta_{ij} x_i x_j + \epsilon \quad (2)$$

Para la mayoría de RSM se pueden emplear uno o ambos modelos. Se utiliza el método de mínimos cuadrados para estimar los parámetros de los modelos (Box & Draper, 2007; Montgomery, 2017).

Los RSM conforman un procedimiento secuencial en el que a menudo se puede iniciar en un punto en la superficie de respuesta que se encuentra lejos del valor óptimo, en estos casos los modelos de primer orden son adecuados, ya que el interés es llevar a la persona experimentadora de forma eficiente y eficaz por una ruta de mejora hacia la vecindad general del óptimo. Una vez que se ha encontrado esta región se suele emplear un modelo de mayor complejidad como el de segundo orden y se puede realizar un análisis para localizar el óptimo. Este procedimiento se resume en la Figura 2.1.

Figura 2.1. La naturaleza secuencial de los RSM



Fuente: Adaptado de Montgomery (2017)

En este sentido, se debe tener en cuenta dos conceptos importantes:

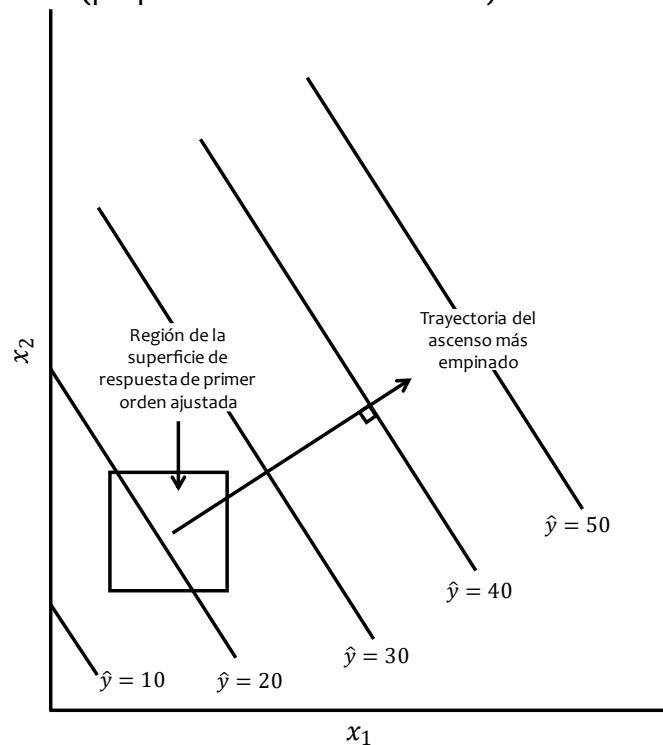
- i. Región experimental o de experimentación: es el espacio delimitado por los rangos de experimentación utilizados con cada factor.
- ii. Región de operación o de operabilidad: es la región delimitada por el conjunto de puntos o condiciones donde el proceso puede ser operado, este espacio es difícil de delimitar con certeza, ya que el rango de operabilidad de un factor individual puede estar supeditado a los rangos de operabilidad de los factores restantes (Gutiérrez Pulido & de la Vara Salazar, 2012).

Para los RSM, en todos los casos, la región experimental es un subconjunto de la región de operación.

Los diseños empleados para ajustar RSM principalmente, son el método de ascenso/descenso empinado, el diseño central compuesto (DCC) y el diseño de Box-Behnken (Box & Draper, 2007; Montgomery, 2017; Myers et al., 2009).

- i. Método de ascenso/descenso empinado: es un procedimiento para moverse secuencialmente en la dirección del máximo incremento/decremento en la respuesta (Box & Draper, 2007; Montgomery, 2017). Un ejemplo se muestra en la Figura 2.2. Generalmente, pero no exclusivamente se emplea con modelos de primer orden, por su naturaleza exploratoria de la región experimental en busca de la región del óptimo.

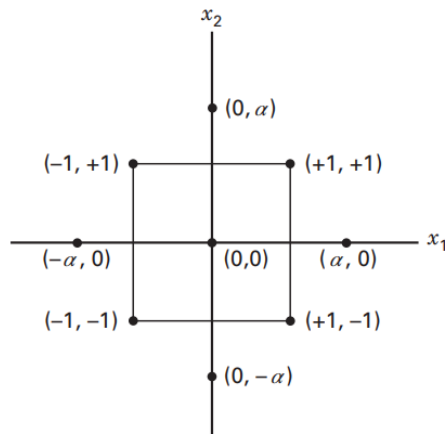
Figura 2.2. Superficie de respuesta de primera orden y ruta del ascenso más empinado (perpendicular a la estimación)



Fuente: Adaptado de Montgomery (2017)

- ii. Diseño central compuesto (DCC): es la clase de diseño más popular para ajustar modelos de segundo orden en RSM. Fue introducido por Box & Wilson (1951). Generalmente un DCC consiste en un experimento factorial completo  $2^k$  (o un fraccionado de resolución V, cuando hay cinco o más factores<sup>7</sup>) con puntos centrales y axiales incrustados, como los de la Figura 2.3. Este experimento nace con la motivación de la experimentación secuencial, ya que usualmente se lleva a cabo cuando después de que un experimento factorial o factorial fraccionado se haya ejecutado y se hayan identificado los factores más importantes (Myers et al., 2009).

<sup>7</sup> Son experimentos en los cuales los efectos principales y las interacciones de dos factores no tienen otros efectos principales o interacciones de dos factores aliadas. En consecuencia, son diseños poderosos que permiten la estimación de efectos principales y de interacción de dos factores, por ello pueden ser empleados para realizar DCC, contribuyendo a la estimación de los efectos lineales y de interacción (Montgomery, 2017; Myers et al., 2009).

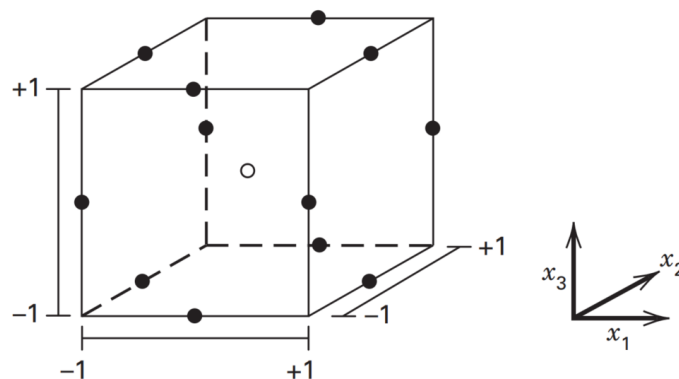
Figura 2.3. Diseño central compuesto para  $k = 2$ 

Fuente: Montgomery (2017)

Nota:  $\alpha$  generalmente varía de 1 a  $\sqrt{k}$  (Myers et al., 2009)

- iii. Diseño Box – Behnken: estos son formados de la combinación de factoriales  $2^k$  con diseños de bloques incompletos, resultando en diseños que por lo general son muy eficientes en términos del número de corridas necesarias (Box & Behnken, 1960). Este es un diseño esférico, que en el caso de  $k = 3$ , todos los puntos están en una esfera de radio  $\sqrt{2}$ , y además no contiene tratamientos en los vértices de la región creada por los límites inferiores y superiores (ver Figura 2.4). Esta característica los hace diseños ventajosos cuando los puntos de esquina representan combinaciones que son onerosamente prohibitivas o imposibles de alcanzar por limitaciones físicas (Montgomery, 2017).

Figura 2.4. Diseño de Box – Behnken para tres factores



Fuente: Montgomery (2017)

### 2.3 Experimentos con mezclas

Son experimentos en los que se asume que la respuesta solo depende de las proporciones relativas de los ingredientes presentes en la cantidad de la mezcla<sup>8</sup>. En un experimento con mezclas si el total permanece constante y el valor de la respuesta cambia cuando se introducen variaciones en las proporciones de los ingredientes, entonces el comportamiento de la respuesta se dice que es una función de la propiedad de la mezcla conjunta de los ingredientes (Cornell, 2002).

En un experimento factorial o en un diseño de superficies de respuesta convencional, se parte de que los niveles de cada factor son independientes de los niveles de los otros factores. En los experimentos con mezclas, que son casos especiales de los RSM (Montgomery, 2017; Montgomery & Voth, 1994; Myers et al., 2009), los factores, ahora llamados componentes, no son independientes. Por ejemplo, si  $x_1, x_2, \dots, x_p$  denotan las proporciones de  $p$  componentes de una mezcla, entonces:

$$0 \leq x_i \leq 1, \quad i = 1, 2, \dots, p \quad (3)$$

$$\sum_{i=1}^p x_i = 1 \quad (4)$$

Las restricciones de la Ecuación (3) y la Ecuación (4) se entienden como que las proporciones tienen que encontrarse entre cero y uno, y que la suma de ellas debe ser uno, por tanto, al aumentar la cantidad de un ingrediente necesariamente se reduce la participación de los componentes restantes en esa misma proporción. En la Ecuación (4) cuando un solo componente cumple con la restricción de ser igual a uno, se conoce como “mezcla pura”.

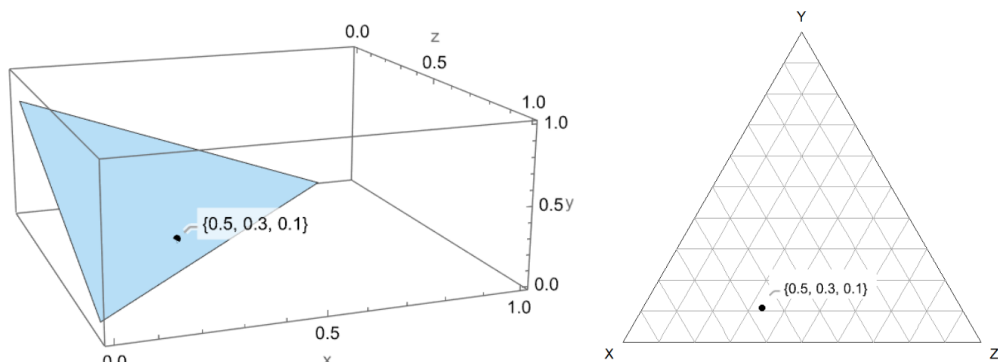
---

<sup>8</sup> Existen experimentos en los que esta suposición no es válida conocidos como experimentos mezcla – cantidad que se definen como aquellos en los que la cantidad de la mezcla varía y la variable de respuesta no solo depende de la cantidad relativa de las proporciones de los ingredientes sino también de la cantidad total de los ingredientes (Cornell, 2002). Estos son abordados con más detalle en secciones posteriores de este documento.

En este sentido, los componentes de una mezcla son o están compuestos por datos composicionales, los cuales son un caso especial en el campo de la estadística y consisten en vectores en los que cada componente es no negativo con la propiedad de que sus valores suman una constante, regularmente estandarizada a uno o 100, y en el que cada componente muestra la importancia relativa de una parte en un total (Jones & Aitchison, 1987; Martínez García, 2022).

Martínez García (2022) menciona que la restricción de suma fija en los datos composicionales [misma que la mostrada en la Ecuación (4)] conduce a una representación geométrica espacial en el espacio del símplex, cuya estructura más simple es un triángulo, conteniendo un dato composicional con tres partes  $x = (x_1, x_2, x_3)$ , donde cada una corresponde con el punto que dista  $x_1, x_2$  y  $x_3$  respectivamente de los lados opuestos a los vértices 1, 2 y 3. En la Figura 2.5 se encuentra un gráfico representando esta situación y la ubicación del punto no estandarizado  $x = (0.5, 0.3, 0.1)$ .

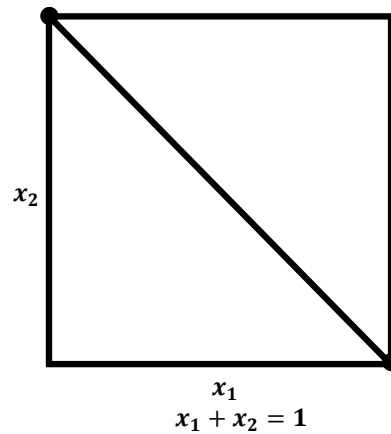
Figura 2.5. Geometría del símplex



Fuente: Martínez García (2022)

Continuando con la geometría símplex, y contrario a lo mencionado por Martínez García (2022), la forma más simple sucede cuando  $p = 2$  (Figura 2.6), ya que el espacio factorial del diseño incluye todos los valores de los dos componentes que se encuentran en el segmento de línea descrita por  $x_1 + x_2 = 1$ , con cada componente delimitado por 0 y 1 (Cornell, 2002; Myers et al., 2009).

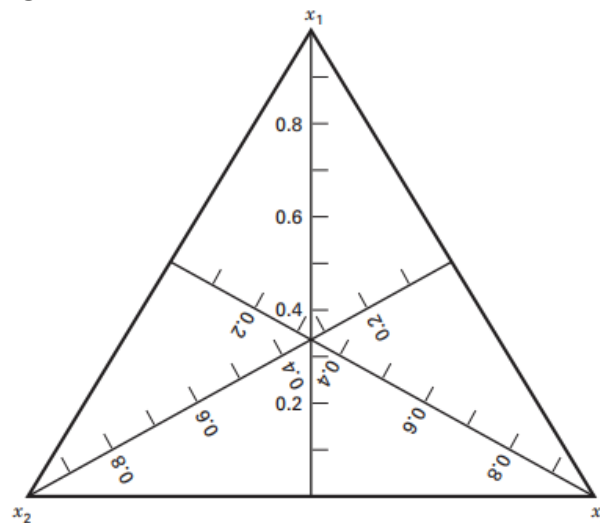
Figura 2.6. Espacio de mezclas para dos componentes



Fuente: Adaptado de Myers et al. (2009)

Para  $p = 3$  el espacio de mezcla es un triángulo equilátero con los vértices correspondientes a formulaciones conocidas como mezclas puras, es decir, mezclas que son el 100 % de un solo componente. Además, con tres componentes de mezcla, la región restringida puede ser convenientemente representada en un espacio de coordenadas trilineales o también un sistema coordenado símplex (Figura 2.7). Para  $p = 4$  el espacio generado es un tetraedro y para más de cuatro componentes la región sería un hipertetraedro (Cornell, 2002; Gutiérrez Pulido & de la Vara Salazar, 2012; Montgomery, 2017; Myers et al., 2009).

Figura 2.7. Sistema coordenado trilineal



Fuente: Montgomery (2017)

Para un experimento con tres componentes, como el de la Figura 2.7, los vértices de un símplex que representan las mezclas puras se denotan por  $x_i = 1, x_j = 0$  para cada  $i, j = 1, 2$  y  $3$  cuando  $i \neq j$ . Los puntos interiores del triángulo corresponden a las mezclas en las que ninguno de los tres componentes está ausente  $x_1 > 0, x_2 > 0$  y  $x_3 > 0$ , y el centroide o centro geométrico corresponde a la mezcla con igual proporción  $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$  para cada ingrediente.

Los experimentos con mezclas trabajan bajo el supuesto de que las características de calidad de la mezcla dependen de las proporciones con las que participan los ingredientes y no de la cantidad absoluta de ellos (Cornell, 2002). Gutiérrez y de la Vara (2012) indican que entre los objetivos de un experimento con mezclas se encuentran:

- i. Determinar cuáles de los ingredientes de la mezcla o interacciones entre ellos tienen mayor influencia sobre una o varias respuestas de interés.
- ii. Modelar las respuestas de interés en función de las proporciones de los componentes de la mezcla.
- iii. Usar dichos modelos para determinar en qué proporción debe participar cada uno de los ingredientes para lograr que la fórmula tenga las propiedades deseadas.

Además, interesa cuantificar la influencia que tienen los diferentes componentes sobre la respuesta, tanto en forma individual como en su acción conjunta con otros componentes. Con este tipo de diseño se trata de modelar la respuesta para predecirla en cualquier formulación posible, y utilizar los modelos con el propósito de encontrar la composición de la mezcla que proporcione mejores resultados (optimizar). Para cumplir con estas tareas es preciso tomar en cuenta las particularidades de los experimentos con mezclas.

En primer lugar, los diseños factoriales tradicionales no son aplicables, ya que su objetivo es medir el cambio en la respuesta cuando se varía el nivel de cada factor

mientras que los otros niveles permanecen constantes (lo que no es posible en un experimento con mezclas). Además, porque los polinomios estándar no son adecuados pues algunos de sus parámetros carecen de sentido, así como su incapacidad para mantener la cantidad total fija (Cornell, 2002; Gutiérrez Pulido & de la Vara Salazar, 2012). Por otra parte, los diseños con mezclas presentan al menos dos diferencias con respecto a los RSM: se requieren tipos especiales de diseños y las formas de los polinomios de mezclas son ligeramente diferentes de los polinomios estándar (Myers et al., 2009).

### 2.3.1 Diseños símplex

Una forma de entender los experimentos con mezclas es mediante la representación de sus regiones experimentales, mediante símplex, es decir, figuras geométricas que permiten dibujar  $p$  componentes utilizando  $p - 1$  dimensiones (Cornell, 2002; Crosier, 1984; McLean & Anderson, 1966; Scheffé, 1958).

Los diseños símplex son utilizados para estudiar los efectos de los componentes de una mezcla sobre la variable de respuesta. Incluido dentro de este marco se encuentran los diseños símplex reticular y símplex centroide. Estos son utilizados para explorar la región de operabilidad completa.

#### 2.3.1.1 Diseños símplex reticular

El primero de los mencionados es también conocido como diseño símplex lattice<sup>9</sup> y se define mediante  $p$  y  $m$ , que son la cantidad de componentes y el grado del retículo, respectivamente. Este diseño puede tener una correspondencia especial con una ecuación polinómica. Es decir, un diseño símplex reticular se puede expresar de esta manera:  $\{p, m\}$ . Si el grado del retículo es uno, entonces solo existirían mezclas puras, si es dos habría mezclas puras y mezclas binarias (interacciones, que en este caso se conocen como términos cuadráticos), y así sucesivamente. A mayor grado de retículo, mayor es la capacidad del diseño para ajustar modelos de orden superior. Por otra parte,

---

<sup>9</sup> Este nombre se utiliza para hacer referencia a una matriz de puntos.

el primer grado del retículo permite ajustar modelos lineales, el segundo grado permite ajustar modelos cuadráticos y así subsecuentemente. Un diseño símplex reticular consiste en los puntos cuyas coordenadas son definidas por las combinaciones de las proporciones de componentes que se muestran en la Ecuación (5):

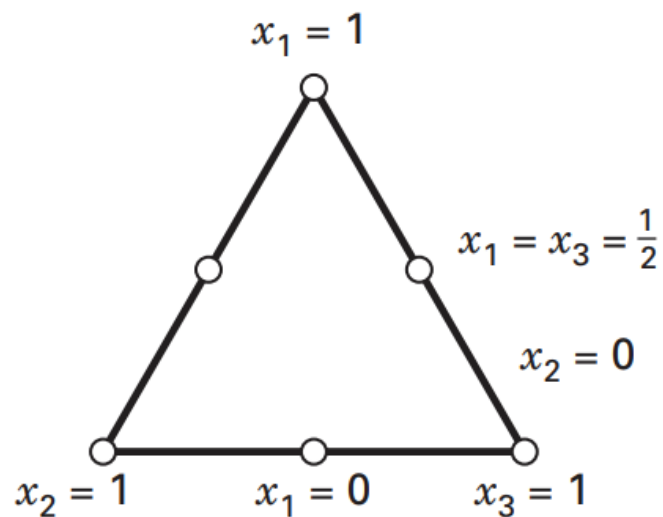
$$x_i = 0, \frac{1}{m}, \frac{2}{m}, \dots, 1, \quad i = 1, 2, \dots, p \quad (5)$$

Por ejemplo, un diseño  $\{3, 2\}$  estaría compuesto por los siguientes puntos de diseño para cada componente ( $x_i$ ) en la mezcla  $x_i = 0, \frac{1}{2}, 1$ . Es decir, por las siguientes coordenadas  $(x_1, x_2, x_3)$ :

$$(x_1, x_2, x_3) \in \left\{ (1, 0, 0), (0, 1, 0), (0, 0, 1), \left(\frac{1}{2}, \frac{1}{2}, 0\right), \left(\frac{1}{2}, 0, \frac{1}{2}\right), \left(0, \frac{1}{2}, \frac{1}{2}\right) \right\}$$

La situación descrita anteriormente se muestra en la Figura 2.8, donde los vértices representan las mezclas puras y los puntos  $x_i = \frac{1}{2}$  son las mezclas binarias de tal forma que  $x_i = x_j = \frac{1}{2}; x_k = 0, k \neq i, j$  (Cornell, 2002). Nótese, además, que cada uno de los puntos del diseño ocurren sobre la frontera (también llamado retículo) de la región.

Figura 2.8. Diseño símplex reticular  $\{3, 2\}$



Fuente: Montgomery (2017).

Cornell (2002) y Montgomery (2017) definen el número de puntos de un diseño símplex reticular según la Ecuación (6).

$$N = \binom{p + m - 1}{m} = \frac{(p + m - 1)!}{m!(p - 1)!} \quad (6)$$

Aplicando dicha ecuación en el diseño de la Figura 2.8 se determina, como se puede observar al contar los puntos, que la cantidad de tratamientos es 6.

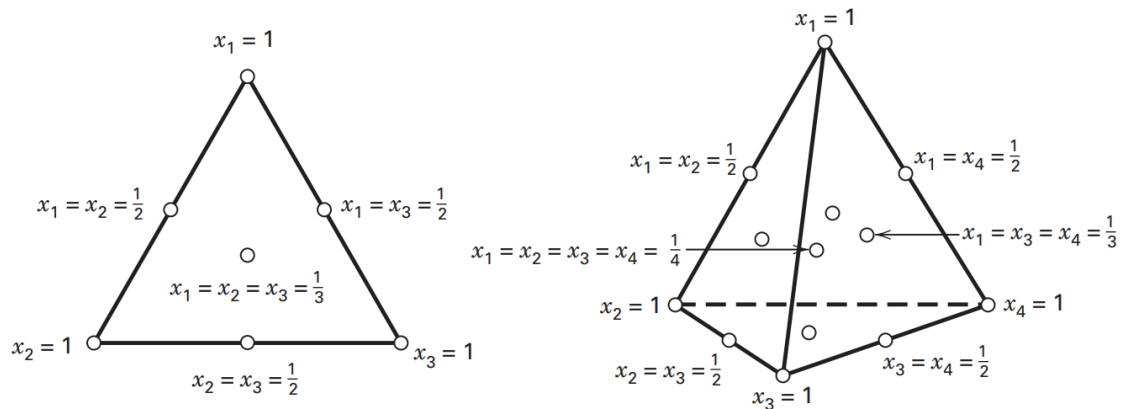
$$N = \frac{(3 + 2 - 1)!}{2!(3 - 1)!} = 6$$

### 2.3.1.2 Diseños símplex centroide

Como alternativa al diseño símplex reticular está el diseño símplex de centroide (Scheffé, 1963), en el cual con  $p$  componentes hay  $2^p - 1$  puntos distintos que corresponden a las  $p$  permutaciones de  $(1, 0, 0, \dots, 0)$  o las  $p$  mezclas puras, las  $\binom{p}{2}$  permutaciones de  $(\frac{1}{2}, \frac{1}{2}, 0, \dots, 0)$  o todas las mezclas binarias, las  $\binom{p}{3}$  permutaciones de  $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0, \dots, 0)$ , ..., y así sucesivamente, y finalmente el centroide global  $(\frac{1}{p}, \frac{1}{p}, \dots, \frac{1}{p})$ . En otras palabras, este diseño consiste en cada subconjunto no vacío de  $p$  componentes, pero solo con las combinaciones en las que los componentes utilizados se encuentran en las mismas proporciones (Cornell, 2002; Montgomery, 2017; Myers et al., 2009; Scheffé, 1963). Por lo que finalmente la cantidad de tratamientos se define como:

$$N = 2^p - 1 \quad (7)$$

En la Figura 2.9 se muestra un diseño símplex centroide para  $p = 3$ , nótese que para este caso particular el diseño es también un símplex reticular  $\{3, 2\}$  ampliado con un punto central. En la misma Figura se encuentra un diseño símplex centroide para  $p = 4$ , pero en este no hay relación con el símplex reticular.

Figura 2.9. Diseño símplex centroide para  $p = 3$  y  $p = 4$ 

Fuente: Montgomery (2017).

Siguiendo la Ecuación (7) se obtienen  $2^3 - 1 = 7$  tratamientos para el primer diseño y  $2^4 - 1 = 15$  para el segundo.

Estos diseños son generalmente empleados cuando la persona experimentadora piensa que algunos términos cúbicos son necesarios en el modelo final. Esto porque para un modelo con  $p = 3$  se puede estimar un modelo cúbico especial y para  $p = 4$  es posible estimar un modelo cúbico especial con un término cuártico adicional (Cornell, 2002; Myers et al., 2009).

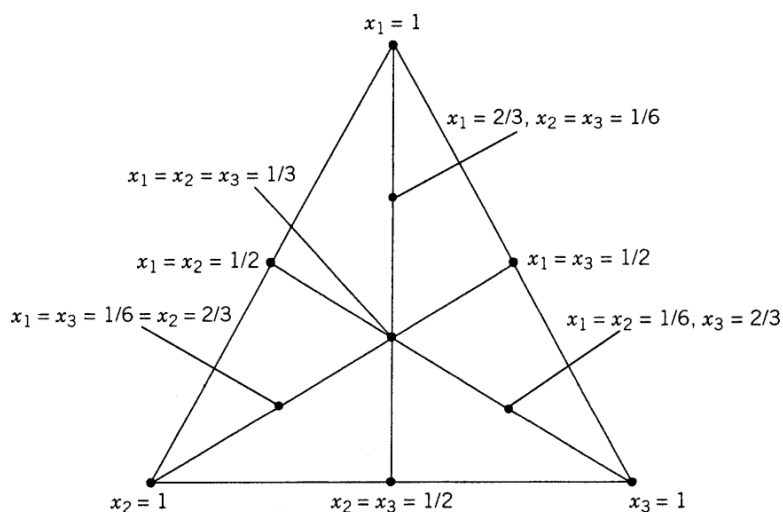
### 2.3.2 Diseños símplex ampliados con puntos axiales

Los diseños símplex reticular y símplex centroide son diseños de frontera, es decir que, con excepción del centroide general, los puntos de diseño se encuentran posicionados sobre los límites (vértices, aristas, caras, entre otros) del espacio símplex y esto fue una crítica usual a los diseños símplex (Cornell, 2002; Myers et al., 2009).

Si el interés es hacer predicciones sobre las propiedades de la mezcla completa, sería deseable aumentar el diseño símplex con puntos en el interior de la mezcla. Se recomienda aumentar los diseños usuales con corridas o puntos axiales y el centroide general, siempre que este no forme parte del diseño original (Myers et al., 2009).

El eje de un componente  $i$  se define formalmente como una línea imaginaria que se extiende desde el punto base  $x_i = 0, x_j = \frac{1}{p-1}$  para todo  $j \neq i$ , hasta el vértice opuesto donde  $x_i = 1, x_j = 0$  para todo  $j \neq i$  (Cornell, 2002; Myers et al., 2009). En la Figura 2.10 se muestra la aplicación de esta definición en un diseño símplex reticular  $\{3, 2\}$ .

Figura 2.10. Diseño símplex reticular  $\{3,2\}$  aumentado con axiales y centroide general



Fuente: Myers et al. (2009).

Los puntos axiales se encuentran a una distancia  $\Delta$  del centroide, el máximo valor de  $\Delta$  es  $\frac{p-1}{p}$ ; se recomienda que las corridas axiales se encuentren en el punto medio entre el centroide y el vértice del eje del componente, eso es  $\Delta = \frac{p-1}{2p}$  por encima del centroide global.

### 2.3.3 Diseños con restricciones de frontera

En ocasiones, cuando no se desea explorar la región símplex completa se puede escoger una región elipsoidal o cuboidal contenida dentro del símplex, pero en estos casos las únicas restricciones reales en la proporción de los componentes son  $0 < x_i < 1$ ,  $i = 1, 2, \dots, p$  y  $x_1 + x_2 + \dots + x_p = 1$ . No obstante, frecuentemente las personas

experimentadoras no son completamente libres de explorar la región simplex completa por restricciones que surgen sobre las proporciones de los componentes.

Dicho de otra forma, ya sea por motivos técnicos o económicos, en algunos problemas con mezclas surgen restricciones sobre los componentes individuales, que se restringen según la Ecuación (8):

$$0 \leq a_i \leq x_i \leq b_i \leq 1, \quad i = 1, 2, \dots, p \quad (8)$$

donde  $a_i$  es la restricción inferior para el componente  $x_i$  de la mezcla, y  $b_i$  es la restricción superior (Cornell, 2002; Gutiérrez Pulido & de la Vara Salazar, 2012; McLean & Anderson, 1966).

En general, las restricciones delimitan la región experimental factible a una zona o subregión del simplex. Esta estrategia confina una región dentro de la región simplex, lo que permite aislar los esfuerzos de diseño y modelado a la subregión, entonces habría una disminución en el costo y el tiempo de experimentación, así como un aumento en la precisión de las estimaciones del modelo (Cornell, 2002). Como se sigue de la Ecuación (8), las restricciones pueden surgir sobre la frontera inferior, la superior o ambas simultáneamente.

### 2.3.3.1 Restricciones sobre la frontera inferior

Cuando solo están presentes restricciones sobre la frontera inferior ( $a_i$ ), la región factible del diseño sigue siendo un diseño simplex, pero que se inscribe dentro de la región del simplex original (Cornell, 2002; Crosier, 1984; Montgomery, 2017).

Para ilustrar esta situación Cornell (2002) propone un ejemplo con tres componentes con las siguientes restricciones sobre la frontera inferior:

$$x_1 \geq 0.35, \quad x_2 \geq 0.20, \quad x_3 \geq 0.15$$

donde la suma de los límites  $x_1 + x_2 + x_3 = 0.70$ , lo cual es menor a 1, por lo que no se forma una mezcla válida.

Para lidiar con esta situación se introducen los pseudocomponentes para la frontera inferior, que redefinen las coordenadas de la subregión de manera similar a como lo hacen las variables codificadas. Los pseudocomponentes se definen como las combinaciones de los componentes originales y se introducen para simplificar la construcción de los diseños y el ajuste de los modelos. El pseudocomponente para la frontera inferior fue introducido por Kurotori (1966) y está definido en la Ecuación (9) como:

$$x'_i = \frac{x_i - a_i}{\left(1 - \sum_{j=1}^p a_j\right)} \quad (9)$$

con  $\sum_{j=1}^p a_j < 1$ . Lo que da como resultado una mezcla válida según la Ecuación (10):

$$x'_i + x'_2 + \dots + x'_p = 1 \quad (10)$$

Por tanto, siguiendo el ejemplo anterior de Cornell (2002), se tiene que al aplicar la Ecuación (9):

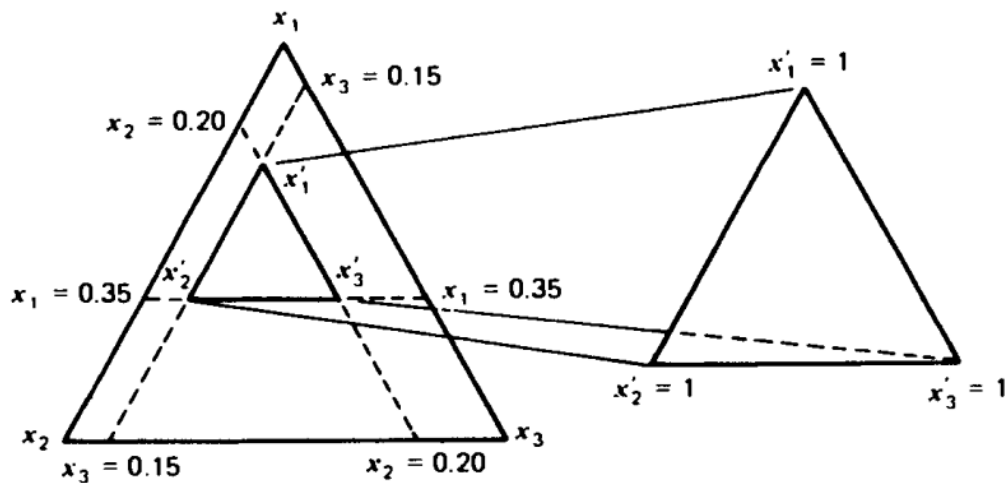
$$x'_1 = \frac{x_1 - 0.35}{0.30}, \quad x'_2 = \frac{x_2 - 0.20}{0.30}, \quad x'_3 = \frac{x_3 - 0.15}{0.30}$$

que produce la región de la Figura 2.11, donde la región de experimentación está reajustada de esta manera, siendo un diseño símplex:

$$0.35 \leq x_1 \leq 0.65, \quad 0.20 \leq x_2 \leq 0.50, \quad 0.15 \leq x_3 \leq 0.30$$

con los pseudocomponentes generando la misma orientación que en el diseño original.

Figura 2.11. Subregión del espacio símplex original redefinida con pseudocomponentes inferiores



Fuente: Cornell (2002).

Los diseños símplex, ya sean reticulares o centroide, se aplican sobre la región parametrizada en pseudocomponentes. El modelo y estadísticos como  $R^2$  y  $R^2_{ajustado}$  se estiman normalmente utilizando dicha parametrización, por ejemplo, ajustando un modelo cuadrático a la situación anterior se tendría:

$$E(y|x) = \gamma_1 x'_1 + \gamma_2 x'_2 + \gamma_3 x'_3 + \gamma_{12} x'_1 x'_2 + \gamma_{13} x'_1 x'_3 + \gamma_{23} x'_2 x'_3$$

Donde  $\gamma_i$  y  $\gamma_{ij}$  son los coeficientes lineales y cuadráticos del modelo estimado con pseudocomponentes. Entonces, al aplicar la Ecuación (9) se obtiene el modelo cuadrático (estudiado más adelante) en las unidades originales ( $\beta_i$  y  $\beta_{ij}$ ).

$$\begin{aligned} E(y|x) = & \gamma_1 \frac{x_1 - a_1}{(1 - \sum_{j=1}^3 a_j)} + \gamma_2 \frac{x_2 - a_2}{(1 - \sum_{j=1}^3 a_j)} + \gamma_3 \frac{x_3 - a_3}{(1 - \sum_{j=1}^3 a_j)} \\ & + \gamma_{12} \frac{x_1 - a_1}{(1 - \sum_{j=1}^3 a_j)} \frac{x_2 - a_2}{(1 - \sum_{j=1}^3 a_j)} + \gamma_{13} \frac{x_1 - a_1}{(1 - \sum_{j=1}^3 a_j)} \frac{x_3 - a_3}{(1 - \sum_{j=1}^3 a_j)} \\ & + \gamma_{23} \frac{x_2 - a_2}{(1 - \sum_{j=1}^3 a_j)} \frac{x_3 - a_3}{(1 - \sum_{j=1}^3 a_j)} \end{aligned}$$

lo que se simplifica de esta manera:

$$\begin{aligned}
E(y|x) = & \gamma_1 \frac{x_1 - a_1}{(1 - \sum_{j=1}^3 a_j)} + \gamma_2 \frac{x_2 - a_2}{(1 - \sum_{j=1}^3 a_j)} + \gamma_3 \frac{x_3 - a_3}{(1 - \sum_{j=1}^3 a_j)} \\
& + \gamma_{12} \frac{(x_1 - a_1)(x_2 - a_2)}{(1 - \sum_{j=1}^3 a_j)^2} + \gamma_{13} \frac{(x_1 - a_1)(x_3 - a_3)}{(1 - \sum_{j=1}^3 a_j)^2} \\
& + \gamma_{23} \frac{(x_2 - a_2)(x_3 - a_3)}{(1 - \sum_{j=1}^3 a_j)^2}
\end{aligned}$$

Es importante tener en cuenta al interpretar los coeficientes, que el modelo es una descripción de la superficie solo en la región restringida (Cornell, 2002). El uso de pseudocomponentes permite utilizar diseños de tipo símplex (sobre la cual se definen directamente los diseños de mezclas usuales) cuando las fronteras inferiores forman parte de la situación experimental y  $\sum_{i=1}^p a_i < 1$ . Caso contrario, al no cumplirse la segunda condición, no se estaría en presencia de un experimento con mezclas, pues con solo los límites inferiores no se cumpliría con la definición provista en la Ecuación (4).

### 2.3.3.2 Restricciones sobre la frontera superior

Si las restricciones están presentes sobre el límite superior ( $x_i \leq b_i$ ) en uno o más componentes, Crosier (1984) sugiere el uso de pseudocomponentes en el caso en el que el diseño sea un símplex; es decir, sí la suma de las  $p - 1$  fronteras superiores son menores o iguales a 1. En otras palabras, se requiere que cumplan con la Ecuación (11):

$$\sum_{i=1}^p b_i - b_{min} \leq 1 \quad (11)$$

donde  $b_{min}$  es el mínimo de las  $p$  restricciones sobre el límite superior. Una transformación en pseudocomponentes que puede ser usada es la de la Ecuación (12):

$$x'_i = \frac{b_i - x_i}{\left(\sum_{j=1}^p b_j - 1\right)} \quad (12)$$

con  $\sum_{j=1}^p b_j > 1$ .

La orientación de estos pseudocomponentes con forma símplex, es opuesta o invertida a la dirección del símplex original; por lo tanto, los pseudocomponentes tienen efectos opuestos a los de los componentes originales (Cornell, 2002), es decir, si la respuesta aumenta con un aumento en  $x_i$ , disminuirá con  $x'_i$ , y esto debe tenerse en cuenta al interpretar los coeficientes obtenidos al ajustar un modelo con los pseudocomponentes.

Siguiendo otro ejemplo de tres componentes propuesto por Cornell (2002) se tienen las siguientes restricciones sobre la frontera superior:

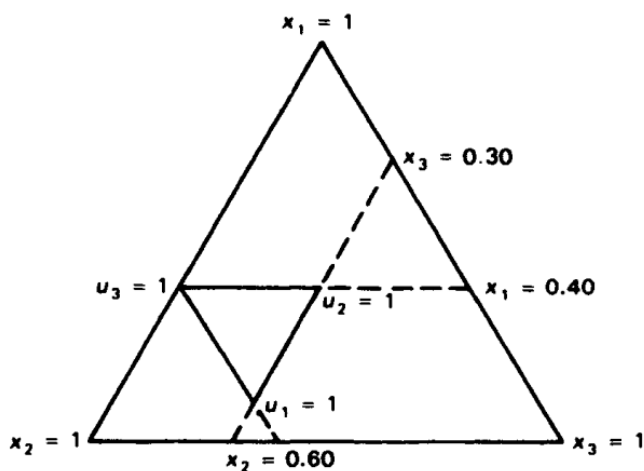
$$x_1 \leq 0.4, \quad x_2 \leq 0.6, \quad x_3 \leq 0.3$$

Aplicando la Ecuación (11) se observa que  $0.4 + 0.6 + 0.3 - 0.3 = 1.0 \leq 1.0$  y, por lo tanto, se puede emplear la transformación en pseudocomponentes de la Ecuación (12):

$$x'_1 = \frac{0.4 - x_1}{0.3}, \quad x'_2 = \frac{0.6 - x_2}{0.3}, \quad x'_3 = \frac{0.3 - x_3}{0.3}$$

de donde se obtiene la nueva región, que se inscribe en la región símplex original (Figura 2.12), siendo a su vez una región símplex invertida.

Figura 2.12. Región símplex invertida dentro de la región símplex original



Fuente: Cornell (2002).  
Nota: entiéndase a  $U_i = b_i$ .

Al igual que con los pseudocomponentes para la frontera inferior, se pueden utilizar los diseños *simplex* centroide o *simplex* reticular para ajustar un modelo y emplear la Ecuación (12) para obtener los coeficientes en las unidades originales.

Para distinguir las transformaciones de las Ecuaciones (9) y (12), Crosier (1984) llama a la primera L – pseudocomponentes y a la segunda U – pseudocomponentes. El mismo autor aclara que muy pocos problemas con restricciones solo en la frontera superior tienen un diseño *simplex* como región experimental.

### 2.3.3.3 *Diseño de vértices extremos*

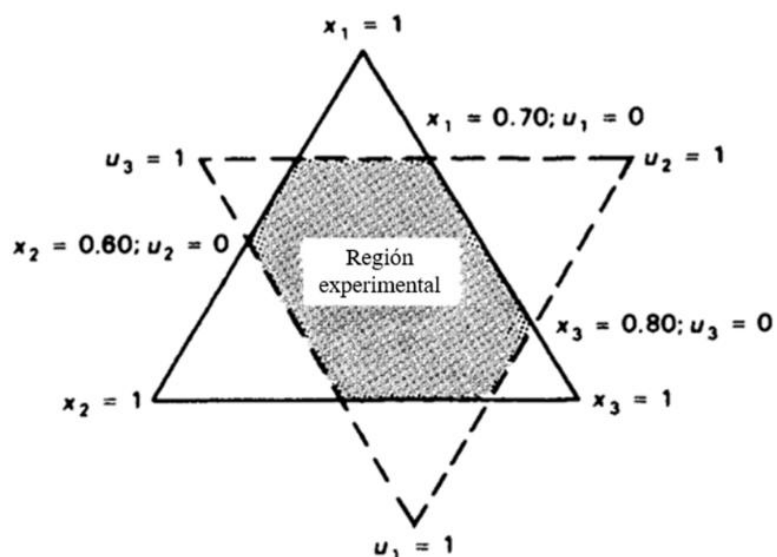
Cuando los componentes tienen restricciones tanto sobre la frontera superior como inferior, la región factible deja de ser un diseño *simplex*; es en cambio un polítopo irregular. Estos diseños son conocidos como de vértices extremos.

Esto también sucede cuando solo existen restricciones sobre la frontera superior, pero no se cumple la condición de la Ecuación (11). Para ilustrar esta situación, Cornell (2002) propone este ejemplo, en el que se toman tres componentes con sus respectivas restricciones superiores:

$$x_1 \leq 0.70, \quad x_2 \leq 0.60, \quad x_3 \leq 0.80$$

Al aplicar la Ecuación (11) se aprecia que no se cumple la restricción ya que  $0.70 + 0.60 + 0.80 - 0.6 = 1.5 \geq 1.0$ , por lo que se obtiene una región experimental que no es un *simplex* (Figura 2.13), sino que compete a la intersección entre la región *simplex* original y la región *simplex* invertida producida por los U – pseudocomponentes.

Figura 2.13. Intersección entre el diseño símplex original y el diseño símplex invertido por pseudocomponentes de frontera superior



Fuente: Adaptado de Cornell (2002).  
Nota: entiéndase a  $L_i = a_i$  y  $U_i = b_i$ .

A menudo en la práctica se cuenta con la presencia, tanto de L – pseudocomponentes como de U – pseudocomponentes para uno o más de los componentes ( $x_i$ ) de la mezcla. Las restricciones definen la región de interés y juegan un rol importante en el diseño y análisis del experimento con mezclas (Piepel, 1983b), por lo que deben seleccionarse con cautela.

Antes de tratar de aplicar un diseño y de seleccionar la transformación o transformaciones a utilizar (L – pseudocomponentes o U – pseudocomponentes) es necesario determinar que las restricciones sobre  $x_i$  sean consistentes, para ello Piepel (1983b) presenta fórmulas para comprobar la consistencia de las restricciones.

#### *Consistencia de las restricciones*

Las restricciones de las Ecuaciones (4) y (8) se dicen que son consistentes cuando luego de enlistar las posibles combinaciones para la región en cuestión, todas y cada una de las proporciones de los componentes (no necesariamente simultáneamente) alcanza

su límite inferior ( $x_i = a_i$ ) y su límite superior ( $x_i = b_i$ ) (Cornell, 2002; Piepel, 1983b). Por ejemplo, las siguientes restricciones son inconsistentes:

$$0.0 \leq x_1 \leq 0.1, \quad 0.1 \leq x_2 \leq 0.2, \quad 0.6 \leq x_3 \leq 0.8$$

porque  $x_3$  no puede ser tan bajo como 0.6, es decir, no hay mezclas en el que  $x_3 = b_3 = 0.6$ . Por el contrario, el valor mínimo de  $x_3$  es  $x_3 = 1 - (b_1 + b_2) = 0.3$ .

Otro tipo de inconsistencia se encuentra en este conjunto de restricciones (Cornell, 2002):

$$0.3 \leq x_1 \leq 0.8, \quad 0.1 \leq x_2 \leq 0.5, \quad 0.2 \leq x_3 \leq 0.6$$

debido a que la frontera superior  $b_1 = 0.8$  no se puede alcanzar, sino que el punto máximo es 0.7.

Para detectar inconsistencias, se calcula el rango para cada componente  $x_i$  con la Ecuación (13).

$$R_i = b_i - a_i, \quad i = 1, 2, \dots, p \quad (13)$$

Entonces  $R'_i$  la Ecuación (14):

$$R'_i = \frac{R_i}{1 - \sum_{i=1}^p a_i}, \quad i = 1, 2, \dots, p \quad (14)$$

representa el rango del L – pseudocomponente  $x'_i$ . Ahora si para cualquier  $i$ ,  $R'_i > 1$ , es decir:

$$R_i > 1 - \sum_{i=1}^p a_i, \quad o, \quad b_i + \sum_{j \neq i}^p a_j > 1 \quad (15)$$

Por lo tanto  $b_i$  es inalcanzable. También,  $a_i$  es inalcanzable si para cualquier  $i$ :

$$\sum_{j \neq i}^p R'_j < 1, \quad o, \quad a_i + \sum_{j \neq i}^p b_j < 1 \quad (16)$$

En el ejemplo en el que  $a_3$  no es alcanzable, se tiene que al aplicar las Ecuaciones (13) y (14):

$$R'_1 + R'_2 = \left( \frac{0.1 - 0.0}{1 - (0.0 + 0.1 + 0.6)} + \frac{0.2 - 0.1}{1 - (0.0 + 0.1 + 0.6)} \right) = \left( \frac{0.1}{0.3} + \frac{0.1}{0.3} = 0.67 < 1 \right)$$

Otra forma de detectar inconsistencia es empleando la transformación en U – pseudocomponentes de la Ecuación (12) y la definición del tamaño lineal del U – símplex, donde:

$$R_b = \sum_{i=1}^p b_i - 1 \quad (17)$$

Entonces  $a_i$  es inalcanzable si para cualquier  $i$ ,  $R_i > R_b$  (Cornell, 2002; Crosier, 1984). Esto aplica de la misma manera al emplear la transformación en L – pseudocomponentes:

$$R_a = 1 - \sum_{i=1}^p a_i \quad (18)$$

Siguiendo el caso de  $R_b$ , si para cualquier  $i$ ,  $R_i > R_a$ , entonces  $b_i$  es inalcanzable.

Cuando las restricciones son inconsistentes, se deben realizar ajustes sobre los límites para hacerlas consistentes; los ajustes se realizan sobre los límites que no son alcanzables. Esto es importante puesto que las restricciones deben ser consistentes para realizar transformaciones en pseudocomponentes que sean adecuadas, seleccionar correctamente los puntos de diseño o ajustar modelos, entre otros.

### Ajuste de restricciones inconsistentes

El conjunto de inconsistencias puede ser ajustado a una forma consistente, pero esto a menudo altera la forma y el tamaño de la región experimental. Por ejemplo, Cornell (2002) ilustra esto con el siguiente conjunto de restricciones:

$$0.20 \leq x_1 \leq 0.40, \quad 0.20 \leq x_2 < 0.60, \quad 0.18 \leq x_3 \leq 0.70$$

Los rangos de los componentes, según la Ecuación (13) son  $R_1 = 0.20$ ,  $R_2 = 0.40$  y  $R_3 = 0.52$ , y, además,  $R_a = 0.42$ , según la Ecuación (18) y  $R_b = 0.7$ , según la Ecuación (17). Entonces, dado que  $R_3 > R_a$ , implica necesariamente que  $b_3 = 0.7$  es inalcanzable y por tanto las restricciones anteriores son inconsistentes.

Estas restricciones deben ajustarse con la Ecuación (19), si es el límite superior:

$$b_i^* = a_i + R_a \quad (19)$$

y con la Ecuación (20) si se trata del límite inferior:

$$a_i^* = b_i - R_b \quad (20)$$

Esto da como resultado que  $b_3^* = 0.18 + 0.42 = 0.60$ , haciendo las restricciones consistentes y reestimando  $R_b = 0.6$ .

También, Cornell (2002) indica que se pueden forzar las restricciones para que sean consistentes alterando los límites inferiores. Para mantener  $b_3 = 0.7$  se debe alterar  $a_1$  y  $a_2$ , disminuyendo el primero a 0.10 o el segundo a 0.10, o ambos en 0.05. La escogencia entre estos tres casos depende del contexto del experimento.

Por medio del criterio de la traza<sup>10</sup> (o criterio de la varianza generalizada), que es uno de varios criterios para seleccionar entre dos o más diseños competentes (Cornell, 2002; Snee, 1975), se determina que se prefiere el diseño con  $a_i = 0.10$ .

Es importante notar que alterar las restricciones también afecta la forma de la región experimental. El diseño con  $b_3 = 0.60$  está compuesto por 5 puntos de diseño. Si se cambia la restricción inferior de  $x_3$  de  $a_3 = 0.18$  a  $a_3 = 0.2$ , el número de puntos de diseño o vértices se reduce de 5 a 4. No obstante, cuando se trabaja con tres o cuatro componentes, Cornell (2002) indica que puede que no valga la pena cualquier intento de reducir el número de vértices. Cabe resaltar que esta recomendación se basa en el ajuste de modelos de orden inferior, puesto que en situaciones en los que se requieran modelos de orden superior (grado de polinomio mayor a 2) o se usan puntos axiales, sí pudiese ser conveniente reducir la cantidad de puntos o corridas en el diseño.

En problemas en los que se involucra un gran número de componentes ( $p \geq 5$ ) ajustar las restricciones en un intento de reducir el número de vértices extremos puede ser altamente deseable (Cornell, 2002). Crosier (1984) demuestra esto con una mezcla de 5 componentes de octano<sup>11</sup>, tomado de Marquardt & Snee (1974), mostrando que una región de 28 vértices podría modificarse en una región con 15 vértices añadiendo 0.01 a las restricciones inferiores y superiores de cada uno de los cinco componentes, es decir, aumentando ligeramente el tamaño de la región de operación.

Finalmente, la escogencia entre las transformaciones de las Ecuaciones (9) y (12) depende de la forma de la región de experimentación. Si  $R_a < R_b$  y, además, el L – simplex está completamente dentro del U – simplex, se selecciona la transformación en L – pseudocomponentes, caso contrario si  $R_b < R_a$  y U – simplex está completamente dentro de L – simplex, se utiliza la transformación en U – pseudocomponentes. Si  $R_a =$

---

<sup>10</sup> Este criterio es mencionado en varias ocasiones, pero es abordado con detalle hasta sección posteriores. La traza de una matriz cuadrada está definida como la suma de los elementos de la diagonal.

<sup>11</sup> Es un hidrocarburo alcano lineal de 8 átomos de carbono.

$R_b$ , la región experimental no es un símplex, se selecciona la transformación en U – pseudocomponentes (Cornell, 2002).

#### 2.3.4 Algoritmos de generación de coordenadas

El diseño de vértices extremos se desarrolla como un procedimiento para la realización de experimentos con mezclas cuando varios componentes tienen restricciones impuestas sobre ellos. Las restricciones así impuestas reducen el tamaño de la región de experimentación que resultaría si los niveles de los componentes se hubiesen encontrado entre el 0 % y el 100 %.

En este tipo de diseños conocer el número y los tipos de fronteras que la región restringida posee permite ser selectivos en la escogencia de un solo subconjunto de ellos para el diseño final. Estos procedimientos pueden ser sencillos cuando la cantidad de componentes es baja ( $p \leq 4$ ), no obstante, existen procedimientos o algoritmos para la generación de las coordenadas de los vértices en la situación contraria, pero que también pueden ser empleados con pocos componentes (Cornell, 2002). A continuación, se presentan algunos de algoritmos más usados para la generación de diseños de vértices extremos.

##### 2.3.4.1 Algoritmo de vértices extremos (EV)

El algoritmo de vértices extremos (EV, por sus siglas en inglés) es también conocido como el algoritmo de McLean y Anderson, ya que ellos fueron sus creadores. La selección de los vértices y los diversos centroides de los politopos resultantes del diseño es un método para determinar un conjunto único de combinaciones de tratamiento. Esta selección está motivada por el deseo de explorar los extremos, así como el centro del espacio factorial (McLean & Anderson, 1966). Los autores proponen la siguiente metodología:

1. Enliste todas las posibles combinaciones de tratamientos de dos niveles, usando los niveles de  $a_i$  y  $b_i$  para todos los componentes excepto para uno de ellos, que se

deja en blanco. Este procedimiento genera  $p \cdot 2^{p-1}$  posibles combinaciones de tratamientos, con un nivel de un componente en blanco en cada uno de ellos.

2. Revise todas las  $p \cdot 2^{p-1}$  posibles combinaciones de tratamientos y complete los espacios en blanco para aquellos componentes en los que es admisible. Por ejemplo, ese nivel, que necesariamente debe estar dentro de los límites de las restricciones para ese componente faltante, que haga que la suma de los niveles de cada combinación de tratamiento sea igual a 1. Cada uno de los tratamientos admisibles es un vértice, empero, algunos vértices pueden aparecer más de una vez.

Luego de que el procedimiento anterior es usado para definir los vértices extremos del politopo, el siguiente paso es definir, si fuese necesario para modelos de segundo orden o superior, la variedad de puntos centrales en las caras y en los bordes, y el centroide. Los puntos centrales en las caras son encontrados aislando todos los vértices que tengan  $p - 3$  niveles idénticos y promediando los niveles de cada uno de los tres componentes restantes.

El centroide en un diseño de vértices extremos puede ser encontrado siguiendo el algoritmo propuesto por Piepel (1983a).

#### 2.3.4.2 Algoritmo XVERT

Un método alternativo al anteriormente nombrado es el algoritmo XVERT, propuesto por Marquardt & Snee (1974). Este algoritmo genera las coordenadas de todos los vértices extremos de la región restringida y además selecciona un subconjunto de vértices extremos que fungen como puntos de diseño para el ajuste de un modelo de primer orden.

El algoritmo XVERT opera bajo el principio de elegir de una lista de candidatos de vértices extremos un diseño cuyos puntos tienen la máxima extensión sobre la región de experimentación. Por ejemplo, al ajustar un modelo de primer orden como el siguiente:

$$y = X\beta + \varepsilon$$

donde  $y$  es el vector  $N \times 1$  de las observaciones y  $X$  es la matriz  $N \times p$  de las proporciones de los componentes correspondientes a las  $N$  mezclas.  $\beta$  es el vector  $p \times 1$  de coeficientes a ser estimados, y  $\varepsilon$  es el vector  $N \times 1$  de los errores aleatorios con media cero y varianza  $\sigma^2$ . El algoritmo XVERT escoge el diseño de una clase de posibles diseños con la máxima extensión, cuya traza de la matriz  $(X'X)^{-1}$  es el valor más bajo (Cornell, 2002). La metodología para encontrar los vértices es la siguiente:

1. Clasifique los componentes en orden ascendente por los rangos  $R_i = b_i - a_i$ . Enliste los componentes como  $X_1, X_2, \dots, X_p$ , donde  $X_1$  es el componente con el menor rango.
2. Establezca una configuración de diseño utilizando los límites inferior y superior de los  $p - 1$  componentes  $X_i$  que tengan los menores rangos. Habría  $2^{p-1}$  combinaciones.
3. Determine el nivel del componente omitido en el paso anterior, con cada uno de los  $2^{p-1}$  combinaciones usando  $X_p = 1.0 - \sum_{i=1}^{p-1} X_i$ .
4. Si el valor de  $X_p$  en el paso 3 toma un valor dentro de la región aceptable, entonces la combinación es un vértice extremo de la región inscrita. Si  $X_p$  no toma un valor dentro de dicha región, entonces ajuste  $X_p$  como el valor del límite inferior o superior, lo que esté más cerca del valor calculado.
5. Para cada punto en donde  $X_p$  está fuera de los límites iniciales, genere puntos adicionales ajustando los niveles de los otros componentes, uno a la vez, por una cantidad igual a la diferencia entre el valor calculado de  $X_p$  en el paso 3 y sustituya  $X_p$  por el valor de la restricción inferior o superior. Por cada punto en el paso 4 que se encuentre fuera de la región de experimentación, se pueden generar al menos  $p - 1$  puntos adicionales, y estos deben ser ajustados según este paso.

El algoritmo está completo cuando se obtienen los  $2^{p-1}$  puntos de diseño. Para modelos de primer orden hay puntos de diseño adicionales que pueden ser utilizados para el cálculo de la falta de ajuste (Draper & Herzberg, 1971). Si no se quieren realizar todas

las corridas o puntos de diseño, se instruye al algoritmo XVERT para buscar el subgrupo de candidatos del tamaño deseado o requerido, que minimice la traza en cada diseño y seleccione el menor.

### 2.3.4.3 Algoritmo XVERT1

Una versión abreviada del algoritmo XVERT es el XVERT1 (Nigam et al., 1983), el cual comienza, como su antecesor, ordenando los componentes de acuerdo con sus rangos  $R_i = b_i - a_i$ , el componente con menor  $R_i$  se denota por  $X_1$  y aquel con el mayor rango con  $X_p$ . Los siguientes pasos del algoritmo XVERT1 son:

1. Enliste las  $2^{p-2}$  combinaciones de los límites inferiores y superiores de los  $p - 2$  componentes con los menores rangos. Esos límites pertenecen a  $X_1, X_2, \dots, X_{p-2}$  y son designados como  $a_i$  y  $b_i$  para  $i = 1, 2, \dots, p - 2$  (nótese el cambio en la nomenclatura de  $a_i$  y  $b_i$ , para distinguirlos de  $a_i$  y  $b_i$ ).
2. Determine a  $C = 1 - \sum_{i=1}^{p-2} X_i$ , donde el valor de  $X_i$  puede ser  $a_i$  o  $b_i$ . Entonces, para cada combinación del paso 1, sustituya en:
  - a. El límite inferior en  $X_{p-1}$  y calcule el valor de  $X_p$  como  $\tilde{X}_p = C - a_{p-1}$
  - b. El límite superior en  $X_{p-1}$  y calcule el valor de  $X_p$  como  $\tilde{X}_p = C - b_{p-1}$
  - c. El límite inferior en  $X_p$  y calcule el valor de  $X_{p-1}$  como  $\tilde{X}_{p-1} = C - a_p$
  - d. El límite superior en  $X_p$  y calcule el valor de  $X_{p-1}$  como  $\tilde{X}_{p-1} = C - b_p$

Las instrucciones desde a. hasta d. generan cuatro combinaciones con cada uno de las  $2^{p-2}$  combinaciones del paso 1, produciendo un total de  $2^p$  puntos de diseño candidatos.

3. Un punto candidato es un vértice extremo de la región restringida si los dos valores calculados  $\tilde{X}_{p-1}$  y  $\tilde{X}_p$  se encuentran dentro de los rangos originales de  $X_{p-1}$  y  $X_p$ , es decir:

$$a_{q-1} \leq \tilde{X}_{q-1} \leq b_{p-1}, \quad y, \quad a_p \leq \tilde{X}_p \leq b_p$$

El algoritmo XVERT1 produce a lo sumo un total de  $2^{p-1}$  puntos posibles distintos, y si el número deseado de puntos de diseño es exactamente  $2^{p-1}$ , entonces XVERT1 debería generar esos puntos más rápido que XVERT (Cornell, 2002) y el algoritmo EV (Nigam et al., 1983), pero también, puede no generar todos los vértices extremos.

#### 2.3.4.4 Algoritmo CONSIM (CONStrained SIMplex)

En adición a las restricciones por límites superiores e inferiores existen las restricciones lineales de la forma:

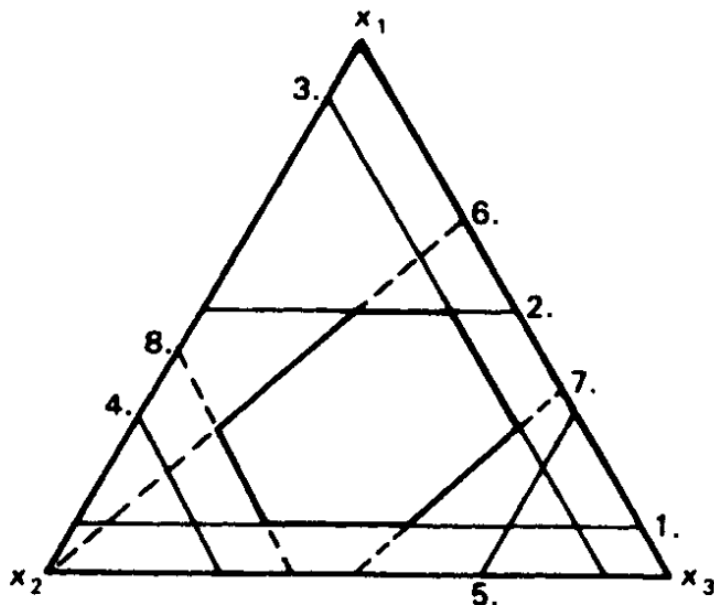
$$C_j \leq A_{1j}x_1 + A_{2j}x_2 + \dots + A_{pj}x_p \leq D_j \quad (21)$$

donde  $A_{ij}$  son constantes escalares. Por ejemplo, Koons (1989) enlista algunas restricciones de frontera y multicomponente, entre ellas la siguiente mezcla que se trata de minerales, deben cumplir:

- $x_1 \leq 0.15$ ,
- $x_2 \leq 0.90$  y
- $\frac{x_1}{(x_1+x_2)} \leq \frac{1}{3}$

La presencia de restricciones multicomponente puede cambiar la forma de la región y por ende el número de vértices extremos, aristas, caras, entre otros, tal cual se muestra en la Figura 2.14, donde las restricciones de frontera se representan con líneas sólidas y numeradas del 1 al 5, mientras que las restricciones multicomponente son representadas con líneas discontinuas y numeradas del 6 al 8.

Figura 2.14. Región doblemente restringida con restricciones de frontera y restricciones multicomponente



Fuente: Cornell (2002).

El algoritmo llamado CONSIM (Snee, 1979) se puede utilizar para generar vértices extremos en una región doblemente restringida cuando las restricciones multicomponente son de la forma de la Ecuación (22):

$$0 \leq A_{0j} + \sum_{i=1}^p A_{ij}x_i, \quad j = 1, 2, \dots, 2h \quad (22)$$

Por ejemplo, la restricción lineal propuesta por Koons (1989) se puede reescribir como  $0 \leq -x_1 + 0.5x_2$ , al resolver la inecuación igualándola a cero. La región estaría definida por un conjunto de puntos extremos  $a_k, k = 1, 2, \dots, n$ . Una vez dado el conjunto de inecuaciones, el algoritmo calcula la combinación lineal  $d_k = a_k A_i - C_{h+1}$ . Si  $d_k < 0$  entonces  $a_k$  no es un punto extremo de la región (Snee, 1979). Cornell (2002) indica que no es necesario trabajar con los pasos de CONSIM dada la existencia de CONVRT.

### 2.3.4.5 Algoritmos CONVRT y CONAEV

Ambos algoritmos están basados en CONSIM (Cornell, 2002; Piepel, 1988) y fueron originalmente escritos en FORTRAN<sup>12</sup>. El algoritmo CONVRT está escrito para generar vértices extremos y CONAEV para generar centroides de varias dimensiones para regiones experimentales descritas por restricciones de la forma mostrada en la Ecuación (22). Estos mejoran a CONSIM en dos formas:

1. CONSIM solo lidia con experimentos de mezclas, es decir, aquellos en los que las  $x_i$  variables son proporciones de componentes en una mezcla y por tanto deben sumar uno. CONVRT y CONAEV son aplicables tanto con variables que forman parte de una mezcla como con aquellas que no (Piepel, 1988).
2. Inconsistencias en las restricciones, así como restricciones innecesarias no fueron abordadas por Snee (1979). Las restricciones inconsistentes dan lugar a regiones experimentales “vacías” mientras que las restricciones innecesarias no son, valga la redundancia, necesarias para describir la región experimental. CONVRT y CONAEV comprueban si las restricciones son innecesarias o inconsistentes, procediendo de forma adecuada si se da esta situación (Piepel, 1988).

## 2.4 Polinomios canónicos

El ajuste de los modelos para experimentos con mezclas se realiza mediante Mínimos Cuadrados Ordinarios (MCO) (Cornell, 2002; Lawson & Willden, 2016; Montgomery, 2017; Myers et al., 2009). El método de MCO selecciona el estimador  $b_i$  para los parámetros desconocidos  $\beta_i, i = 1, 2, \dots, p$  que minimiza la suma de cuadrados de las diferencias entre las respuestas observadas y las estimadas (Cornell, 2002).

---

<sup>12</sup> Lenguaje de programación, cuyo nombre es la contracción del inglés The IBM Mathematical Formula Translating System

Las propiedades de los estimadores dependen de las propiedades de la distribución de los errores aleatorios ( $\varepsilon_i$ ). Se asume que los errores no están correlacionados  $E(\varepsilon_i \varepsilon_j) = 0$ , y que están idénticamente distribuidos con media cero  $E(\varepsilon_i) = 0$ , y con varianza común  $E(\varepsilon_i^2) = \sigma^2$ , con  $\varepsilon_i \neq \varepsilon_j$ , y  $\varepsilon_i, \varepsilon_j = 1, 2, \dots, N$ . Además, se asume que los errores se distribuyen normalmente [ $\varepsilon \sim N(0, \sigma^2)$ ].

La Ecuación (23) expresa la esperanza condicional de la respuesta con la forma de un polinomio  $m$  - éximo grado.

$$E(y|x) = \beta_0 + \sum_{i=1}^p \beta_i x_i + \sum_{i<j}^p \sum_{i<j}^p \beta_{ij} x_i x_j + \sum_{i<j<k}^p \sum_{i<j<k}^p \beta_{ijk} x_i x_j x_k + \dots \quad (23)$$

Pero para un experimento con mezclas sus términos solo tienen significado bajo la restricción  $\sum x_i = 1$ , por lo que los parámetros  $\beta_i, \beta_{ij}, \beta_{ijk} \dots$  asociados con los términos no son únicos. No obstante, substituyendo  $x_p = 1 - \sum_{i=1}^{p-1} x_i$  en la Ecuación (23) se elimina así la dependencia entre los términos  $x_i$  y no se afecta el grado del polinomio.

Si se toma un polinomio de primer grado:

$$E(y|x) = \beta_0 + \sum_{i=1}^p \beta_i x_i$$

y se multiplica el término  $\beta_0$  por la restricción de la Ecuación (4) da como resultado la Ecuación (24):

$$E(y|x) = \beta_0 \left( \sum_{i=1}^p x_i \right) + \sum_{i=1}^p \beta_i x_i = \sum_{i=1}^p \beta_i^* x_i \quad (24)$$

donde  $\beta_i^* = \beta_0 + \beta_i$  para todo  $i = 1, 2, \dots, p$ . Por tanto, la ausencia del término  $\beta_0$  se debe a la restricción de la Ecuación (4). Se sigue de igual forma para los demás grados de polinomio, en adelante por practicidad y convencionalismo, se remueven los asteriscos

de los términos  $\beta$ . Las formas estándar o canónicas [también conocidas como formas o polinomios de Scheffé (1958)] de los modelos de mezclas son:

- Lineal

$$E(y|x) = \sum_{i=1}^p \beta_i x_i \quad (25)$$

- Cuadrático

$$E(y|x) = \sum_{i=1}^p \beta_i x_i + \sum_{i<j}^p \beta_{ij} x_i x_j \quad (26)$$

Note que en la Ecuación (26) el término cuadrático “convencional”  $\beta_i^2$  de los RSM para modelos cuadráticos [ver Ecuación (2)] no está presente, sino que en estos modelos se le denomina cuadrático a los términos  $\beta_{ij}$  (Cornell, 2002; Montgomery, 2017; Montgomery & Voth, 1994; Myers et al., 2009; Scheffé, 1958). Ocurre de forma similar para los términos cúbicos especiales  $\beta_{ijk}$  y completos  $\delta_{ijk}$  de las Ecuaciones (27) y (28).

- Cúbico especial

$$E(y|x) = \sum_{i=1}^p \beta_i x_i + \sum_{i<j}^p \beta_{ij} x_i x_j + \sum_{i<j<k}^p \beta_{ijk} x_i x_j x_k \quad (27)$$

- Cúbico completo

$$E(y|x) = \sum_{i=1}^p \beta_i x_i + \sum_{i<j}^p \beta_{ij} x_i x_j + \sum_{i<j}^p \delta_{ij} x_i x_j (x_i - x_j) + \sum_{i<j<k}^p \beta_{ijk} x_i x_j x_k \quad (28)$$

Para un diseño simplex lattice el número de términos en el polinomio es función de  $m$ . El efecto de un componente  $x_i$  se define como el cambio en la respuesta a lo largo del eje que une este vértice con el centro del lado opuesto del simplex. A lo largo de este eje la proporción con la que participa el ingrediente  $x_i$  cambia de 1 a 0, mientras que el resto de los ingredientes se mantienen en proporciones iguales, cuya suma cambia de 0 a 1.

Los términos en este modelo, aunque algunos distintos a los usuales en los RSM tienen interpretaciones relativamente simples. El parámetro  $\beta_i$  representa la respuesta esperada por la mezcla pura  $x_i = 1$  y  $x_j = 0$  cuando  $i \neq j$ . El primer término en todas las ecuaciones anteriores es llamado la porción de mezcla lineal.

Cuando la curvatura surge de la mezcla no lineal entre pares de componentes, los parámetros  $\beta_{ij}$  representan el coeficiente cuadrático (de segundo orden o de interacción) sinérgico cuando es positivo, o antagónico, cuando es negativo, no lineal de la mezcla binaria.

Por otro lado  $\delta_{ij}$  es el coeficiente cúbico de sinergismo o antagonismo no lineal de la mezcla binaria y finalmente  $\beta_{ijk}$  representa la mezcla ternaria entre los componentes  $i, j$  y  $k$  en el interior del politopo.

Los parámetros estimados son contrastados mediante la hipótesis nula de que estos sean iguales a cero ( $H_0: \beta = 0$ , es decir, la variable de respuesta no depende del componente de la mezcla), excepto el estimador de la porción de la mezcla lineal ( $\beta_i$ ), ya que aun cuando el parámetro fuera igual a cero esto no implicaría que el componente  $x_i$  no tenga efecto individual o lineal, sino que su efecto es un valor constante ( $\beta_0$ ). Pero, las hipótesis  $H_0: \beta_{ij} = 0$  sobre los coeficientes de segundo orden o interacciones si proveen información acerca del efecto combinado de los componentes  $x_i$  y  $x_j$ , de igual forma para otros coeficientes de orden superior (Cornell, 2002).

Es necesario acotar que, en los diseños de mezclas, por lo general, los términos de orden superior son necesarios porque el fenómeno estudiado es complejo, y la región de experimentación es frecuentemente del mismo tamaño la región de operabilidad completa, y por lo tanto es grande y requiere un modelo elaborado.

No obstante, uno de los supuestos básicos que se hace cuando se emplean métodos de superficie de respuesta es que la superficie puede ser aproximada razonablemente bien dentro del rango de los datos con un modelo de orden inferior (de primero o de segundo orden) (Cornell, 2002).

#### 2.4.1 Colinealidad

La presencia de estimaciones inestables de los coeficientes surge de lo que se conoce como colinealidad (o multicolinealidad). Esta es una condición entre el conjunto de las  $q$  variables regresoras  $x_1, x_2, \dots, x_q$  en el modelo, donde existe una dependencia lineal aproximada (Cornell, 2002). Dicho de otra forma, es una condición que ocurre cuando algunas o todas las variables predictoras están correlacionadas con otras variables predictoras.

Existen varias técnicas para detectar la colinealidad. Según Cornell (2002), dos de las más usadas son:

- i. Calcular el valor del factor de Inflación de la varianza (VIF, por sus siglas en inglés) de los coeficientes estimados en el modelo. Por lo general se esperan valores menores a 10 (Marquardt, 1970).
- ii. Calcular los autovalores (eigenvalues)  $[\lambda_j]$  de la matriz escalada  $X'X$  y buscar valores pequeños de autovalores ( $\lambda_j < 0.001$ ), donde la matriz  $X$  es de tamaño  $N \times q$ , ( $q \geq p$ ) y representa a los predictores.

St. John (1984) discute la presencia de colinealidad entre los términos de los modelos de mezclas de Scheffé y los problemas que esto conlleva, sobre todo cuando la región se encuentra altamente restringida; a la vez sugiere el uso de la regresión Ridge

como una medida para estabilizar los coeficientes estimados en el modelo, ya que este método de estimación se puede emplear en escenarios donde las variables independientes están altamente correlacionadas (McDonald, 2009).

#### 2.4.2 Reducción del modelo

Es posible reducir el modelo para hacerlo lo más simple posible, bajo la premisa de que formas más simples del modelo son más fáciles de interpretar y entender. En la búsqueda de formas más simples del modelo, se debe recordar, además, que, a mayor simpleza, mayor es la limitación en el campo de la utilidad del modelo (Cornell, 2002).

La reducción del modelo puede lograrse de diferentes formas, una de ellas es simplemente sumando términos, por ejemplo, sumando las proporciones de  $x_i$  y  $x_j$ . Otra forma es la de eliminar términos del modelo en lugar de combinarlos. El enfoque más evidente para la reducción del modelo es el de remover los términos de orden superior que no son significativos (Cornell, 2002), respetando principios como el de jerarquía (Montgomery, 2017).

#### 2.4.3 Falta de ajuste de los modelos

Un enfoque de análisis que puede ser usado en el desarrollo de una forma apropiada del modelo es la de inicialmente ajustar uno de primer orden y evaluar si dicho modelo se ajusta adecuadamente a las respuestas observadas. En este enfoque se asume inicialmente que los componentes se mezclan linealmente y que al ajustarse el modelo de primer grado se cuestiona si hay evidencia que sea contraria a esta suposición. Si hay evidencia que sugiera que este primer modelo no es adecuado, entonces se deben llevar a cabo nuevos experimentos que puedan mejorar el ajuste o cambiar la forma del modelo (Cornell, 2002). Esto se lleva a cabo de forma similar a la naturaleza secuencial de los RSM (Montgomery, 2017; Myers et al., 2009).

Cuando se toman observaciones replicadas en uno o más puntos del diseño, siempre que el número de puntos distintos de diseño exceda el número de términos en

el modelo ajustado, la suma de cuadrados (SC) del error puede ser dividida en dos partes: la SC de la falta de ajuste del modelo, y la SC de error puro, que es calculada usando las réplicas (Cornell, 2002; Draper & Herzberg, 1971). Con la falta de ajuste se prueba la hipótesis nula de que el modelo es adecuado al conjunto de datos, de rechazarse implicaría la necesidad de ajustar modelos de orden superior o inferior.

## 2.5 Otras formas de polinomios

Existen otras formas de polinomios, además de los canónicos, que fueron introducidos por Draper & St. John (1977) y que en algunas situaciones pueden ser más apropiados. Una de estas técnicas es la inclusión de términos inversos en el polinomio. Estos se incluyen para modelar un cambio extremo en el comportamiento de la respuesta ya que el valor de uno o más componentes tiende a un límite de la región simplex. Estos se muestran en las Ecuaciones (29), (30), (31) y (32):

- Lineal con término inverso

$$E(y|x) = \sum_{i=1}^p \beta_i x_i + \sum_{i=1}^p \beta_{-i} x_i^{-1} \quad (29)$$

- Cuadrático con término inverso

$$E(y|x) = \sum_{i=1}^p \beta_i x_i + \sum_{i < j}^p \beta_{ij} x_i x_j + \sum_{i=1}^p \beta_{-i} x_i^{-1} \quad (30)$$

- Cúbico especial con término inverso

$$E(y|x) = \sum_{i=1}^p \beta_i x_i + \sum_{i < j}^p \beta_{ij} x_i x_j + \sum_{i < j < k}^p \beta_{ijk} x_i x_j x_k + \sum_{i=1}^p \beta_{-i} x_i^{-1} \quad (31)$$

- Cúbico completo con término inverso

$$\begin{aligned}
 E(y|x) = & \sum_{i=1}^p \beta_i x_i + \sum_{i<j}^p \sum_{i<j}^p \beta_{ij} x_i x_j + \sum_{i<j}^p \sum_{i<j}^p \delta_{ij} x_i x_j (x_i - x_j) \\
 & + \sum_{i<j<k}^p \sum_{i<j<k}^p \beta_{ijk} x_i x_j x_k + \sum_{i=1}^p \beta_{-i} x_i^{-1}
 \end{aligned} \tag{32}$$

Estos modelos son los polinomios canónicos con los términos adicionales de la forma  $x_i^{-1}$  incluida para tomar en cuenta el posible cambio en la respuesta como cuando  $x_i$  se aproxima a cero. La necesidad de incluir este término en los modelos es más probable que se descubra durante un programa de experimentación secuencial. Este a menudo comienza con un polinomio canónico de primer orden y se avanza, si es necesario, hasta modelos de tercer orden (Cornell, 2002).

Respecto a esto, Draper & St. John (1977) mencionan que asumen que un experimentador con frecuencia seguiría un enfoque secuencial en el ajuste de los polinomios de Scheffé. De esta manera, un experimentador ajustaría un modelo de primer orden para luego juzgar la adecuación del modelo; si el modelo se considera inadecuado, se ajustaría un modelo de segundo orden, que, si también se considera inapropiado, se continuaría aumentando el grado del polinomio ajustado.

## 2.6 Criterios para la selección entre diseños competentes

En algunos diseños con mezclas se puede encontrar que un punto de diseño o un subconjunto de ellos pueden ser influyentes de forma desproporcionada en el ajuste por MCO del modelo. Esto da como resultado que las estimaciones de los parámetros o las predicciones con el modelo puedan depender más de la influencia del subconjunto de puntos que de la mayoría de los datos (Cornell, 2002).

Un punto de diseño o varios de ellos puede ser influyente por su localización en la región símplex o los valores de la respuesta observada en esos puntos, o ambos. La influencia de un subconjunto particular de puntos de diseño está directamente

relacionada por la cantidad de apalancamiento (leverage) que el punto o puntos tienen en la estimación y ajuste por MCO, en otras palabras, entre más alto el apalancamiento, más grande la influencia (Montgomery & Voth, 1994).

Además, la varianza de la predicción del modelo ajustado es directamente proporcional al apalancamiento en un punto, tal que, si la distribución de este a través de la región experimental no es uniforme, la varianza tampoco lo es (Cornell, 2002), razón por la cual diseños con valores altos de apalancamiento no son deseables (Montgomery & Voth, 1994).

Existen varios criterios de optimalidad de diseño que se pueden utilizar para elegir un conjunto de puntos. Para explicar los mismos, se escribe el modelo de regresión lineal general de Scheffé de forma compacta y en notación matricial para  $N$  observaciones y  $q$  variables predictoras en la Ecuación (33):

$$y = X\beta + \varepsilon \quad (33)$$

donde la matriz  $X$  es de tamaño  $N \times q$ , ( $q \geq p$ ).

La configuración del diseño de los  $p$  componentes para los  $N$  ensayos o corridas experimentales son especificados por las primeras columnas de la matriz  $X$ . Si  $q > p$ , las restantes  $q - p$  columnas de  $X$  corresponden a los términos cruzados entre las proporciones de los componentes.

Cuando los elementos del vector  $N \times 1$  de los errores ( $\varepsilon$ ) son independientes e idénticamente distribuidos con  $E(\varepsilon) = 0$  y con varianza común  $\sigma^2$ , la estimación por MCO de  $\beta$ , de tamaño  $q \times 1$ , es decir, los elementos de  $b$  y, además, la matriz de varianza – covarianza de los elementos de  $b$ , están dados por:

$$b = (X'X)^{-1}X'y, \quad var(b) = (X'X)^{-1}\sigma^2 \quad (34)$$

donde la matriz  $q \times q$  se considera de rango completo<sup>13</sup>. La respuesta predicha en un punto de la composición  $x$  es  $\hat{y}(x) = x'b$  y entonces:

$$\text{var}[\hat{y}(x)] = x'(X'X)^{-1}x\sigma^2 \quad (35)$$

La matriz “sombrero” ( $\hat{H}$ ) es, según la Ecuación (36):

$$\hat{H} = X(X'X)^{-1}X' \quad (36)$$

El apalancamiento ( $h_{ii}$ ) es la diagonal de la matriz “sombrero”. Valores de  $h_{ii}$  ampliamente diferentes o distantes indican que la calidad de la predicción del modelo puede diferir considerablemente sobre la región de interés. Es decir, que un valor grande de  $h_{ii}$  implica que la varianza de la respuesta predicha  $\text{var}(\hat{y}_i)$  es grande en ese punto de diseño (Montgomery & Voth, 1994).

Una regla general tomada de la literatura para regresión no relacionada con mezclas para considerar un punto de  $h_{ii}$  como un valor con alto apalancamiento es  $h_{ii} > \frac{2q}{N}$ , pero se sugiere no utilizarla en este contexto, pues en los experimentos con mezclas a menudo  $q$  es grande relativo a  $n$ , resultando en que  $\frac{2q}{N} > 1$ . Dicho esto, Montgomery & Voth (1994) sugieren utilizar la regla:  $h_{ii} > 0.75$ .

La replicación reduce los valores de  $h_{ii}$ , por lo que se recomienda su uso en puntos en los que los valores de apalancamiento sean altos, haciendo la distribución de esta medida más uniforme (Montgomery & Voth, 1994). Particularmente cuando el objetivo de experimentación recae en utilizar la ecuación de regresión para describir la forma de la superficie de respuesta de mezclas.

Algunos de los criterios de optimalidad de diseños son el ADGV (Cornell, 2002; Lawson, 2014; Montgomery, 2017; Myers et al., 2009; Snee, 1975):

---

<sup>13</sup> Es decir, cada elemento del codominio  $y$  es en realidad una salida de la función.

- i. A – optimalidad: es cuando la traza de  $(X'X)^{-1}$  es minimizada. En cuyo caso la varianza promedio de los elementos de  $b$  es minimizada.
- ii. D – optimalidad: es cuando el determinante  $\det|X'X|$  es maximizado o bien cuando  $\det|X'X|^{-1}$  es minimizado; con ello se minimiza la varianza general de los elementos de  $b$ . Si los errores están normalmente distribuidos, el diseño D – óptimo minimiza el volumen del elipsoide de confianza para los parámetros desconocidos en la Ecuación (33).
- iii. G – optimalidad: busca minimizar la varianza máxima de predicción,  $\max \{d = x'(X'X)^{-1}x\}$ , sobre un conjunto especificado de puntos de diseño. Es decir, se prefieren valores mínimos del resultado del cociente del apalancamiento promedio entre el apalancamiento máximo.
- iv. V – optimalidad: busca minimizar el valor promedio de  $d$  sobre un conjunto especificado de puntos de diseño. Es decir, lo que se quiere minimizar es el apalancamiento promedio. Este criterio es también conocido por varios sinónimos: I-, IV-, Q- y V- (Martin et al., 1999; Mee, 2007; Myers et al., 2009).

En resumen, los criterios A y D son dirigidos a las varianzas de los parámetros estimados, mientras que G y V son dirigidos a la varianza de la ecuación de predicción. La conveniencia de un diseño incrementa a medida que los valores de la traza de la matriz  $(X'X)^{-1}$ ,  $\det|X'X|^{-1}$ ,  $\max \{d\}$  y el valor promedio de  $d$  descienden (Cornell, 2002; Smith, 2005).

Cornell (2002) indica que, en la mayoría de las aplicaciones prácticas estas cantidades disminuirían en valor a medida que se amplíen los rangos de los componentes  $p$ . Sin embargo, en el caso donde la región experimental es una región restringida dentro de la región simplex de los  $p$  componentes, la localización del centroide de la región restringida en relación con el centroide de la región simplex completa podría tener un efecto en los valores anteriores.

Además, se sabe que diseños óptimos bajo un criterio no lo son necesariamente para los otros. Por otro lado, el cálculo matricial tiene dependencia del orden en el que

se encuentran los puntos candidatos. Por ejemplo, en el Cuadro 2.1 se muestran dos diseños con cinco tratamientos, idénticos, pero colocados en diferente orden.

Cuadro 2.1. Diferencia en el cálculo del criterio de A – Optimalidad y D – Optimalidad con base en las diferencias en el orden de las corridas experimentales

Corrida	Orden A			Orden B		
	$x_1$	$x_2$	$x_3$	$x_1$	$x_2$	$x_3$
1	0.40	0.42	0.18	0.22	0.60	0.18
2	0.40	0.20	0.40	0.10	0.20	0.70
3	0.10	0.20	0.70	0.40	0.20	0.40
4	0.10	0.60	0.30	0.10	0.60	0.30
5	0.22	0.60	0.18	0.40	0.42	0.18
<b>A – Óptimo</b>	<b>11.70</b>			<b>17.73</b>		
<b>D – Óptimo</b>	<b>0.07</b>			<b>0.04</b>		

Fuente: Elaboración propia a partir de los datos de Cornell (2002)

En el ejemplo anterior se obtienen valores de diferentes de A – óptimo y D – óptimo. Tanto V – óptimo como G – óptimo dependen del apalancamiento cuyo valor máximo y promedio son respectivamente 0.87 y 0.60, sin importar el orden.

Asociados a estos criterios se encuentran medidas de eficiencia que toman en cuenta la cantidad de corridas necesarias y el número de términos a estimar en el modelo. Estas son una medida relativa, para comparar un diseño con otro, donde eficiencias inferiores al 100 % pueden ser perfectamente satisfactorias (Goos & Jones, 2011).

- i. A – Eficiencia: es una medida relacionada con el criterio de A – Optimalidad, cuya fórmula se define según la Ecuación (37):

$$A_{eff} = \frac{q}{n \cdot tr[(X'X)^{-1}]} \quad (37)$$

- ii. D – Eficiencia: compara el determinante de la matriz de información de un diseño con un “determinante ideal” denotado por  $n^q$  (Goos & Jones, 2011). De tal forma que D – Eficiencia se define como [Ecuación (38)]:

$$D_{eff} = \left( \frac{\det|X'X|^{\frac{1}{q}}}{n^q} \right) = \frac{\det|X'X|^{\frac{1}{q}}}{n} \quad (38)$$

- iii. G – Eficiencia: está, desde luego, relacionada con la G – Optimalidad, y por tanto emplea el máximo valor de  $d$  (Cornell, 2002; Snee, 1975) y se define como [Ecuación (39)]:

$$G_{eff} = \frac{q}{n \cdot d} \quad (39)$$

- iv. V – eficiencia: minimizar la varianza promedio de predicción es un criterio relevante. En este caso, no se depende de  $n$  o  $q$ , tal cual se muestra en la Ecuación (40), sino que  $P_x$  representa la varianza promedio de predicción de un diseño cualquiera (Goos et al., 2016).

$$V_{eff} = \frac{P_2}{P_1} \quad (40)$$

Si el valor obtenido es mayor que uno, quiere decir que el diseño ( $P_1$ ) es mejor que el diseño 2 ( $P_2$ ). En este sentido, basta con comparar las varianzas promedio en términos absolutos, por lo cual esta medida no es empleada como punto de comparación.

## 2.7 Función de deseabilidad

En ocasiones, los experimentos se llevan a cabo y miden múltiples variables de respuesta de forma simultánea<sup>14</sup>. Por otro lado, como ya ha sido abordado, en los experimentos con mezclas suelen ser necesarios polinomios más complejos al momento de estimar una superficie de respuesta, lo que puede dificultar la búsqueda de la mejor formulación posible. Para ello, hay varias técnicas de programación no lineal para formular y resolver el problema de la optimización de múltiples respuestas.

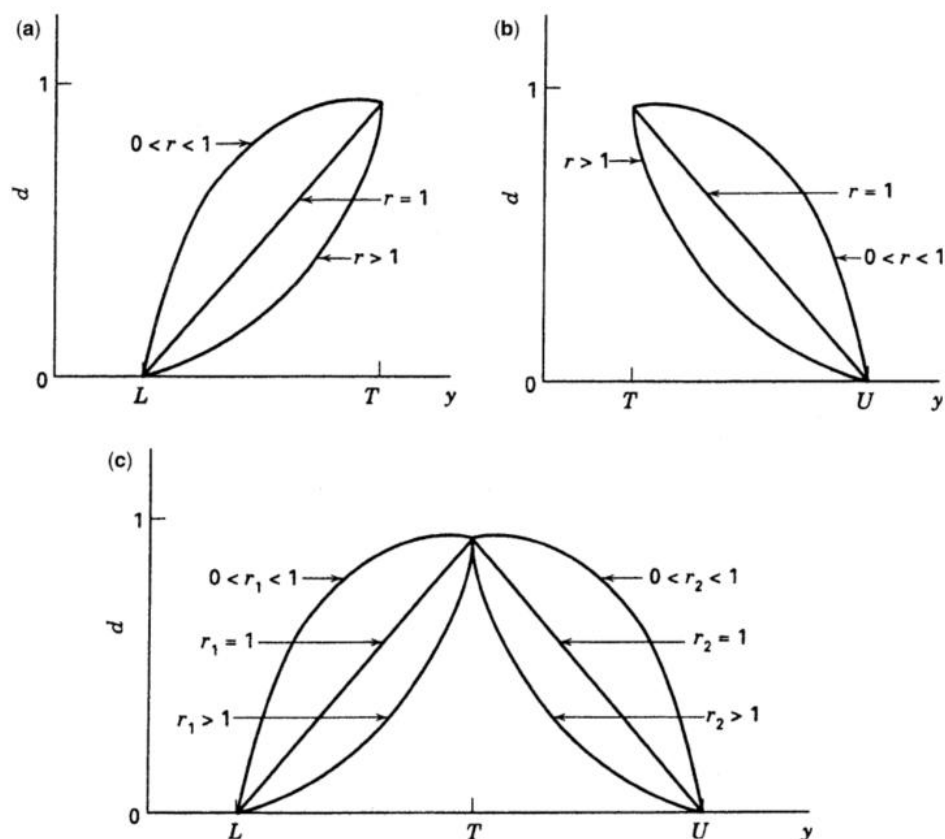
---

<sup>14</sup> Es el caso del ejemplo práctico que se desarrolla en esta investigación y el motivo por el cual se incluye este concepto dentro del marco teórico.

Un enfoque común utilizado y recomendado en la optimización simultánea de variables de respuesta en el DdE es el de utilizar la función de deseabilidad (García-Goñi, 2024; Montgomery, 2017; Myers et al., 2009) propuesta por Derringer & Suich (1980).

Esta función consiste en primero traducir cada respuesta  $y_i$  en una función de deseabilidad individual  $d_i$  que varía sobre el rango  $0 \leq d_i \leq 1$ , donde si la respuesta  $y_i$  está en su objetivo o meta, entonces  $d_i = 1$ , caso contrario si está fuera,  $d_i = 0$ . Las variables de diseño son escogidas para maximizar la deseabilidad compuesta o global  $D = (d_1 \cdot d_2 \cdots d_m)^{\frac{1}{m}} = \prod_{i=1}^m d_i^{\frac{1}{m}}$ , donde  $m$  es la cantidad de respuestas (Myers et al., 2009). Las funciones de deseabilidad individuales están detalladas en la Figura 2.15.

Figura 2.15. Funciones de deseabilidad individuales para optimización simultánea



Fuente: Myers et al. (2009).

Nota: El objetivo en (a) es maximizar  $y$ . El objetivo en (b) es minimizar  $y$ . El objetivo en (c) es que  $y$  se acerque lo más posible a la meta ( $T$ ).

Si el objetivo  $T$  para la respuesta es un valor máximo se emplea la Ecuación (41):

$$d = \begin{cases} 0, & y < L \\ \left(\frac{y-L}{T-L}\right)^r, & L \leq y \leq T \\ 1, & y > T \end{cases} \quad (41)$$

donde  $r$  es un peso, y cuando  $r = 1$ , la función de deseabilidad es lineal. Escoger  $r > 1$  le da más énfasis a acercarse al valor objetivo  $T$ , y al seleccionar  $0 < r < 1$  lo hace menos importante (Myers et al., 2009).

Si la meta  $T$  es un valor mínimo se recurre a la Ecuación (42):

$$d = \begin{cases} 1, & y < T \\ \left(\frac{U-y}{U-T}\right)^r, & T \leq y \leq U \\ 0, & y > U \end{cases} \quad (42)$$

Por su parte, la deseabilidad de dos vías, mostrada en la Figura 2.15(c) asume que la meta  $T$  está localizada entre el límite inferior ( $L$ ) y superior ( $U$ ) y está definida en la Ecuación (43):

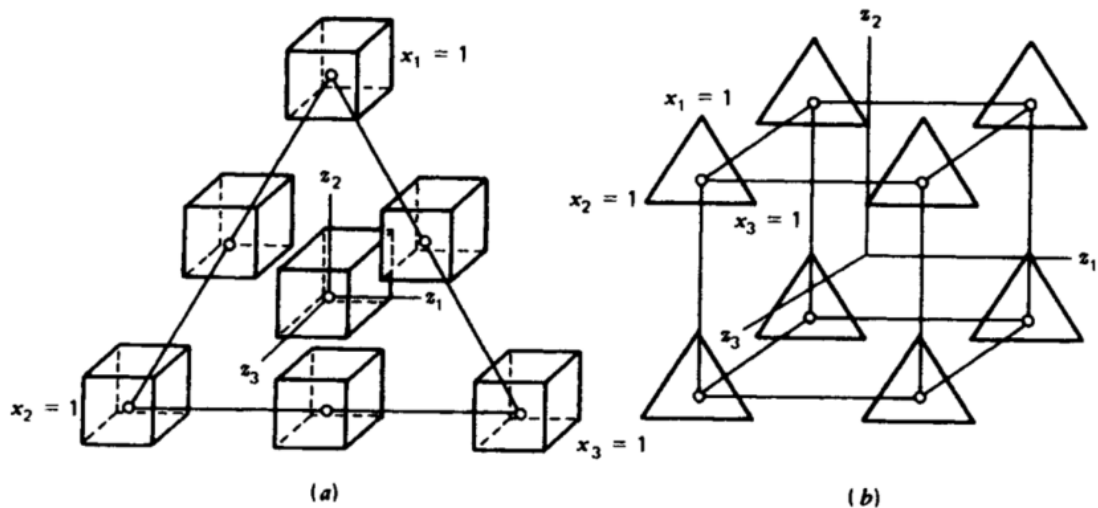
$$d = \begin{cases} 0, & y < L \\ \left(\frac{y-L}{T-L}\right)^{r_1}, & L \leq y \leq T \\ \left(\frac{U-y}{U-T}\right)^{r_2}, & T \leq y \leq U \\ 0, & y > U \end{cases} \quad (43)$$

## 2.8 Experimentos con mezclas y variables de proceso

Las variables de proceso son factores en un experimento con mezclas que no forma ninguna porción de la mezcla, pero que sus niveles, cuando son cambiados, pueden afectar las propiedades de la mezcla de los ingredientes. Estos factores reciben el nombre de variables de proceso ( $z_i$ ) del experimento, en contraposición con el término componentes ( $p$ ) que está reservado para las porciones de la mezcla (Cornell, 2002).

La inclusión de variables de proceso decanta en la necesidad de utilizar modelos jerárquicos. Con las variables de proceso es posible realizar arreglos factoriales de diversos tipos, incluyendo fraccionamiento (Cornell & Gorman, 1984). Por ejemplo, en la Figura 2.16(a) se muestra un diseño en el que el arreglo factorial está anidado dentro del experimento símplex centroide, en contraposición en la Figura 2.16(b) donde el experimento símplex está anidado dentro del arreglo factorial  $2^3$ .

Figura 2.16. Experimento símplex – centroide de tres componentes con un arreglo factorial completo  $2^3$



Fuente: Cornell (2002)

La elección entre estos arreglos se hace por conveniencia de la persona experimentadora en virtud de que la aleatorización completa de las corridas es virtualmente imposible, siguiendo la estrategia de parcelas divididas (Cornell, 2002).

Cornell (2002) recomienda diseñar estos experimentos con cautela, ya que la cantidad de corridas experimentales aumenta considerablemente con la inclusión de variables de proceso. Asumiendo el segundo arreglo, Figura 2.16(b), el modelo general para este tipo de diseño está dado por la Ecuación (44):

$$y_{ijk} = \mu + R_i + Z_j + RZ_{ij} + X_k + ZX_{jk} + \varepsilon_{ijk}, \quad \begin{cases} i = 1, 2, \dots, r \\ j = 1, 2, \dots, P \\ k = 1, 2, \dots, m \end{cases} \quad (44)$$

donde:

$\mu$  = la media general

$R_i$  = efecto aleatorio de la  $i$ -ésima réplica;  $R_i \sim NID(0, \sigma_R^2)$

$Z_j$  = efecto fijo de la  $j$ -ésimo tratamiento de la parcela completa

$RZ_{ij}$  = efecto aleatorio de la réplica por la interacción de los tratamientos de la parcela completa;  $RZ_{ij} \sim NID(0, \sigma_{RZ}^2)$

$X_k$  = efecto fijo del  $k$ -ésimo tratamiento de la parcela dividida.

$ZX_{jk}$  = efecto fijo de la interacción de la parcela completa y la parcela dividida

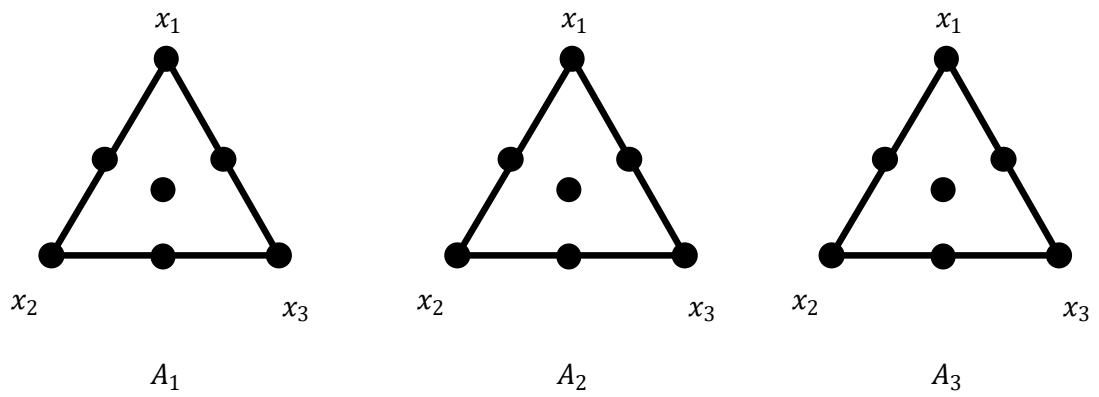
$\varepsilon_{ijk}$  = error aleatorio asociado con  $y_{ijk}$ ;  $\varepsilon_{ijk} \sim NID(0, \sigma_e^2)$

El error de parcela completa es denotado como  $\sigma_{RZ}^2$ , mientras que el error de la parcela dividida es  $\sigma_e^2$ .

## 2.9 Experimentos mezcla – cantidad

Es un experimento que se diseña en dos o más niveles del total de la mezcla. La respuesta se asume que depende de las proporciones en las que participan los ingredientes y, además, del total de la mezcla (Cornell, 2002). La meta u objetivo de los experimentos mezcla – cantidad es aprender acerca de los efectos en la variable de respuesta al variar estas proporciones y la cantidad total de la mezcla; estos modelos son estimados de la misma forma que cuando se incluyen variables de proceso, pero siempre que sea solo una de ellas (Piepel & Cornell, 1985, 1987), como se aprecia en la Figura 2.17.

Figura 2.17. Diseño símplex – centroide de tres componentes para tres niveles ( $A_i$ ) del total de una mezcla



Fuente: Adaptado de Piepel & Cornell (1987)

### 3. ABORDAJE METODOLÓGICO

El aporte de esta investigación consiste en la creación de un algoritmo para la reducción de DdE con mezclas de vértices extremos que hayan sido creados únicamente por restricciones de frontera superior. Este último elemento es un delimitador del alcance de este trabajo. La figura geométrica producida por la aplicación de este algoritmo es un diseño L – Símplex que se encuentra dentro de la región de operabilidad y a la vez dentro de la región de experimentación definida por los límites de los vértices extremos.

La intención original de esta investigación consistía en encontrar una región símplex que se encontrara dentro de la región de experimentación, pero, durante el desarrollo de esta se da con un algoritmo adicional que produce una región que puede tener una pequeña parte fuera de dicha región. Y por tanto se decide también estudiarlo.

Se busca que sea L – Símplex pues en un diseño con esta forma se espera una menor cantidad de corridas experimentales que en el DdE con restricciones originales, ya que la geometría es más simple.

Se sabe del marco teórico que las restricciones de frontera se pueden forzar para reducir la cantidad de tratamientos experimentales, siempre que valga la pena este intento, es decir, cuando el número de componentes sea alto, el modelo a ajustar sea complejo, se esté en presencia de experimentos jerárquicos o una combinación de las condiciones mencionadas. Ejemplos de estos esfuerzos de disminución de corridas experimentales se muestran en los trabajos de Cornell (2002) y Crosier (1984).

Es, sobre todo, pero no exclusivamente, en etapas avanzadas de programas basados en la experimentación secuencial que se espera que esta estrategia de reducción de corridas experimentales cobre su mayor relevancia, pues el reducir la región a una dentro de la geometría símplex completa (el diseño sin restricciones)

permite aislar los esfuerzos de diseño y modelado, disminuyendo costos y tiempo de experimentación.

Además, se mejora la precisión de las estimaciones del modelo (Cornell, 2002), lo cual es útil siempre y cuando el DdE tenga como objetivo modelar las respuestas de interés en función de las proporciones de los componentes de la mezcla y usar dichos modelos para determinar en qué porcentaje debe participar cada uno de los ingredientes para lograr que la fórmula tenga las propiedades deseadas.

Una vez creado o diseñado el algoritmo deseado se sigue con la evaluación, mediante los criterios de optimalidad (A, D, V y G), ya que según Cornell (2002), se pueden utilizar para la selección de uno entre varios diseños. Esto se realiza con el apoyo de simulaciones en las que se varía:

- a. El grado del polinomio a estimar (es decir la cantidad de puntos necesarios para la estimación del modelo y no la obtención del modelo per se)
- b. El número total de réplicas
- c. La cantidad de componentes ( $p$ ) involucrados
- d. La cantidad y el tamaño de las restricciones superiores en cada componente ( $p$ ).

Esta estrategia de reducción de puntos de diseño busca brindar a la investigación, a la industria y a cualquier interesado/a la posibilidad de ejecutar DdE con mezclas en cualesquiera de las etapas de un programa de experimentación secuencial de forma más eficiente. Además, se brindan las nociones de las implicaciones teórico – prácticas de la utilización de ésta, incluyendo las desventajas plausibles.

Para ello y a modo de cierre se aborda el desarrollo de un caso práctico, empleado únicamente a modo de ejemplo, en el cual se aplica el algoritmo creado en al menos una parte del programa de DdE.

### 3.1 Diseño del algoritmo

Se toma como inspiración a la forma de llamar el algoritmo CONSIM (CONStrained SIMplex) de Snee (1979) y se nombra a este como CONVEXSIM, cuyo acrónimo tiene significación tanto en inglés (**CON**Vert **EX**treme vertices to **SIM**plex) como en español (**CON**vertir **V**értices **EX**tremos a **SIM**plex).

Como primer requisito, la aplicación de este algoritmo está reservada para cuando el diseño de vértices extremos esté compuesto por una o varias restricciones de frontera superior. Es decir, el tipo de diseño de vértices extremos debe generarse sin restricciones de frontera inferior ( $a_i = 0$ ) y cuando sus restricciones de frontera superior ( $b_i$ ) no cumplan con la Ecuación (11).

Con esto mencionado, se proponen dos algoritmos, uno que cumple con la promesa de esta pesquisa y otro que puede generar una región con una porción fuera de la región de experimentación original. Estos son:

- i. CONVEXSIM – I: el cual produce un diseño L – símplex que se encuentra totalmente dentro o a lo interno (I) de la región de experimentación original restringida (la de los vértices extremos).
- ii. CONVEXSIM – S: con este se produce igualmente un diseño L – Símplex, pero que no necesariamente se encuentra íntegramente dentro de la región de experimentación original, sino que una parte, por lo general menor, puede estar fuera, es decir, el diseño se encuentra superpuesto (S) al original. En ciertas condiciones este genera una región igual a la de CONVEXSIM – I.

A continuación, se muestra la secuencia de pasos para cada uno de ellos, a la vez que se ilustra su uso en un ejemplo con  $p = 3$  y uno con  $p = 4$ .

## 3.1.1 CONVEXSIM – I

La secuencia de pasos para el algoritmo CONVEXSIM – I consiste en:

- i. Revise que solo haya restricciones de frontera superiores. Es decir, que en la Ecuación (8) el valor de  $a_i = 0$ .

$$0 \leq a_i \leq x_i \leq b_i \leq 1, \quad i = 1, 2, \dots, p$$

Esto implica que  $a_i = 0$  debe ser alcanzable, para ello se obtiene  $R_i$  de la Ecuación (13) y  $R_b$  de la Ecuación (17), si  $R_i > R_b$  entonces  $a_i$  no es alcanzable y no se puede aplicar este algoritmo.

- ii. Calcule el resultado de la Ecuación (11), si no se cumple con esta condición de frontera superior, es necesario un diseño de vértices extremos:

$$\sum_{i=1}^p b_i - b_{min} \leq 1, \quad i = 1, 2, \dots, p$$

De cumplirse la ecuación anterior este algoritmo no es necesario, pues la región ya es símplex.

- iii. Determine el valor mínimo ( $b_g$ ) entre todas las restricciones impuestas sobre la frontera superior [Ecuación (45)].

$$b_g = \min \{b_i\}, \quad i = 1, 2, \dots, p \quad (45)$$

En caso de empates, seleccione  $b_g$  aleatoriamente o por conveniencia en función del contexto del experimento. Ahora, obtenga el valor de  $a_i$ ,  $i \in \{1, 2, 3, \dots, p\} \setminus \{g\}$ , empleando la Ecuación (46):

$$a_i = \frac{1 - b_g}{p - 1}, \quad i \in \{1, 2, 3, \dots, p\} \setminus \{g\} \quad (46)$$

Si  $a_i \geq b_i$ , el diseño no tendría sentido por su propia definición [Ecuación (8)]. Por tanto, se toma a  $a_i$  como la mitad de  $b_i$ , con base en la Ecuación (47):

$$a_i = \frac{b_i}{2}, \quad i \in \{1, 2, 3, \dots, p\} \setminus \{g\} \quad (47)$$

manteniendo a  $a_g$  como cero. No obstante, cabe resaltar que este puede generar restricciones inválidas.

- iv. Calcule los rangos  $R_i$  empleando la Ecuación (13), excepto para el componente seleccionado con subíndice  $g$ .

$$R_i = b_i - a_i, \quad i \in \{1, 2, 3, \dots, p\} \setminus \{g\}$$

Determine el valor mínimo de  $R_i$ , siguiendo la Ecuación (48):

$$R_{min} = \min\{R_i\}, \quad i \in \{1, 2, 3, \dots, p\} \setminus \{g\} \quad (48)$$

Seguidamente obtenga  $R_a$  con base en la Ecuación (18):

$$R_a = 1 - \sum_{i=1}^p a_i, \quad i = 1, 2, \dots, p$$

y determine  $a_g$  con la Ecuación (49):

$$a_g = \begin{cases} R_a - R_{min}, & \text{si } R_a - R_{min} > 0 \\ 0, & \text{si } R_a - R_{min} < 0 \\ 0, & \text{si } R_a - R_{min} = b_g \end{cases} \quad (49)$$

Si  $R_a - R_{min} = b_g$ , ajuste a  $a_i$  con la Ecuación (50):

$$a_i^* = a_i + \frac{R_{min}}{p-1}, \quad i \in \{1, 2, 3, \dots, p\} \setminus \{g\} \quad (50)$$

Además, si  $R_a = 0$ , el algoritmo no es aplicable.

- v. Revise y ajuste las restricciones superiores. Para ello, primero recalculé  $R_a$  y utilice la Ecuación (19), de tal forma que la nueva región sea un símplex:

$$b_i^* = a_i + R_a, \quad i = 1, 2, \dots, p$$

si  $b_i^* > b_i$  el algoritmo es CONVEXSIM – S. Posteriormente, puede verificar que las nuevas restricciones impuestas sean consistentes empleando los criterios de  $R_a$  y  $R_b$ .

- vi. Finalmente, aplique la transformación en L – pseudocomponentes  $x'_i$  de la Ecuación (9) y verifique que la región generada es un L – Símplex.

### 3.1.2 CONVEXSIM – S

De la creación de CONVESIM – I se desprende una variante que también genera una región L – Símplex, pero ésta, si bien también debe estar contenida dentro de la región de operabilidad, puede no encontrarse en su totalidad dentro de la región de experimentación, sino que una parte, generalmente menor, se puede encontrar fuera de la misma.

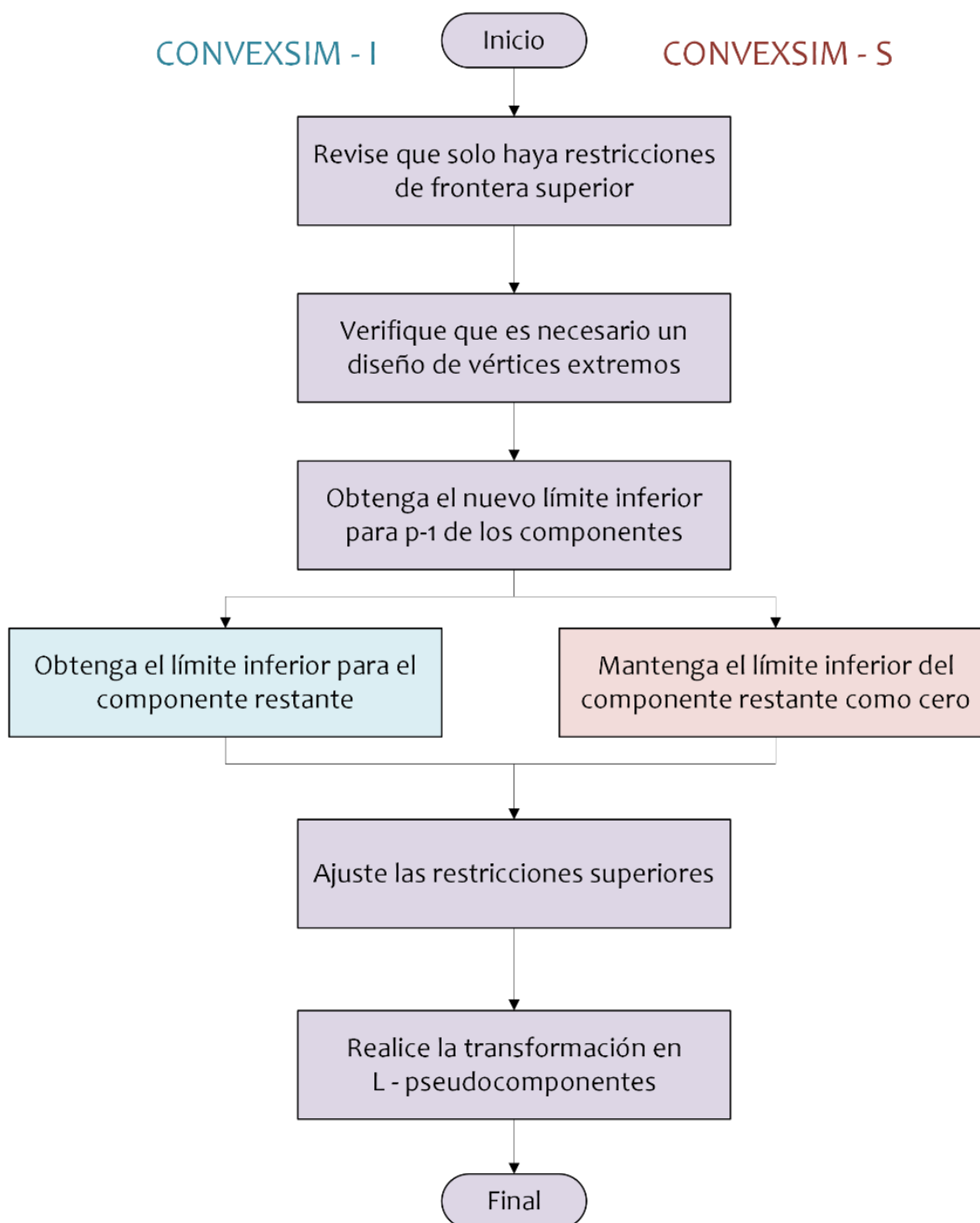
Esto se debe a que con este algoritmo no necesariamente se respetan los límites superiores establecidos, sino que para algunos componentes se estaría sobrepasando en una cantidad igual a la determinada por  $a_g$  en el paso v. Por lo que si  $a_g = 0$  ambos algoritmos generarían siempre exactamente la misma región. Esto sucede, por ejemplo, cuando solo hay una restricción sobre la frontera superior. También se da cuando al aplicar CONVEXSIM – I y se obtiene que  $b_i^* > b_i$ .

En resumen, para aplicar CONVEXSIM – S, la secuencia de pasos a seguir es la misma que en CONVEXSIM – I, exceptuando el paso iv, que se sustituye, dejando que  $a_g = 0$ , y realizando el ajuste sobre los límites superiores con la Ecuación (19):

$$b_i^* = a_i + R_a$$

Finalmente, el algoritmo concluye con la transformación en L - pseudocomponentes. De esta manera, la aplicación de ambos algoritmos se esquematiza en la Figura 3.1.

Figura 3.1. Resumen de la secuencia de pasos para ambos algoritmos



### 3.1.3 Ejemplo de aplicación de los algoritmos para $p = 3$

A continuación, se muestra un ejemplo de la aplicación de ambos algoritmos en una situación en la que  $p = 3$

#### 3.1.3.1 CONVEXSIM - I

*Paso i.*

Se utiliza un diseño propuesto por Cornell (2002), que presenta las siguientes restricciones de frontera, donde efectivamente  $a_i = 0$ :

$$x_1 \leq 0.70, \quad x_2 \leq 0.60, \quad x_3 \leq 0.80$$

*Paso ii.*

Se verifica con la Ecuación (11) que las restricciones impuestas sobre la frontera superior generan un diseño de vértices extremos:

$$\sum_{i=1}^3 b_i - b_{min} = 0.70 + 0.60 + 0.80 - 0.60 = 1.5 \geq 1$$

*Paso iii.*

Se obtiene el valor mínimo de las restricciones superiores según la Ecuación (45):

$$b_g = \min\{0.70, 0.60, 0.80\} = 0.60$$

Nótese que en este caso  $g = 2$  y está asociado a  $x_2$ . El valor de  $a_i$  para todo  $i$ , excepto para cuando  $i = g$ , se determina con la Ecuación (46):

$$a_i = \frac{1 - 0.60}{2} = \frac{0.40}{2} = 0.20$$

Los resultados obtenidos hasta este paso se tabulan en el Cuadro 3.1. Se resalta con colores los cambios introducidos, respecto al diseño original, sobre las restricciones de los componentes.

Cuadro 3.1. Resultados de la aplicación del paso iii. de CONVEXSIM – I sobre un diseño de vértices extremos con  $p = 3$

Restricción inferior ( $a_i$ )	Componente	Restricción superior ( $b_i$ )
0.20	$x_1$	0.70
0.00	$x_2$	0.60
0.20	$x_3$	0.80

Fuente: Elaboración propia a partir del diseño propuesto por Cornell (2002)

Paso iv.

Se calculan los rangos  $R_i$  con la Ecuación (13), excepto para el componente que se asocia con el subíndice  $g$ . Estos se muestran en el Cuadro 3.2.

Cuadro 3.2. Rangos ( $R_i$ ) obtenidos en el paso iv. de CONVEXSIM – I sobre un diseño de vértices extremos con  $p = 3$

Restricción inferior ( $a_i$ )	Componente	Restricción superior ( $b_i$ )	Rangos ( $R_i$ )
0.20	$x_1$	0.70	0.50
0.00	$x_2$	0.60	–
0.20	$x_3$	0.80	0.60

Fuente: Elaboración propia a partir del diseño propuesto por Cornell (2002)

El valor mínimo según la Ecuación (48) es  $R_{min} = \min\{0.50, 0.60\} = 0.50$ . Mientras que, con la Ecuación (18) se define el valor de  $R_a = 1 - (0.20 + 0.00 + 0.20) = 0.60$ . Por lo que, finalmente al aplicar la Ecuación (49), se consigue  $a_g$ .

$$a_g = 0.60 - 0.50 = 0.10$$

En el Cuadro 3.3 se muestran los resultados obtenidos de este paso, y de igual manera se resalta con color el cambio realizado respecto al Cuadro 3.2.

Cuadro 3.3. Resultados de la aplicación del paso iv. de CONVEXSIM – I sobre un diseño de vértices extremos con  $p = 3$

Restricción inferior ( $a_i$ )	Componente	Restricción superior ( $b_i$ )
0.20	$x_1$	0.70
0.10	$x_2$	0.60
0.20	$x_3$	0.80

Fuente: Elaboración propia a partir del diseño propuesto por Cornell (2002)

Paso v.

El algoritmo sigue al revisar y ajustar las restricciones superiores, se comienza recalculando  $R_a = 1 - (0.20 + 0.10 + 0.20) = 0.50$  y luego se emplea la Ecuación (19):

$$b_i^* = \begin{cases} 0.20 + 0.50 = 0.70, & \text{para } i = 1 \\ 0.10 + 0.50 = 0.60, & \text{para } i = 2 \\ 0.20 + 0.50 = 0.70, & \text{para } i = 3 \end{cases}$$

Estos valores sustituyen a las restricciones superiores anteriores. Los resultados se observan en el Cuadro 3.4, donde en todos los componentes se cumple que  $R_i < R_a$  y por tanto  $b_i$  es alcanzable. Además, se estima con la Ecuación (17) a  $R_b = (0.70 + 0.70 + 0.80) - 1 = 1.2$ , lo que cumple con  $R_i < R_b$ , haciendo también a  $a_i$  como alcanzable.

Cuadro 3.4. Resultados de la aplicación del paso v. de CONVEXSIM – I sobre un diseño de vértices extremos con  $p = 3$

Restricción inferior ( $a_i$ )	Componente	Restricción superior ( $b_i$ )	Rangos ( $R_i$ )
0.20	$x_1$	0.70	0.50
0.10	$x_2$	0.60	0.60
0.20	$x_3$	0.70	0.50

Fuente: Elaboración propia a partir del diseño propuesto por Cornell (2002)

Paso vi.

Se concluye con la aplicación del algoritmo al usar la transformación en L – Pseudocomponentes de la Ecuación (9). Se sabe que el diseño generado es un L – Símplex si  $0 \leq x'_i \leq 1$ . Esto se muestra en el Cuadro 3.5.

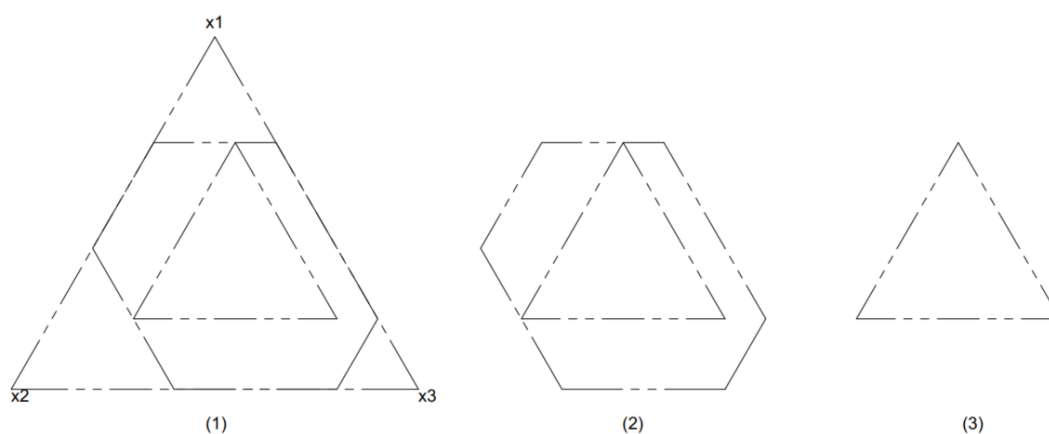
Cuadro 3.5. Transformación en pseudocomponentes del diseño generado por el algoritmo CONVEXSIM – I, sobre un diseño de vértices extremos con  $p = 3$

Restricción inferior ( $a_i$ )	Pseudocomponente	Restricción superior ( $b_i$ )
0	$x'_1$	1
0	$x'_2$	1
0	$x'_3$	1

Fuente: Elaboración propia a partir del diseño propuesto por Cornell (2002)

El resultado geométrico de la aplicación del algoritmo se presenta en la Figura 3.2 donde se muestra (1) la región de operabilidad, que es la región símplex completa, (2) la región de vértices extremos del diseño original y (3) la nueva geometría generada.

Figura 3.2. Resultados geométricos de la aplicación de CONVEXSIM – I sobre un diseño de vértices extremos con  $p = 3$



Fuente: Elaboración propia a partir del diseño propuesto por Cornell (2002)

## 3.1.3.2 CONVEXSIM – S

Se toman los resultados obtenidos hasta el paso iii, mostrados en el Cuadro 3.1. Como ya se ha especificado, en CONVEXSIM – S se sustituye el paso iv y se realiza el ajuste en la frontera superior con la Ecuación (19). Se obtiene el valor de  $R_a = 1 - (0.20 + 0.00 + 0.20) = 0.60$  y se agrega a todos los valores de  $a_i$  produciendo los resultados del Cuadro 3.6. Se resaltan los cambios en comparación con los obtenidos con CONVEXIM – I del Cuadro 3.4.

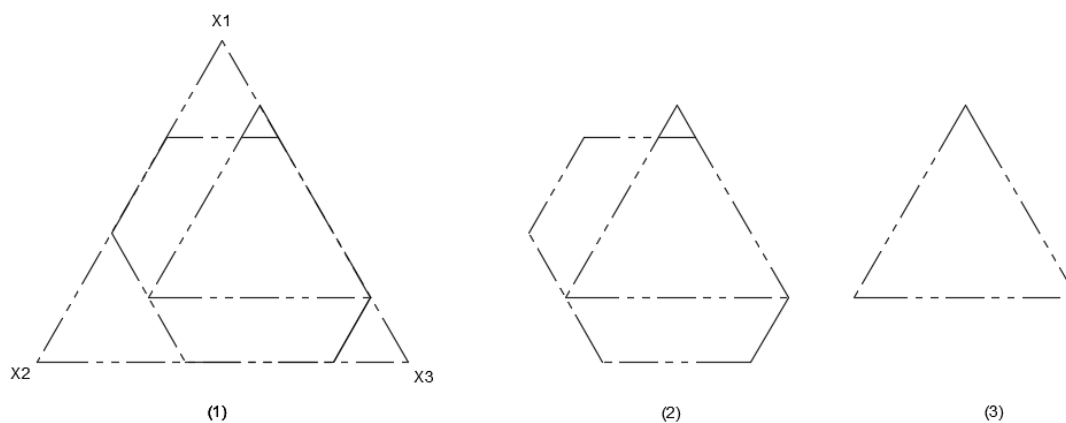
Cuadro 3.6. Resultados de la aplicación de CONVEXSIM – S sobre un diseño de vértices extremos con  $p = 3$

Restricción inferior ( $a_i$ )	Componente	Restricción superior ( $b_i$ )
0.20	$x_1$	0.80
0.00	$x_2$	0.60
0.20	$x_3$	0.80

Fuente: Elaboración propia a partir del diseño propuesto por Cornell (2002)

La geometría resultante se muestra en la Figura 3.3, donde es apreciable (1) la región símplex completa, (2) la superposición del diseño símplex sobre el diseño original generado por las restricciones de frontera y (3) el diseño símplex creado.

Figura 3.3. Resultados geométricos de la aplicación de CONVEXSIM – S sobre un diseño de vértices extremos con  $p = 3$



Fuente: Elaboración propia a partir del diseño propuesto por Cornell (2002)

Además, la transformación en L – pseudocomponentes genera un diseño que cumple con la condición  $0 \leq x'_i \leq 1$ .

### 3.1.4 Ejemplo de aplicación de los algoritmos para $p = 4$

Este ejemplo es desarrollado por Cornell (2002), donde se ha formulado una bebida tropical con sandía ( $x_1$ ), naranja ( $x_2$ ), piña ( $x_3$ ) y uva ( $x_4$ ), con la restricción de que no más del 80 % de la bebida sea de sandía, debido a los costos de adquisición de ésta, de tal forma que las restricciones son las siguientes:

$$0.00 \leq x_1 \leq 0.80, \quad 0.00 \leq x_2 \leq 1.00, \quad 0.00 \leq x_3 \leq 1.00, \quad 0.00 \leq x_4 \leq 1.00$$

#### 3.1.4.1 CONVEXSIM – I

Se verifica que el diseño no cuenta con restricciones inferiores, es decir  $a_i = 0$ . A la vez, se comprueba que es necesario un diseño de vértices extremos.

$$\sum_{i=1}^4 b_i - b_{min} = 0.80 + 1.00 + 1.00 + 1.00 - 0.80 = 3 \geq 1$$

El valor mínimo de las restricciones superiores se determina como  $b_g = \min\{0.80, 1.00, 1.00, 1.00\} = 0.80$ , por lo que en este caso  $g = 1$ . Luego se obtienen y sustituyen los límites inferiores para todo  $i \neq g$ .

$$a_i = \frac{1 - 0.80}{3} = \frac{0.20}{3} = 0.067$$

Se obtiene  $R_{min} = \min\{0.933, 0.933, 0.933\} = 0.933$ , obviando el valor asociado a  $g$ , y se define  $R_a = 1 - (0.00 + 0.067 + 0.067 + 0.067) = 0.80$ , con lo que se calcula la frontera inferior para el componente asociado con  $g$ :

$$a_g = 0.80 - 0.933 = -0.133 \rightarrow a_g = 0$$

De esta forma se recalcula  $R_a = 0.80$  y se utiliza para ajustar las restricciones superiores:

$$b_i^* = \begin{cases} 0.000 + 0.800 = 0.800, & \text{para } i = 1 \\ 0.067 + 0.800 = 0.867, & \text{para } i = 2 \\ 0.067 + 0.800 = 0.867, & \text{para } i = 3 \\ 0.067 + 0.800 = 0.867, & \text{para } i = 4 \end{cases}$$

El resultado de la aplicación de CONVEXSIM – I se muestra en el Cuadro 3.7.

Cuadro 3.7. Resultados de la aplicación de CONVEXSIM – I sobre un diseño de vértices extremos con  $p = 4$

Restricción inferior ( $a_i$ )	Componente	Restricción superior ( $b_i$ )
0.000	$x_1$	0.800
0.067	$x_2$	0.867
0.067	$x_3$	0.867
0.067	$x_4$	0.867

Fuente: Elaboración propia a partir del diseño propuesto por Cornell (2002)

#### 3.1.4.2 CONVEXSIM – S

Este algoritmo produce el mismo resultado anterior (Cuadro 3.7), en tanto que  $a_g = 0$  y por ende el cálculo de  $R_a$  es equivalente al de CONVEXSIM – I. La transformación en L – pseudocomponentes, para el resultado obtenido de ambos algoritmos, se muestra en el Cuadro 3.8.

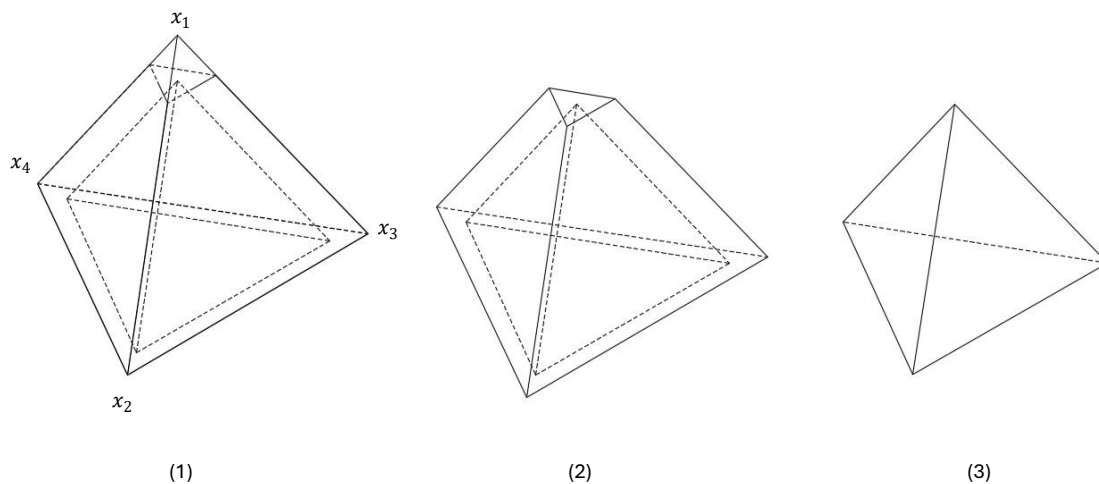
Cuadro 3.8. Transformación en pseudocomponentes del diseño generado por el algoritmo CONVEXSIM – I y CONVEXSIM – S, sobre un diseño de vértices extremos con  $p = 4$

Restricción inferior ( $a_i$ )	Pseudocomponente	Restricción superior ( $b_i$ )
0	$x'_1$	1
0	$x'_2$	1
0	$x'_3$	1
0	$x'_4$	1

Fuente: Elaboración propia a partir del diseño propuesto por Cornell (2002)

Ambos algoritmos generan la misma geometría, la cual se muestra en la Figura 3.4, que representa (1) la región de operabilidad completa, (2) el diseño de vértices extremos y (3) el diseño generado por CONVEXSIM.

Figura 3.4. Resultados geométricos de la aplicación de CONVEXSIM – I o CONVEXSIM – S sobre un diseño de vértices extremos con  $p = 4$



Fuente: Elaboración propia a partir del diseño propuesto por Cornell (2002)

### 3.2 Diseño de la simulación

La simulación está compuesta por diversos escenarios con los que se busca comparar el diseño generado por las restricciones transformadas por CONVEXSIM – I y CONVEXSIM – S contra el diseño de vértices extremos del que deviene, y a su vez con un diseño símplex reticular en la región de operabilidad completa.

Estos escenarios son creados al variar la cantidad de componentes ( $p$ ), la cantidad y el tamaño de las restricciones sobre la frontera superior ( $b_i$ ), el grado del polinomio a estimar ( $m$ ), la presencia o ausencia de puntos axiales y, finalmente, la cantidad total de réplicas.

En la Tabla 3.1 se muestra el rango en el que se varían dichos parámetros, así como la descripción de la variable y la razón de la inclusión en el diseño de la simulación.

Tabla 3.1. Resumen, rango de variación y descripción de las variables de entrada utilizadas en el diseño del estudio de simulación

Variables	Rango de variación	Detalles
Cantidad de componentes ( $p$ )	Entre 3 y 12 componentes, ambos inclusive.	<p>La dimensión de un diseño es una variable que depende de la cantidad de parámetros:</p> $\text{Dimensión} = p - 1$ <p>Además, los experimentos con mezclas solo se definen para <math>p \geq 2</math>. Cuando <math>p = 2</math> la región producida es una línea recta, por lo que, sin importar las restricciones impuestas, su geometría no se ve afectada.</p> <p>Por otro lado, <math>p = 12</math> es el máximo valor posible que se puede ingresar en la función <math>Xvert^{15}</math> de <math>\{mixexp\}</math> (Lawson &amp; Willden, 2016).</p> <p>Finalmente, de las Ecuaciones (6) y (7) se desprende que la cantidad de corridas experimentales es una función que depende de <math>p</math>.</p>
Cantidad de restricciones sobre la frontera superior	Entre 1 y $p$ , ambos inclusive	Se debe tomar en cuenta que los algoritmos CONVEXSIM solo son aplicables cuando la geometría generada por las restricciones no es símplex, por lo que se excluye del diseño de la simulación aquellos diseños que no sean de vértices extremos. Además, es imperativo que los diseños tengan sentido, porque lo que solo se utilizan diseños con restricciones alcanzables.
Tamaño de las restricciones sobre la	Entre 0.1 y 1.0, ambos inclusive. El valor incremental en	

<sup>15</sup> Esta función crea un diseño de vértices extremos en un espacio de mezclas restringido (Lawson & Willden, 2016). Esta llama a la función *crvtave* que es el algoritmo CONVRT de Piepel (1988).

Tabla 3.1. Resumen, rango de variación y descripción de las variables de entrada utilizadas en el diseño del estudio de simulación

Variables	Rango de variación	Detalles
frontera superior	la secuencia entre 0.1 y 1 es variable en función de $p$ : De 3 $\rightarrow$ 0.05 De 4 a 5 $\rightarrow$ 0.10 De 6 $\rightarrow$ 0.15 De 7 $\rightarrow$ 0.25 De 8 a 12 $\rightarrow$ 0.40	La cantidad y el tamaño de las restricciones afectan la geometría del diseño generado y se hipotetiza que formas más complejas o irregulares (es decir, con más restricciones) dan a lugar a diseños con mayor cantidad de tratamientos.  Por otro lado, el valor incremental en la secuencia de 0.05 a 1 en el tamaño de las restricciones es variable ya que la cantidad de diseños generados crece a una tasa exponencial y se agota la capacidad computacional. Los casos son generados mediante combinaciones.
Grado del polinomio ( $m$ ) máximo que se podría estimar.	Entre 1 y 3, ambos inclusive	Como punto de comparación se emplean diseños simplex reticular o lattice, cuya función en $\{mixexp\}$ es $SLD^{16}$ , pues estos brindan la posibilidad de escoger la cantidad de puntos de diseño que son necesarios para estimar un modelo particular, aumentando el grado del retículo. Esto no ocurre con el diseño simplex centroide ( $SCD^{17}$ en $\{mixexp\}$ ) que el máximo modelo a estimar es fijo en función de $p$ .

<sup>16</sup> Esta función crea diseños simplex lattice en espacios de mezclas no restringidos (Lawson & Willden, 2016).

<sup>17</sup> Esta función crea diseños simplex centroide en espacios de mezclas no restringidos (Lawson & Willden, 2016).

Tabla 3.1. Resumen, rango de variación y descripción de las variables de entrada utilizadas en el diseño del estudio de simulación

Variables	Rango de variación	Detalles
Puntos axiales	Con o sin puntos axiales	<p>Por otro lado, <i>Xvert</i> permite lo mismo que <i>SLD</i> (genera puntos en el retículo) y genera diseños equivalentes hasta <math>m = 2</math>, además, esta función no recibe valores mayores, a diferencia de <i>SLD</i> que permite <math>m = 5</math>.</p> <p>La inclusión de puntos axiales aumenta en un grado el polinomio máximo que se puede estimar, es decir, si <math>m = 1</math> se puede estimar un modelo lineal, y al agregar los puntos axiales se podría estimar un modelo cuadrático.</p> <p>Además, los diseños ampliados con puntos axiales son siempre recomendados para eludir la crítica usual de que los puntos de diseño solo se encuentran sobre los límites de la región de experimentación; y al incluir estos tratamientos, indefectiblemente se aumenta la cantidad de corridas experimentales.</p>
Total de réplicas	Entre 1 y 4, ambos inclusive	<p>Según Montgomery &amp; Voth (1994) la replicación reduce los valores de apalancamiento, por lo que se recomienda el uso de estas en los puntos que tengan valores altos.</p> <p>Se incluye esta variable para estudiar si su uso generalizado (es decir, en todos los puntos de diseño) tiene un efecto sobre los valores del</p>

Tabla 3.1. Resumen, rango de variación y descripción de las variables de entrada utilizadas en el diseño del estudio de simulación

Variables	Rango de variación	Detalles
		apalancamiento y los criterios de optimalidad y eficiencia.

Una vez definidas las variables de entrada que se varían en esta simulación, en la Tabla 3.2 se describen las variables de salida que se emplean para medir la pertinencia de los diseños generados por CONVEXSIM.

Tabla 3.2. Variables de salida que son empleadas para medir la eficacia y la eficiencia de los diseños generados por CONVEXSIM

Medida	Descripción
Cantidad total de corridas experimentales	<p>Es la meta principal de esta investigación el reducir la cantidad de tratamientos y corridas experimentales, procurando un diseño eficiente y eficaz.</p> <p>Esta medición se compara contra el diseño de vértices extremos que da origen a CONVEXSIM, y a su vez que con el diseño símplex reticular de la región de operabilidad mediante tasa de la Ecuación (51). Donde CE es el acrónimo de Corridas Experimentales:</p> $T_n = \frac{CE_{Diseño\ original} - CE_{CONVEXSIM}}{CE_{Diseño\ original}} \quad (51)$ <p>El diseño original es uno de los siguientes: el diseño de vértices extremos o el diseño símplex de la región completa. La hipótesis de trabajo es que cuando la comparación se hace con el primero se esperan valores positivos y que entre más altos sean, mejor. Cuando se</p>

Tabla 3.2. Variables de salida que son empleadas para medir la eficacia y la eficiencia de los diseños generados por CONVEXSIM

Medida	Descripción
	<p>aplique la ecuación con el segundo diseño, se espera un valor de cero, pues el diseño CONVEXSIM debe generar la misma cantidad de corridas (no los mismos puntos de diseño) que el símplex reticular de la región de operabilidad, excepto cuando <math>m = 3</math>, en este caso se esperan valores positivos.</p>
<p>Distancia entre centroides</p>	<p>Cornell (2002) indica que en los casos en los que la región de experimentación es una región restringida dentro de la región <math>p</math> – símplex, la localización del centroide en relación con el centroide global también tendrá un efecto en los valores de optimalidad.</p> <p>La diferencia entre la localización de los centroides se realiza con la distancia euclidiana [Ecuación (52)]:</p> $d_E = \sqrt{\sum_{i=1}^p (c_{Diseño\ original_i} - c_{CONVEXSIM_i})^2} \quad (52)$ <p>La conceptualización del “Diseño original” sigue la misma idea anterior.</p> <p>En este caso no se espera un valor particular, sino que se busca evaluar o determinar si existe variación en los criterios de optimalidad en función de la distancia entre centroides.</p>
<p>Tamaño de la región de experimentación</p>	<p>Una consecuencia de esta pesquisa es que, al reducir los puntos de diseño, se reduce la región de experimentación.</p>

Tabla 3.2. Variables de salida que son empleadas para medir la eficacia y la eficiencia de los diseños generados por CONVEXSIM

Medida	Descripción
	<p>En función del contexto del experimento particular, esto puede ser una desventaja, por ejemplo: el óptimo global puede encontrarse dentro de la región de experimentación original pero fuera de la nueva región y por tanto este valor podría no ser localizado.</p> <p>Para evaluar la reducción de la región de experimentación se propone medir el promedio de los rangos <math>R_i</math> calculados con la Ecuación (13), para los <math>p</math> componentes, siguiendo la Ecuación (53):</p> $\bar{R}_i = \frac{\sum_{i=1}^p R_i}{p} \quad (53)$ <p>Estos rangos se comparan entre el diseño de vértices extremos y el diseño generado por CONVEXSIM. Entre menor sea la diferencia entre ellos, mejor.</p> <p>Por otro lado, Cornell (2002) indica que los valores de optimalidad disminuyen a medida que los rangos de los componentes se amplían y se desea evaluar esta afirmación.</p>
<p>Criterios de optimalidad y de eficiencia</p>	<p>Estos criterios (A, D, G y V) se utilizan para la selección de diseños. La conveniencia de un diseño aumenta conforme los valores de estos criterios disminuyan.</p> <p>El orden de las corridas experimentales se mantiene como una variable controlada, debido a que los valores de A – Optimalidad y D – Optimalidad pueden variar en función de estas variables (ver Cuadro 2.1 para más detalles). El orden</p>

Tabla 3.2. Variables de salida que son empleadas para medir la eficacia y la eficiencia de los diseños generados por CONVEXSIM

Medida	Descripción
	<p>empleado para cada réplica es el siguiente: vértices, puntos en la frontera, centroide global y axiales.</p> <p>El cálculo de estos criterios se realiza tanto en proporciones como en pseudocomponentes, para el diseño símplex de la región de operabilidad completa, el diseño de vértices extremos y el diseño L – Símplex generado por CONVEXSIM (I o S).</p> <p>Además, se incluye el cálculo de los criterios de eficiencia asociados, tomando en cuenta la cantidad de términos en el modelo (<math>q</math>) y el número de corridas necesarias (<math>n</math>).</p>

Para realizar el análisis de resultados se hace uso de la estadística descriptiva y la visualización de datos para extraer conclusiones y recomendaciones con base en el comportamiento observado en la simulación.

### 3.3 Caso práctico

El propósito de este caso práctico es únicamente ilustrar la aplicación de CONVEXSIM en un DdE real; con énfasis en la reducción de corridas experimentales. No se pretende expandir o desarrollar teoría en relación con el ejemplo utilizado, aunque si se presenta un pequeño abordaje teórico que explica la motivación para la inclusión de algunos ingredientes en el diseño experimental y que da contexto a la persona lectora.

#### 3.3.1 Contexto del caso práctico

Durante el III Ciclo del año 2023 en el curso II-1117 Diseño de Experimentos Avanzado de la Carrera de Ingeniería Industrial (CII) de la Universidad de Costa Rica

(UCR) impartido en la Sede Interuniversitaria de Alajuela (SIUA), fue asignado por la persona docente<sup>18</sup> como parte de los rubros de evaluación del curso, el diseño y confección de un producto cosmético con ingredientes naturales empleando el DdE. Contextualmente, a las personas estudiantes se les fue brindada esta información, que se presenta a continuación de forma resumida:

Un producto cosmético es una preparación que se aplica en diversas partes del cuerpo, como la piel, el cabello, los labios, las uñas, con el fin de limpiar, perfumar, proteger, alterar su apariencia o corregir olores corporales. Puede estar compuesto por ingredientes naturales o sintéticos y suele someterse a pruebas de calidad y seguridad. Ejemplos de productos cosméticos incluyen jabones, champús, tintes para el cabello, cremas faciales, entre otros (Arroyo Figueroa et al., 2016; Farfán et al., 2022). Su comercialización puede basarse en componentes naturales, lo que atrae a un segmento específico de consumidores (Flores Martínez, 2019). Además, se dice que los productos cosméticos naturales ofrecen una serie de beneficios para la piel y el cuidado personal, incluyendo propiedades antibacterianas, antifúngicas, antioxidantes y nutrientes, así como el cuidado sostenible y la personalización según las necesidades individuales (Farfán et al., 2022; Garcillán López Rúa, 2014; Gómez-Fierro et al., 2022; Quizphe Sagasti, 2018).

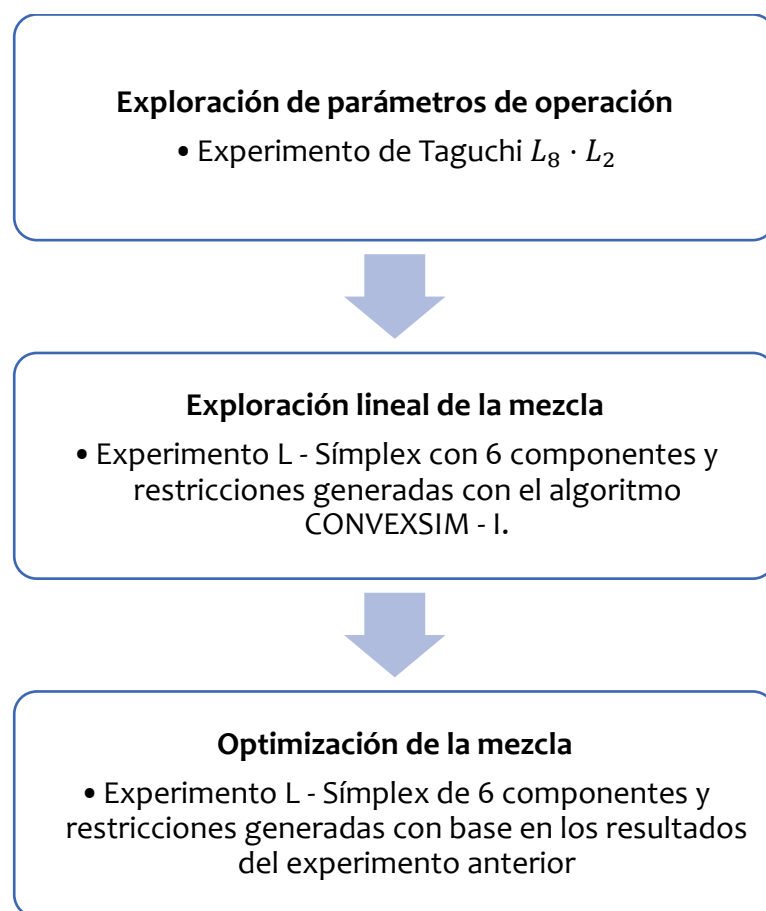
A las personas estudiantes, en los respectivos subgrupos de trabajo, se les solicita el desarrollo de un producto cosmético de libre elección, que tenga como mínimo cuatro ingredientes, entre los cuales, al menos uno debe ser de origen natural. Este debe ser funcional, por lo que no debe contener agentes perjudiciales para la salud y, además, ninguno de los ingredientes se debe encontrar presente en la formulación final en más de un 75 %.

---

<sup>18</sup> El autor de este documento

Con esta información, un subgrupo de estudiantes<sup>19</sup> realiza una propuesta en la que se emplean ingredientes sub – utilizados, donde buscan diseñar una receta de un cosmético que contenga ingredientes naturales e inusuales. Para ello se emplean subproductos del trigo (germen y salvado) como ingredientes exóticos en la elaboración de un exfoliante que además contiene miel, azúcar, aceite y jabón. Para lograr este producto, plantean un programa de experimentación que está compuesto por los siguientes pasos (Figura 3.5):

Figura 3.5. Esquematación del programa de experimentación para la producción de un exfoliante con subproductos de trigo



<sup>19</sup> Conformado por Francisco Javier Garro Charpentier, Andrés González Fuentes, Ana Paula Madrigal Wachong, Allison Paola Mendoza Rodríguez, Luis Carlos Mesén Jiménez y Fabiola Sandoval Atán; a quienes agradezco su tiempo y dedicación con el curso para la obtención de los datos que son utilizados en esta tesis como caso de aplicación.

Como se aprecia, es en el segundo experimento en el que se aplica la metodología diseñada y por tanto donde se centra el desarrollo del caso de aplicación. Este tiene como objetivo: “Determinar la proporción de los componentes en el exfoliante que mejoren de forma simultánea las variables de respuesta, para el establecimiento de una mezcla adecuada a las características de calidad esperadas del producto”.

### 3.3.2 Componentes de la mezcla

Son los ingredientes que se investigan en el experimento para observar cómo afectan o influyen en la variable de respuesta (Gutiérrez Pulido & de la Vara Salazar, 2012).

- i. Azúcar blanca ( $x_1 \leq 0.75$ ): los cristales de azúcar actúan como partículas exfoliantes y eliminan las células muertas y las impurezas de la piel (Carranza Chusán, 2019).
- ii. Aceite de coco ( $x_2 \leq 0.75$ ): el aceite puede afectar la consistencia y la sensación exfoliante sobre la piel. Además, evita que la piel se reseque a pesar del proceso de limpieza profunda, sin dejarla grasosa o aceitosa (Carranza Chusán, 2019).
- iii. Salvado de trigo ( $x_3 \leq 0.75$ ): también actúa como exfoliante, se aconseja que las partículas tengan una dureza media con bordes sin aristas para evitar la irritación.
- iv. Germen de trigo ( $x_4 \leq 0.75$ ): es la parte más nutritiva del trigo, e igual que con el componente anterior ( $x_3$ ), puede afectar la textura y sensación del exfoliante.
- v. Miel de abeja ( $x_5 \leq 0.75$ ): recomendable por sus propiedades hidratantes y capacidad antimicrobiana, antioxidantes y antiinflamatorias (González-Pérez et al., 2024).

- vi. Jabón ( $x_6 \leq 0.75$ ): aporta propiedades exfoliantes adicionales al producto y ayuda a limpiar la piel con profundidad (Espino Pérez et al., 2019).

### 3.3.3 Variables controladas

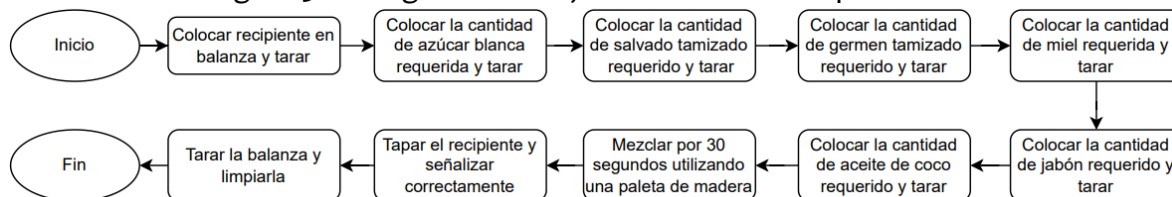
- Son variables de proceso o características de los materiales que se pueden fijar en un nivel dado (Gutiérrez Pulido & de la Vara Salazar, 2012). Es con el experimento Taguchi  $L_8 \cdot L_2$ , de la Figura 3.5, que fueron determinadas las variables de operación necesarias para la confección del producto, entre las cuales se destacan (Tabla 3.3):

Tabla 3.3. Variables de operación determinadas a partir del experimento de Taguchi

<b>Variable de operación</b>	<b>Resultado</b>
Tipo de azúcar	Blanca
Tipo de aceite	Coco
Tipo de salvado	Tamizado
Tipo de germen	Tamizado
Tiempo de mezclado	30 segundos
Instrumento de mezclado	Paleta de madera
Orden de mezclado	Sólidos primero

Esta es la combinación de factores y niveles más beneficiosa para las variables de respuesta. Estas variables, si bien no forman parte de la mezcla, si pueden afectar las propiedades de la combinación de esta y por tanto se mantienen controladas durante la ejecución del experimento, así como el orden en el que se añaden los ingredientes sólidos y líquidos. En el diagrama de flujo de la Figura 3.6 se representa de forma visual el proceder de cada corrida.

Figura 3.6. Diagrama de flujo de una corrida experimental



Además, bajo el supuesto general de los experimentos con mezclas, el total de exfoliante a producir en cada corrida se mantiene constante en 50 g.

### 3.3.4 Variables no controladas

Son variables que no se pueden controlar durante el experimento o bien, durante la operación normal del proceso (Gutiérrez Pulido & de la Vara Salazar, 2012). Entre ellas se incluyen:

- i. Condiciones ambientales: está fuera del alcance del experimento poder controlar las condiciones ambientales en su totalidad, ya que, si bien las condiciones de temperatura pueden ajustarse mediante el uso del aire acondicionado, esta es susceptible a fallas y variación. Además, no se pueden regular la humedad, la iluminación, entre otras condiciones ambientales. No obstante, se realizan acciones para mitigar el impacto de esta variable, como mantener la habitación cerrada y el aire acondicionado en funcionamiento durante la ejecución del experimento.
- ii. Estado de ánimo de los colaboradores: se considera fuera del alcance el controlar el estado de ánimo de las personas involucradas en la realización del experimento, puesto que, no es posible determinar o condicionar esta variable.
- iii. Estado de la piel de las personas: puesto que el producto a realizar debe ser probado por distintas personas, no se puede asegurar que estas van a tener condiciones de piel equivalentes.

- iv. Rendimiento de los instrumentos electrónicos: al trabajar con instrumentos electrónicos estos son susceptible a fallar o disminuir su rendimiento en cualquier momento debido a cambios en el voltaje o variaciones propias de su funcionamiento.
- v. Calidad de la materia prima: aunque se emplea producto del mismo proveedor, está fuera de control que la calidad del producto durante todo el programa de experimentación sea constante.

### 3.3.5 Variables de respuesta

A través de estas variables se conoce el efecto o los resultados de cada prueba experimental, por lo que pueden ser características de calidad de un producto o variables que miden el desempeño de un proceso. El objetivo de muchos estudios experimentales es encontrar la forma de mejorar las variables de respuesta (Gutiérrez Pulido & de la Vara Salazar, 2012).

Para este experimento, el de exploración lineal de la mezcla en la Figura 3.5, se miden de forma simultánea cinco variables de respuesta, de las cuales se analizan en este caso de aplicación solamente dos de ellas a modo de ejemplo<sup>20</sup>:

- i. Densidad: Para medir la densidad del exfoliante se emplea la fórmula que la compone, es decir masa entre volumen (g/mL). En este caso se espera que la densidad de un producto para exfoliar la piel se encuentre entre 0.9 y 1.01, por lo que el objetivo de optimización es nominal para un valor de 1.00. Para esto, se toma una muestra de 20 mL en cada corrida en un recipiente graduado para volumen; para la masa se utiliza una balanza con resolución de 1 gramo, en donde, primero se toma la masa del recipiente vacío y luego del recipiente lleno de exfoliante y se restan para obtener la

---

<sup>20</sup> El análisis y explicación de los gráficos de contorno puede ser excesivamente extenso y no es per se el objetivo de esta investigación, sino que, lo que se busca es ilustrar el uso de CONVEXSIM en un ejemplo real.

masa del contenido. Además, el recipiente en donde se realizan las muestras se limpia luego cada corrida ejecutada para que el sobrante de la corrida anterior no influya el resultado en las demás corridas.

- ii. Oleosidad: se refiere a la cualidad que tiene un objeto o fluido de ser aceitoso. Es de interés para este experimento medir esta variable de respuesta, puesto que se desea que se perciba al exfoliante como áspero, pero no en exceso para evitar incomodidades al aplicarse el producto. Para medir la oleosidad, se le aplica una muestra de cada corrida a tres personas en la mano y se utiliza una escala del 1 al 5, donde se evalúa ¿Qué tan oleoso (grasoso) es el exfoliante?, siendo 1, muy poco oleoso y 5, bastante oleoso. Se desea un valor nominal, en 3, para esta variable de respuesta. La variable de respuesta es el promedio de la calificación brindada por las tres personas.

### 3.3.6 Selección del diseño experimental

Acorde al objetivo de este experimento se selecciona un diseño experimental con mezclas. Dicho experimento resulta especialmente útil ya que las variables de respuesta que se le miden al exfoliante dependen de las proporciones de los diferentes componentes incluidos en la mezcla. Se pretende un diseño mediante el cual se pueda estimar un modelo lineal, ya que este es útil cuando el objetivo es la exploración de la región; el cual se utiliza en las primeras fases de una investigación, sobre todo cuando se tienen muchos componentes y se desea revelar los más importantes (Ortega-Pérez et al., 2015). Además, se decide replicar dos veces y agregar puntos centrales, por lo ya estudiado en el marco teórico.

## 3.4 Herramientas de generación y análisis de los datos

Tanto para el estudio de simulación como para el caso práctico se hace uso del lenguaje de programación R, versión 4.1.1 “Race for Your Life” (R Core Team, 2024) en el Entorno de Desarrollo Integrado (IDE) RStudio, versión 2023.6.0.421 “Mountain

Hydrangea” (Posit team, 2023), en conjunto con los siguientes paquetes: `{tidyverse}` versión 2.0.0 (Wickham et al., 2019) que es empleado para el tratamiento de los datos, el diseño de la simulación y la confección de gráficos, `{mixexp}` versión 1.2.7 (Lawson & Willden, 2016) con sus dependencias `{daewr}` versión 1.2.11 (Lawson & Krennrich, 2023) y `{lattice}` versión 0.22-6 (Sarkar, 2008) que se emplean para el diseño y análisis de experimentos con mezclas, así como para la representación de los gráficos de contorno.

Además, el paquete `{desirability}` versión 2.1 (Kuhn, 2016) con la cual se realiza la función de deseabilidad, `{car}` versión 3.1-2 (Fox & Weisberg, 2019) para el cálculo de estadísticos como el factor de inflación de la variancia, `{gtools}` versión 3.9.5 (Warnes et al., 2023) que genera combinaciones y permutaciones, `{gridExtra}` versión 2.3 (Auguie, 2017) para realizar gráficos conjuntos, `{matlib}` versión 0.9.8 (Friendly et al., 2024) con el que se realizan operaciones matriciales para los criterios de optimalidad y eficiencia y `{openxlsx}` versión 4.6.2.1 (Schauberger & Walker, 2024) para la carga de los datos del caso práctico en formato xlsx.

## 4. RESULTADOS DEL ESTUDIO DE SIMULACIÓN

Para el logro del segundo objetivo específico se realiza un estudio de simulación determinística, con el cual se compara la eficacia y la eficiencia de los diseños en diferentes situaciones que incluyen variaciones en la cantidad de componentes ( $p$ ), la cantidad y el tamaño de las restricciones sobre la frontera superior, el grado máximo del polinomio (en función de  $m$ ) que se puede estimar, incluyendo la presencia o no de puntos axiales y el total de réplicas del experimento.

Los códigos empleados en el diseño de esta simulación y el análisis de esta, así como algunos de los resultados obtenidos, se encuentran disponibles en Zenodo (García-Goñi, 2025).

### 4.1 Detalle de la simulación

Como primer paso se crean las restricciones de frontera superior que generan diseños de vértices extremos para  $3 \leq p \leq 12$ , donde a su vez se varían la cantidad y el rango de las restricciones. Esto se consigue mediante las combinaciones de los niveles que se desean probar para cada cantidad de componentes.

Luego se eliminan aquellas restricciones que produzcan diseños símplex o L – símplex y también las que contienen puntos de diseño que no son alcanzables, de tal forma que se producen 4 132 diseños válidos, segregados como se observa en el Cuadro 4.1.

Cuadro 4.1. Cantidad de restricciones generadas para cada número de componentes variados en la simulación

Cantidad de componentes	Rango de variación <sup>1</sup>	Restricciones válidas generadas
3	Cada 0.05	493
4	Cada 0.10	548
5	Cada 0.15	415
6	Cada 0.15	884
7	Cada 0.25	320
8	Cada 0.40	160
9	Cada 0.40	215
10	Cada 0.40	281
11	Cada 0.40	362
12	Cada 0.40	454

<sup>1</sup> Entre 0.1 y 1.0, ambos inclusive.

Luego se obtienen las nuevas restricciones inferiores y superiores devenidas de aplicar CONVEXSIM – I y CONVEXSIM – S. Del abordaje metodológico se sabe que ambos algoritmos pueden generar diseños inscritos (I) o superpuestos (S) a la región, que son equivalentes entre sí. Dada esta condición, se filtran solo aquellos casos que sean exclusivos de cada uno de ellos. La cantidad de restricciones válidas generadas se muestran en el Cuadro 4.2.

Cuadro 4.2. Cantidad de restricciones inferiores y superiores válidas que fueron generadas por CONVEXSIM – I y CONVEXSIM – S

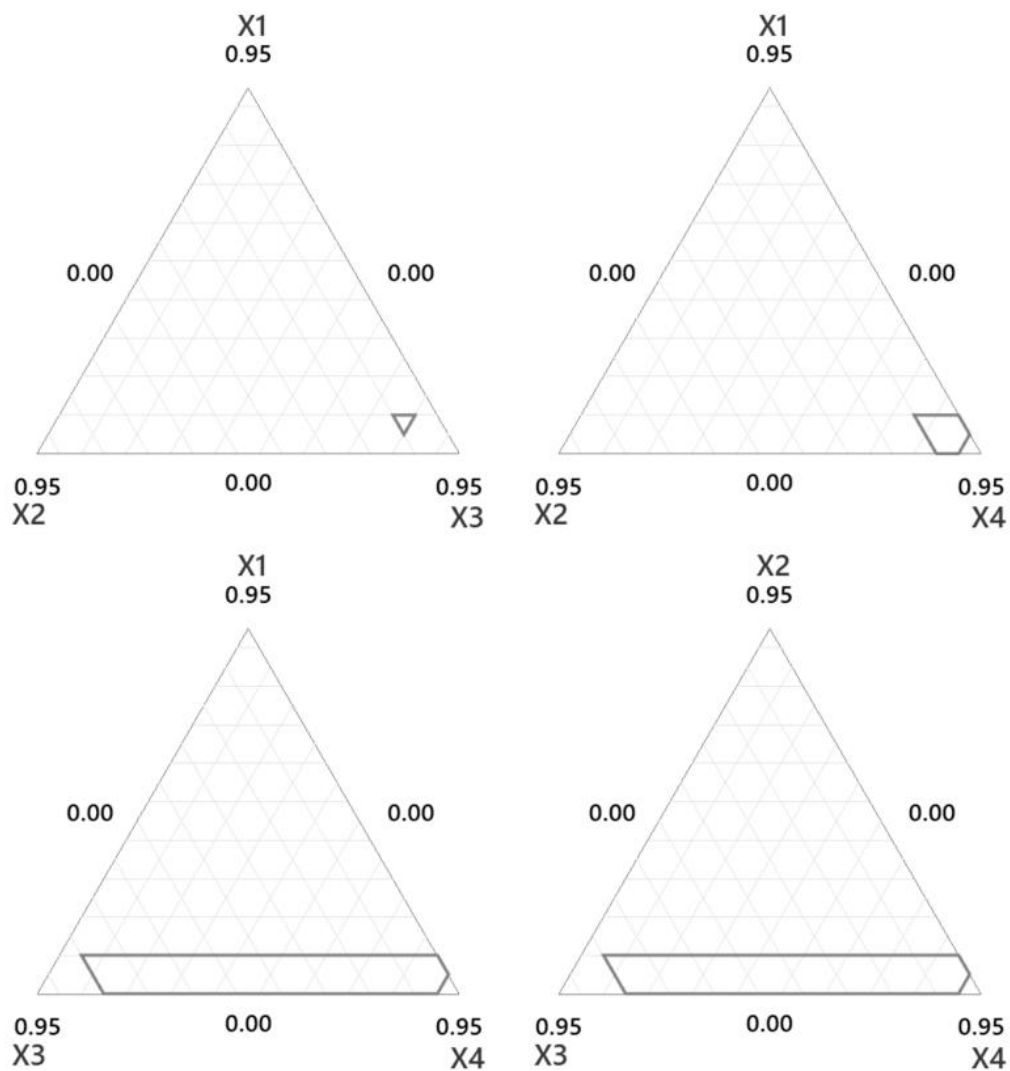
Cantidad de componentes	Cantidad de restricciones originales	Restricciones válidas generadas	
		CONVEXSIM – I	CONVEXSIM – S
3	493	493	248
4	548	518	279
5	415	375	218
6	884	724	402
7	320	208	92
8	160	82	36
9	215	99	45
10	281	120	55
11	362	362	284
12	454	454	363
<b>Total</b>	<b>4 132</b>	<b>3 435</b>	<b>2 022</b>

La diferencia encontrada entre la cantidad de original y de las nuevas restricciones se debe a dos motivos:

- i. Se producen restricciones superpuestas en el caso de CONVEXSIM – I o viceversa, es decir, inscritas en CONVEXSIM – S.
- ii. En el tercer paso de CONVEXSIM, si  $a_i \geq b_i$  la región de experimentación es particularmente reducida en función de una o varias restricciones de frontera superior y esto genera que algunas restricciones creadas por los algoritmos CONVEXSIM no sean válidas por alguna de las siguientes razones: 1) al menos una de las restricciones inferiores es igual a las superiores o 2) las restricciones inferiores son menores a cero ( $a_i < 0$ ).

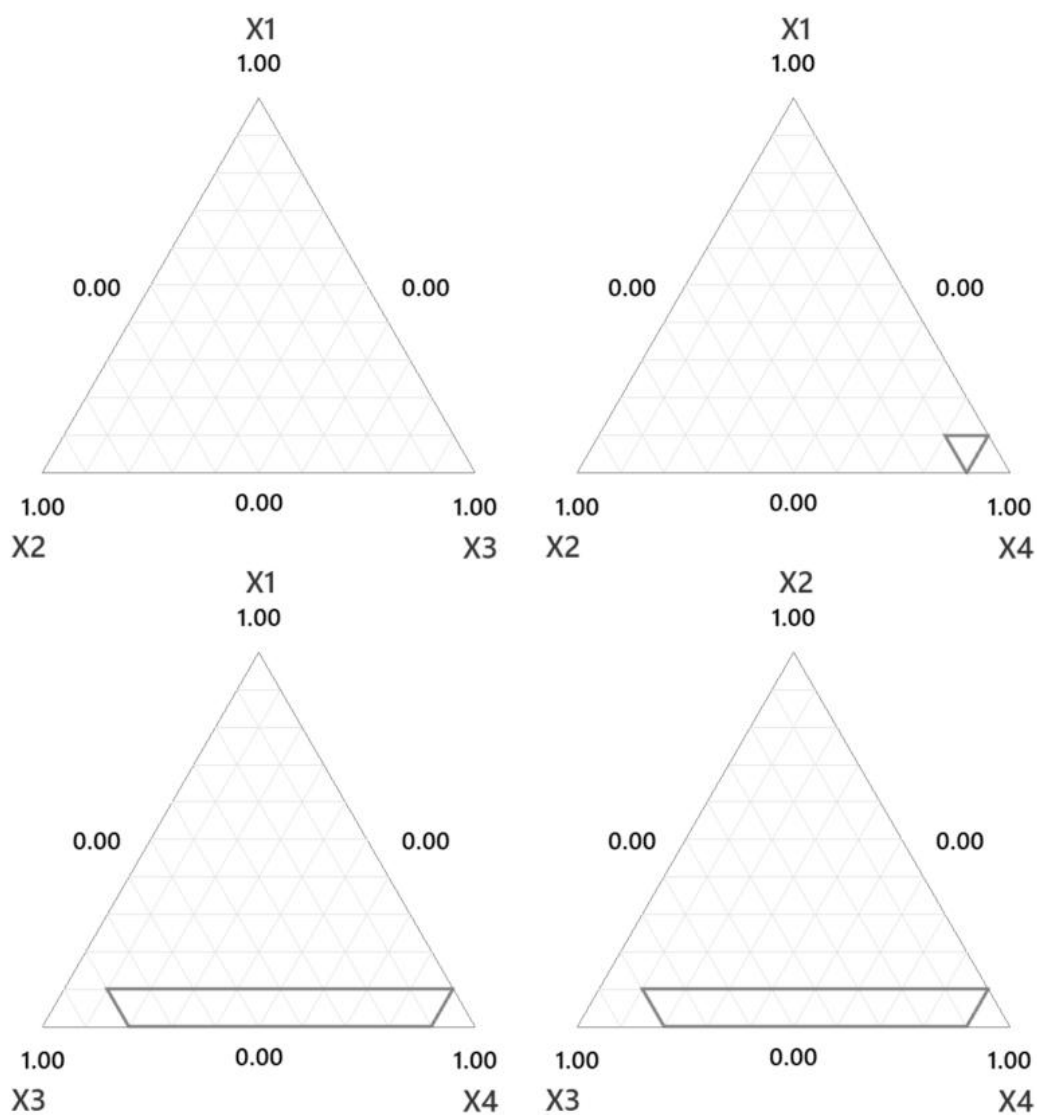
Un ejemplo de las razones que provocan este segundo motivo se puede visualizar en la Figura 4.1, donde la región de experimentación está acotada por las siguientes restricciones  $b_1 = 0.10$ ,  $b_2 = 0.10$ ,  $b_3 = 0.80$  y  $b_4 = 0.90$ .

Figura 4.1. Representación en grupos de triadas, cuando el componente restante es  $x_i = 0.05$ , de una región de experimentación reducida en la que no aplican los algoritmos CONVEXSIM



Es notable que dicha región es muy reducida en comparación con la región de operación y, además, su proyección en  $\{x_1, x_2, x_3\}$ , cuando  $x_4 = 0$ , es un único punto dado por  $\{0.1, 0.1, 0.8\}$ , tal como se visualiza en la triada de la esquina superior izquierda de la Figura 4.2, donde no se muestra ninguna forma geométrica.

Figura 4.2. Representación en grupos de triadas de una región de experimentación reducida en la que no aplican los algoritmos CONVEXSIM



De esta manera, en el Cuadro 4.3 se muestra la cantidad de restricciones inválidas que fueron generadas por cada algoritmo debido a que la región de experimentación es el caso particular mostrado en la Figura 4.1 y en la Figura 4.2; por lo tanto, al no tomar en cuenta las restricciones inválidas se cuenta con 5 457 restricciones válidas. Se prescinde, además, de las restricciones superpuestas (S) que produce CONVEXSIM – I, restricciones inscritas (I) generadas por CONVEXSIM – S con el fin de evitar resultados duplicados.

Cuadro 4.3. Cantidad de restricciones inferiores y superiores inválidas que fueron generadas por CONVEXSIM – I y CONVEXSIM – S

Cantidad de componentes	Cantidad de restricciones originales	Restricciones inválidas generadas	
		CONVEXSIM - I	CONVEXSIM - S
3	493	0	0
4	548	10	0
5	415	30	0
6	884	146	0
7	320	108	0
8	160	77	0
9	215	115	0
10	281	160	0
11	362	0	0
12	454	0	0
<b>Total</b>	<b>4 132</b>	<b>874</b>	<b>0</b>

Seguidamente, se toman en cuenta las variables de entrada restantes: 1) la presencia o no de puntos axiales, 2) el grado del polinomio a estimar y 3) la cantidad total de réplicas. Se generan 130 968 escenarios ( $5\,457 \cdot 2 \cdot 3 \cdot 4$ ) de los cuales se obtienen los valores de salida descritos en la Tabla 3.2.

Cabe resaltar que a pesar de que la documentación de los paquetes empleados indica que es posible generar diseños de vértices extremos hasta para  $p = 12$  (*nfac* en la función *Xvert* de *{mixexp}*), al momento de ejecutar la misma se encuentra con el siguiente error en la función *crtave*, necesaria en la función *Xvert* [traducción propia]: “Demasiados vértices, esta función solo funciona cuando (*#vértices* + *#centroides*)  $\leq 1000$ ” (Lawson & Willden, 2016) como es el caso del siguiente conjunto de restricciones sobre la frontera superior, cuando  $m = 3$ :

$$\begin{array}{cccccc}
 x_1 \leq 0.1, & x_2 \leq 0.1, & x_3 \leq 0.1, & x_4 \leq 0.1, & x_5 \leq 0.1, & x_6 \leq 0.1, \\
 & x_7 \leq 0.1, & x_8 \leq 0.5, & x_9 \leq 0.5 & & 
 \end{array}$$

que genera más de 3 000 puntos de diseño<sup>21</sup>. Es importante hacer notar que al emplear las restricciones generadas con CONVEXSIM el diseño se reduce a 139 tratamientos, obteniéndose un valor de  $T_n \geq 0.95$ , según la Ecuación (51).

También ocurre que, durante la simulación, sobre todo cuando  $p = 11$  y  $p = 12$ , se interrumpe la sesión con la siguiente leyenda [traducción propia]: “Sesión de R interrumpida. R encontró un error fatal. La sesión fue finalizada” (Posit team, 2023) esto es debido al tamaño del DdE generado y al cálculo de la matriz inversa<sup>22</sup>.

A partir de esto se decide descartar los casos asociados a estas cantidades de componentes, por lo que el total de casos se reduce a 93 228; el detalle de los casos desechados se muestra en el Cuadro 4.4.

Cuadro 4.4. Cantidad de casos desechados por error en la función  $X_{vert}$  de  $\{mixexp\}$  o interrupción de R y RStudio

Cantidad de componentes	Cantidad de casos generados	Cantidad de casos descartados
3	17 784	0
4	19 128	0
5	14 232	0
6	27 024	0
7	7 200	0
8	2 832	0
9	3 456	1 160 <sup>a</sup>
10	4 200	1 408 <sup>a</sup>
11	15 504	15 504 <sup>b</sup>
12	19 608	19 608 <sup>b</sup>
<b>Total</b>	<b>130 968</b>	<b>37 680</b>

<sup>a</sup> Ocho de los casos descartados ocurren cuando  $m = 2$ , los demás ocurren cuando  $m = 3$

<sup>b</sup> Todos los casos son descartados

<sup>21</sup> Estos diseños fueron generados en el software Minitab® 21.1, donde igual que con  $\{mixexp\}$  se obtiene un error al exceder, en este caso, 3 000 corridas sin réplicas.

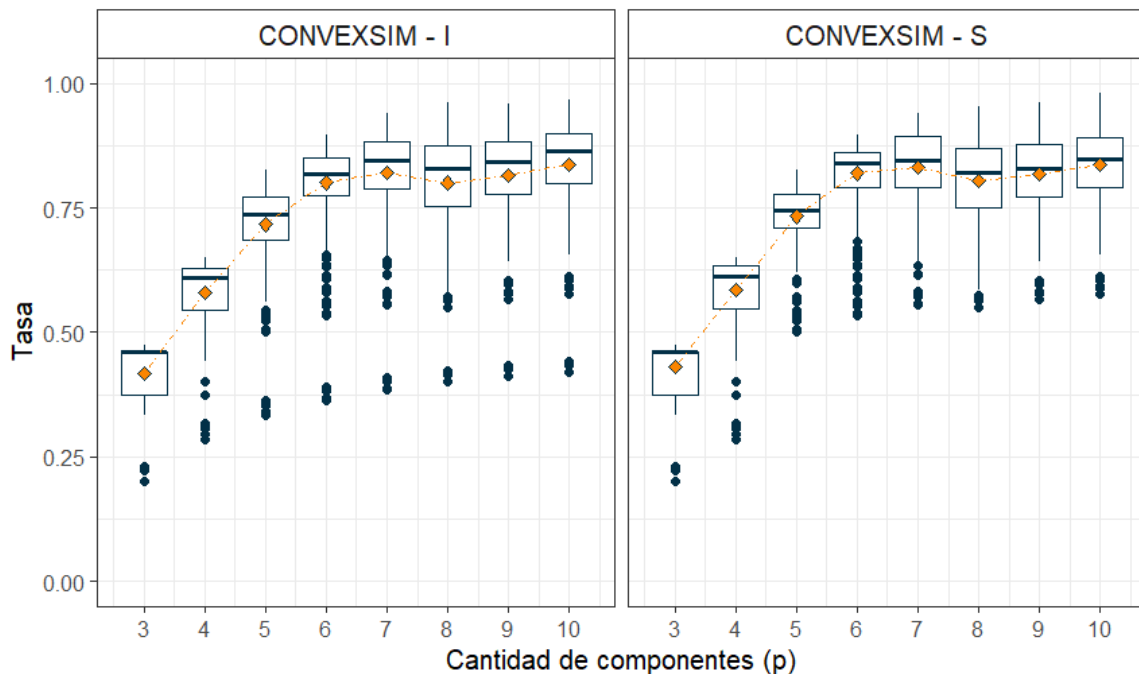
<sup>22</sup> La operación de calcular la matriz inversa puede ser computacionalmente costosa debido a varios factores, especialmente cuando se trabaja con matrices grandes, incluyendo la complejidad algorítmica, ya que a medida que el tamaño de la matriz crece, el número de operaciones necesarias aumenta cúbicamente.

## 4.2 Cantidad de corridas experimentales

Como ya ha sido mencionado, reducir la cantidad total de corridas experimentales es la meta principal de esta investigación, determinada en el objetivo general. Esta reducción es medida como una tasa ( $T_n$ ), del diseño creado con las restricciones CONVEXSIM contra el de vértices extremos y el simplex reticular, dado por la Ecuación (51).

En la Figura 4.3 se aprecia lo que ya se sospechaba desde la teoría, donde Cornell (2002) menciona que para  $p = 3$  y  $p = 4$  puede que no valga la pena cualquier intento para reducir la cantidad de corridas experimentales. No obstante, a partir de lo aquí dilucidado se sugiere que esta decisión quede supeditada a las personas investigadoras y al contexto del experimento, incluyendo sus costos asociados.

Figura 4.3. Reducción de corridas experimentales entre el diseño de vértices extremos y la alternativa con CONVEXSIM

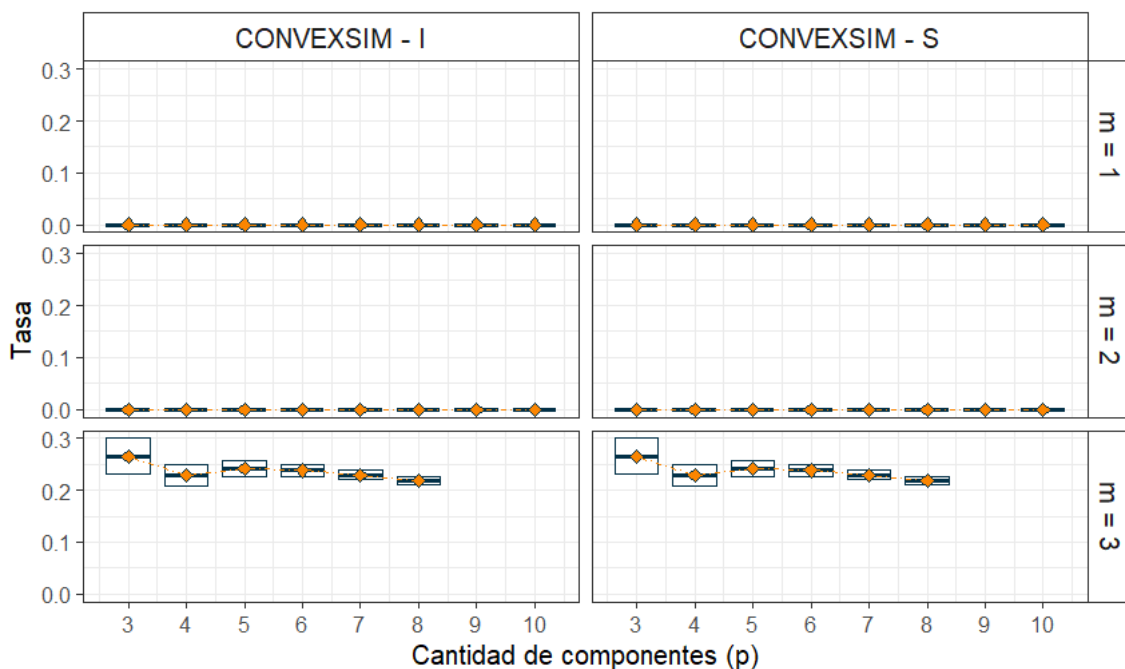


Las reducciones cuando  $p = 3$  es menor a 50 % en ambos casos, pero a partir de  $p = 5$  se esperan reducciones promedio superiores al 70 % en la cantidad de corridas

experimentales tanto para CONVEXSIM – I como para CONVEXSIM – S. Otro hallazgo esperado es que los resultados entre ambas variantes de CONVEXSIM sean prácticamente equivalentes, ya que ambos producen regiones con forma L – simplex.

De la misma forma, la comparación contra el diseño simplex reticular, mostrada en la Figura 4.4, produce resultados esperados, ya que los diseños son equivalentes cuando  $m = 1$  y  $m = 2$  y se muestran valores positivos para cuando  $m = 3$  con tendencia decreciente conforme aumenta la cantidad de componentes.

Figura 4.4. Reducción de corridas experimentales entre el diseño de simplex reticular y la alternativa con CONVEXSIM, en función del grado del polinomio ( $m$ )



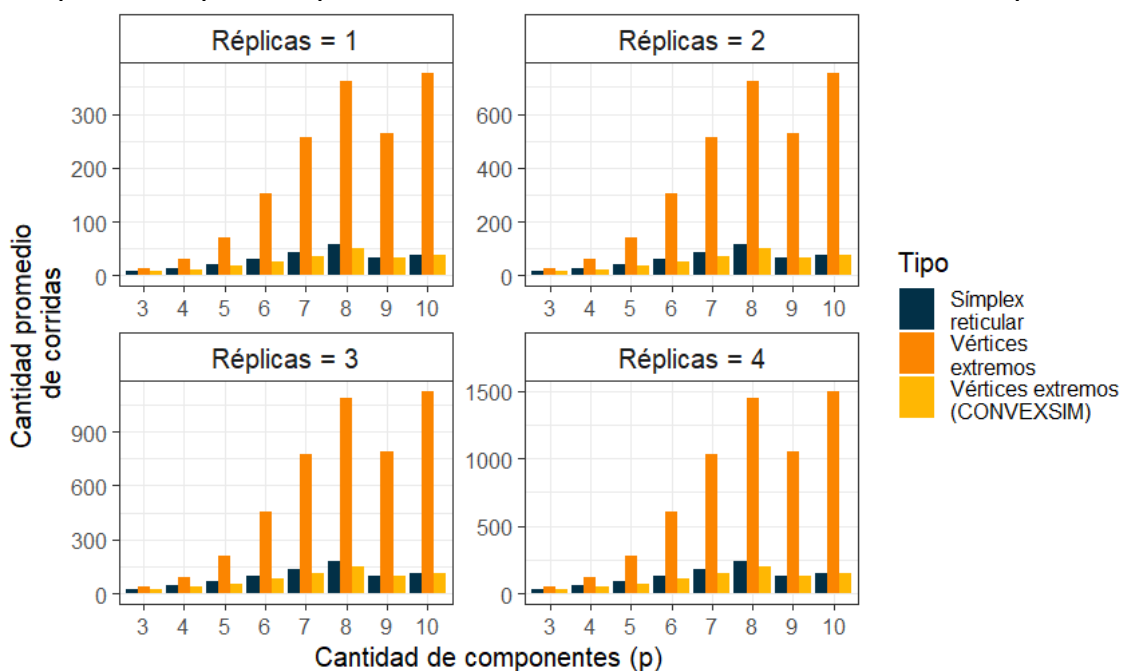
En el Cuadro 4.4 se detalla que fueron descartados los casos con  $m = 3$  cuando  $p \geq 9$ , razón por la cual no se muestran valores en el gráfico; también, para estos valores de  $p$ , en la Figura 4.3 se muestra una disminución en el promedio de la tasa de reducción de corridas experimentales ( $T_n$ ), que se explica por la limitación detallada previamente del paquete `{mixexp}` al no crear experimentos con más de 1 000 puntos de diseño. Además, como la Ecuación (51) se puede reescribir de la siguiente manera, sacando a factor común la cantidad de réplicas ( $n$ ):

$$T_n = \frac{n \cdot CE_{\text{Diseño original}} - n \cdot CE_{\text{CONVEXSIM}}}{n \cdot CE_{\text{Diseño original}}}$$

se puede concluir que el valor de  $T_n$  no se ve afectado por la cantidad de réplicas. Por esta razón no se muestra ninguna visualización.

La Figura 4.5 permite dimensionar la diferencia en términos absolutos que existe entre los tres tipos de experimentos analizados. De esta manera se comprueba el aumento vertiginoso en la cantidad de tratamientos experimentales que se produce con los diseños de vértices extremos con restricciones que generan formas de politopo irregular, a su vez que le da peso a la afirmación de Cornell (2002) en los casos en los que  $p \leq 4$ .

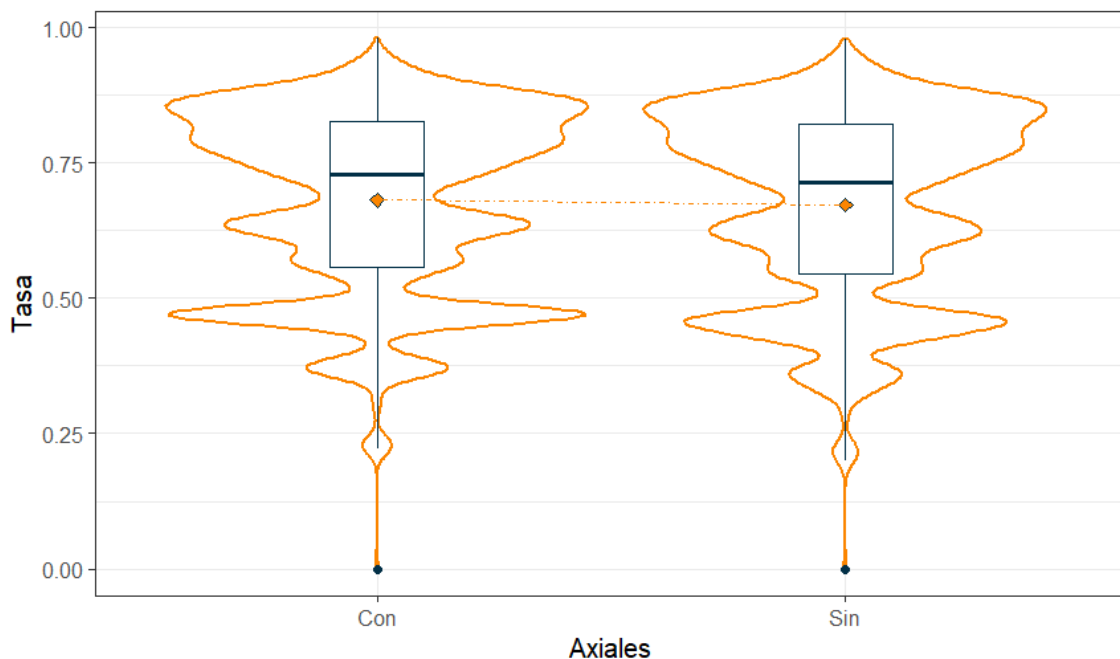
Figura 4.5. Diferencia entre los promedios de la cantidad de corridas experimentales por cada tipo de experimento analizado en función de la cantidad de réplicas



La caída en  $p = 9$  y  $p = 10$  se explica por las razones previamente detalladas. Por otro lado, la diferencia entre los valores de los diseños simplex reticular y CONVEXSIM se debe a que en el primero, cuando  $m = 3$ , la cantidad de tratamientos es ligeramente mayor.

Ahora bien, aunque es indiscutible que la presencia o no de puntos axiales aumenta la cantidad de tratamientos, se estima que este no afecta en gran medida a la diferencia entre las tasas de reducción de corridas para los tipos de diseños analizados (Figura 4.6), debido a que la cantidad de puntos axiales depende de la cantidad de vértices presentes en el diseño. No obstante, si se constata que el valor de  $T_n$  es en promedio someramente mayor cuando hay puntos axiales.

Figura 4.6. Reducción de corridas experimentales entre el diseño de vértices extremos y la alternativa con CONVEXSIM, en función de la presencia o no de puntos axiales

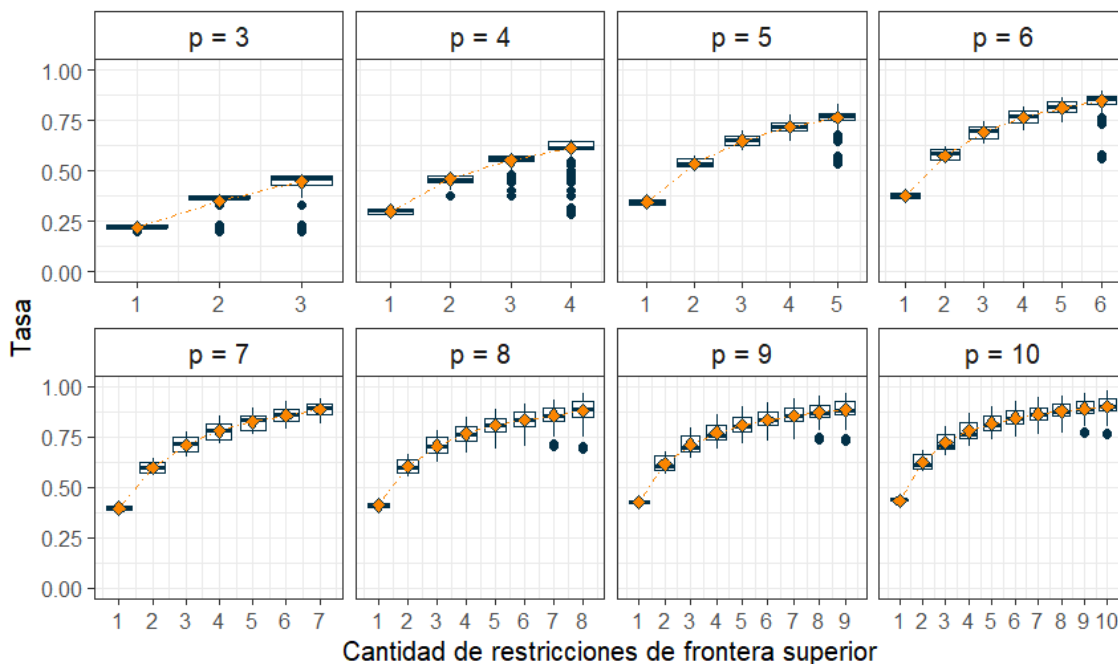


Los gráficos de violín alrededor de los gráficos de caja con bigotes de la Figura 4.6 muestra la frecuencia de los datos en cada región mediante curvas de densidad, de tal forma que se infiere que en este aspecto ambas tasas de reducción tienen, también, comportamientos similares.

Una de las hipótesis de este trabajo se relaciona con la cantidad de restricciones sobre la frontera superior del diseño original que decanta en la necesidad de utilizar el algoritmo CONVEXSIM. Dado que el problema es geométrico, se intuye que formas más irregulares dan lugar a diseños de vértices extremos con mayor cantidad de

tratamientos. Con la Figura 4.7 se muestra evidencia en favor de la veracidad de la anterior información.

Figura 4.7. Reducción de corridas experimentales entre el diseño de vértices extremos y la alternativa con CONVEXSIM en función de la cantidad de restricciones impuestas sobre la frontera superior en el diseño de vértices extremos



Por otro lado, el tamaño promedio de las restricciones originales impuestas sobre la frontera superior tiene un coeficiente de correlación lineal de Pearson de 0.39 y de Spearman de 0.45; mostrando que conforme se aumente el tamaño de estas, se aumenta moderadamente la tasa de reducción de la cantidad de corridas; siendo consistente con los hallazgos presentados hasta ahora, pues estaría parcialmente relacionado con la forma restringida en el espacio símplex.

Ahora bien, la cantidad de restricciones sobre la frontera superior, así como el tamaño de éstas, no tiene impacto alguno, como es lógico, en la cantidad de corridas experimentales del diseño símplex reticular, como se puede notar de la Ecuación (6). Sino que este valor está supeditado a la cantidad de componentes; tal cual quedó plasmado en la Figura 4.4.

$$N = \binom{p + m - 1}{m} = \frac{(p + m - 1)!}{m! (p - 1)!}$$

### 4.3 Eficacia

Los criterios de optimalidad son empleados para la evaluación de la eficacia de los diseños generados. Estos criterios se utilizan para la selección entre diseños competentes, a menor valor en estos criterios, mayor es la conveniencia del diseño.

Es importante resaltar que esta comparación se realiza tanto en proporciones como en pseudocomponentes, ya que la estimación con MCO de los modelos puede realizarse en cualquiera de los dos espacios: el original y el parametrizado. Se realiza la acotación de que para el diseño símplex reticular, al no tener restricciones de frontera no requiere transformación en pseudocomponentes y que de aplicarse resultaría en la misma matriz de diseño.

En este punto se trae a colación un error que está presente en la función Xvert de `{mixexp}` ya que produce diseños que son inadecuados. Para ilustrar esta situación se recurre al siguiente ejemplo para  $p = 3$  y las restricciones a continuación:

$$0.00 \leq x_1 \leq 0.10, \quad 0.45 \leq x_2 \leq 0.55, \quad 0.45 \leq x_3 \leq 0.55$$

con  $m = 3$ , e indiferente de la presencia de axiales o de la replicación, lo cual produce la matriz de diseño del Cuadro 4.5.

Cuadro 4.5. Ejemplo de matriz de diseño errónea para el conjunto de restricciones empleadas cuando  $p = 3$  y  $m = 3$

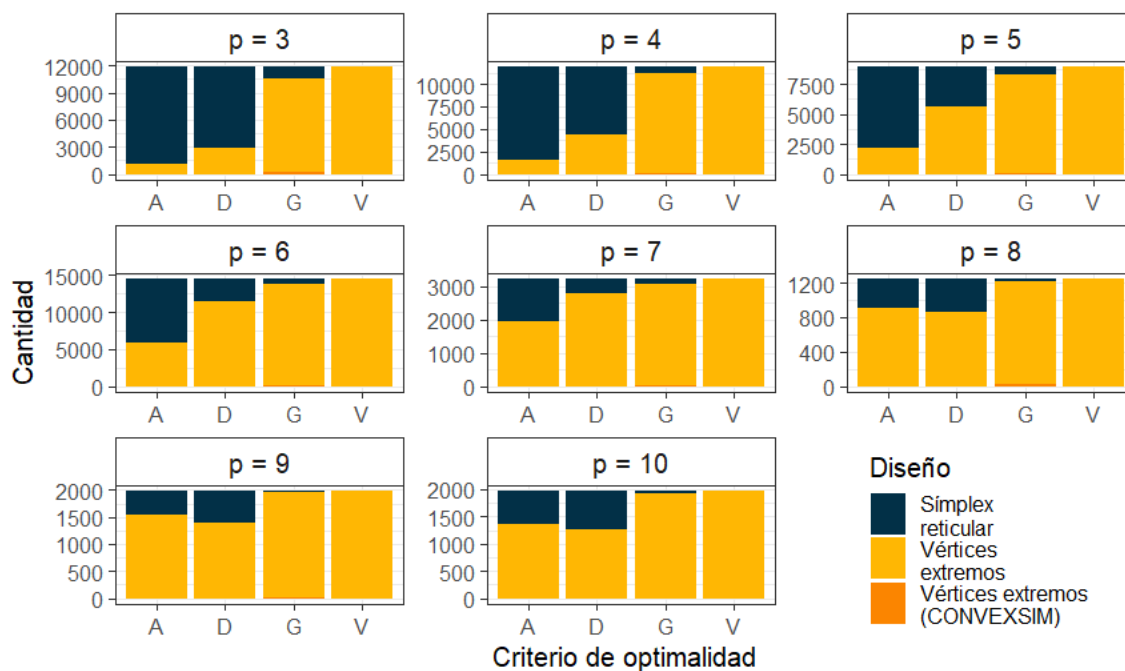
Corrida	$x_1$	$x_2$	$x_3$
1	0.000	0.450	0.550
2	0.100	0.450	0.450
3	0.000	0.550	0.450
4	0.100	0.550	0.350
5	0.000	0.550	0.450
6	0.100	0.550	0.350
7	0.033	0.483	0.483
8	0.017	0.467	0.517
9	0.067	0.467	0.467
10	0.017	0.517	0.467

Puede notarse cómo en las corridas o tratamientos resaltados se encuentran puntos de diseño que no deberían existir, ya que  $0.45 \leq x_3 \leq 0.55$ . Al momento de publicación de este trabajo no se ha encontrado el motivo por el cual ocurre esto, no obstante, se verifica que este error solo ocurre ocasionalmente cuando  $m = 3$  y que solo cambia el valor del punto de diseño, mas no la cantidad total de corridas, por lo que el análisis anterior permanece válido.

De este modo, por lo ya mencionado, sumado a la pequeña diferencia entre los puntos de diseño del simplex reticular y de vértices extremos cuando  $m = 3$  se decide emplear en este análisis solo aquellos diseños que sean consistentes con la teoría, es decir, que se espera que al aplicar la transformación en L – pseudocomponentes el diseño simplex reticular y el de vértices extremos con el algoritmo CONVEXSIM sean equivalentes. Finalmente, permanecen 55 537 casos por estudiar.

Retomando los criterios de optimalidad, en la Figura 4.8 se muestra como el diseño simplex reticular y el de vértices extremos son los amplios ganadores de la comparativa. El primero de ellos es la mayoría de las veces A – optimal y D – optimal siempre que  $p \leq 4$ , mientras que el segundo lo es en los criterios G – optimal y V – optimal, sin importar el valor de  $p$  y cuando  $p \geq 7$  es también A – optimal y D – optimal.

Figura 4.8. Cantidad de diseños en proporciones que cumplen con cada uno de los criterios de optimalidad según la cantidad de componentes

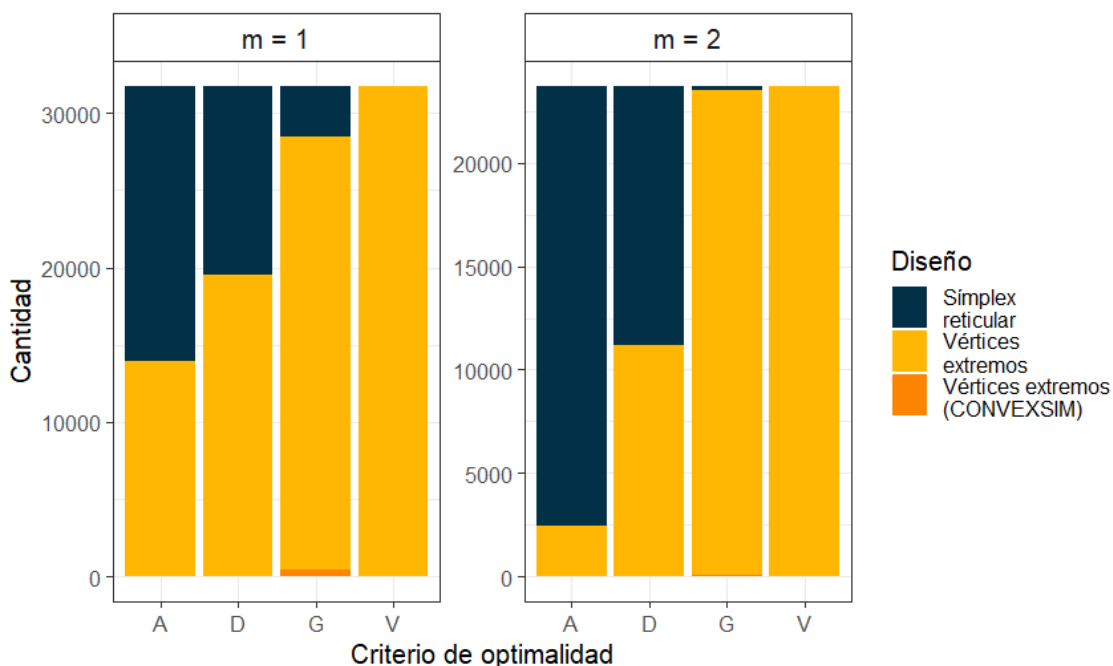


Lo visto en la Figura 4.8 es consistente con la teoría, pues los criterios A y D están relacionados con la disminución de la varianza de los parámetros estimados, mientras que G y D están asociados con la reducción de la varianza de la ecuación de predicción.

Otro resultado destacable son las pocas ocasiones en las que el diseño de vértices extremos con el algoritmo CONVEXSIM produce un diseño optimal y solo en el caso de G – optimalidad. La situación descrita en la Figura 4.8 se espera que se vea revertida al hacer el análisis en pseudocomponentes, pues los puntos de diseño para CONVEXSIM y simplex reticular son equivalentes.

Estos resultados son similares al segregarlos por el grado del polinomio ( $m$ ) que se puede estimar, tal cual se muestra en la Figura 4.9. Donde resalta el detalle que gran parte de los casos en los que los diseños de vértices extremos con CONVEXSIM son G – optimal ocurren cuando  $m = 1$  y muy pocos cuando  $m = 2$ .

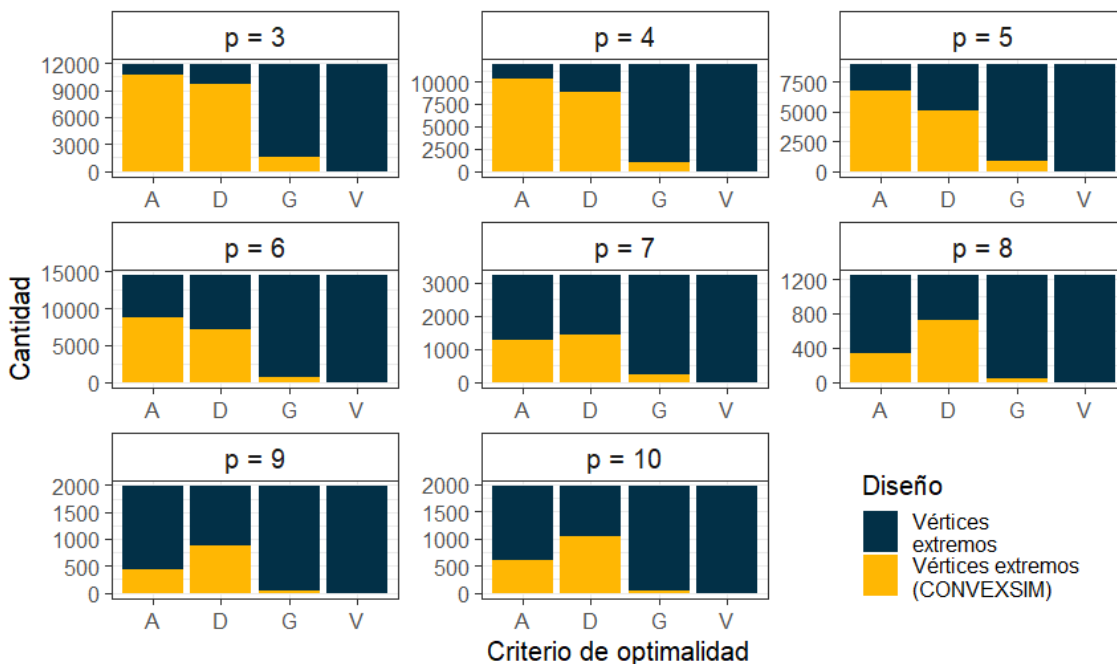
Figura 4.9. Cantidad de diseños en proporciones que cumplen con cada uno de los criterios de optimalidad según el grado del polinomio ( $m$ )



En pseudocomponentes, el diseño *simplex reticular* y de *vértices extremos* con restricciones CONVEXSIM son equivalentes, por tanto, para la construcción de la Figura 4.10 se toma únicamente al diseño de *vértices extremos* con restricciones CONVEXSIM y al diseño de *vértices extremos* con el fin de disminuir el ruido en la visualización de datos.

En este se puede apreciar como varios diseños con CONVEXSIM son A y D – optimal, aunque disminuye gradualmente conforme aumenta la cantidad de componentes. Es decir, ya sea por D, sin importar la cantidad de componentes o con A principalmente cuando  $p \leq 6$  se obtienen diseños que reducen la varianza en la estimación de los parámetros, pero que no lo son de esa forma para la ecuación de predicción.

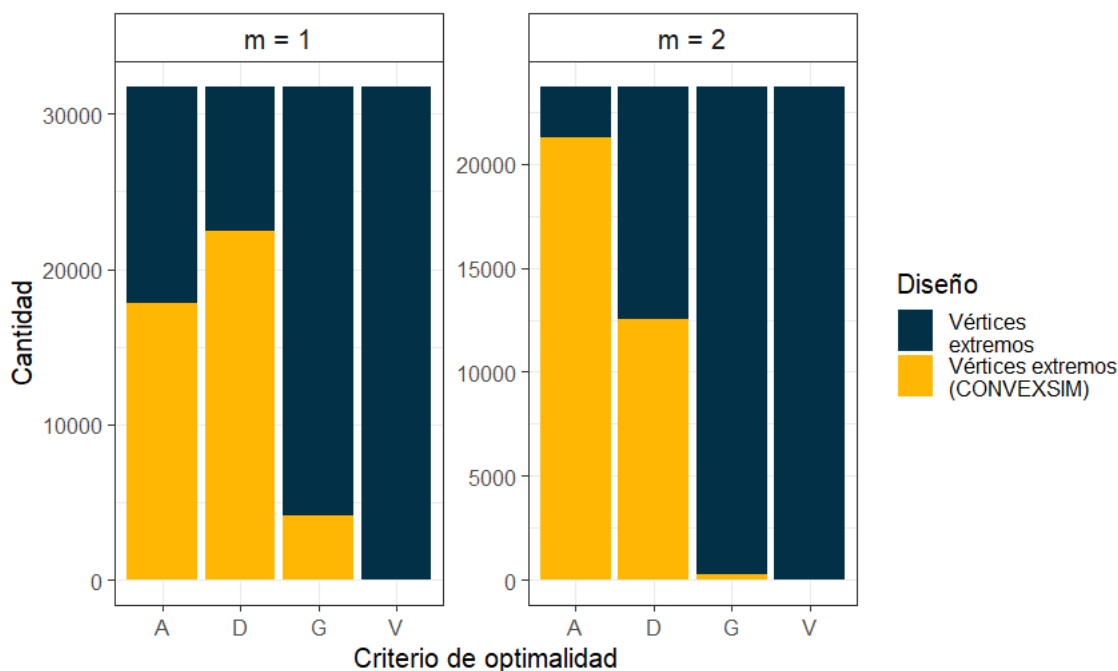
Figura 4.10. Cantidad de diseños en pseudocomponentes que cumplen con cada uno de los criterios de optimalidad según la cantidad de componentes



El hecho de que varios diseños CONVEXSIM sean A y D optimal, en pseudocomponentes, es un hallazgo importante. Esto porque tanto el modelo como los estadísticos de bondad de ajuste se pueden estimar empleando la región parametrizada en pseudocomponentes, como fue detallado en el marco teórico. Por lo que hay variadas situaciones en las que el diseño de vértices extremos con restricciones CONVEXSIM produce, principalmente, una matriz de puntos que es A o D – optimal, lo cual puede ser ventajoso.

En la Figura 4.11 se muestra como la D – optimalidad, para el caso CONVEXSIM, ocurre principalmente cuando  $m = 1$ , mientras que la A – optimalidad cuando  $m = 2$ . Por otro lado, ningún diseño es V – optimal y solo en pocas ocasiones es G – optimal y con más ahínco cuando  $m = 1$ . También se muestra como cuando  $m = 1$  hay más diseños D – optimal, mientras que cuando  $m = 2$  los más son A – optimal. Estos resultados son congruentes con los observados en los homólogos para proporciones mostrados en la Figura 4.8 y Figura 4.9.

Figura 4.11. Cantidad de diseños en pseudocomponentes que cumplen con cada uno de los criterios de optimalidad según el grado del polinomio ( $m$ )



Del análisis tanto en proporciones como en pseudocomponentes se extrae que en todas de las ocasiones los diseños de vértices extremos producen matrices que son V – optimal, sin importar la cantidad de componentes o el grado del polinomio. Una situación similar ocurre con la G – optimalidad, ya que son pocos los escenarios en los que este tipo de diseño no resulta ganador. Este comportamiento es razonable, pues ambos criterios son los que están relacionados con la varianza de la ecuación de predicción.

En proporciones, prácticamente ningún caso de vértices extremos con restricciones generadas por el algoritmo CONVEXSIM resulta optimal. Situación que se ve invertida al emplear pseudocomponentes, donde a menor  $p$  (Pearson = -0.59 y -0.68, Spearman = -0.63 y -0.78) mayor cantidad de escenarios están relacionadas con la reducción de la varianza de los parámetros (A y D optimalidad respectivamente).

Como paso final se contrastan las siguientes hipótesis de trabajo:

- i. La localización del centroide en relación con el centroide global tendrá un efecto sobre los valores de optimalidad (Cornell, 2002). Para ello se emplea la distancia euclidiana entre el centroide del diseño de vértices extremos con restricciones CONVEXSIM y los diseños de vértices extremos y símplex reticular.
- ii. Los valores de optimalidad disminuyen a medida que los rangos de los componentes se amplían (Cornell, 2002).
- iii. Se sabe que la replicación en puntos específicos reduce los valores de apalancamiento (Montgomery & Voth, 1994). Pero se desea determinar si la replicación generalizada tiene un efecto sobre los valores de optimalidad.

Para realizar este análisis de una forma más simple se emplea el mapa de relaciones de la Figura 4.12 con las correlaciones lineales (Pearson) y monótonas (Spearman) entre las variables de interés y los valores de optimalidad de la Figura 4.12 para proporciones y el de la Figura 4.13 para pseudocomponentes.

Figura 4.12. Correlaciones entre las hipótesis de trabajo y los criterios de optimalidad para proporciones

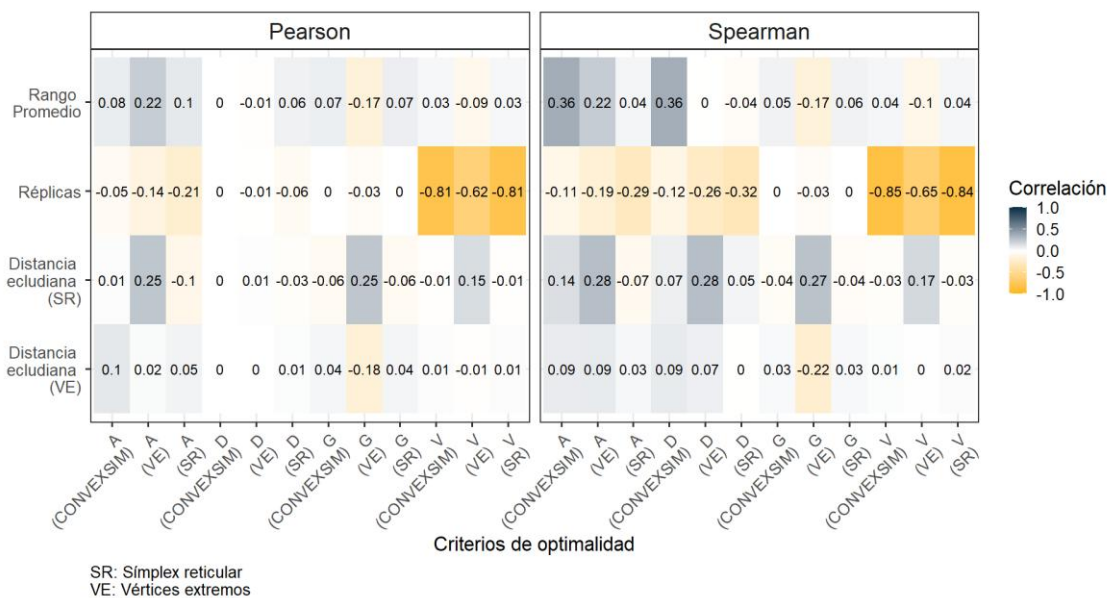
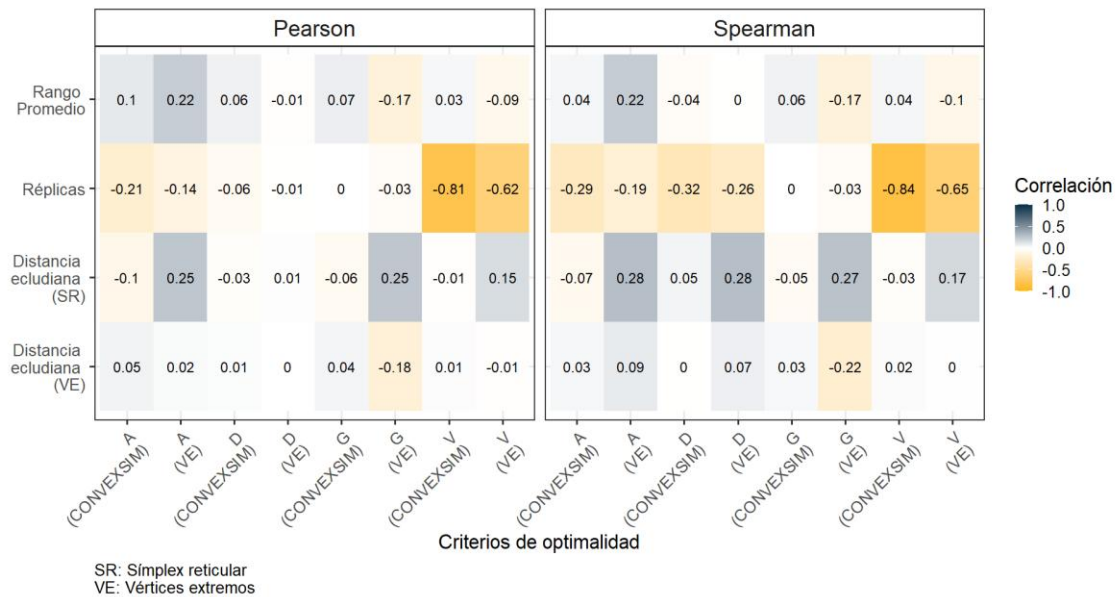


Figura 4.13. Gráfico de correlaciones entre las hipótesis de trabajo y los criterios de optimalidad para pseudocomponentes



De ambas figuras se infiere que:

- i. La localización del centroide, medido como la distancia euclidiana del diseño de vértices extremos con restricciones CONVEXSIM con respecto al centroide global  $x_i = 0, i = 1, 2, 3 \dots$  tiene una relación lineal positiva moderada con respecto a los criterios de A – optimalidad y G – Optimalidad en los diseños de vértices extremos tanto en proporciones como pseudocomponentes, y en menor medida con la V – optimalidad. Respecto al criterio de Spearman se aprecia una relación monótona positiva e igualmente moderada, pero con el criterio de D – optimalidad. Por lo que se infiere que efectivamente la localización del centroide global si tiene un efecto sobre tres de los criterios de optimalidad empleados en el diseño de vértices extremos. Es decir que, a mayor distancia de su centroide respecto al global, mayor es el valor de la optimalidad del diseño de vértices extremos (y, por tanto, menos optimal).

- ii. La replicación tiene un efecto negativo sobre los valores de  $V$  – optimalidad en proporciones y pseudocomponentes, sin importar el tipo de diseño, es decir, la replicación aumenta la  $V$  – optimalidad. Lo que significa que el apalancamiento promedio disminuye. Del marco teórico se desprende que este criterio está dirigido a la varianza de la ecuación de predicción y, además, que el principio de replicación tiene dos propiedades importantes, entre la que destaca que permite obtener una mejor estimación del error experimental. Por otro lado, Myers et al. (2009) indica que este criterio puede ser escrito en función de  $N$  (cantidad de observaciones), por lo que resulta lógico este resultado.
- iii. Conforme aumenta el rango promedio, aumenta de forma lineal y moderada el valor de  $A$  – optimalidad para el diseño de vértices extremos en pseudocomponentes y en proporciones. De igual forma para el diseño con restricciones CONVEXSIM se obtiene una relación monótona moderada positiva para el criterio de  $A$  – optimalidad y  $D$  – optimalidad únicamente en proporciones. Esto contradice lo expuesto por Cornell (2002), excepto para  $G$  – optimalidad en el diseño de vértices extremos en ambos espacios, el original y el parametrizado.

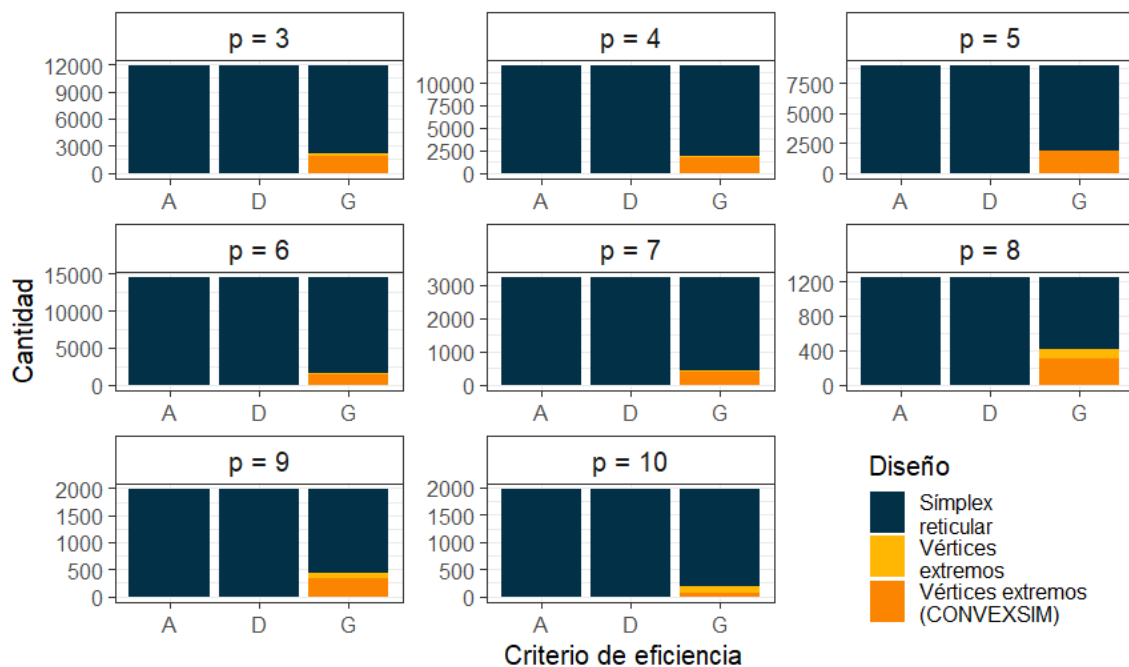
#### 4.4 Eficiencia

Los valores de eficiencia son una medida relativa para seleccionar un diseño. Esta medida, al contrario de la optimalidad, indica que entre más alto sea, mejor. Estas medidas toman en cuenta la cantidad de corridas experimentales ( $n$ ), así como la cantidad de términos en el modelo que se desea estimar ( $q$ ). Tal como fue señalado en el marco teórico al citar a Goos & Jones (2011), un valor de 100 % o cercano a este no tiene un significado particular, y resultados diferentes a éstos pueden ser totalmente aceptables.

La eficiencia es determinada mediante tres criterios: A, D y G, que están relacionados con sus homónimos en la optimalidad y, por tanto, los resultados también se ven afectados por lo analizado previamente tanto para la eficacia como para la cantidad de corridas experimentales.

En la Figura 4.14 se muestra la cantidad de casos que resultan eficientes según el criterio, cuyo comportamiento era lo esperado, ya que al reducirse la cantidad de corridas experimentales el valor de la eficiencia aumenta.

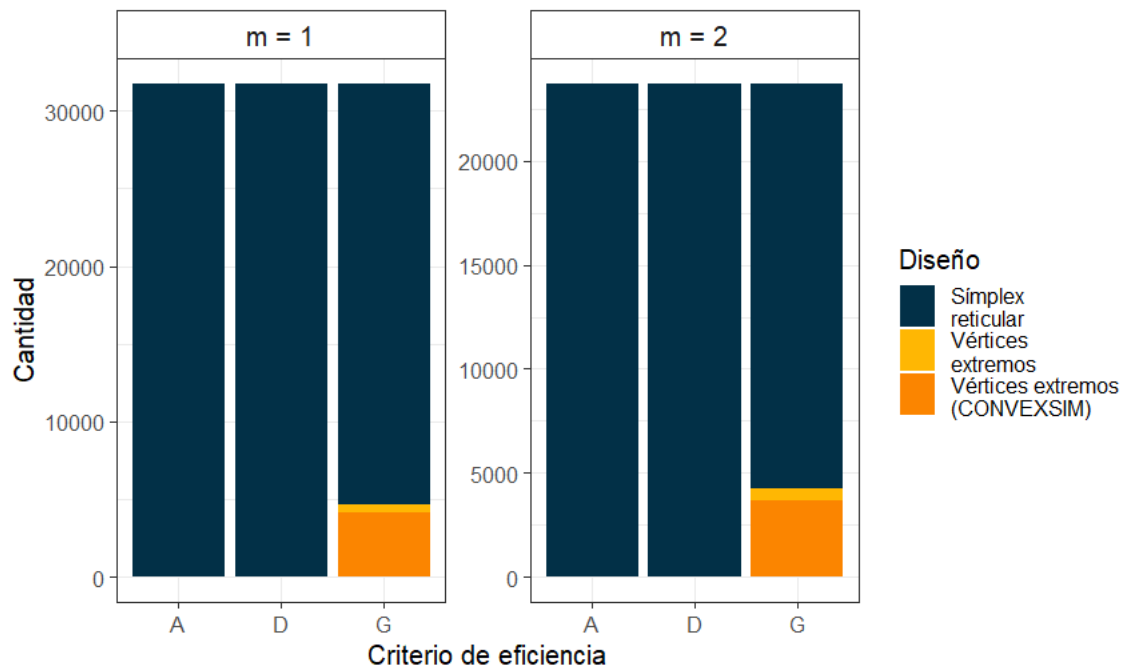
Figura 4.14. Cantidad de diseños en proporciones que cumplen con cada uno de los criterios de eficiencia según la cantidad de componentes



Por lo tanto, los diseños de vértices extremos con restricciones CONVEXSIM, así como los diseños simplex reticular son los ganadores de esta comparación. No obstante, en algunas ocasiones, con la menor frecuencia, el diseño de vértices extremo resultaría más eficiente, pero solo bajo el criterio G. Resalta el hecho de que sin importar la cantidad de componentes ( $p$ ) los diseños simplex reticular son siempre A y D eficientes en comparación con el resto, y en la mayoría de las ocasiones son G eficientes, seguido por el diseño de vértices extremos con restricciones CONVEXSIM.

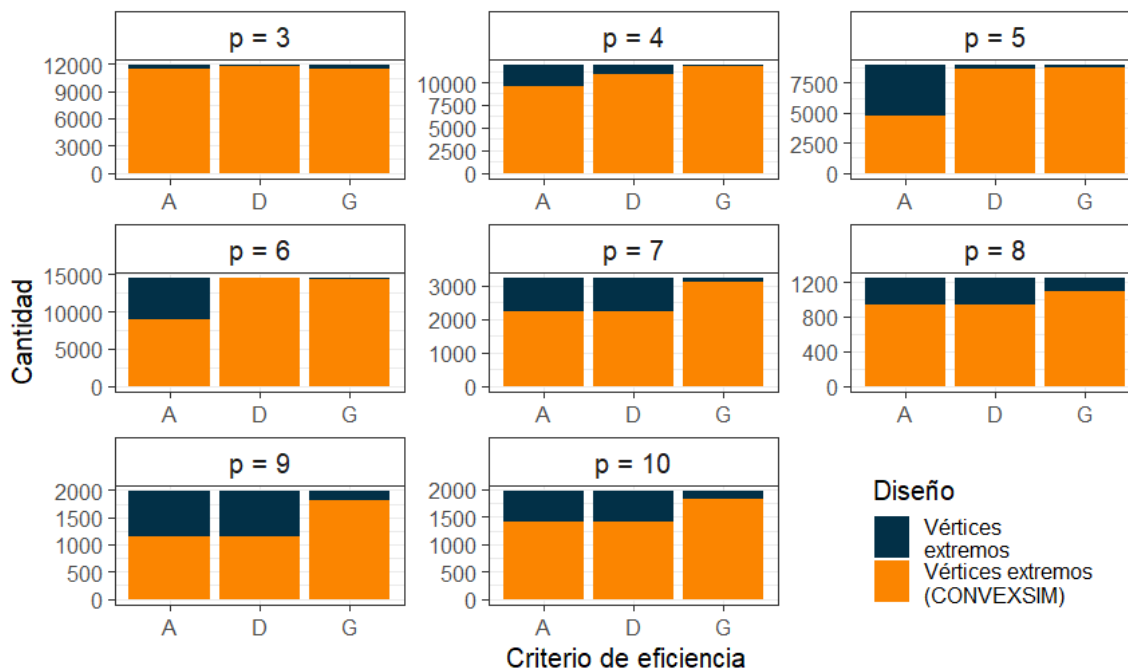
A continuación, se muestra el efecto sobre los valores de eficiencia que tiene el grado máximo que tiene el polinomio (Figura 4.15). Se aprecia que cuando  $m = 2$  aumenta proporcionalmente la cantidad de casos del diseño de vértices extremos con CONVEXSIM que cumplen con el criterio de G – eficiencia.

Figura 4.15. Cantidad de diseños en proporciones que cumplen con cada uno de los criterios de eficiencia según el grado del polinomio ( $m$ )



Al igual que con la optimalidad, se espera que el resultado de la Figura 4.14 se revierta en cierto grado al emplear los valores en pseudocomponentes (Figura 4.16).

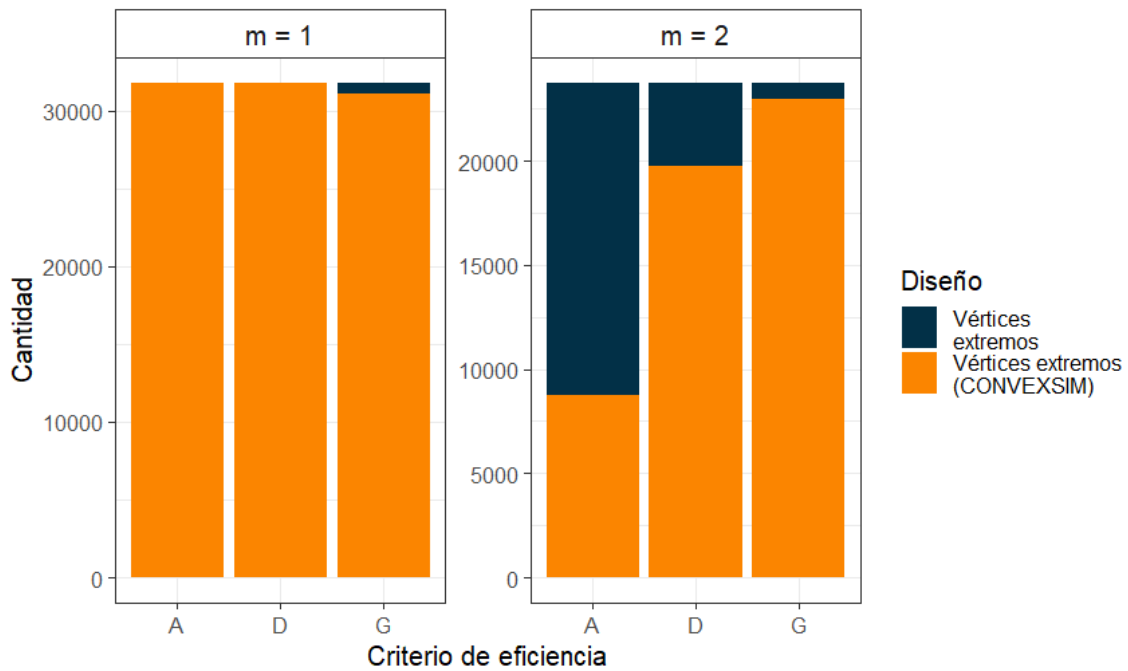
Figura 4.16. Cantidad de diseños en pseudocomponentes que cumplen con cada uno de los criterios de eficiencia según la cantidad de componentes



Los resultados que se muestran en la Figura 4.16 son valiosos en el tanto reafirman que los algoritmos CONVEXSIM – I y CONVEXSIM – S producen restricciones que en conjunto con un diseño de vértices extremos no solo reducen la cantidad de corridas experimentales, sino que al parametrizar los diseños en L – pseudocomponentes los vuelve en A, D y G eficientes en la mayoría de las ocasiones. Además, se observa que la G – eficiencia se alcanza en la mayoría de las veces, mientras que A y D varían someramente al cambiarse la cantidad de componentes.

Por otro lado, en pseudocomponentes, en la Figura 4.17, se muestra que cuando  $m = 2$  la cantidad de diseños de vértices extremos A y D eficientes aumenta, sobresaliendo el criterio A. No obstante, en el caso  $m = 1$  el diseño de vértices extremos CONVEXSIM es siempre, en estos casos, A y D – eficiente y la mayoría de las veces G – eficiente.

Figura 4.17. Cantidad de diseños en pseudocomponentes que cumplen con cada uno de los criterios de eficiencia según el grado del polinomio ( $m$ )



De esta manera, luego de exponer los resultados obtenidos del estudio de simulación se comprueba que los mismos son una alternativa viable a considerar para la reducción en la cantidad de corridas experimentales en los casos en los que las personas experimentadoras lo consideren conveniente al sopesar los costos y los beneficios de los mismos.

## 5. CASO PRÁCTICO

Los datos recolectados, así como el código empleado para el análisis de datos se encuentran disponibles en Zenodo (García-Goñi et al., 2024). Este experimento se analiza en concordancia con su objetivo, donde el interés recae en obtener la proporción de cada ingrediente o componente que logre la mejora simultánea de las dos variables de respuesta utilizadas. Como ya se ha detallado, este primer experimento con mezclas es de carácter exploratorio y sirve para encontrar cuáles de los ingredientes de la mezcla tienen mayor influencia sobre la variable de respuesta, así como para determinar en qué porcentaje debe participar cada ingrediente.

### 5.1 Eficacia y eficiencia del diseño

Dada las restricciones de frontera ( $x_i \leq 0.75$ ,  $i = 1, 2, \dots, 6$ ), la cantidad de tratamientos del diseño de vértices extremos es 31, de los cuales uno es el punto central, el resto son vértices; por lo que el diseño, al final consiste en 62 corridas experimentales. Ahora, al aplicar el algoritmo CONVEXSIM – I se obtienen nuevas restricciones de frontera ( $0.05 \leq x_i \leq 0.75$ ,  $i = 1, 2, \dots, 6$ ) que tienen forma símplex, por lo que se puede emplear un diseño reticular  $\{6, 1\}$  que tiene, según la Ecuación (9) un total de 6 tratamientos:

$$N = \binom{6 + 1 - 1}{1} = \frac{(6 + 1 - 1)!}{1!(6 - 1)!} = 6$$

Este número de tratamientos se convierte en 7 al agregar el punto central y que decanta en 14 corridas al replicar el experimento. Esto implica, en comparación con el diseño original, una reducción en la cantidad de corridas de un  $\frac{(62-14)}{62} * 100 = 77.42 \%$ , haciendo asequible la consecución de este experimento y su objetivo. Comparando ambos diseños bajo los criterios de optimalidad y eficiencia se obtienen los resultados del Cuadro 5.1.

Cuadro 5.1. Valores de optimalidad y eficiencia para el diseño simplex de la región de operabilidad completa, el diseño de vértices extremos original y el diseño creado con CONVEXSIM en proporciones y pseudocomponentes

Diseño	Métrica						
	Optimalidad				Eficiencia		
	A	D	G	V	A	D	G
<b>Símplex</b>							
Proporciones	2.928	0.013	0.878	0.428	0.146	0.147	0.488
Pseudocomponentes	2.928	0.013	0.878	0.428	0.146	0.147	0.488
<b>Vértices extremos</b>							
Proporciones	1.006	1.92x10 <sup>-5</sup>	0.973	0.097	0.096	0.099	0.099
Pseudocomponentes	1.006	1.92x10 <sup>-5</sup>	0.973	0.097	0.096	0.099	0.099
<b>CONVEXSIM</b>							
Proporciones	5.531	0.474	0.878	0.428	0.077	0.081	0.488
Pseudocomponentes	2.928	0.013	0.878	0.428	0.146	0.147	0.488

Se concluye que, en este caso particular, la aplicación de CONVEXSIM, hace al diseño obtenido G – optimal; sin embargo, no lo es así para los criterios A, D y V en términos de optimalidad. Pero, por otra parte, es un diseño eficiente para los criterios A, D y G, donde es equivalente a un diseño símplex sin restricciones. Respecto al apalancamiento, para los vértices se obtiene un valor de 0.49 y para el centroide es de 0.07, tanto en proporciones como en pseudocomponentes.

## 5.2 Análisis de resultados

Se retoman los componentes que conforman la mezcla del experimento diseñado, en el Abordaje metodológico se encuentran las razones de la inclusión de estos ingredientes para observar cómo influyen en las variables de respuesta.

- i. Azúcar blanca ( $x_1$ )
- ii. Aceite de coco ( $x_2$ )
- iii. Salvado de trigo ( $x_3$ )
- iv. Germen de trigo ( $x_4$ )
- v. Miel de abeja ( $x_5$ )
- vi. Jabón ( $x_6$ )

Estos componentes deben cumplir con la Ecuación (3), es decir, tienen que estar entre 0 y 1, y en este caso particular se encuentran todos entre 0.05 y 0.75. Se obtienen y se analizan los coeficientes de regresión que representan el efecto de diferentes componentes en una mezcla sobre la variable que está siendo explicada. Para la variable de respuesta densidad se obtiene el siguiente modelo, según la Ecuación (25):

$$E(y|x)_{densidad} = 0.63x_1 + 0.95x_2 - 0.22x_3 + 0.45x_4 + 1.50x_5 + 1.19x_6$$

con un valor de la media del error residual ( $s$ ) de 0.1122 g/mL.

A continuación, en el Cuadro 5.2 se muestran los resultados obtenidos para la variable de respuesta densidad en términos de los coeficientes, sus errores estándar y el Factor de Inflación de la Varianza (FIV).

Cuadro 5.2. Coeficientes de regresión, error estándar y FIV para la variable de respuesta densidad

<b>Término</b>	<b>Coefficiente</b>	<b>EE. del coeficiente</b>	<b>FIV</b>
$x_1$	0.63	0.11	1.11
$x_2$	0.95	0.11	1.11
$x_3$	-0.22	0.11	1.11
$x_4$	0.45	0.11	1.11
$x_5$	1.50	0.11	1.11
$x_6$	1.19	0.11	1.11

EE: Error Estándar

FIV: Factor de Inflación de la Varianza

De esta manera, el coeficiente de determinación múltiple ( $R^2$ ) es de 98.97 % y 98.20 %. Esto es un indicativo de que la capacidad explicativa del modelo es alta. Por otro lado, del marco teórico se desprende que a los coeficientes lineales de la porción de la mezcla lineal ( $\beta_i$ ) no se les prueba la hipótesis nula de que estos sean iguales a cero ( $H_0: \beta_i = 0$ ) como consecuencia de lo mostrado en la Ecuación (24), donde  $\beta_i^* = \beta_0 + \beta_i$  y la característica colineal de los componentes de la mezcla.

Esto quiere decir que aun cuando el parámetro fuera igual a cero, esto no implicaría que el componente  $x_i$  no tenga un efecto individual, sino que su efecto es un valor constante ( $\beta_0$ ).

Dicho esto, la interpretación de los coeficientes obtenidos es la siguiente, por ejemplo, para densidad el término sinérgico<sup>23</sup> qué más aporta de forma individual a la variable de respuesta es el asociado con  $x_5$ : Miel, con el cual se espera una respuesta de media de 1.50 g/ml para la mezcla pura, es decir cuando  $x_5 = 1$ , es decir 100 %, y los demás componentes sean cero.

Este es un resultado esperado en el tanto se toma en cuenta que la densidad de la miel varía dependiendo de varios factores, pero en general se encuentra entre 1.37 g/ml y 1.42 g/ml (Torres Mejía et al., 2023) y que la región de experimentación no incluye las mezclas puras por la restricción de frontera superior ( $x_i \leq 0.75$ ), por lo que este valor está siendo extrapolado.

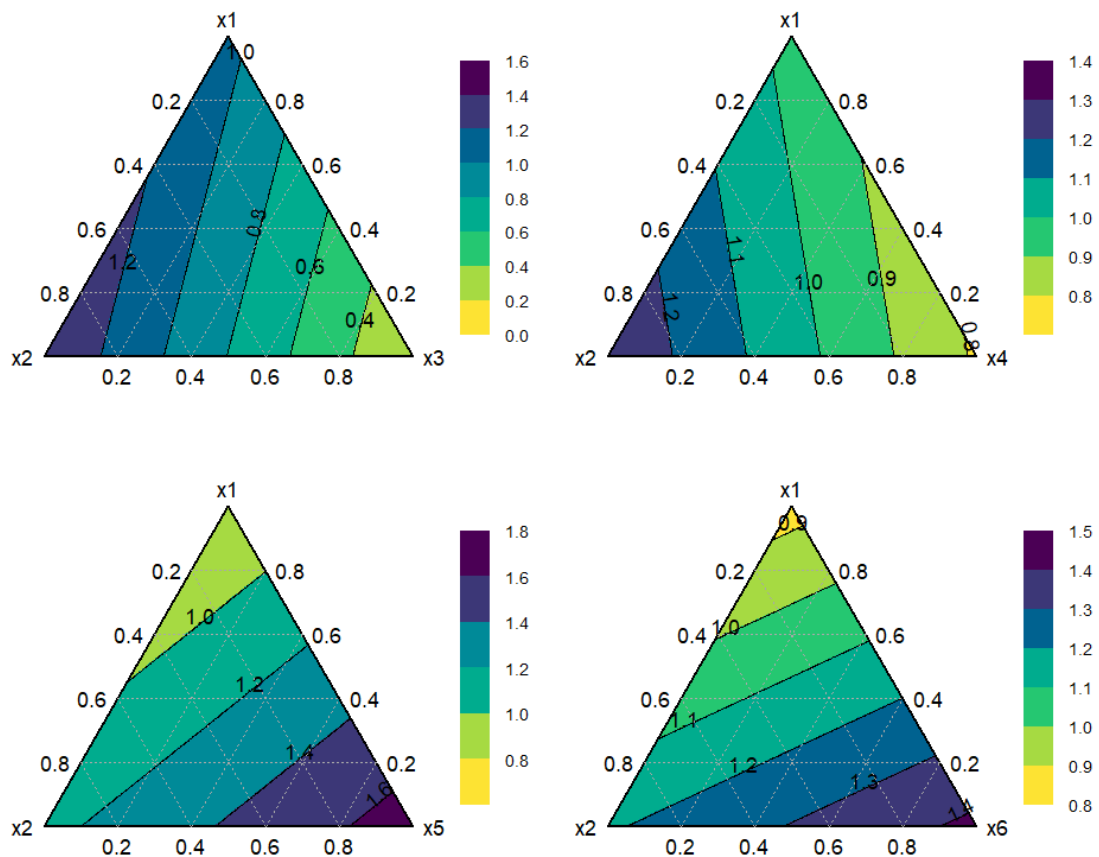
Siguiendo otro ejemplo, el coeficiente asociado con  $x_6$ : Jabón, con el cual el resultado medio esperado cuando la mezcla es pura es de 1.19 g/ml. Este comportamiento puede observarse mediante gráficos de contorno. A modo ilustrativo<sup>24</sup>, se presentan solo algunos grupos de gráficos de contorno para cada variable de respuesta. Por ejemplo, en la Figura 5.1 se muestra como el valor de densidad aumenta conforme aumenta la proporción de  $x_5$  y  $x_6$ .

---

<sup>23</sup> Cuando un término es positivo se dice que aporta a la variable de respuesta y se le llama sinérgico, por el contrario, cuando es negativo se dice que es antagónico. Los coeficientes resultan comparables para el mismo modelo en el tanto todos son parte de la misma proporción y en este caso, además, sus errores estándar son iguales.

<sup>24</sup> Cabe resaltar que como la representación gráfica solo puede realizarse en triadas y hay seis componentes, se cuenta con un total de  $6C3 = 20$  configuraciones posibles, que se agrupan en cuartetos, por lo que se generan  $\frac{20}{4} = 5$  grupos de gráficos de contorno por variable de respuesta. Todos los gráficos están disponibles en Zenodo (García-Goñi et al., 2024).

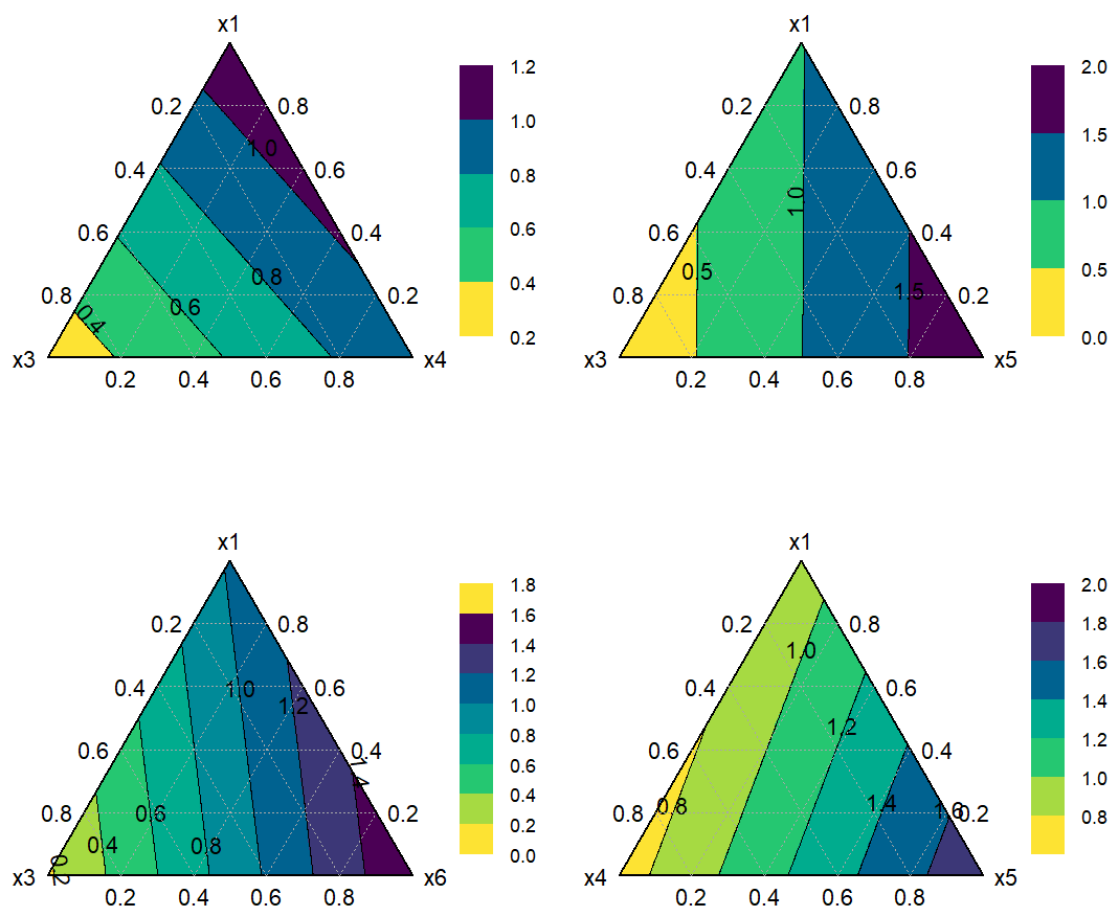
Figura 5.1. Primer grupo de gráficos de contorno lineales para la variable de respuesta densidad



El único término antagónico para este modelo se encuentra en  $x_3$ : Salvado de trigo, con el cual se esperaría un valor de  $-0.22$  g/ml cuando la mezcla sea pura, no obstante, este carece de interpretación práctica, pues es incongruente al fenómeno físico que estima, ya que la densidad es una variable siempre positiva, en el tanto la masa y el volumen que lo conforman lo son también.

A pesar de la inconsistencia, que se puede deber a la ausencia del punto de diseño de la mezcla pura debido a las restricciones, lo que se espera con este término es que, al agregar más de este ingrediente, se reduzca la densidad del producto final. Esto es lo esperado, pues la densidad del salvado de trigo tamizado es baja, así se observa en la Figura 5.2.

Figura 5.2. Segundo grupo de gráficos de contorno lineales para la variable de respuesta densidad



El modelo obtenido para la variable de respuesta oleosidad es el siguiente:

$$E(y|x)_{oleosidad} = 2.36x_1 + 5.69x_2 + 0.93x_3 + 0.45x_4 + 5.2x_5 + 4.50x_6$$

con un valor de  $s = 0.4039$ .

Ahora bien, en el Cuadro 5.3, se muestran los resultados obtenidos homólogos de la regresión para la variable de respuesta oleosidad.

Cuadro 5.3. Coeficientes de regresión, error estándar y FIV para la variable de respuesta oleosidad

<b>Término</b>	<b>Coeficiente</b>	<b>EE. del coeficiente</b>	<b>FIV</b>
$x_1$	2.36	0.39	1.11
$x_2$	5.69	0.39	1.11
$x_3$	0.93	0.39	1.11
$x_4$	0.45	0.39	1.11
$x_5$	5.21	0.39	1.11
$x_6$	4.50	0.39	1.11

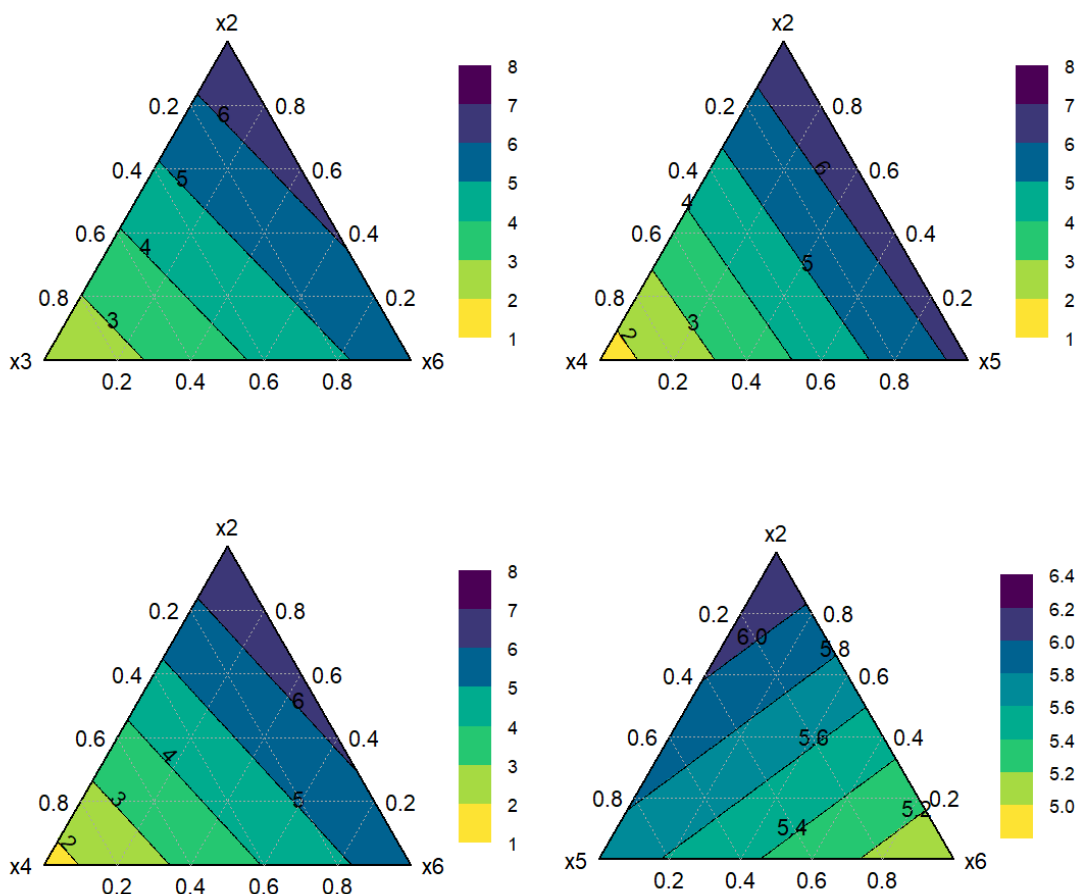
EE: Error Estándar

FIV: Factor de Inflación de la Varianza

Para esta variable de respuesta se obtiene un valor de  $R^2 = 99.23 \%$ . Por lo que se llega a la misma conclusión que con la variable de respuesta anterior. Por otro lado, el FIV para ambos modelos es de  $1.11 < 10$ , por lo que cumple con el criterio establecido por Marquardt (1970).

En esta variable solo hay términos sinérgicos, es decir, todos aportan en cierta medida a la variable de respuesta pues son positivos. Los que más aporte individual tienen para la variable de respuesta, en orden descendente son  $x_2$ : Aceite de coco,  $x_5$ : Miel y  $x_6$ : Jabón, de los cuales se espera una respuesta media, al encontrarse como mezcla pura en un 100 % de 5.69, 5.21 y 4.50 respectivamente. Esto se puede observar en la Figura 5.3, donde también resalta que  $x_3$ : Salvado de trigo y  $x_4$ : Germen de trigo son los coeficientes sinérgicos que menos aportan a la variable de respuesta. Este comportamiento es similar para ambas variables de respuesta.

Figura 5.3 Tercer grupo de gráficos de contorno lineales para la variable de respuesta oleosidad



### 5.3 Optimización simultánea de las variables de respuesta

La función de deseabilidad es utilizada para obtener una solución simultánea para ambas variables de respuesta, en las cuales se busca alcanzar valores nominales o metas (T) de 1.00 g/mL para la densidad y de 3.00 para la oleosidad. Las funciones de deseabilidad individuales generadas responden a las explicadas en la Figura 2.15 (c) y en la Ecuación (43). En el Cuadro 5.4 se muestran los valores de deseabilidad individual obtenidos.

Cuadro 5.4. Deseabilidad individual y valor esperado para cada variable de respuesta

Variable de respuesta	Deseabilidad individual	Valor esperado
Densidad	0.98	1.00 g/mL
Oleosidad	0.84	3.31

Se resalta que el valor deseado para la densidad es alcanzado y que en consecuencia se obtiene una estimación alta de deseabilidad individual, no obstante, para el caso de la oleosidad se obtiene un valor que se encuentra 0.31 por encima de lo requerido, y así es reflejado por la deseabilidad asociada, empero, en el contexto de este experimento dicha estimación se considera apropiada. Con estos resultados la deseabilidad compuesta alcanzada es de  $(0.98 \cdot 0.84)^{\frac{1}{2}} = 0.91$ .

En el Cuadro 5.5 se muestra la proporción en la que debe participar cada componente para obtener los resultados descritos con anterioridad, a la vez que se compara con la proporción original utilizada en el experimento Taguchi  $L_8 \cdot L_2$ .

Cuadro 5.5. Proporción en la que debe participar cada componente en la mezcla del producto cosmético

Componente	Proporción del primer experimento	Proporción necesaria para mejorar las variables de respuesta
$x_1$ Azúcar	8.0 %	7.7 %
$x_2$ Aceite de coco	30.0 %	5.4 %
$x_3$ Salvado de trigo	6.0 %	5.6 %
$x_4$ Germen de trigo	6.0 %	28.0 %
$x_5$ Miel de abeja	10.0 %	48.3 %
$x_6$ Jabón	40.0 %	5.0 %

Los nuevos valores son relevantes para la siguiente etapa del programa de experimentación, pues acotan la región de experimentación, permitiendo concentrar los esfuerzos de optimización alrededor de estos. Queda así ejemplificado que el algoritmo CONVEXSIM puede generar un diseño experimental adecuado para estimar un modelo que es capaz de cumplir los objetivos de un programa de experimentación a la vez que requiere de una menor cantidad de corridas experimentales.

## 6. CONCLUSIONES Y LIMITACIONES

### 6.1 Conclusiones

El objetivo general de esta tesis fue diseñar un algoritmo para la reducción de tratamientos en diseños experimentales con mezclas de vértices extremos generados por restricciones en la frontera superior, basado en la recuperación de la forma símplex, reduciendo a su vez el tamaño de la región de experimentación.

La motivación de esta investigación se basa en el aumento en la cantidad de tratamientos experimentales que deviene de incrementar la cantidad de componentes y el grado del polinomio máximo que se puede estimar. Por otro lado, se sabe que como consecuencia de la aplicación de los principios del DdE clásico que la ejecución de diseños experimentales resulta en una actividad onerosa en la mayoría de las ocasiones, por lo que cualquier intento de reducir los costos de ejecución puede resultar beneficioso en algunos contextos.

Los DdE con mezclas cuentan con varias particularidades que los separan de otros RSM. Una de estas y la de más interés en esta investigación es que al variar la forma del politopo, cambia la cantidad total de corridas que se deben ejecutar. Es decir, que una consecuencia de la forma símplex es que al restringirse se cambia la geometría y por ende se requerirían mayor cantidad de tratamientos para cubrir la región de experimentación.

Para lidiar con ello se propuso un algoritmo que recupera la forma de politopo regular en los diseños experimentales con mezclas de vértices extremos generados por restricciones de frontera superior, haciendo posible aplicar un diseño símplex o L – Símplex (de forma regular), este es nombrado como CONVEXSIM y cuenta con dos variantes (I y S).

Estos algoritmos se contrastaron midiendo la reducción en la cantidad de corridas y criterios de optimalidad y eficiencia; esto variando la cantidad de componentes, el

grado del polinomio máximo que se podría estimar, la cantidad de restricciones sobre la frontera superior y el tamaño de estas, así como la presencia o no de puntos axiales.

Como efecto, se halló que ambos algoritmos proveen resultados equivalentes, además, se obtuvo evidencia en favor de la hipótesis de que formas geométricas más irregulares decantan en DdE de vértices extremos con mayor cantidad de tratamientos y se concluye que cuando  $p \geq 5$  puede resultar valioso tratar de reducir la cantidad de corridas ajustando la forma de la región de experimentación ya que se podrían obtener reducciones superiores en promedio al 70 %.

Se resalta el hecho de que el objetivo de esta investigación no pretende suplantar la pericia y experticia de las personas experimentadoras, sino que busca fungir como una guía para la toma de decisiones en contextos de experimentación clásica con mezclas.

Por otro lado, se comparó la eficiencia y eficacia de los diseños, tanto en proporciones como en pseudocomponente para la región simple original, la región generada por vértices extremos y la región generada con el algoritmo propuesto.

Se concluye que al aplicar el algoritmo CONVEXSIM es más ventajoso realizar la estimación de los modelos en la región de experimentación parametrizada en pseudocomponentes y luego regresar las estimaciones a las unidades originales; ya que los diseños de vértices extremos con restricciones de frontera generadas por el algoritmo propuesto crearon diseños competentes y capaces de cumplir con los criterios de optimalidad estudiados. Con respecto a la eficiencia se comprobó que en L – pseudocomponentes en la mayoría de los casos los diseños resultaron ser A, D y G eficientes, un hallazgo que es concordante con la reducción mostrada en la cantidad de corridas experimentales.

De la aplicación del caso práctico, realizado a modo de ejemplo del uso del algoritmo CONVEXSIM, se concluye que la reducción del tamaño de la región de experimentación no afecta a la consecución de los objetivos, permitiendo resultados

razonables en relación con el contexto del experimento y la optimización simultánea de las variables de respuesta.

Por ejemplo, en el caso aplicado, es más económico emplear experimentación secuencial y hacer un segundo experimento en otra región de experimentación si fuese necesario al no alcanzar los resultados deseados, que aplicar el experimento original de vértices extremos. Ya que dos experimentos secuenciales con CONVEXSIM tomarían  $14 \cdot 2 = 28$  corridas en contraposición de las 64 corridas del diseño de vértices extremos.

En resumen, la gran conclusión de este trabajo es que hace notar que la recuperación de la forma símpex de los diseños de vértices extremos conduce a un ahorro importante en la cantidad de tratamientos, pudiendo producir diseños eficaces, en términos de la optimalidad, y eficientes en función de la cantidad de corridas experimentales y de términos a estimar en el modelo.

No obstante, se enfatiza que este algoritmo no es la panacea, y que su principal desventaja recae en que al reducir la región de experimentación el óptimo global puede quedar fuera y nunca ser estudiado. Es por tanto que su uso en la práctica queda sujeto al contexto del experimento y a la perspicacia y experiencia de las personas experimentadoras para moverse dentro de la región de operabilidad mediante el uso de la experimentación secuencial.

## 6.2 Limitaciones e investigaciones futuras

Esta tesis tuvo como pilar para los cálculos y la simulación realizada el uso del lenguaje de programación R y el IDE RStudio, enfocado principalmente en el paquete `{mixexp}` para la creación de los DdE con mezclas. La utilización de este paquete, único para este propósito, generó dos problemas principalmente: 1) la incapacidad de generar diseños con más 1 000 tratamientos, lo que provocó que se tuviera que desechar los casos con 11 y 12 componentes, pese a que la documentación avala su uso y 2) los fallos en la generación de diseños cuando  $m = 3$ , que incluían puntos que incumplían con las

restricciones de frontera impuestas. A partir de esto se tuvieron que desechar varios casos.

Otra limitación digna de mención es el aumento en la necesidad de recursos de hardware conforme aumentaba tanto la cantidad de componentes como el grado del polinomio a estimar, la razón principal de esto es la operación de matriz inversa, que resulta computacionalmente costosa. Esta condición es la que impidió probar otros escenarios más específicos, ya que el tiempo para generar todos los escenarios de simulación, usando dos computadores simultáneamente fue de aproximadamente 5 días.

Finalmente, el algoritmo diseñado se centra únicamente en la reducción de tratamientos experimentales en diseños de vértices extremos generados por restricciones en la frontera superior. No formó parte de los objetivos de esta investigación los diseños de vértices extremos creados a partir de restricciones sobre la frontera inferior y superior simultáneamente; empero, se pueden extrapolar algunas de las conclusiones, como por ejemplo que al recuperar la forma símplex regular se puede reducir considerablemente en función del contexto la cantidad de recursos de experimentación.

Como líneas de investigación futuras se sugiere la adaptación del algoritmo a DdE de mezclas de vértices extremos con restricciones de frontera tanto inferiores como superiores. Así como el estudio de criterios de optimalidad y eficiencia en modelos lineales generalizados.

**BIBLIOGRAFÍA**

- Arroyo Figueroa, G., Herrera Méndez, C. H., Dzul Cauch, J. G., Vargas Rodríguez, L., & Peña Caballero, V. (2016). Color in cosmetics made with cochinea. *Acta Universitaria*, 26(NE-1), 3–7. <https://doi.org/10.15174/au.2016.836>
- Auguie, B. (2017). *gridExtra: Miscellaneous Functions for “Grid” Graphics*. <https://CRAN.R-project.org/package=gridExtra>
- Box, G. E. P. (1999). Statistics as a Catalyst to Learning by Scientific Method Part II—A Discussion. *Journal of Quality Technology*, 31(1), 16–29. <https://doi.org/10.1080/00224065.1999.11979890>
- Box, G. E. P., & Behnken, D. W. (1960). Some New Three Level Designs for the Study of Quantitative Variables. *Technometrics*, 2(4), 455–475. <https://doi.org/10.1080/00401706.1960.10489912>
- Box, G. E. P., & Draper, N. R. (2007). *Response Surfaces, Mixtures, and Ridge Analyses*. Wiley. <https://doi.org/10.1002/0470072768>
- Box, G. E. P., & Wilson, K. B. (1951). On the Experimental Attainment of Optimum Conditions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 13(1), 1–38. <https://doi.org/10.1111/j.2517-6161.1951.tb00067.x>
- Carranza Chusán, S. I. (2019). *Propuesta para la creación de una empresa productora de exfoliantes naturales de terrón de azúcar*. [Trabajo de titulación, Universidad Católica de Santiago de Guayaquil]. <http://repositorio.ucsg.edu.ec/handle/3317/13854>
- Claringbold, P. J. (1955). Use of the Simplex Design in the Study of Joint Action of Related Hormones. *Biometrics*, 11(2), 174. <https://doi.org/10.2307/3001794>
- Cornell, J. A. (2002). *Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data* (3rd ed.). Wiley. <https://doi.org/10.1002/9781118204221>

- Cornell, J. A., & Gorman, J. W. (1984). Fractional Design Plans for Process Variables in Mixture Experiments. *Journal of Quality Technology*, 16(1), 20–38. <https://doi.org/10.1080/00224065.1984.11978883>
- Crosier, R. B. (1984). Mixture Experiments: Geometry and Pseudocomponents. *Technometrics*, 26(3), 209. <https://doi.org/10.2307/1267546>
- Derringer, G., & Suich, R. (1980). Simultaneous Optimization of Several Response Variables. *Journal of Quality Technology*, 12(4), 214–219. <https://doi.org/10.1080/00224065.1980.11980968>
- Draper, N. R., & Herzberg, A. M. (1971). On Lack of Fit. *Technometrics*, 13(2), 231–241. <https://doi.org/10.1080/00401706.1971.10488779>
- Draper, N. R., & St. John, R. C. (1977). A Mixtures Model with Inverse Terms. *Technometrics*, 19(1), 37–46. <https://doi.org/10.1080/00401706.1977.10489496>
- Espino Pérez, K. A., Zamora Rosales, J. D., Velázquez Martínez, M. V., & Ortiz Ciéneiga, J. B. (2019). Determinación de las propiedades de un surfactante para manos. *JÓVENES EN LA CIENCIA*, 5(1), 1–5. <https://www.jovenesenlaciencia.ugto.mx/index.php/jovenesenlaciencia/article/view/3055>
- Farfán, C., Garofalo Cedeño, E., Egresado, B., & García Muentes, E. (2022). Caracterización del propóleo del cantón Rocafuerte y su aplicación en la elaboración de un producto cosmético. *MQRInvestigar*, 6(4), 438–454. <https://doi.org/10.56048/MQR20225.6.4.2022.438-454>
- Fisher, R. (1971). *The Design of Experiments* (9th ed.). Collier Macmillan Publishers.
- Flores Martínez, R. (2019). *Modelo de negocio para la comercialización de un producto cosmético con base en componentes naturales* [Universidad Autónoma del Estado de Morelos]. <http://riaa.uaem.mx/xmlui/handle/20.500.12055/897>

- Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression (Third)*. Sage. <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>
- Friendly, M., Fox, J., & Chalmers, P. (2024). *matlib: Matrix Functions for Teaching and Learning Linear Algebra and Multivariate Statistics*. <https://CRAN.R-project.org/package=matlib>
- García-Goñi, S. (2024). Diseño de parcelas divididas para el desarrollo de una harina mejorada y funcional de granos germinados de lenteja o arveja. *Revista de Iniciación Científica*, 10(1), 36–47. <https://doi.org/10.33412/rev-ric.v10.1.4001>
- García-Goñi, S. (2025). *Algoritmo CONVEXSIM (V.1.1)*. Zenodo. <https://doi.org/10.5281/zenodo.14702664>
- García-Goñi, S., Garro Charpentier, F., González Fuentes, A., Madrigal Wachong, A. P., Mendoza Rodríguez, A. P., Mesén Jiménez, L. C., & Sandoval Atán, F. (2024). *Caso de aplicación de CONVEXSIM*. Zenodo. <https://doi.org/10.5281/zenodo.13334804>
- Garcillán López Rúa, M. de. (2014). La psicología del consumidor a la hora de comprar un producto cosmético. *Revista de Análisis Transaccional y Psicología Humanista*, ISSN 0212-9876, N°. 71, 2014, Págs. 589-600, 71, 589–600. <https://dialnet.unirioja.es/servlet/articulo?codigo=5614095>
- Gómez-Fierro, W. A., Del, M., Aroca-Osorio, M., & Díaz-Medina, D. (2022). Eficacia de Moringa oleífera para la elaboración de productos cosméticos. *Pädi Boletín Científico de Ciencias Básicas e Ingenierías Del ICBI*, 10(Especial7), 134–138. <https://doi.org/10.29057/ICBI.V10IESPECIAL7.9697>
- González-Pérez, L. M., Figueredo-Urbina, C. J., Luna-Rodríguez, L., Robles Ortiz, D., & Medina-Pérez, G. (2024). Una breve revisión de la composición y valor nutracéutico de la miel de Apis mellifera. *Boletín de Ciencias Agropecuarias Del ICAP*, 10(20), 1–9. <https://doi.org/10.29057/icap.v10i20.12886>

- Goos, P., & Jones, B. (2011). *Optimal Design of Experiments*. Wiley.  
<https://doi.org/10.1002/9781119974017>
- Goos, P., Jones, B., & Syafitri, U. (2016). I-Optimal Design of Mixture Experiments. *Journal of the American Statistical Association*, 111(514), 899–911.  
<https://doi.org/10.1080/01621459.2015.1136632>
- Gutiérrez Pulido, H., & de la Vara Salazar, R. (2012). *Análisis y diseño de experimentos* (3ra ed.). McGraw-Hill.
- Jones, M. C., & Aitchison, J. (1987). The Statistical Analysis of Compositional Data. *Journal of the Royal Statistical Society. Series A (General)*, 150(4), 396.  
<https://doi.org/10.2307/2982045>
- Koons, G. F. (1989). Effect of Sinter Composition on Emissions: A Multi-Component, Highly-Constrained Mixture Experiment. *Journal of Quality Technology*, 21(4), 261–267. <https://doi.org/10.1080/00224065.1989.11979184>
- Kuhn, M. (2016). *desirability: Function Optimization and Ranking via Desirability Functions*.  
<https://cran.r-project.org/package=desirability>
- Kurotori, I. S. (1966). Experiments with mixtures of components having lower bounds. *Industrial Quality Control*, 22(11), 592–596.
- Lawson, J. (2014). *Design and Analysis of Experiments with R*. Chapman and Hall/CRC.  
<https://doi.org/10.1201/b17883>
- Lawson, J., & Krennrich, G. (2023). *daewr: Design and Analysis of Experiments with R*.  
<https://CRAN.R-project.org/package=daewr>
- Lawson, J., & Willden, C. (2016). Mixture Experiments in R Using **mixexp**. *Journal of Statistical Software*, 72(Code Snippet 2), 1–20. <https://doi.org/10.18637/jss.v072.c02>

- Marquardt, D. W. (1970). Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation. *Technometrics*, 12(3), 591. <https://doi.org/10.2307/1267205>
- Marquardt, D. W., & Snee, R. D. (1974). Test Statistics for Mixture Models. *Technometrics*, 16(4), 533. <https://doi.org/10.2307/1267604>
- Martin, R. J., Bursnall, M. C., & Stillman, E. C. (1999). Efficient designs for constrained mixture experiments. *Statistics and Computing*, 9(3), 229–237. <https://doi.org/10.1023/A:1008978116760>
- Martínez García, S. (2022). *Datos Composicionales: Geometría de Aitchison*. <http://hdl.handle.net/10366/151477>
- McDonald, G. C. (2009). Ridge regression. *WIREs Computational Statistics*, 1(1), 93–100. <https://doi.org/10.1002/wics.14>
- McLean, R. A., & Anderson, V. L. (1966). Extreme Vertices Design of Mixture Experiments. *Technometrics*, 8(3), 447–454. <https://doi.org/10.1080/00401706.1966.10490377>
- Mee, R. W. (2007). Optimal Three-Level Designs for Response Surfaces in Spherical Experimental Regions. *Journal of Quality Technology*, 39(4), 340–354. <https://doi.org/10.1080/00224065.2007.11917700>
- Montgomery, D. (2017). *Design and analysis of experiments* (9th ed.). John Wiley & Sons, Inc.
- Montgomery, D., & Voth, S. R. (1994). Multicollinearity and Leverage in Mixture Experiments. *Journal of Quality Technology*, 26(2), 96–108. <https://doi.org/10.1080/00224065.1994.11979510>

- Myers, R., Montgomery, D., & Anderson-Cook, C. (2009). *Response Surface Methodology: Process and Product Optimization Using Designed Experiments* (D. Balding & N. Cressie, Eds.; 3rd ed.). John Wiley & Sons, Inc.
- Nigam, A. K., Gupta, S. C., & Gupta, S. (1983). A New Algorithm for Extreme Vertices Designs for Linear Mixture Models. *Technometrics*, 25(4), 367. <https://doi.org/10.2307/1267858>
- Ortega-Pérez, D. C., Bustamante-Rúa, M. O., Gutiérrez-Rôa, D. F., & Correa-Espinal, A. A. (2015). Mixture experiments in industrial formulations. *DYNA*, 82(189), 149–156. <https://doi.org/10.15446/dyna.v82n189.42785>
- Piepel, G. F. (1983a). Calculating Centroids in Constrained Mixture Experiments. *Technometrics*, 25(3), 279. <https://doi.org/10.2307/1268613>
- Piepel, G. F. (1983b). Defining Consistent Constraint Regions in Mixture Experiments. *Technometrics*, 25(1), 97. <https://doi.org/10.2307/1267731>
- Piepel, G. F. (1988). Programs for Generating Extreme Vertices and Centroids of Linearly Constrained Experimental Regions. *Journal of Quality Technology*, 20(2), 125–139. <https://doi.org/10.1080/00224065.1988.11979093>
- Piepel, G. F., & Cornell, J. A. (1985). Models for Mixture Experiments When the Response Depends on the Total Amount. *Technometrics*, 27(3), 219–227. <https://doi.org/10.1080/00401706.1985.10488046>
- Piepel, G. F., & Cornell, J. A. (1987). Designs for Mixture-Amount Experiments. *Journal of Quality Technology*, 19(1), 11–28. <https://doi.org/10.1080/00224065.1987.11979029>
- Posit team. (2023). *RStudio: Integrated Development Environment for R*. <http://www.posit.co/>

- Quenouille, M. H. (1953). *The Design and Analysis of Experiments*. Charles Griffin and Company.
- Quenouille, M. H. (1959). Experiments with Mixtures. *Journal of the Royal Statistical Society: Series B (Methodological)*, 21(1), 201–202. <https://doi.org/10.1111/j.2517-6161.1959.tb00330.x>
- Quispe, A. M., Alvarez-Valdivia, M. G., & Loli-Guevara, S. (2020). Metodologías Cuantitativas 2: Sesgo de confusión y cómo controlar un confusor. *Revista Del Cuerpo Médico Del HNAAA*, 13(2), 205–212. <https://doi.org/10.35434/rcmhnaaa.2020.132.675>
- Quizphe Sagasti, J. A. (2018). *Análisis del comportamiento de compra del producto cosmético “bases de maquillaje” de los clientes de MYA Makeup Studio*. [Universidad Católica de Santiago de Guayaquil]. <http://repositorio.ucsg.edu.ec/handle/3317/9970>
- R Core Team. (2024). *R: A Language and Environment for Statistical Computing*. <https://www.R-project.org/>
- Sarkar, D. (2008). *Lattice: Multivariate Data Visualization with R*. Springer. <http://lmdvr.r-forge.r-project.org>
- Schauberger, P., & Walker, A. (2024). *openxlsx: Read, Write and Edit xlsx Files*. <https://CRAN.R-project.org/package=openxlsx>
- Scheffé, H. (1958). Experiments with Mixtures. *Journal of the Royal Statistical Society: Series B (Methodological)*, 20(2), 344–360. <https://doi.org/10.1111/j.2517-6161.1958.tb00299.x>
- Scheffé, H. (1963). The Simplex-Centroid Design for Experiments with Mixtures. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 25(2), 235–251. <https://doi.org/10.1111/j.2517-6161.1963.tb00506.x>

- Shapiro, S. (2008). Causation, bias and confounding: a hitchhiker's guide to the epidemiological galaxy.: Introductory comments. *Journal of Family Planning and Reproductive Health Care*, 34(2), 81–82. <https://doi.org/10.1783/jfp.34.2.81>
- Smith, W. F. (2005). *Experimental Design for Formulation*. Society for Industrial and Applied Mathematics. <https://doi.org/10.1137/1.9780898718393>
- Snee, R. D. (1975). Experimental Designs for Quadratic Models in Constrained Mixture Spaces. *Technometrics*, 17(2), 149. <https://doi.org/10.2307/1268345>
- Snee, R. D. (1979). Experimental designs for mixture systems with multicomponent constraints. *Communications in Statistics - Theory and Methods*, 8(4), 303–326. <https://doi.org/10.1080/03610927908827762>
- St. John, R. C. (1984). Experiments with Mixtures, Ill-Conditioning, and Ridge Regression. *Journal of Quality Technology*, 16(2), 81–96. <https://doi.org/10.1080/00224065.1984.11978895>
- Taguchi, G. (1987). *System of experimental design; engineering methods to optimize quality and minimize costs*.
- Torres Mejía, F., Torres Mejía, J. A., Bautista Cruz, M. D., & Pérez Licon, E. (2023). Análisis fisicoquímico de miel de tres especies de abejas en el Oriente de Honduras. *Ciencia Latina Revista Científica Multidisciplinar*, 7(1), 10691–10713. [https://doi.org/10.37811/cl\\_rcm.v7i1.5241](https://doi.org/10.37811/cl_rcm.v7i1.5241)
- Warnes, G. R., Bolker, B., Lumley, T., Magnusson, A., Venables, B., Ryodan, G., & Moeller, S. (2023). *gtools: Various R Programming Tools*. <https://CRAN.R-project.org/package=gtools>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., ... Yutani, H. (2019). Welcome to the

Tidyverse. *Journal of Open Source Software*, 4(43), 140-1686.  
<https://doi.org/10.21105/joss.01686>