

Diseño de experimentos y análisis de conjunto

Autor: Christian Rosero Barzola, Ph.D.

DISEÑO DE EXPERIMENTOS Y ANÁLISIS DE CONJUNTO.

AUTOR:

Christian Xavier Rosero Barzola, Ph.D.

2021

TÍTULO

DISEÑO DE EXPERIMENTOS Y ANÁLISIS DE CONJUNTO.

AUTOR:

Christian Xavier Rosero Barzola, Ph.D.

AÑO

2021

EDICIÓN

Econ. César Augusto Pozo Estupiñán - Departamento de Publicaciones
Universidad ECOTEC

ISBN

978-9942-960-59-7

No. PÁGINAS

151

LUGAR DE EDICIÓN

Samborondón - Ecuador

DISEÑO DE CARÁTULA

Ing. Annabell Esperanza Aguilar Muñoz - Departamento de Relaciones
Públicas y Marketing.
Universidad ECOTEC

NOTA EDITORIAL:

El trabajo de investigación que conforman los capítulos del presente libro son resultado de investigación por parte del autor en busca de contribuir con el desarrollo académico de los estudiantes de grado y posgrado, tributando a la Línea de Investigación de "Teoría y Desarrollo económico", de la Universidad ECOTEC. El autor de esta obra tuvo la responsabilidad de seleccionar dichas investigaciones científicas, tomando en consideración el impacto y relevancia de la información, en virtud de la difusión del conocimiento.

CONTENIDO

DATOS DEL AUTOR.....	5
PRÓLOGO.....	6
PRESENTACIÓN.....	7
CAPÍTULO 1: EXPERIMENTACIÓN.....	9
1.1. Principios básicos del diseño experimental.....	10
1.2. Realización de réplicas.....	10
1.3. Aleatorización.....	10
1.4. Formación de bloques.....	11
1.5. Estrategias de experimentación.....	11
1.9. Pautas generales para diseñar experimentos.....	13
1.10. Conclusiones y recomendaciones.....	17
1.11. Referencias bibliográficas.....	18
CAPÍTULO 2: DISEÑOS FACTORIALES.....	19
2.1. Ventaja de los diseños factoriales.....	19
2.2. Comparaciones múltiples.....	20
2.3. Elección del tamaño de la muestra.....	20
2.4. El supuesto de no interacción en un modelo de dos factores.....	21
2.5. Una observación por celda.....	21
2.6. Diseño factorial general.....	23
2.7. Ajuste de curvas y superficies de respuesta.....	24
2.8. Formación de bloques en un diseño factorial.....	24
2.9. El diseño 2^2	25
2.10. El diseño 2^3	25
2.11. El diseño general 2^k	27
2.12. Una sola réplica del diseño 2^k o diseño factorial no replicado.....	28
2.13. Adición de puntos centrales en el diseño 2^k	29
2.14. Formación de bloques de un diseño factorial 2^k con réplicas.....	30
2.15. Confusión del diseño factorial 2^k	31
2.16. Confusión del diseño factorial 2^k en dos bloques.....	31
2.17. Métodos adicionales para la construcción de bloques.....	32
2.18. Estimación del error.....	32
2.19. La fracción un medio del diseño 2^k	35
2.20. Resolución del diseño.....	35
2.21. Construcción de fracciones un medio.....	36
2.22. Secuencias de diseños factoriales fraccionados.....	36
2.23. La fracción un cuarto del diseño 2^k	37

Diseño de Experimentos y Análisis de Conjunto

2.24.	El diseño factorial fraccionado 2^{k-p} general	37
2.25.	Análisis de los diseños factoriales fraccionados 2^{k-p}	38
2.26.	Proyección del diseño factorial fraccionado 2^{k-p}	38
2.27.	Diseños de resolución III	39
2.28.	La relación de definición para un diseño de doblez	40
2.29.	Diseños de Plackett-Burman.....	40
2.30.	Diseños de resolución IV y V	40
2.31.	Diseños factoriales y factoriales fraccionados con tres niveles y con niveles mixtos. 41	
2.32.	El diseño 3^2	42
2.33.	El diseño 3^3	42
2.34.	El diseño general 3^k	43
2.35.	Confusión en el diseño factorial 3^k en tres bloques	44
2.36.	El diseño factorial 3^k en nueve bloques.....	45
2.37.	El diseño factorial 3^k en 3^p bloques.....	46
2.38.	La fracción un tercio del diseño factorial 3^k	46
2.39.	Otros diseños factoriales fraccionados 3^{k-p}	46
2.40.	Modelos de regresión lineal	48
2.41.	Estimación de los parámetros en modelos de regresión lineal	48
2.42.	Estimación de σ^2	49
2.43.	Propiedades de los estimadores	49
2.44.	Prueba de significación de la regresión	50
2.45.	Pruebas de los coeficientes de regresión individuales y de grupos de coeficientes 50	
2.46.	Intervalos de confianza para los coeficientes de regresión individuales.....	50
2.47.	Predicción de nuevas observaciones de la respuesta	51
2.48.	Residuales estandarizados y studentizados	51
2.49.	Residuales PRESS	52
2.50.	R- student.....	52
2.51.	Diagnóstico de influencia.....	53
2.52.	Referencias bibliográficas.....	53
CAPÍTULO 3: ANÁLISIS CONJUNTO.....		54
3.1	Modelo ANOVA	54
3.2	Experimentación de elección discreta	57
3.3	Modelos de utilidad aleatorios	58
3.4	El modelo logístico.....	58
3.5	Principios del diseño experimental	62
3.6	Diseño experimental versus tratamiento	63
3.7	Diseños de bloques incompletos equilibrados y diseños 3	64
3.8	Experimentos factoriales	65

3.9	Matrices de Hadamard y matrices ortogonales.....	68
3.10	Diseños Foldover	71
3.11	Experimentos de mezcla.....	72
3.12	Estimación	75
3.13	Modelos de probabilidad	75
3.14	Estimación lineal: Mínimos cuadrados y mínimos cuadrados ponderados.....	76
3.15	Estimación de máxima verosimilitud	78
3.16	Pruebas de hipótesis lineales	79
3.17	Diseños con atributos ordenados	79
3.18	Efectos lineales, cuadráticos y cúbicos.....	80
3.19	Componentes de interacción: Lineal y Cuadrática	82
3.20	Diseños óptimos de Pareto	84
3.21	Inferencias sobre efectos principales	87
3.22	Teorema 4.1.....	87
3.23	Inferencias sobre los efectos principales en experimentos de 2^m	88
3.24	Teorema 4.2.....	88
3.25	Polinomios ortogonales.....	89
3.26	Tasa de sustitución de los atributos	90
3.27	Referencias bibliográficas.....	91
CAPITULO 4: CASOS.....		92
4.1	Caso 1: Evaluación de los métodos aplicados en el análisis conjunto	92
4.2	Caso 2: Optimización conjunta, un algoritmo exacto de ramificación y límite para el problema de compartir de elección.....	98
4.3	Caso 3: Diseño de un sistema de análisis interactivo conjunto basado en web....	111
4.4	Caso 4: Justificación y diseño de experimentos de elección moderada: una aplicación de precios estratégica.....	116
4.5	Caso 5: Procesamiento de información en experimentos conjuntos basados en la elección.....	123
4.6	Caso 6: Segmentación de Mercado con Análisis Conjunto basado en la Elección	141

Índices de Figuras

Figura 1. Pautas para el diseño experimental 14
Figura 2. Pasos para el diseño general de 2^k 28

Índices de Tablas

Tabla 1. Fórmulas para la elección del tamaño de la muestra 21
Tabla 2. Observaciones por celda 22
Tabla 3. Corridas del diseño 2^3 26

DATOS DEL AUTOR

CHRISTIAN XAVIER ROSERO BARZOLA, Ph.D.

Economista con mención en Gestión Empresarial especialización Finanzas, Escuela Superior Politécnica del Litoral.

Magister en Economía y Dirección de Empresas, Escuela Superior Politécnica del Litoral.

Doctor en Administración Estratégica de Empresas, Pontificia Universidad Católica del Perú.

Diploma Superior en Economía del Ecuador y del Mundo, Universidad de Guayaquil.

Diploma Superior en Economía Internacional, Universidad de Guayaquil.

Diploma Superior en Gestión Empresarial Internacional, Universidad de Guayaquil.

Diploma Superior en Diseño Gestión y Evaluación de Proyectos de Desarrollo, Facultad Latinoamericana de Ciencias Sociales.

Áreas de Especialidad:

- Modelos financieros y de riesgos.
- Diseño y gestión de proyectos de inversión y mejoramiento en los sectores públicos y privados.
- Diseño modelos matemáticos y econométricos.
- Diagnósticos del sector empresarial y Metodología Six Sigma y procesos (CRM, RFS, CSCM, Black Belt Six Sigma).
- Experiencia en el sector público y privado como especialista en planificación, finanzas, gestión de proyectos e investigación a nivel nacional e internacional.
- Catedrático universitario de pregrado y posgrado

PRÓLOGO

La experimentación se utiliza para revelar el comportamiento de un proceso o sistema en particular, puede definirse como una prueba o una serie de pruebas en las que se efectúan cambios deliberados en las variables de entrada para identificar las razones de las variaciones que se puedan observar en la respuesta de la variable de salida. El objetivo principal del diseño de un experimento es desarrollar un proceso robusto, es decir, que no pueda ser afectado por los factores externos y poder observar directamente los resultados que se proponen en una investigación. El principal propósito de este libro es abordar teóricamente los diferentes diseños experimentales y corroborar su aplicación, con el fin de tener una guía que asegure la validez y confiabilidad de los procesos de investigación.

El análisis conjunto es una técnica de investigación multivariada en donde el principal objetivo es comprender de qué manera las personas toman decisiones y lo que valoran de los productos y/o servicios. A este método también se lo conoce como estimación discreta de elección o investigación de preferencias declaradas. Esta herramienta permite presentar diferentes opciones a las personas y luego analizar cuáles fueron los impulsores de sus decisiones, también es denominado modelo de descomposición porque puede desagregar las preferencias para determinar el valor de cada atributo del factor. Esta técnica difiere de otras herramientas multivariantes en tres áreas distintas: (a) su naturaleza de técnica de descomposición, (b) el hecho de que las estimaciones pueden hacerse a nivel individual, y (c) su flexibilidad en términos de relaciones entre variables dependientes e independientes.

A lo largo del libro se desarrolla un análisis de la fundamentación teórica y la aplicabilidad de los diseños experimentales y la aplicación en una herramienta de análisis multivariado, como es el análisis de conjunto, para desarrollar y exponer una guía que pueda ser utilizada en la investigación experimental.

PRESENTACIÓN

Los investigadores han utilizado los experimentos para poder llevar a cabo estudios y poder obtener conclusiones. El experimento o prueba se lo practica para descubrir algo acerca de un proceso o sistema particular. Este puede definirse como una prueba o una serie de pruebas en las que realizan cambios deliberados en las variables de entrada de un proceso o sistema para identificar las razones de las variaciones que se puedan observar en la respuesta de salida.

El análisis conjunto es una técnica avanzada de investigación de mercado en donde el principal objetivo es comprender de qué manera las personas toman decisiones y lo que valoran en productos y servicios. A este método también se lo conoce como estimación discreta de elección o investigación de preferencia declarada. El análisis conjunto implica presentar a las personas con opciones y luego analizar cuáles fueron los impulsores de sus elecciones.

El capítulo 1 trata acerca de los principios básicos de los experimentos, de la planeación, realización y del análisis de los resultados con el objetivo de poder obtener conclusiones válidas. Se pueden contemplar varias ventajas de la utilización de los experimentos en especial en las áreas de ingeniería y las ciencias físicas y químicas. La experimentación es de gran utilidad para la realización de procesos robustos, es decir en donde el proceso se vea afectado de forma mínima por fuentes de variabilidad externas.

El capítulo 2 muestra las definiciones y principios básicos de los diseños factoriales. Debido a que en muchos experimentos intervienen los efectos de dos o más factores, tratar este tema es de gran importancia. Los diseños factoriales son los más eficientes para este tipo de experimentos. El punto principal del diseño factorial es poder investigar todas las combinaciones posibles en cada ensayo. Es decir, si el factor a tiene a niveles y el factor b tiene b niveles, cada réplica va a contener todas las ab combinaciones de los tratamientos.

El tema que aborda el capítulo 3 es el de análisis conjunto. El análisis conjunto y la experimentación de elección discreta, también conocido como análisis conjunto basado en la elección, son técnicas desarrolladas desde 1960 para evitar los ejemplos notables de estudios de preferencias reveladas. La estrategia con estos enfoques es hacer inferencias sobre la parte de los niveles de atributo de los encuestados declaró preferencias o decisiones declaradas.

El capítulo 4 muestra diferentes casos aplicados en investigaciones acerca del tema de análisis conjunto. El objetivo de este capítulo es que el lector pueda obtener información de la metodología que se implementa en las investigaciones utilizando el análisis conjunto y que puedan observar los resultados obtenidos en las diferentes investigaciones y las respectivas recomendaciones.

El Autor

CAPÍTULO 1: EXPERIMENTACIÓN

En investigación se llevan a cabo experimentos para descubrir el funcionamiento de un proceso o sistema, por lo que se consideran pruebas que evalúan la causalidad de las variables que se están estudiando. Contemplado desde una perspectiva más formal, un experimento puede ser una prueba o un conjunto de pruebas en las que se realizan cambios en las variables de entrada de un proceso o sistema para observar e identificar las razones de las variaciones que pudieran observarse en la respuesta de salida. El objetivo principal sería desarrollar un proceso robusto, es decir, un proceso el cual no sea afectado de gran manera por factores externos. La forma en la que se recolectan los datos influye directamente en los resultados y conclusiones de un experimento.

El proceso puede visualizarse como una mezcla de personas, máquinas, métodos, u otros recursos que transforman cierta entrada (con frecuencia un material) en una salida que tiene una o más respuestas observables. Algunas variables del proceso x_1, x_2, \dots, x_p son controlables, mientras que otras z_1, z_2, \dots, z_q son no controlables (aunque pueden serlo para los fines de una prueba). El objeto principal del experimentador es determinar la influencia que tienen estos factores sobre la respuesta de salida del sistema, generalmente se utiliza un enfoque conocido como estrategia de experimentación.

Entre los objetivos principales de un experimento comprenden los siguientes:

1. Determinar cuáles son las variables que tienen mayor influencia sobre la respuesta y (variable dependiente o respuesta).
2. Determinar cuál es el ajuste de las x que tiene mayor influencia para que y esté casi siempre cerca del valor nominal deseado.
3. Determinar cuál es el ajuste de las x que tiene mayor influencia para que la variabilidad de y sea reducida.
4. Determinar cuál es el ajuste de las x que tiene mayor influencia para que los efectos de las variables no controlables z_1, z_2, \dots, z_q sean mínimos.

1.1. Principios básicos del diseño experimental

El diseño estadístico de experimentos se define como el proceso para llevar a cabo la planeación del experimento para poder recolectar los datos adecuados que puedan analizarse con métodos estadísticos que producirán conclusiones objetivas y válidas. El enfoque estadístico del diseño experimental es necesario si se necesitan conclusiones significativas de los datos. Si el problema contiene datos sujetos a errores experimentales, la metodología estadística es el único enfoque objetivo de análisis. De tal modo que cualquier problema experimental incluye dos aspectos: el diseño del experimento y el análisis estadístico de los datos.

1.2. Realización de réplicas

Se conoce como realización de réplicas a la repetición del experimento básico. Esta posee dos propiedades. Primera, el experimentador puede obtener una estimación del error experimental. La cual se transforma en una unidad de medición básica para establecer si las diferencias que se observaron en los datos no son estadísticamente iguales. Segunda, si se usa la media muestral (\bar{y}) para estimar el efecto de un factor en el experimento, la realización de réplicas permite al experimentador obtener una estimación más precisa de este efecto. Por ejemplo, si σ^2 es la varianza de una observación individual y hay n réplicas, la varianza de la media muestral es

$$\sigma_y^2 = \frac{\sigma^2}{n} \quad (1)$$

1.3. Aleatorización

La aleatorización es la base en la que se fundamenta el uso de los métodos estadísticos en el diseño experimental. Se conoce a la aleatorización mediante la asignación del material experimental como el orden en que se realizarán las pruebas o ensayos individuales del experimento determinados al azar. Uno de los requerimientos de los métodos estadísticos es que las observaciones o errores sean variables aleatorias con distribuciones independientes. La aleatorización por lo general valida este supuesto.

Ayuda también a sacar el promedio de los efectos de los factores extraños que pudieran estar presentes. Es muy común el uso de software para auxiliar a los experimentadores a seleccionar y construir diseños experimentales. Mediante estos programas se pueden presentar las corridas del diseño experimental de manera aleatoria utilizando un generador de números aleatorios. Los programas más utilizados en este campo son SPSS, Minitab, Stata.

1.4. Formación de bloques

Esta es una técnica de diseño que se utiliza para mejorar la precisión de las comparaciones que se hacen entre los factores de interés. Se emplea esta técnica para poder reducir o eliminar la variabilidad transmitida por variables perturbadoras; aquellos factores que pueden intervenir en la respuesta experimental, pero en los que no se tiene un interés específico.

Agruparemos las unidades experimentales en bloques "homogéneos" donde todos los niveles del factor principal están igualmente representados. El análisis de varianza de un diseño de bloque de control aleatorizado divide el término residual de un factor único equivalente a diseño completo aleatorizado en bloques y componentes residuales. Cabe señalar, sin embargo, que este último componente tiene menos grados de libertad que en los diseños de factor único, lo que conduce a mayores estimaciones para

$$MS_{residual} = \frac{SS_{Residual}}{df_{MS_{Residual}}} = \frac{SS_{Residual}}{df} \quad (2)$$

1.5. Estrategias de experimentación

En el marco de la experimentación se presentan distintas estrategias o enfoques para poder llegar a conclusiones precisas y objetivas. Estos enfoques son los siguientes: enfoque de la mejor conjetura, enfoque de un factor a la vez y el experimento factorial.

1.6. Enfoque de la mejor conjetura

El enfoque de la mejor conjetura consiste en seleccionar una combinación arbitraria de los factores, probarlos y ver qué ocurre. Es una estrategia común, utilizada principalmente por ingenieros y científicos. Funciona correctamente si los experimentadores cuentan con una gran cantidad de conocimientos teóricos o técnicos del sistema en estudio y una vasta experiencia práctica. Sin embargo, este enfoque presenta dos desventajas. Primera, supóngase que la conjetura inicial no produce los resultados deseados, entonces el experimentador tiene que hacer otra conjetura acerca de la combinación correcta de los niveles de los factores. Esto podría continuar, sin garantía alguna de éxito. Segunda, supóngase que la mejor conjetura inicial produce un resultado satisfactorio. Entonces, el experimentador se ve tentado a suspender las pruebas, aun cuando no hay ninguna garantía de que se ha encontrado la mejor solución.

1.7. Enfoque de un factor a la vez

El enfoque de un factor a la vez consiste en seleccionar un punto de partida o línea base de los niveles para cada factor en su rango manteniendo siempre constantes las variables restantes en el nivel base y es una estrategia muy común en la práctica. Luego de haber realizado todas las pruebas se construye una serie de gráficas en las que se muestra la forma en que la variable de respuesta es afectada al variar cada factor manteniendo los demás factores constantes.

La desventaja principal de la estrategia de un factor a la vez es que no puede tomar en consideración cualquier posible interacción entre los factores. Hay una interacción cuando uno de los factores no produce el mismo efecto en la respuesta con niveles diferentes de otro factor. Las interacciones entre factores son muy comunes y, en caso de existir, la estrategia de un factor a la vez casi siempre producirá resultados deficientes. Los experimentos de un factor a la vez siempre son menos eficientes que otros métodos basados en un enfoque estadístico del diseño experimental.

1.8. Experimento factorial

El experimento factorial es recomendado si se quiere trabajar con varios factores. Esta es una estrategia experimental en la que los factores se hacen variar en conjunto, en lugar de uno a la vez. Además, existe el experimento factorial fraccionado el cual es una variación del diseño factorial básico en donde se realiza un subconjunto de las corridas.

Típicamente, hay muchos factores como género, genotipo, dieta, condiciones de vivienda, protocolos experimentales, interacciones sociales y edad que pueden influir en el resultado de un experimento. Estos a menudo necesitan ser investigados con el fin de determinar la generalidad de una respuesta. Puede ser importante saber si una respuesta sólo se ve en, por ejemplo, las mujeres, pero no los hombres. Una manera de hacer esto sería hacer experimentos separados en cada sexo. Este enfoque de una variable a la vez es, sin embargo, un derroche de recursos científicos. Una alternativa mucho mejor es incluir ambos sexos o más de una cepa, etc. en un solo experimento "factorial". Tales diseños pueden incluir varios factores sin utilizar un número excesivo de sujetos experimentales.

1.9. Pautas generales para diseñar experimentos

El gráfico 1 muestra las pautas generales para el diseño de experimentos. Los siguientes pasos representan una importante metodología para un funcionamiento eficaz del enfoque estadístico en el diseño y análisis del experimento de los cuales se debe de conocer desde un principio que es lo que se va a estudiar, de qué forma se recolectarán los datos, y una comprensión cualitativa de la forma en que van a ser analizados. Los pasos son los siguientes: identificar el problema, elegir los factores, los niveles y los rangos, seleccionar la variable de respuesta, elegir el diseño experimental, realizar el experimento, analizar estadísticamente los datos, realizar las respectivas conclusiones.

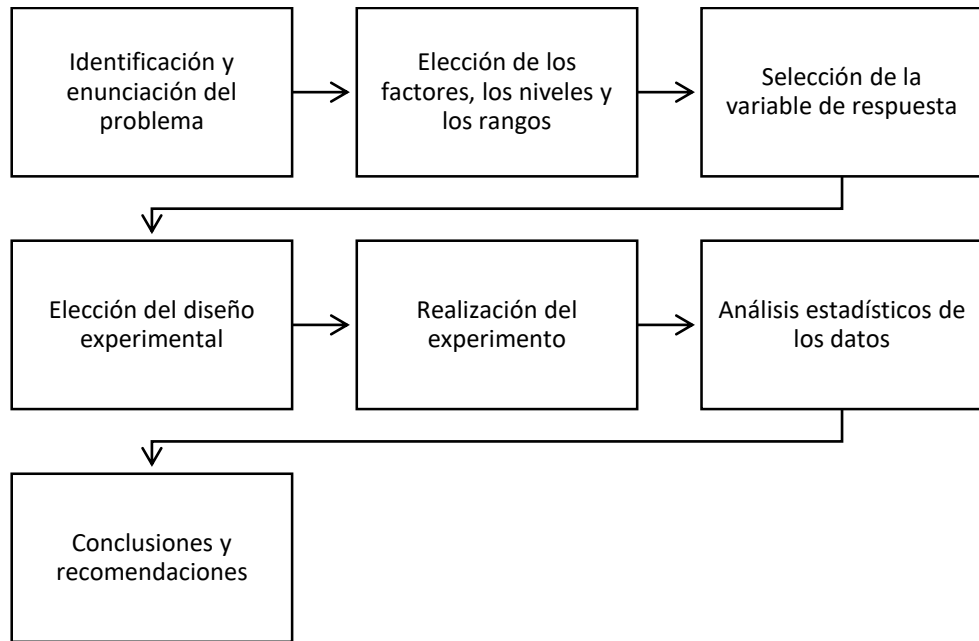


Figura 1. Pautas para el diseño experimental.

1.9.1. Identificación y enunciación del problema.

Es necesario desarrollar todas las ideas acerca de los objetivos del experimento, es importante solicitar aportaciones de todas las áreas involucradas: ingeniería, aseguramiento de calidad, manufactura, mercadotecnia, administración, el cliente y el personal de operación (el cual usualmente conoce a fondo el proceso pero que es frecuentemente ignorado). Por esta razón, se recomienda un enfoque de equipo para diseñar experimentos.

Usualmente es conveniente realizar una lista de los problemas o las preguntas específicas que van a abordarse en el experimento. Se debe de enunciar de forma clara el problema debido a que esto contribuye sustancialmente para alcanzar una mejor comprensión de los fenómenos bajo estudio y la solución final del problema.

1.9.2. Elección de los factores, los niveles y los rangos.

Los factores pueden ser clasificados como factores potenciales del diseño o como factores perturbadores. Los factores potenciales del diseño son aquellos que el experimentador posiblemente quiera hacer variar en el experimento.

Algunas clasificaciones útiles son factores del diseño, factores que se mantienen constantes y factores a los que se permite variar.

Los factores del diseño son los que se seleccionan realmente para estudiarlos en el experimento. Los factores que se mantienen constantes son variables que pueden tener cierto efecto sobre la respuesta, pero que para los fines del experimento en curso no son de interés, por lo que se mantendrán fijos en un nivel específico.

Por otro lado, los factores perturbadores pueden tener efectos considerables que deben tomarse en consideración, a pesar de que no haya interés en ellos en el contexto del experimento en curso. Los factores perturbadores suelen clasificarse como factores controlables, no controlables o de ruido. Un factor perturbador controlable es aquel cuyos niveles pueden ser ajustados por el experimentador. Por ejemplo, el experimentador puede seleccionar lotes diferentes de materia prima o diversos días de la semana para conducir el experimento. Si un factor perturbador no es controlable en el experimento, pero puede medirse, muchas veces puede usarse el procedimiento de análisis denominado análisis de covarianza para compensar este efecto.

Una vez que se han seleccionado los factores del diseño, se debe elegir los rangos en los que hará variar estos factores, así como los niveles específicos con los que se realizarán las corridas. También deberá pensarse cómo van a controlarse estos factores en los valores deseados y cómo van a medirse. También se requiere del conocimiento del proceso.

Cuando el objetivo del experimento es el tamizado de los factores o caracterización del proceso, por lo general es mejor mantener reducido el número de niveles de los factores. En general, dos niveles funcionan bastante bien en los estudios de tamizado de factores. Elegir la región de interés también es importante. En el tamizado de factores, la región de interés deberá ser relativamente grande; es decir, el rango en el que se hacen variar los factores deberá ser amplio. Conforme se sepa más acerca de las variables que son importantes y de los niveles que producen los mejores resultados, la región de interés se hará por lo general más estrecha.

1.9.3. Selección de la variable de respuesta

Para poder seleccionar la variable de respuesta, se debe de estar seguro de que esta variable proporciona la información útil acerca del proceso que está en estudio. Por lo general, el promedio o la desviación estándar de la característica medida será la variable de respuesta. No son la excepción las respuestas múltiples. Si la eficiencia de los instrumentos de medición es inadecuada, el experimentador sólo detectará los efectos relativamente grandes de los factores o quizá sean necesarias réplicas adicionales. En algunas situaciones en que la eficiencia de los instrumentos de medición es pobre, el experimentador puede decidir medir varias veces cada unidad experimental y usar el promedio de las mediciones repetidas como respuesta observada. Suele ser de importancia determinante identificar los aspectos relacionados con la definición de las respuestas de interés y cómo van a medirse antes de llevar a cabo el experimento.

1.9.4. Elección del diseño experimental

Este paso implica la consideración del tamaño de la muestra (número de réplicas), la selección de un orden de las pruebas para los ensayos experimentales y establecer si entran en juego o no la formación de bloques u otras restricciones sobre la aleatorización. Al seleccionar el diseño, es importante tener en mente los objetivos experimentales.

1.9.5. Realización del experimento

Se debe de monitorear el proceso de experimentación para poder asegurarse de que todo se esté haciendo de acuerdo a la planeación. Poner en un primer plano la planeación es crucial para el éxito. Se argumenta que antes de llevar a cabo el experimento, es muy importante en ciertas ocasiones llevar a cabo algunas corridas pilotos o de prueba. De esta forma se podrá observar la consistencia del material experimental, comprobación del sistema de medición, una idea aproximada del error experimental y la oportunidad de poner en práctica la técnica experimental global.

1.9.6. Análisis estadísticos de los datos

Se deben de utilizar métodos estadísticos para analizar los datos para que los resultados y las conclusiones no sean apreciativos sino totalmente objetivos. Si el experimento se lo ha diseñado de forma correcta y se lo ha manejado en concordancia con el diseño, los métodos estadísticos utilizados no deben de ser complejos. Debido a que muchas de las preguntas que el experimentador quiere responder pueden insertarse en el marco de la prueba de hipótesis, los procedimientos para probar hipótesis y estimar intervalos de confianza son muy útiles en el análisis de datos de un experimento diseñado. Muchas veces es muy útil también presentar los resultados de varios experimentos en términos de un modelo empírico, es decir, mediante una ecuación derivada de los datos que expresa la relación entre la respuesta y los factores importantes del diseño. El análisis residual y la verificación de la adecuación del modelo son también técnicas de análisis importantes.

Los métodos estadísticos no pueden demostrar que un factor (o factores) poseen un efecto particular, sólo proporcionan pautas generales en cuando a la confiabilidad y la validez de los resultados. Aplicados en forma correcta, los métodos estadísticos no permiten la demostración experimental de nada, pero si sirven para medir el error posible en una conclusión o asignar un nivel de confianza a un enunciado. La ventaja principal de los métodos estadísticos es que agregan objetividad al proceso de toma de decisiones.

1.10. Conclusiones y recomendaciones

Los métodos gráficos suelen ser útiles en esta etapa, en particular para presentar los resultados. También deberían de realizarse corridas de seguimiento o pruebas de confirmación para validar las conclusiones del experimento. A lo largo del proceso completo es importante tener presente que la experimentación es una parte esencial del proceso de aprendizaje, en la que se formulan hipótesis tentativas acerca de un sistema, se realizan experimentos para investigar estas hipótesis y se formulan nuevas hipótesis con base en los resultados, y así sucesivamente. Esto sugiere que la experimentación es iterativa.

Por lo general, es un gran error diseñar un solo experimento comprensivo y extenso al principio de un estudio. Un experimento exitoso requiere conocer los factores importantes, los rangos en los que deberán hacerse varios estos factores, el número apropiado de niveles que deberán usarse y las unidades de medición apropiadas para estas variables. Por consiguiente, la experimentación se hace en forma secuencial y, como regla general, no deberá invertirse más de 25% de los recursos disponibles en el primer experimento. Con esto se asegurará que se contará con los recursos suficientes para realizar las corridas de confirmación y que se alcanzará en última instancia el objetivo final del experimento.

1.11. Referencias bibliográficas

Montgomery, D. C. (2015). *Diseño y análisis de experimentos* (Segunda ed.). México: Limusa.

Pérez López, C. (2004). *Técnicas de análisis multivariante de datos. Aplicaciones con SPSS*. Madrid, Universidad Complutense de Madrid.: Pearson Educación.

Raghavarao, D., Wiley, J. B., & Chitturi, P. (2011). *Choice-Based Conjoint Analysis Models and Designs*. New York: CRC Press Taylor & Francis Group.

CAPÍTULO 2: DISEÑOS FACTORIALES

En varios experimentos interviene el estudio de los efectos de dos o más factores. Usualmente, los diseños factoriales tienden a ser los más eficientes para este tipo de experimentos. Se conoce como diseño factorial, a cada réplica completa del experimento que se investigan, tomando todas las posibles combinaciones de los niveles de los factores. Se define como efecto de un factor al cambio en la respuesta producido por un cambio en el nivel del factor. A este generalmente se le conoce como efecto principal porque se refiere a los factores de interés primario en el experimento.

Se puede presentar el concepto de interacción de otra manera. Si los dos factores del diseño tratado son cuantitativos, entonces una representación con un modelo de regresión del experimento factorial de dos factores puede representarse como

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_{12}x_1x_2 + \varepsilon \quad (3)$$

Donde y es la respuesta, las β son parámetros cuyos valores deben determinarse, x_1 es una variable que representa al factor A , x_2 es una variable que representa al factor B , y ε es un término del error aleatorio. Las variables x_1 y x_2 se definen en una escala codificada de -1 a +1 (los niveles bajo y alto de A y B), y x_1x_2 representa la interacción entre x_1 y x_2 . Las estimaciones de los parámetros en este modelo de regresión resultan estar relacionadas con las estimaciones de los efectos. Las estimaciones de los parámetros obtenidas de esta manera para el diseño factorial en el que todos los factores tienen dos niveles resultan ser estimaciones de mínimos cuadrados. Por lo tanto, una interacción es una forma de curvatura en el modelo de superficie de respuesta fundamental del experimento. Este modelo de superficie de respuesta de un experimento es de gran importancia y utilidad.

2.1. Ventaja de los diseños factoriales

Se conoce que los diseños factoriales pueden ser más eficientes que los experimentos de un factor a la vez. Asimismo, se prefiere utilizar el diseño factorial cuando se presentan interacciones con el fin de evitar obtener conclusiones incorrectas. Los diseños factoriales permiten la estimación de los efectos de un factor con diferentes niveles de los factores restantes, estableciendo conclusiones válidas para un rango de condiciones experimentales.

Las observaciones de un experimento factorial se pueden describir como un modelo. Este modelo puede escribirse de diferentes formas.

2.2. Comparaciones múltiples

Se utiliza la herramienta de comparaciones múltiples cuando el análisis de varianza revela que las medias de las columnas o renglones son diferentes, por lo cual se realizan comparaciones entre las medias individuales de los renglones o columnas de modo que se puedan identificar las diferencias. Usualmente, este tipo de comparaciones se las realiza a través de la prueba Tukey.

2.3. Elección del tamaño de la muestra

El experimentador puede determinar el tamaño de la muestra (número de réplicas) ideal para un diseño factorial de dos factores, apoyándose en las curvas de operación característica que se visualizan en la parte inferior.

Para el empleo de estas curvas se necesita encontrar el valor menor de ϕ^2 la cual pertenece a una diferencia especificada entre las medias de dos tratamientos cualesquiera. Por ejemplo, si la diferencia en las medias de dos renglones da como resultado D, entonces el valor mínimo de ϕ^2 es el siguiente

$$\phi^2 = \frac{nbD^2}{2a\sigma^2} \quad (4)$$

Mientras que, si la diferencia en las medias de dos columnas es D, el calor mínimo de ϕ^2 es

$$\phi^2 = \frac{naD^2}{2b\sigma^2} \quad (5)$$

Y si el valor mínimo de ϕ^2 que corresponde a una diferencia D entre dos efectos de interacción cualesquiera es

$$\phi^2 = \frac{nD^2}{2\sigma^2[(a-1)(b-1)+1]} \quad (6)$$

Tabla 1. Fórmulas para la elección del tamaño de la muestra

Factor	Φ^2	Grados de libertad del numerados	Grados de libertad del denominador
A	$\frac{bn \sum_{i=1}^a \tau_i^2}{a\sigma^2}$	$a - 1$	$ab(n - 1)$
B	$\frac{an \sum_{j=1}^b \beta_j^2}{b\sigma^2}$	$b - 1$	$ab(n - 1)$
AB	$\frac{n \sum_{i=1}^a \sum_{j=1}^b (\tau\beta)_{ij}^2}{\sigma^2[(a - 1)(b - 1) + 1]}$	$(a - 1)(b - 1)$	$ab(n - 1)$

Nota: Resumen de la elección del tamaño de la muestra. Tomado del libro *Diseño y análisis de experimentos* (Segunda ed.) por Montgomery, D. C. (2015). México: Limusa.

2.4. El supuesto de no interacción en un modelo de dos factores

En ciertas ocasiones, el experimentador decide que es apropiado utilizar un modelo de dos factores sin interacción como se muestra a continuación.

$$y_{ijk} = \mu + \tau_i + \beta_j + \varepsilon_{ijk} \quad \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \\ k = 1, 2, \dots, n \end{cases} \quad (7)$$

Para este tipo de modelos el experimentador debe de tener mucho cuidado al no utilizar los términos de interacción, debido a que la presentación de una interacción significativa puede tener un impacto dramático sobre la interpretación de los datos.

2.5. Una observación por celda

En ciertas ocasiones se encuentran experimentos de dos factores que cuentan con una sola réplica, es decir que solo tienen una observación por celda. Cuando sucede esto, el modelo de los efectos es el siguiente

$$y_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ij} \quad \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \end{cases} \quad (8)$$

En este caso se puede observar que la varianza del error σ^2 es no estimable; es decir, que el efecto de la interacción de los dos factores $(\tau\beta)_{ij}$ y el error experimental no se pueden separar de manera obvia.

Lo cual quiere decir que no hay pruebas suficientes para determinar los efectos principales a menos que el efecto de la interacción sea cero. Si no existe una interacción, entonces $(\tau\beta)_{ij} = 0$ para todas las i, j , y un modelo posible es

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \quad \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \end{cases} \quad (9)$$

Tabla 2. Observaciones por celda

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	Cuadrado medio esperado
Renglones (A)	$\sum_{i=1}^a \frac{y_i^2}{b} - \frac{y^2}{ab}$	$a - 1$	MS_A	$\sigma^2 + \frac{b \sum \tau_i^2}{a - 1}$
Columnas (B)	$\sum_{j=1}^b \frac{y_j^2}{a} - \frac{y^2}{ab}$	$b - 1$	MS_B	$\sigma^2 + \frac{a \sum \beta_j^2}{b - 1}$
Residual o AB	Sustracción	$(a - 1)(b - 1)$	$MS_{Residual}$	$\sigma^2 + \frac{\sum \sum (\tau\beta)_{ij}^2}{(a - 1)(b - 1)}$
Total	$\sum_{i=1}^a \sum_{j=1}^b y_{ij}^2 - \frac{y^2}{ab}$	$ab - 1$		

Nota: Resumen de las observaciones por celda. Tomado del libro *Diseño y análisis de experimentos* (Segunda ed.) por Montgomery, D. C. (2015). México: Limusa.

Para este tipo de experimentos se puede utilizar una prueba desarrollada por Tukey y de esta forma poder establecer si está presente una interacción. Se entiende que la interacción tiene una forma particularmente simple, como la siguiente

$$(\tau\beta)_{ij} = \gamma\tau_i\beta_j \quad (10)$$

Y es una constante desconocida. Al definirse de esta forma el término de interacción, puede usarse un enfoque de regresión para probar la significación del término de la interacción. Lo siguiente es realizar una prueba en donde se haga la partición de la suma de cuadrados de los residuales en un componente con un solo grado de libertad debido a la no aditividad y un componente del error con $(a - 1)(b - 1) - 1$ grados de libertad. Lo cual hace referencia al siguiente cálculo.

$$SS_N = \frac{\left[\sum_{i=1}^a \sum_{j=1}^b y_{ij} y_i y_j - y_{..} (SS_A + SS_B + \frac{y^2}{ab}) \right]^2}{ab SS_A SS_B} \quad (11)$$

Con un grado de libertad, y

$$SS_{error} = SS_{Residual} - SS_N \quad (12)$$

con $(a - 1)(b - 1) - 1$ grados de libertad. Para probar la presencia de una interacción, se calcula

$$F_0 = \frac{SS_N}{SS_{Error}/[(a-1)(b-1)-1]} \quad (13)$$

Si $F_0 > F_{\alpha, 1, (a-1)(b-1)-1}$, debe rechazarse la hipótesis de que no hay ninguna interacción.

2.6. Diseño factorial general

Los resultados del diseño factorial de dos factores pueden ser mayores al caso general en que hay α niveles del factor A , b niveles del factor B , c niveles del factor C , etc., dispuestos en un experimento factorial. Es decir, que habrá n observaciones en total si se realizan n de todo el experimento. Cabe destacar que se necesitan aproximadamente de un mínimo de dos réplicas para poder establecer una suma de cuadrados debida al error si todas las interacciones posibles están incluidas en el modelo.

Se hace más sencillo formular y probar hipótesis acerca de los efectos principales y las interacciones cuando todos los factores del experimento son fijos. En cuanto a un modelo con efectos fijos se refiere, los estadísticos de prueba para cada efecto principal e interacción pueden construirse dividiendo el cuadrado medio correspondiente del efecto o interacción por el cuadrado medio del error. Todas las pruebas F serán de una cola superior. Cabe destacar que el número de grados de libertad de cualquier efecto principal es el número de niveles del factor menos uno, y el número de grados de libertad de una interacción es el producto del número de grados de libertad asociados con los componentes individuales de la interacción.

2.7. Ajuste de curvas y superficies de respuesta

Se pudo apreciar que ajustar una curva de respuesta a los niveles de un factor cuantitativo es de mucha utilidad. Esto beneficia al experimentador para que cuente con una ecuación que relaciona la respuesta con el factor. La ecuación puede manejarse para hacer interpolaciones, es decir, para predecir la respuesta en niveles intermedios entre los factores, respecto de los que se emplearon realmente en el experimento. Asimismo, se puede ajustar una superficie de respuesta para predecir y con varias combinaciones de los factores del diseño, siempre y cuando dos de los factores sean cuantitativos. Usualmente, se utilizan métodos de regresión lineal para que estos modelos se ajusten a los datos experimentales.

2.8. Formación de bloques en un diseño factorial

En algunos casos no es recomendable o práctico hacer la aleatorización completa de todas las corridas de un diseño factorial. Puede suceder que la presencia de un factor perturbador provoque que el experimento se corra en bloques. Si se toma como análisis un experimento factorial con dos factores (A y B) y n réplicas. El modelo estadístico lineal de este diseño es el siguiente:

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk} \quad \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \\ k = 1, 2, \dots, n \end{cases} \quad (14)$$

Donde τ_i , β_j y $(\tau\beta)_{ij}$ representan los efectos de los factores A , B y la interacción AB , respectivamente. Suponga que para la puesta en marcha del experimento se necesita utilizar materia prima que sólo está disponible en lotes cuyo tamaño no es suficiente para permitir que se corran todas las abn combinaciones de los tratamientos de igual lote. Pero, si existiera un lote que posea material suficiente para hacer ab observaciones, entonces un diseño alternativo es correr cada una de las n réplicas utilizando un lote separado de materia prima.

Por lo tanto, los lotes de materia prima constituyen una limitación sobre la aleatorización o un bloque y se corre una sola réplica de un experimento factorial completo dentro de cada bloque. El modelo de los efectos para este diseño es:

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \delta_k + \varepsilon_{ijk} \quad \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \\ k = 1, 2, \dots, n \end{cases} \quad (15)$$

Donde δ_k es el efecto del bloque k -ésimo. Desde luego, dentro de un bloque el orden en que se corren las combinaciones de los tratamientos está completamente aleatorizado.

En la ecuación anterior se supone que la interacción entre los bloques y los tratamientos es insignificante. El mismo supuesto también se aplicó anteriormente en el análisis de diseños de bloques aleatorizados. Si estas interacciones existen, no pueden separarse del componente del error. El término error se combina en realidad de las interacciones $(\tau\delta)_{ik}$, $(\beta\delta)_{jk}$ y $(\tau\beta\delta)_{ijk}$.

2.9. El diseño 2^2

El diseño principal de la serie 2^k solo presenta dos factores, A y B ; cada uno se corre a dos niveles. Este tipo de diseño toma el nombre de diseño factorial 2^2 . Los niveles de los factores existentes pueden llamarse arbitrariamente bajo y alto. Estos niveles pueden ser cuantitativos, como dos valores de temperatura, presión o tiempo, o también cualitativos, como dos máquinas, dos operadores, los niveles alto y bajo de un factor, o quizá la presencia o ausencia de un factor.

2.10. El diseño 2^3

En el diseño 2^3 se considera que tres factores, A , B y C , cada uno con dos niveles, son de interés. En este caso la representación geométrica de las ocho combinaciones de tratamientos puede hacerse con un cubo. De la misma forma, se pueden utilizar la notación "+" y "-" para representar los niveles alto y bajo de los factores. A esta figura se la conoce ocasionalmente como la matriz del diseño. Asimismo, las combinaciones de los tratamientos en el orden estándar se escriben como (1), a , b , ab , c , ac , bc y abc . Estos símbolos constituyen también el total de las n observaciones hechas con esa combinación de tratamientos particular.

Constan tres notaciones diferentes para las corridas del diseño 2^3 que son de uso general. Se tiene en primer lugar la notación + y -, llamada frecuentemente notación geométrica. En segundo lugar, se utilizan las etiquetas en letras minúsculas para identificar las combinaciones de los tratamientos. La tercera y última notación utiliza 1 y 0 para denotar los niveles alto y bajo, respectivamente, de los factores, en lugar de + y -. Estas notaciones se presentan para el diseño 2^3 a continuación.

Tabla 3. Corridas del diseño 2^3

Corrida	A	B	C	Etiquetas	A	B	C
1	-	-	-	(1)	0	0	0
2	+	-	-	a	1	0	0
3	-	+	-	b	0	1	0
4	+	+	-	ab	1	1	0
5	-	-	+	c	0	0	1
6	+	-	+	ac	1	0	1
7	-	+	+	bc	0	1	1
8	+	+	+	abc	1	1	1

Nota. Resumen de las corridas del diseño 2^3 considerando los factores A, B y C. Tomado del libro *Diseño y análisis de experimentos* (Segunda ed.) por Montgomery, D. C. (2015). México: Limusa.

Existen siete grados de libertad entre las ocho combinaciones de tratamientos del diseño 2^3 . Tres grados de libertad se asocian con los efectos principales de A, B y C. Cuatro grados de libertad se asocian con las interacciones; uno con cada una de las interacciones AB, AC y BC y uno con la interacción ABC.

De este modo se debe de considerar la estimación del principal efecto A, el efecto de A cuando B y C están en el nivel bajo es $[a - (1)]/n$. De manera similar, el efecto de A cuando B está en el nivel alto y C está en el nivel bajo es $[ab - b]/n$. El efecto de A cuando C está en el nivel alto y B se encuentra en un nivel bajo es $[ac - c]/n$. Por último, el efecto de A cuanto tanto B como C están en el nivel alto es $[abc - bc]/n$. Por lo cual, el efecto promedio de A es sólo el promedio de estos cuatro efectos.

2.11. El diseño general 2^k

Los métodos presentados hasta ahora se pueden generalizar para el caso de un diseño factorial 2^k , lo cual quiere decir que es un diseño con K factores que tienen dos niveles cada uno.

El modelo estadístico para este tipo de diseño incluiría k efectos principales, $\binom{k}{2}$ interacciones de dos factores, $\binom{k}{3}$ interacciones de tres factores y una interacción de K factores. Por lo tanto, para un diseño 2^k el modelo completo contendrá $2^k - 1$ efectos. Del mismo modo, se utiliza la notación introducida anteriormente para las combinaciones de los tratamientos.

Por ejemplo, en un diseño 2^5 , abd denota la combinación de tratamientos con los factores A, B y D en el nivel alto y los factores C y E en el nivel bajo. Las combinaciones de los tratamientos pueden escribirse en orden estándar introduciendo los factores uno a la vez y combinando sucesivamente cada nuevo factor con los que lo preceden. Por ejemplo, el estándar de un diseño 2^4 es (1), a, b, ab, c, ac, bc, abc, d, ad, bd, abd, cd, acd, bcd, y abcd.

El enfoque general para el análisis estadístico del diseño 2^k es el siguiente. En primer lugar, se debe de estimar los efectos de los factores y explorar sus signos y magnitudes. De esta forma el experimentador puede proveerse de información preliminar respecto de los factores y las interacciones que puedan ser importantes y en que direcciones se puedan ajustar los factores para mejorar la respuesta. Luego se debe de formar el modelo inicial del experimento que usualmente se selecciona el modelo completo, es decir, todos los efectos principales y las interacciones, siempre que se haya hecho una réplica de al menos uno de los puntos del diseño. A continuación, se realizan las pruebas estadísticas correspondientes. El cuarto paso consiste en refinar el modelo, es decir eliminar las variables no significativas del modelo. El quinto paso será analizar los residuales para verificar la adecuación del modelo y los supuestos. Por último, se interpretan los resultados mediante gráficas.

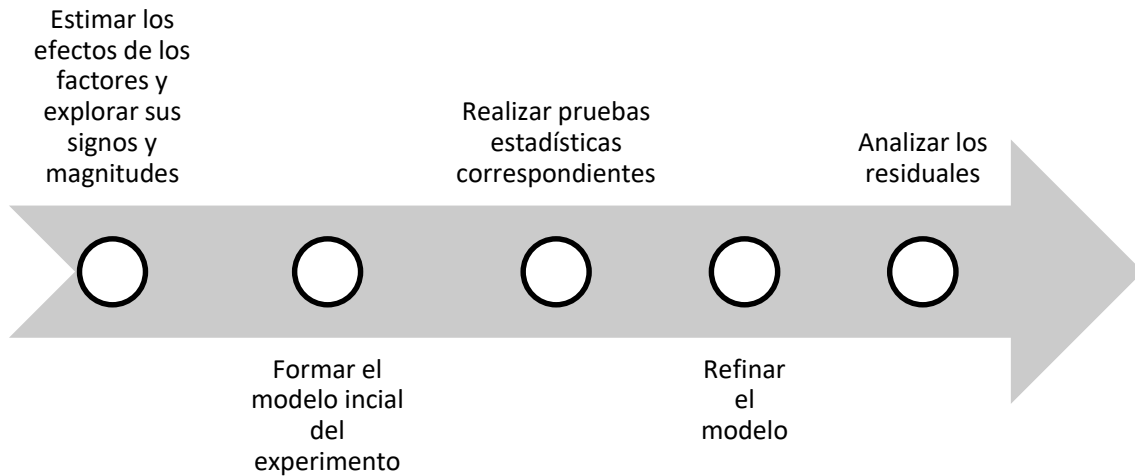


Figura 2. Pasos para el diseño general de 2^k

2.12. Una sola réplica del diseño 2^k o diseño factorial no replicado

Se considera que el número total de combinaciones de tratamientos existentes en un diseño factorial 2^k es grande, inclusive para un moderado de factores. Por ejemplo, sea un diseño 2^5 tiene 32 combinaciones de tratamientos, un diseño 2^6 tiene 64 combinaciones de tratamientos, etc. Puesto que ocasionalmente los recursos son limitados, el número de réplicas que el experimentador puede emplear quizás esté restringido. Frecuentemente, los recursos disponibles permiten hacer únicamente una sola réplica del diseño, a menos que el experimentador esté dispuesto a omitir algunos de los factores originales.

El riesgo muy común que se corre al momento de realizar un experimento que tiene una sola corrida para cada combinación de prueba es que el modelo puede ajustarse al ruido. Lo que quiere decir, es que, si la respuesta es demasiado variable, pueden resultar conclusiones engañosas del experimento. Por otro lado, si hay menos variabilidad en la respuesta, la posibilidad de una conclusión errónea será más reducida. Asimismo, se puede incrementar la distancia entre los niveles bajo (-) y alto (+) del factor para asegurarse de que se obtienen estimaciones confiables. Es muy común utilizar este tipo de estrategia de una sola réplica en los experimentos de exploración cuando

hay un número relativamente grande de factores bajo consideración. Puesto que, en estos experimentos nunca se puede tener la certeza absoluta de que el error experimental es pequeño, se recomienda para una buena práctica que se separen los niveles de los factores de manera agresiva.

Con una sola réplica, no se cuenta con ninguna estimación interna del error (error puro). Una forma de abordar este análisis de un diseño factorial no replicado consiste en suponer que ciertas interacciones de orden superior son insignificantes y combinar sus cuadrados medios para estimar el error. Esto es un requerimiento al principio de efectos esparcidos; es decir, la mayoría de los sistemas están dominados por algunos de los efectos principales y las interacciones de orden inferior, y la mayor parte de las interacciones de orden superior son insignificantes.

Ciertas veces ocurren interacciones de orden superior reales puesto que se analizan datos de diseños factoriales no replicados. El empleo de un cuadrado medio del error que se obtiene agrupando las interacciones de orden superior no es apropiado en estos casos. Una forma simple de resolver este problema es examinando una gráfica de probabilidad normal de las estimaciones de los efectos. Los efectos considerados como insignificantes siguen una distribución normal, con media cero y varianza, estos se ubicarán sobre una línea recta en esta gráfica, mientras que los efectos significativos poseerán medias distintas a cero y no se van a localizar sobre la línea recta. De esta forma, el modelo preliminar se especificará de tal modo que posea que contenga aquellos efectos que aparentemente son diferentes de cero, con base en la gráfica de probabilidad normal. Los efectos que aparecen como insignificantes se combinan como una estimación del error.

2.13. Adición de puntos centrales en el diseño 2^k

Al momento de utilizar los diseños factoriales de dos niveles, existe la preocupación del supuesto de la linealidad de los efectos de los factores. Cabe destacar, que no es necesaria la linealidad perfecta, y el sistema 2^k funcionará bastante bien incluso cuando el supuesto de linealidad sea válido sólo de manera muy aproximada. De tal manera, se ha indicado ya que, si se agregan los términos de interacción a un modelo de los efectos principales o de primer orden, de donde se obtiene

$$y = \beta_0 + \sum_{j=1}^k \beta_j x_j + \sum \sum \beta_{ij} x_i x_j + \epsilon \quad (16)$$

Entonces se presenta un modelo con la capacidad de plasmar cierta curvatura en la función de respuesta. Esta curvatura, es el resultado del torcimiento del plano inducido por los términos de interacción $\beta_{ij} x_i x_j$. En ciertas ocasiones la curvatura de la función de respuesta no estará modelada adecuadamente por la ecuación presentada anteriormente por lo cual se presenta a continuación un modelo lógico

$$y = \beta_0 + \sum_{j=1}^k \beta_j x_j + \sum \sum \beta_{ij} x_i x_j + \sum_{j=1}^k \beta_{jj} x_j^2 + \epsilon \quad (17)$$

Las β_{jj} representan efectos cuadráticos o de segundo orden puros. La ecuación anterior lleva el nombre de modelo de superficie de respuesta de segundo orden. Al momento de realizar un experimento factorial de dos niveles, usualmente se anticipa el ajuste del modelo de primer orden, pero se deberá estar atento ante la posibilidad de que el modelo de segundo orden sea en realidad más apropiado. Se conoce que hay un método que permite realizar una réplica de ciertos puntos de un diseño factorial 2^k que ofrecerá protección contra la curvatura de los efectos de segundo orden a la vez que permitirá una estimación independiente del error que va a obtenerse. En este método hay que añadir dos puntos centrales en el diseño 2^k . Consisten en n réplicas que se corren en los puntos $x_i = 0$ ($i = 1, 2, \dots, k$). Se agregan réplicas de las corridas en el centro del diseño debido a que los puntos centrales no afectan las estimaciones usuales de los efectos en un diseño 2^k . Al momento de añadir los puntos centrales, se supone que los k factores son cuantitativos.

2.14. Formación de bloques de un diseño factorial 2^k con réplicas

Suponga que se han corrido n réplicas del diseño factorial 2^k . Entonces, si existen n réplicas cada conjunto de condiciones no homogéneas define un bloque, y cada réplica se corre en uno de los bloques. Estas corridas de cada bloque (o réplica) se las realizará de manera aleatoria. El análisis del diseño es similar al de cualquier experimento factorial separado en bloques.

2.15. Confusión del diseño factorial 2^k

En ciertas ocasiones se visualizan problemas en los que no es posible realizar una réplica completa de un diseño factorial en un bloque, por lo cual se utilizan diferentes técnicas para corregir esta situación. La confusión o mezclado es una de estas técnicas de diseño mediante la cual un experimento factorial completo se distribuye en bloques, donde el tamaño del bloque es menor que el número de combinaciones de los tratamientos de una réplica. La técnica hace que la información acerca de ciertos efectos de los tratamientos sea indistinguible de los bloques o esté confundida con los bloques. La estructura especial del sistema factorial 2^k permite un método de análisis simplificado aún cuando los diseños que se presentan son diseños de bloques incompletos.

Se considera la construcción y el análisis del diseño factorial 2^k en 2^p bloques incompletos, donde $p < k$. Por consiguiente, estos diseños pueden correrse en dos bloques, en cuatro bloques, en ocho bloques, etc.

2.16. Confusión del diseño factorial 2^k en dos bloques

Para poder comprender de mejor manera se establece el siguiente ejemplo. Suponga que se quiere correr una sola réplica del diseño 2^2 . Cada una de las $2^2=4$ combinaciones de los tratamientos demanda cierta cantidad de materia prima, y cada lote de materia prima solo alcanza para probar dos combinaciones de tratamientos. Es decir, se requieren aproximadamente dos lotes de materia prima. Si se consideran a los lotes de materia prima como bloques, entonces deben asignarse a cada bloque dos de las cuatro combinaciones de tratamientos. Suponga que las combinaciones de tratamientos localizadas en diagonales opuestas de una figura cuadrada se asignan a bloques diferentes. Al mismo tiempo, asuma que el bloque 1 contiene las combinaciones de los tratamientos (1) y ab y que el bloque 2 contiene a y b . Es de suponer que el orden en que se corren las combinaciones de los tratamientos dentro de un bloque se determina aleatoriamente. De igual manera, se decidirá aleatoriamente cuál de los bloques se correrá primero. Suponga que los principales efectos de A y B se calculan como si no se hubiera hecho la formación de bloques. De esta forma se obtienen las siguientes ecuaciones.

$$A = \frac{1}{2}[ab + a - b - (1)] \quad (18)$$

$$B = \frac{1}{2}[ab + b - a - (1)] \quad (19)$$

Se puede visualizar que la formación de bloques no afecta a A ni B. puesto que en cada estimación hay una combinación de un tratamiento positivo y uno negativo de cada bloque. Es decir, cualquier diferencia entre el bloque 1 y el bloque 2 se cancela. Considerando ahora la interacción AB

$$AB = \frac{1}{2}[ab + (1) - a - b] \quad (20)$$

Debido a que las dos combinaciones de tratamientos con signo positivo [ab y (1)] están en el bloque 1 y las dos con signo negativo (a y b) están en el bloque 2, el efecto de los bloques y la interacción AB son idénticos. Lo que quiere decir que AB está confundido o mezclado con los bloques.

2.17. Métodos adicionales para la construcción de bloques

Un método adicional para la construcción de bloques utiliza la combinación lineal

$$L = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_k x_k \quad (21)$$

En donde x_i es el nivel del factor i -ésimo que aparece en una combinación de tratamientos particular y α_i es el exponente que aparece en el factor i -ésimo para el efecto que va a confundirse. En el sistema 2^k se cuenta con $\alpha_i = 0$ o 1 y $x_i = 0$ (nivel bajo) o $x_i = 1$ (nivel alto). A la ecuación anterior se la conoce como definición de contrastes. Las combinaciones de tratamientos que ocasionan el mismo valor de L se colocarán en el mismo bloque. Debido a que los valores posibles de L son 0 y 1, con esto las 2^k combinaciones de tratamientos se asignarán a exactamente dos bloques.

2.18. Estimación del error

Se puede estimar el error realizando réplicas del experimento, siempre y cuando el número de variables sea pequeño, es decir $k= 2$ o 3 . Suponga que un diseño factorial 2^3 debe correrse en dos bloques con ABC confundido, y el experimentador decide hacer cuatro réplicas del diseño.

Se obtienen 32 observaciones y 31 grados de libertad. Adicionalmente, debido a que existen ocho bloques, siete grados de libertad deben asociarse con estos bloques. La suma de cuadrados del error se compone de las interacciones de dos factores entre las réplicas, y cada uno de los efectos. Usualmente, es seguro considerar que las interacciones son cero y tratar el cuadrado medio resultante como una estimación del error. Se puede observar que el cuadrado medio del bloque o ABC podría compararse con el error del cuadrado medio ABC , el cual en realidad es réplicas \times bloques.

Se recomienda utilizar un método diferente para diseñar los bloques encada réplica, siempre y cuando se cuente con los recursos suficientes. Este enfoque se fundamenta en confundir un efecto diferente en cada réplica para obtener cierta información sobre todos los efectos. El procedimiento toma el nombre de confusión o mezclado parcial. Sin embargo, si k es moderadamente grande, solo será posible realizar una réplica.

2.18.1 Confusión del diseño factorial 2^k en cuatro bloques

Se puede construir diseños factoriales 2^k liados en cuatro bloques con 2^{k-2} observaciones cada uno. Este tipo de diseños son de mucha utilidad cuando el número de factores es moderadamente grande, por ejemplo, $k \geq 4$ y el tamaño de los bloques es relativamente pequeño. Suponga un ejemplo, un diseño 2^5 .

Puesto que cada bloque tendrá ocho corridas, entonces se deberá de emplear cuatro bloques. Se construye este bloque de una forma directa, es decir que se seleccionan dos efectos para confundirlos con los bloques, por ejemplo, ADE y BCE . Este tipo de efectos poseen las dos definiciones de contrastes asociadas con ellos.

$$L_1 = x_1 + x_4 + x_5 \quad (22)$$

$$L_2 = x_2 + x_3 + x_5 \quad (23)$$

Entonces cada combinación de tratamientos producirá un par particular de valores de L_1 y L_2 , es decir, cualquiera de $(L_1, L_2) = (0,0), (0,1), (1,0)$ o bien $(1,1)$. Las combinaciones de tratamientos que producen los mismos valores de (L_1, L_2) se asignan al mismo bloque.

2.18.2 Confusión del diseño factorial 2^k en 2^p bloques

Los métodos explicados anteriormente pueden extenderse a la elaboración de un diseño factorial 2^k mezclado en 2^p bloques ($p < k$), en donde cada bloque contiene exactamente 2^{k-p} corridas. Se deben de escoger p efectos independientes que van a confundirse, en donde se entiende por independiente que ninguno de los efectos elegidos es la interacción generalizada de los demás. Los bloques pueden ser creados mediante el uso de las p definiciones de contrastes L_1, L_2, \dots, L_p asociadas con estos efectos. De la misma forma, se confundirán otros $2^p - 1$ efectos con los bloques, siendo estos las interacciones generalizadas de los p efectos independientes elegidos inicialmente. Se debe de prestar atención al momento de seleccionar los efectos que van a confundirse para que no se sacrifique información sobre los efectos que pueden ser de interés potencial.

Para este tipo de diseños las sumas de cuadrados de todos los efectos se estiman como si no se hubiera hecho la formación de bloques y el análisis estadístico es directo. La suma de cuadrados de los bloques se la calcula sumando las sumas de cuadrados de todos los efectos confundido con los bloques. La estructura del mezclado o confusión del diseño depende directamente de la elección de los p efectos usados para producir el bloque.

2.18.3 Confusión parcial

Se le llama confusión parcial cuando solo pueden obtenerse tres cuartas partes de la información de las interacciones porque no están confundidas en sólo tres réplicas. Se presenta a esta relación $\frac{3}{4}$ como la información relativa de los efectos confundidos y por lo cual se argumenta que este tipo de diseño está parcialmente confundido o mezclado. En este diseño sólo se utilizan los datos de las réplicas en las que no está confundida una interacción para de esta forma poder calcular las sumas de cuadrados de las interacciones. La suma de cuadrados del error se la obtiene de la multiplicación entre las sumas de cuadrados de réplicas y las sumas de cuadrados de efecto principal, sumándole la multiplicación de cuadrados de réplicas por las sumas de cuadrados de interacción para cada réplica en la que esa interacción no está confundida. Dentro de este diseño existen siete grados de libertad entre los ocho bloques. Se puede realizar la distribución de tres grados de libertad para las réplicas y cuatro grados de libertad para los bloques dentro de las réplicas.

2.19. La fracción un medio del diseño 2^k

Suponga una situación en la que se tiene tres factores, cada uno cuenta con dos niveles y son de interés, pero los experimentadores no están en posición de correr las $2^3 = 8$ combinaciones de tratamientos. Pero pueden llevar a cabo cuatro corridas. En esta situación se plantea una fracción un medio de un diseño 2^3 . Debido a que el diseño posee $2^{3-1} = 4$ combinaciones de tratamientos, se lo puede llamar diseño 2^{3-1} a una fracción un medio del diseño 2^3 .

2.20. Resolución del diseño

Al diseño 2^{3-1} anterior se le conoce como diseño de resolución III. Los efectos principales de este tipo de diseños son alias de las interacciones de dos factores. Se entiende como resolución R cuando ningún efecto del factor p es alias de otro efecto que contiene menos de $R-p$ factores. Usualmente, se utiliza un subíndice con un número romano para indicar la resolución del diseño. De esta forma, se puede concluir que la fracción un medio del diseño 2^3 con la relación de definición $I = ABC$ (o $I = -ABC$) es un diseño 2^{3-1}_{III} . Las definiciones de estos diseños se presentan a continuación.

2.20.1. Diseños de resolución III

Son diseños en los que ninguno de los efectos principales es alias de ningún otro efecto principal, pero los principales efectos son alias de las interacciones de dos factores. Además, algunas de las interacciones de dos factores pueden ser alias entre sí.

2.20.2. Diseños de resolución IV

Son diseños en los cuales ninguno de los efectos principales es alias de ningún otro efecto principal ni de las interacciones de dos factores, pero estas interacciones entre factores son alias entre sí mismos. Un diseño 2^{4-1} con $I = ABCD$ es un diseño de resolución IV (2^{4-1}_{IV})

2.20.3. Diseños de resolución V

En este grupo entran los diseños en los que ninguno de los efectos principales ni de las interacciones de dos factores son alias de otro efecto principal o interacción de dos factores, pero las interacciones de dos factores son alias de las interacciones de tres factores.

La resolución de un diseño factorial fraccionado de dos niveles será igual al menor número de letras en cualquier palabra de la relación de definición. Consecuentemente, los anteriores diseños podrían llamarse diseños de tres, cuatro y cinco letras, correspondientemente. Por lo tanto, es mejor trabajar con diseños fraccionados que tengan la resolución más alta posible que sea consistente con el grado de fraccionamiento requerido. Entre más alta sea la resolución, menos restrictivos serán los supuestos que se piden respecto de cuales de las interacciones son insignificantes para obtener una completa y única interpretación de los datos.

2.21. Construcción de fracciones un medio

Se puede construir una fracción un medio del diseño 2^k mediante la aplicación del diseño básico, que cuenta de las corridas de un diseño factorial 2^{k-1} completo y añadiéndole después el factor k -ésimo identificando sus niveles positivo y negativo con los signos positivo y negativo de la interacción ABC ($K-1$) del orden más alto. Mediante esta forma, se puede obtener el diseño factorial fraccionado 2_{III}^{3-1} apuntando el diseño 2^2 completo como diseño básico e igualando después el factor C con la interacción AB. Se puede obtener la fracción alterna realizando la igualdad del factor C con la interacción $-AB$.

2.22. Secuencias de diseños factoriales fraccionados.

Usualmente, la utilización de los diseños factoriales fraccionados conlleva a una gran economía y eficiencia en la experimentación, particularmente si se pueden realizar las corridas secuencialmente. Suponga que se está explorando $k=4$ factores ($2^4=16$ corridas). Se aconseja correr un diseño fraccionado 2_{IV}^{4-1} es decir ocho corridas, examinar los resultados y luego decidir cuál es la mejor serie de corridas que deberá correrse después. En el caso que se requiera resolver ambigüedades, siempre puede correrse la fracción alterna y completar el diseño 2^4 . Al momento de utilizar este método para completar el diseño, las dos fracciones un medio simbolizan bloques del diseño

completo con las interacciones de orden superior confundidas con los bloques. Entonces, el resultado de esta experimentación secuencial es la pérdida de información sólo en la interacción de orden más alto. Una de las ventajas es que se puede sacar información suficiente de la fracción un medio para proceder a la siguiente etapa de la experimentación, y esto podría significar la incorporación o eliminación de factores, variación de algunos de los factores en nuevos rangos o el cambio de las respuestas.

2.23. La fracción un cuarto del diseño 2^k

Frecuentemente, se utilizan fracciones menores del diseño 2^k para un número moderadamente grande de factores. Este diseño se lo conoce como diseño factorial fraccionado 2^{k-2} y posee 2^{k-2} corridas. Este diseño se lo construye utilizando un diseño básico compuesto por las corridas asociadas con un diseño factorial completo $k-2$ factores y asociando después las dos columnas adicionales con las interacciones seleccionadas apropiadamente que incluyan los primeros $k-2$ factores. Además, este diseño posee dos generadores. Se les llama las relaciones generadoras del diseño a $I = P$ e $I = Q$ y P y Q representan los generadores escogidos. Los signos de P y Q determinan cuál de las fracciones un cuarto se produce. Las cuatro fracciones asociadas con la elección de los generadores $\pm P$ y $\pm Q$ pertenecen a la misma familia. La fracción en donde P y Q son positivas es la fracción principal.

Las columnas que son iguales a la columna identidad I componen la relación de definición completa. Éstas contarán con P , Q y su interacción generalizada PQ ; lo que quiere decir que la relación de definición es $I=P=Q=PQ$. A los elementos P, Q y PQ de la relación de definición se les denomina palabras. Para poder obtener los alias de cualquier efecto se lo realiza mediante la multiplicación de la columna de ese efecto por cada palabra de la relación de definición. Cabe destacar que cada efecto posee tres alias. El experimentador debe de prestar atención al momento de escoger los generadores para que los efectos potencialmente importantes no sean alias entre sí.

2.24. El diseño factorial fraccionado 2^{k-p} general

Al diseño factorial fraccionado 2^k que posee 2^{k-p} corridas se le conoce como fracción $1/2^p$ del diseño 2^k o, también, como diseño factorial fraccionado 2^{k-p} . Para estos tipos de diseños se deben de seleccionar p generadores independientes.

La relación de definición de este diseño se compone de los p generadores elegidos inicialmente y sus 2^p-1 interacciones generalizadas.

La multiplicación de la columna de cada efecto por la relación de definición da como resultado la estructura de los alias. Se deben de elegir de forma correcta los generadores para que de esta forma los efectos de interés potencial no sean alias entre sí. Cada efecto posee 2^p-1 alias. Si se obtienen valores moderadamente grandes de k , es común suponer que las interacciones de órdenes superiores son insignificantes, con esto se simplifica de gran manera la estructura de los alias. Así también, es importante elegir los p generadores de un diseño factorial fraccionado 2^{k-p} de tal manera que los resultados sean las mejores relaciones de los alias posibles. Una forma razonable es escogiendo los generadores para que el diseño 2^{k-p} resultante obtenga la resolución más alta posible. En ciertas ocasiones la resolución por sí sola no es suficiente para diferenciar entre los diseños.

2.25. Análisis de los diseños factoriales fraccionados 2^{k-p}

El diseño puede analizarse de dos formas, mediante software o de la forma manual recurriendo a los principios básicos; el efecto i -ésimo se estima con la siguiente ecuación

$$l_i = \frac{2(\text{Contraste}_i)}{N} = \frac{\text{Contraste}_i}{\binom{N}{2}} \quad (24)$$

El Contraste_i utiliza los signos positivos y negativos de la columna i y donde $N=2^{k-p}$ es el número de observaciones totales. El diseño 2^{k-p} permite solo el cálculo de $2^{k-p}-1$ efectos y sus alias.

2.26. Proyección del diseño factorial fraccionado 2^{k-p}

El diseño 2^{k-p} se reduce a un factorial completo o a un factorial fraccionado en cualquier subconjunto de $r \leq k - p$ de los factores originales. Los subconjuntos de factores que producen diseños factoriales fraccionados son subconjuntos que aparecen como palabras en la relación de definición completa. Este tipo de diseño es muy útil al aplicarlo a experimentos de tamizados cuando se sospecha desde el inicio del experimento que la mayoría de los factores originales tendrán efectos pequeños.

El diseño factorial fraccionado 2^{k-p} puede ser presentado en un factorial completo, por ejemplo, en los factores de mayor interés. Las conclusiones que se obtengan mediante el uso de este tipo de diseños deberán de considerarse tentativas y deben de ser sometidas a un análisis más profundo. Ocasionalmente, es posible que se encuentren explicaciones alternativas de los datos que intervienen en interacciones de órdenes superiores. Ciertas veces, se presentan diseños factoriales fraccionados que necesitan tantas corridas que no es posible llevarlas a cabo todas bajo condiciones homogéneas. En este tipo de situaciones, los diseños factorial fraccionados pueden mezclarse o confundirse en bloques.

2.27. Diseños de resolución III

El uso secuencial de los diseños factoriales fraccionado es de mucha utilidad, llevando con frecuencia a una gran economía y eficiencia de la experimentación. Se pueden construir diseños de resolución III para investigar hasta $k = N-1$ factores en sólo N corridas, en donde N será múltiplo de 4. Estos diseños que requieren 4 corridas para hasta 3 factores, 8 corridas para hasta 7 factores y 16 corridas para hasta 15 factores, son muy importantes. Si se encuentra que $k = N-1$, se interpreta que el diseño factorial fraccionado está saturado.

Un diseño adicional factorial fraccionado que es de mucha utilidad es el diseño para analizar siete factores en ocho corridas, es decir, el diseño 2_{III}^{7-4} . Este diseño representa una fracción un dieciseisavo del diseño 2^7 . Se puede construir apuntando primero los niveles positivos y negativos de un diseño 2^3 completo en A , B y C como el diseño básico, y asociando después los niveles de cuatro factores adicionales con las interacciones de los tres factores originales de la siguiente manera: $D = AB$, $E = AC$, $F = BC$ y $G = ABC$. Los generadores de este diseño serán $I = ABD$, $I = ACE$, $I = BCF$ e $I = ABCG$.

Para poder obtener la relación de definición completa de este diseño se debe de multiplicar entre si los cuatro generadores ABD , ACE , BCF y $ABCG$ de dos en dos, de tres en tres y los cuatro a la vez. Por otro lado, para poder encontrar el alias de cualquier efecto, se debe de multiplicar el efecto por cada palabra de la relación de definición. Se pueden utilizar los siete grados de libertad de este diseño para estimar los siete efectos principales.

Se puede conseguir una simplificación considerable de la estructura de los alias, cuando se observa que las interacciones de tres o más factores son insignificantes.

Es posible aislar de manera sistemática los efectos de interés potencial mediante las combinaciones de diseños factoriales en los que se han intercambiado ciertos signos. Este tipo de experimento secuencial lleva el nombre de doblez o plegado del diseño original. La estructura de los alias de cualquier fracción que posean los signos de uno o más de los factores invertidos se los obtiene realizando el cambio de signo de manera apropiada en los factores de la estructura de los alias de la fracción original.

2.28. La relación de definición para un diseño de doblez

Usualmente, el experimentador se interesa en conocer la relación de definición del diseño combinado, este puede establecerse con facilidad. Cada una de las fracciones separadas tendrá $L+U$ palabras usadas como generadores. L palabras con el mismo signo y U palabras con signos diferentes. Para el diseño combinado se usarán $L+U-1$ palabras como generadores. De esta forma se identificarán las L palabras de igual signo y las $U-1$ palabras que constan de productos pares independientes de las palabras que tienen signos distintos.

2.29. Diseños de Plackett-Burman

Estos son diseños factoriales fraccionados que cuentan con dos niveles para estudiar $k = N-1$ variables en N corridas, en donde N se presenta como múltiplo de 4. Estos diseños se pueden considerar idénticos a los que se presentaron, si N es una potencia de 2. Sin embargo, si $N = 12, 20, 24, 28$ y 36 , los diseños de Plackett-Burman en ocasiones son de interés. Estos tipos de diseños también toman el nombre de diseños no geométricos, puesto que no pueden representarse en un cubo. Además, los diseños no geométricos para $N = 12, 20, 24, 28$ y 36 poseen estructuras de los alias muy intrincadas.

2.30. Diseños de resolución IV y V

Se considera un diseño de resolución IV a un diseño factorial fraccionado 2^{k-p} en donde los efectos estén separados de las interacciones de dos factores y ciertas interacciones de dos factores son alias entre sí.

Por ende, al suprimirse las interacciones de tres factores y de órdenes superiores, los principales efectos pueden estimarse directamente en un diseño 2_{IV}^{k-p} . Cualquiera de este tipo de diseños debe incluir al menos $2k$ corridas. A los diseños de resolución IV que contienen exactamente $2k$ corridas se les conoce como diseños mínimos. Este tipo de diseños pueden obtenerse a partir de diseños de resolución III por el proceso de doblado. Para poder realizar este tipo de proceso, se debe de agregar a la fracción original una segunda fracción con todos los signos invertidos. De esta forma, los signos positivos en la columna identidad I de la primera fracción podrían intercambiarse en la segunda fracción, y el factor $(k+1)$ -ésimo podría asociarse con esta columna.

También se puede realizar el doblado de diseños de resolución IV para separar las interacciones de dos factores que son alias entre sí. Un experimentador puede tener distintos objetivos al momento de realizar el doblado de un diseño de resolución IV, estos se presentan a continuación. 1) romper cadenas de alias de interacciones de dos factores como sea posible, 2) romper las interacciones de dos factores en una cadena de alias en particular, 3) romper las interacciones de dos factores que incluyen un factor en particular. Otra forma de realizar doblado de un diseño de resolución IV es corriendo una segunda fracción en la que se invierte el signo de todos los generadores del diseño que tienen un número impar de letras.

2.31. Diseños factoriales y factoriales fraccionados con tres niveles y con niveles mixtos.

El diseño factorial 3^k es simplemente un arreglo factorial de k factores que posee tres niveles cada uno. Para esto se utilizarán letras mayúsculas para distinguir los factores y las interacciones. Se hará referencia a tres niveles de los factores como bajo, intermedio y alto. Se pueden utilizar varias notaciones para representar estos niveles de los factores; una posibilidad es representar los niveles de los factores con los dígitos 0 (bajo), 1 (intermedio) y 2 (alto). Cada una de las combinaciones de tratamientos del diseño 3^k se representará por k dígitos, donde el primer dígito indica el nivel del factor K . Suponga un diseño 3^2 , 00 denota la combinación de tratamientos correspondiente a A y B ambos en el nivel bajo, y 01 corresponde a la combinación de tratamientos correspondientes a A en el nivel bajo y B en el nivel intermedio.

Este tipo de sistema de notación también pudo haberse aplicado en los diseños 2^k presentados anteriormente, utilizando 0 y 1 en lugar de -1 negativo y el 1 positivo, respectivamente. En el diseño 2^k se utilizó la notación ± 1 porque se hace más fácil la vista geométrica del diseño y a la vez puede aplicarse directamente al modelado de regresión, la separación en bloques y la construcción de factoriales fraccionados. En el caso de los diseños 3^k , cuando los factores son cuantitativos, se representan a los niveles como bajo, intermedio y alto con -1, 0 y +1, respectivamente. De modo que se haga más fácil el ajuste de un modelo de regresión que relaciona la respuesta con los niveles de los factores.

2.32. El diseño 3^2

El diseño más simple 3^k es el diseño 3^2 , este posee dos factores, cada uno con tres niveles. Se observa que hay ocho grados de libertad entre estas combinaciones de tratamientos, debido a que hay $3^2 = 9$ combinaciones de tratamientos. Los efectos principales de A y B poseen dos grados de libertad cada uno y la interacción AB posee cuatro grados de libertad. Si existe n réplicas, habrá $n3^2 - 1$ grados de libertad totales y $3^2 (n-1)$ grados de libertad del error. Cada efecto principal puede representarse con un componente lineal y uno cuadrático, cada uno con un grado de libertad y esto solo puede realizarse si el factor es cuantitativo.

Se pueden aplicar dos métodos para la partición de la interacción de dos factores AB . El primer método consiste en subdividir AB en los cuatro componentes con un solo grado de libertad que corresponden a $AB_{L \times L}$, $AB_{L \times Q}$, $AB_{Q \times L}$, y $AB_{Q \times Q}$. Esto se lo puede realizar ajustando los términos $\beta_{12}x_1x_2$, $\beta_{122}x_1x_2^2$, $\beta_{112}x_1^2x_2$ y $\beta_{1122}x_1^2x_2^2$, respectivamente. Los cuadrados latinos ortogonales es el segundo método que se puede aplicar, se les llama así debido a que si uno de los cuadrados se superpone en el otro, cada letra del primer cuadrado aparecerá exactamente una vez con cada letra del segundo cuadrado.

2.33. El diseño 3^3

Imagine que existen tres factores (A , B , y C) en estudio, y cada factor posee tres niveles dispuestos en un experimento factorial. Las 27 combinaciones de tratamientos tienen 26 grados de libertad. Cada efecto principal posee 2 grados de libertad, cada interacción de dos factores posee 4 grados de libertad y la interacción de tres factores tiene 8 grados

de libertad. Si se realizan n réplicas, existen n^3-1 grados de libertad totales y $3^3 (n-1)$ grados de libertad del error.

Las sumas de cuadrados pueden estimarse empleando los métodos estándares para los diseños factoriales. Es posible realizar la partición de los efectos principales en un componente lineal y uno cuadrático, cada uno con un solo grado de libertad, siempre y cuando los factores sean cuantitativos. Las interacciones de dos factores pueden descomponerse en efectos lineal x lineal, lineal x cuadrático, cuadrático x lineal y cuadrático x cuadrático. Además, puede realizarse la partición de la interacción de tres factores ABC en ocho componentes con un solo grado de libertad que corresponden a lineal x lineal x lineal, lineal x lineal x cuadrático, etc. Este tipo de descomposición por lo general no es de gran utilidad.

Es posible también realizar la partición de las interacciones de dos factores en sus componentes I y J . Estos se designarían AB , AB^2 , AC , AC^2 , BC y BC^2 , y cada componente tendría dos grados de libertad, pero estos componentes no tienen significación física. También, se presenta la posibilidad de realizar la partición de la interacción de tres factores ABC en cuatro componentes ortogonales con dos grados de libertad, a los que suele denominarse los componentes W , X , Y y Z de la interacción. De la misma forma, se hace referencia a ellos como los componentes AB^2C^2 , AB^2C , ABC^2 y ABC de la interacción ABC , respectivamente.

$$W(ABC) = AB^2C^2 \quad (25)$$

$$X(ABC) = AB^2C \quad (26)$$

$$Y(ABC) = ABC^2 \quad (27)$$

$$Z(ABC) = ABC \quad (28)$$

Se puede observar que ninguna de las primeras letras puede tener un exponente distinto de 1. Al igual que los componentes I y J , los componentes W , X , Y y Z no tienen ninguna interpretación práctica, pero sirven para construir diseños más complejos.

2.34. El diseño general 3^k

Los conceptos utilizados anteriormente en los diseños 3^2 y 3^3 pueden extenderse de inmediato al caso de k factores, cada uno con tres niveles, a un diseño factorial 3^k .

Para esto se debe de utilizar la notación digital usual para las combinaciones de tratamientos, por lo que 0120 representa una combinación de tratamientos en un diseño 3^4 con A y D en los niveles bajos, B en el nivel intermedio y C en el nivel alto. Se presentan 3^k combinaciones de tratamientos, con $3^k - 1$ grados de libertad entre ellas. Estas combinaciones de tratamientos permiten determinar las sumas de cuadrados de k factores principales, cada uno con dos grados de libertad. $\binom{k}{2}$ interacciones de dos factores, cada una con cuatro grados de libertad y una interacción de k factores con 2^k grados de libertad.

En general, una interacción de h factores tiene 2^h grados de libertad. Si se realizan n réplicas, hay $n3^k - 1$ grados de libertad totales y $3^k (n-1)$ grados de libertad del error.

Las sumas de cuadrados de los efectos y las interacciones se calculan con los métodos usuales para los diseños factoriales. No se hace ninguna descomposición adicional de las interacciones de tres factores y de órdenes superiores. Sin embargo, cualquier interacción de h factores tiene 2^{h-1} componentes ortogonales con dos grados de libertad. Estos componentes de la interacción no tienen ninguna interpretación física, pero son útiles para construir diseños complejos.

2.35. Confusión en el diseño factorial 3^k en tres bloques

Es necesario realizar la confusión o mezclado en bloques, puesto que cuando se considera una sola réplica del diseño 3^k , esta requiere tantas corridas que es imposible que puedan hacerse las 3^k corridas bajo condiciones uniformes. Por ejemplo, se necesita confundir el diseño 3^k en tres bloques incompletos. Estos tres bloques poseen tres grados de libertad entre ellos; por lo tanto, debe haber dos grados de libertad confundidos con los bloques. Cabe recalcar que en la serie factorial 3^k cada efecto principal tiene dos grados de libertad. Además, cada interacción de dos factores posee cuatro grados de libertad y puede descomponerse en dos componentes de la interacción, cada uno con dos grados de libertad; cada interacción de tres factores posee ocho grados de libertad y puede descomponerse en cuatro componentes de la interacción, cada uno con dos grados de libertad y así sucesivamente. Es conveniente confundir un componente de interacción con los bloques. Se construye una definición de contrastes.

$$L = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_k x_k \quad (29)$$

α_i representa el exponente del factor i -ésimo en el efecto que va a confundirse y x_i es el nivel del factor i -ésimo en una combinación de tratamientos particular. Para la serie 3^k se tiene $\alpha_i = 0, 1$ o 2 , donde la primera α_i diferente de cero es la unidad, y $x_i = 0$ (nivel bajo), 1 (nivel intermedio) o 2 (nivel alto). Las combinaciones de tratamientos del diseño 3^k se asignan a bloques con base en el valor de L . Debido a que L sólo puede asumir los valores $0, 1$ o 2 , tres bloques están definidos de manera única. Las combinaciones de tratamientos que satisfacen $L = 0$ constituyen el bloque principal. Este bloque incluirá siempre la combinación de tratamientos $00\dots 0$.

2.36. El diseño factorial 3^k en nueve bloques

En ciertas situaciones experimentales puede ser necesario confundir el diseño 3^k en nueve bloques. Debido a esto, ocho grados de libertad se confundirán con los bloques. Para la construcción de estos diseños se eligen dos componentes de interacción y, como resultado, dos más se confundirán automáticamente, produciendo los ocho grados de libertad requeridos. Estos dos son las interacciones generalizadas de los dos efectos elegidos originalmente. En el sistema 3^k , las interacciones generalizadas de dos efectos (es decir, P y Q) se definen como PQ y PQ^2 (o P^2Q).

Los dos componentes de interacción elegidos inicialmente producen dos definiciones de contrastes

$$L_1 = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_k x_k = u \quad u = 0, 1, 2 \quad (30)$$

$$L_2 = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k = h \quad h = 0, 1, 2 \quad (31)$$

donde $\{\alpha_i\}$ y $\{\beta_j\}$ son los exponentes de la primera y la segunda interacciones generalizadas, respectivamente, con la convención de que las primeras α_i y β_j diferentes de cero son la unidad. Los nueve bloques pueden construirse utilizando las definiciones de contrastes y la propiedad de la teoría de grupos del bloque principal. Dentro del diseño 3^k en nueve bloques habrá cuatro componentes de interacción confundidos. Los otros componentes de estas interacciones que no están confundidos pueden determinarse restando la suma de cuadrados del componente confundido de la suma de cuadrados de la interacción completa.

2.37. El diseño factorial 3^k en 3^p bloques

El diseño factorial 3^k puede confundirse en 3^p bloques con 3^{k-p} observaciones cada uno, donde $p < k$. La metodología consiste en seleccionar p efectos independientes que habrán de confundirse con los bloques. Como resultado, otros $(3^p - 2p - 1)/2$ efectos se confunden de manera automática. Estos efectos son las interacciones generalizadas de los efectos elegidos originalmente.

2.38. La fracción un tercio del diseño factorial 3^k

La fracción un tercio que contiene 3^{k-1} corridas es la fracción más grande del diseño 3^k . Por lo tanto, se hace referencia a él como el diseño factorial fraccionado 3^{k-1} . Al momento de construir un diseño factorial fraccionado 3^{k-1} se elige un componente de interacción con dos grados de libertad y se realiza la partición del diseño 3^k completo en tres bloques. Cada uno de los tres bloques resultantes es un diseño fraccionado 3^{k-1} y puede seleccionarse cualquiera de los bloques para usarlo. Si $AB^{\alpha_2}C^{\alpha_3} \dots K^{\alpha_k}$ es el componente de interacción utilizado para definir los bloques, entonces a $I = AB^{\alpha_2}C^{\alpha_3} \dots K^{\alpha_k}$ se le llama la relación de definición del diseño factorial fraccionado. Cada efecto principal o componente de interacción estimado a partir del diseño 3^{k-1} tiene dos alias, los cuales pueden encontrarse multiplicando el efecto tanto por I como por I^2 .

2.39. Otros diseños factoriales fraccionados 3^{k-p}

Para contener los valores grandes de k , se desea un fraccionamiento todavía mayor del diseño 3^k . En general, se puede construir una fracción $(1/3)^p$ del diseño 3^k para $p < k$, donde la fracción contiene 3^{k-p} corridas. Este diseño toma el nombre de diseño factorial fraccionado 3^{k-p} . Por ende, un diseño 3^{k-2} es una fracción un noveno, un diseño 3^{k-3} es una fracción un veintisieteavo, etcétera. Para poder construir un diseño factorial fraccionado se debe de escoger p componentes de interacciones y usar estos efectos para hacer la partición de las 3^k combinaciones de tratamientos en 3^p bloques. Por lo tanto, cada bloque es un diseño factorial fraccionado 3^{k-p} . La relación de definición I de cualquier fracción consta de los p efectos inicialmente seleccionados y sus $(3^p - 2^p - 1)/2$ interacciones generalizadas. Para poder obtener el alias de cualquier efecto principal o componente de interacción se multiplica módulo 3 del efecto por I e I^2 .

Por otro lado, también se pueden generar las corridas que definen un diseño factorial fraccionado 3^{k-p} anotando primero las combinaciones de tratamientos de un diseño factorial 3^{k-p} completo e introduciendo después los p factores adicionales igualándolos con los componentes de las interacciones. El proceso se lo realizará construyendo un diseño 3^{4-2} , es decir, una fracción un noveno del diseño 3^4 . Sean AB^2C y BCD los dos componentes de interacciones elegidos para construir el diseño. Sus interacciones generalizadas son $(AB^2C)(BCD) = AC^2D$ y $(AB^2C)(BCD)^2 = ABD^2$. Por lo tanto, la relación de definición de este diseño es $I = AB^2C = BCD = AC^2D = ABD^2$, y el diseño es de resolución III.

Las nueve combinaciones de tratamientos del diseño se encuentran apuntando un diseño 3^2 en los factores A y B, y agregando después dos nuevos factores haciendo esto equivalente a utilizar AB^2C y BCD para hacer la partición del diseño 3^4 completo en nueve bloques y luego seleccionar uno de estos bloques como la fracción deseada. El diseño consta con ocho grados de libertad que pueden usarse para estimar cuatro efectos principales y sus alias.

2.39.1 Factores con dos y tres niveles

Los diseños en donde algunos factores poseen dos niveles y otros tres niveles pueden provenir de la tabla de signos positivos y negativos del diseño 2^k usual. Imagine que se tienen dos variables, donde A tiene dos niveles y X tres. El factor X tiene dos grados de libertad, y si el factor es cuantitativo, es posible realizar su partición en un componente lineal y uno cuadrático, con cada componente teniendo un grado de libertad.

2.39.2 Factores con dos y cuatro niveles

El procedimiento para hacerlo implica el uso de dos factores con dos niveles para representar el factor con cuatro niveles. Por ejemplo, se tiene un factor de cuatro niveles y dos factores de dos niveles y es necesario estimar todos los efectos principales y las interacciones en las que intervienen estos factores. Esto se lo puede llevar a cabo con un diseño de 16 corridas.

2.40. Modelos de regresión lineal

Este modelo describe un plano en espacio bidimensional en el cual utiliza dos variables independientes, también conocidas como variables independientes predictoras o regresores. La intersección del plano con el eje de las ordenadas se define por el parámetro β_0 . Los coeficientes de regresión parcial se denominan β_1 y β_2 puesto que β_1 determina el cambio esperado en γ para cada cambio unitario de x_1 cuando x_2 se mantiene constante y β_2 mide el cambio esperado de γ para cada cambio unitario de x_2 cuando x_1 se mantiene constante.

El modelo de regresión lineal múltiple con k regresores relaciona la variable de respuesta γ con k regresores

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (32)$$

Este modelo describe un hiperplano en el espacio de k dimensiones de los regresores x_j . El parámetro β_j muestra el cambio esperado en la respuesta γ para un cambio unitario en x_j .

2.41. Estimación de los parámetros en modelos de regresión lineal

El modelo más utilizado para la estimación de los coeficientes de regresión de un modelo de regresión lineal múltiple es el modelo de mínimo cuadrados, el cual consiste en escoger las β de la ecuación. Imagine que $n > k$ observaciones de la variable de respuesta. Por ejemplo, y_1, y_2, \dots, y_n . Junto con cada respuesta observada y_i se tendrá una observación de cada uno de los regresores, y que x_{ij} denote la observación o nivel i -ésimo de la variable x_j . El término del error ε del modelo tiene $E(\varepsilon) = 0$ y $V(\varepsilon) = \sigma^2$ y que las $\{\varepsilon_i\}$ son variables aleatorias no correlacionadas. La ecuación del modelo puede escribirse de la siguiente manera

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i \quad (33)$$

$$= \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \varepsilon_i \quad i = 1, 2, \dots, n \quad (34)$$

2.42. Estimación de σ^2

El siguiente paso es necesario estimar σ^2 y para poder desarrollar un estimador de este parámetro, se toma en cuenta la suma de cuadrados de los residuales como

$$\begin{aligned} SS_E &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ &= \sum_{i=1}^n e_i^2 \quad (35) \\ &= e' e \end{aligned}$$

La ecuación final queda de la siguiente manera y se le conoce como suma de cuadrados residual o del error, esta posee $n - p$ grados de libertad asociados con ella.

$$E(SS_E) = \sigma^2(n - p) \quad (36)$$

Y el estimador insesgado de σ^2 está determinado por la siguiente ecuación

$$\sigma^2 = \frac{SS_E}{n-p} \quad (37)$$

2.43. Propiedades de los estimadores

El método de mínimos cuadrados produce un estimador insesgado del parámetro β del modelo de regresión lineal. Esto puede demostrarse fácilmente tomando el valor esperado de $\hat{\beta}$ de la siguiente manera:

$$E(\hat{\beta}) = E[(X'X)^{-1}X'y] \quad (38)$$

$$= E[(X'X)^{-1}(X\beta + \varepsilon)] \quad (39)$$

$$= E[(X'X)^{-1}X'X\beta + (X'X)^{-1}X'\varepsilon] \quad (40)$$

$$= \beta \quad (41)$$

La propiedad de la varianza del beta estimado se presenta en la matriz de covarianza

$$Cov(\hat{\beta}) = E\{[\hat{\beta} - E(\hat{\beta})][\hat{\beta} - E(\hat{\beta})]'\} \quad (42)$$

La matriz de covarianza de $\hat{\beta}$ es

$$Cov(\hat{\beta}) = \sigma^2(X'X)^{-1} \quad (43)$$

2.44. Prueba de significación de la regresión

La prueba de significación de la regresión es un procedimiento que determina la relación lineal entre la variable de respuesta y y un subconjunto de los regresores x_1, x_2, \dots, x_k . Las hipótesis utilizadas son

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0 \quad (44)$$
$$H_1: \beta_j \neq 0 \quad \text{para al menos una } j \quad (45)$$

El rechazo de H_0 implica que al menos uno de los regresores x_1, x_2, \dots, x_k contribuye de manera significativa al modelo.

2.45. Pruebas de los coeficientes de regresión individuales y de grupos de coeficientes

Las siguientes pruebas son de mucha utilidad para establecer el valor de cada uno de los regresores del modelo de regresión. Puede darse el caso en el que el modelo podría ser más eficaz con la inclusión de variables adicionales o quizá con la eliminación de una o más de las variables que están ya en el modelo. Al momento de añadir una variable al modelo de regresión ocasiona siempre que la suma de cuadrados de regresión se incremente y que la suma de cuadrados del error decrezca. Se debe de decidir si el incremento de la suma de cuadrados de regresión es suficiente para garantizar el uso de la variable adicional en el modelo. Además, agregar una variable no importante al modelo puede incrementar el cuadrado medio del error, reduciéndose la utilidad del modelo.

2.46. Intervalos de confianza para los coeficientes de regresión individuales

Debido a que el estimador de mínimos cuadrados es una combinación lineal de las observaciones, se sigue que β posee una distribución normal con vector medio β y matriz de covarianza $\sigma^2(X'X)^{-1}$ por lo tanto cada uno de los estadísticos se distribuye como t con $n - p$ grados de libertad, donde C_{jj} es el elemento (jj) -ésimo de la matriz $(X'X)^{-1}$, y σ^2 es la estimación de la varianza del error. Debido a esto un intervalo de confianza de $100(1-\alpha)$ por ciento para el coeficiente de regresión es

$$\hat{\beta} - t_{\frac{\alpha}{2}, n-p} se(\hat{\beta}_j) \leq \beta_j \leq \hat{\beta}_j + t_{\frac{\alpha}{2}, n-p} se(\hat{\beta}_j) \quad (46)$$

2.47. Predicción de nuevas observaciones de la respuesta

Es posible utilizar un modelo de regresión para predecir observaciones futuras de las respuestas y que corresponden a valores particulares de los regresores, por ejemplo $x_{01}, x_{02}, \dots, x_{0k}$. Si $x'_0 = [1, x_{01}, x_{02}, \dots, x_{0k}]$, por lo tanto una estimación puntual de la observación futura y_0 en el punto $x_{01}, x_{02}, \dots, x_{0k}$ se calcula con la siguiente ecuación

$$\hat{y}(x_0) = x'_0 \hat{\beta} \quad (47)$$

Y un intervalo de predicción para esta observación futura es

$$\hat{y}(x_0) - t_{\frac{\alpha}{2}, n-p} \sqrt{\sigma^2 (1 + x'_0 (x'x)^{-1} x_0)} \leq y_0 \quad (48)$$

$$\leq \hat{y}(x_0) + t_{\frac{\alpha}{2}, n-p} \sqrt{\sigma^2 (1 + x'_0 (x'x)^{-1} x_0)} \quad (49)$$

Es necesario tener cuidado para no realizar una extrapolación fuera de la región que posee las observaciones originales. Usualmente, cuando un modelo se ajusta bien en la región de los datos originales entonces deja de hacerlo fuera de la región.

2.48. Residuales estandarizados y studentizados

Experimentadores de modelos trabajan con residuales escalado en vez de los residuales de mínimos cuadrados ordinarios. Estos residuales escalados transmiten con frecuencia más información que los residuales ordinarios. El residual estandarizado es un tipo de residual escalado.

$$d_i = \frac{e_i}{\hat{\sigma}} \quad i = 1, 2, \dots, n \quad (50)$$

Este tipo de residuales estandarizados poseen media cero y varianza aproximadamente unitaria; por lo cual, son muy útiles para encontrar puntos atípicos. Usualmente, los residuales estandarizados deberán de localizarse en el intervalo $-3 \leq d_i \leq 3$ y cualquier observación con un residual estandarizado que esté fuera de este intervalo es

potencialmente inusual con respecto a su respuesta observada. Los puntos atípicos deberán examinarse con atención, debido a que pueden representar algo tan simple como un error al registrar los datos o algo que sea motivo de mayor preocupación, como una región del espacio del regresor, en donde el modelo ajustado es una aproximación pobre de la verdadera superficie de respuesta.

2.49. Residuales PRESS

PRESS (Prediction Error Sum of Squares) o también llamada en español como la suma de cuadrados del error de predicción expresa una útil escalación de los residuales. El método de cálculo de PRESS es seleccionar una observación, por ejemplo, la i . Se ajusta el modelo de regresión a las $n - 1$ observaciones restantes y se usa esta ecuación para predecir la observación que se apartó y_i . Al denotar este valor predicho $\widehat{y}_{(i)}$, puede encontrarse el error de predicción del punto i como $e_{(i)} = y_i - \widehat{y}_{(i)}$. Al error de predicción suele llamársele el residual PRESS i -ésimo. Este procedimiento se repite para cada observación $i = 1, 2, \dots, n$, produciéndose un conjunto de n residuales PRESS $e_{(1)}, e_{(2)}, \dots, e_{(n)}$. Entonces el estadístico PRESS se define como la suma de cuadrados de los n residuales PRESS como en

$$PRESS = \sum_{i=1}^n e_{(i)}^2 = \sum_{i=1}^n [y_i - \widehat{y}_{(i)}]^2 \quad (51)$$

PRESS utiliza cada subconjunto posible de $n - 1$ observaciones como un conjunto de datos de estimación, y se utiliza una observación a la vez para formar un conjunto de datos de predicción.

2.50. R- student

Los experimentadores comúnmente consideran al residual studentizado r_i comentado antes como el diagnóstico de un punto atípico. Se acostumbra usar MS_E como una estimación de σ^2 en el cálculo de r_i . Se hace referencia a este enfoque como la escalación interna del residual, ya que MS_E es una estimación de σ^2 generada internamente que se obtiene del ajuste del modelo a las n observaciones. Un enfoque adicional sería utilizar una estimación de σ^2 basada en un conjunto de datos en el que se elimina la observación i -ésima. La estimación de σ^2 así obtenida se denota por $S_{(i)}^2$.

Se puede demostrar que la estimación de la siguiente ecuación se usa en lugar de MS_E para producir un residual studentizado externamente, al que es común llamar R-student.

$$S_{(i)}^2 = \frac{(n-p)MS_E - e_i^2 / (1-h_{ii})}{n-p-1} \quad (52)$$

$$t_i = \frac{e_i}{\sqrt{S_{(i)}^2(1-h_{ii})}} \quad i = 1, 2, \dots, n \quad (53)$$

En ciertas ocasiones habrá una ligera diferencia entre t_i y el residual studentizado r_i . Sin embargo, si la observación i -ésima es influyente, entonces $S_{(i)}^2$ puede diferir significativamente de MS_E , y por lo tanto la R-student será más sensible a este punto. Además, bajo los supuestos usuales, t_i tiene una distribución t_{n-p-1} . Por lo tanto, la R-student ofrece un procedimiento más formal para detectar puntos atípicos a través de la prueba de hipótesis.

2.51. Diagnóstico de influencia

Ocurren situaciones en las que se encuentra un subconjunto pequeño de los datos que ejerce una influencia desproporcionada sobre el modelo de regresión ajustado. Es decir, las estimaciones o predicciones de los parámetros pueden depender más del subconjunto influyente que de la mayoría de los datos. Sería más conveniente localizar estos puntos influyentes y valorar su impacto en el modelo. Si estos puntos influyentes son valores malos, entonces deben de eliminarse. Si estos puntos controlan propiedades clave del modelo, sería deseable saberlo, puesto que podría afectar el uso del modelo.

2.52. Referencias bibliográficas

Montgomery, D. C. (2015). *Diseño y análisis de experimentos* (Segunda ed.). México: Limusa.

Pérez López, C. (2004). *Técnicas de análisis multivariante de datos. Aplicaciones con SPSS*. Madrid, Universidad Complutense de Madrid.: Pearson Educación.

CAPÍTULO 3: ANÁLISIS CONJUNTO

Simon (1957), Hoffman (1960, 1968) y Churchman (1961) sugirieron por primera vez que podríamos inferir, o capturar, las políticas y valores de los tomadores de decisiones observando sus decisiones sobre circunstancias suficientes. Durante los últimos años 1960 y 1970, los psicólogos que trabajaron en una variedad de problemas aparentemente no relacionados y desarrollaron un paradigma por el cual las políticas de los encargados de tomar decisiones pudieron inferirse.

Estos investigadores estaban principalmente interesados en determinar las reglas de composición utilizadas por los tomadores de decisiones para combinar información en juicios generales. Uno de estos grupos de investigadores estaba interesado en la integración de la información, la teoría del cambio de actitud, la percepción de la persona y la teoría de la decisión. Otro grupo de investigadores estaba interesado en cómo los médicos combinaban la información para llegar a un juicio diagnóstico general. El vínculo con la economía y la valoración de los objetos fue provisto por Lancaster (1966), quien propuso que una utilidad de los consumidores para un bien podía entenderse como una función de la utilidad para los componentes del bien.

Es decir, los psicólogos proporcionaron la opinión que la tarea de entender cómo la gente elige podría ser entendida en términos de cómo combinaron la información sobre el objeto de la opción, mientras que Lancaster proporcionó la penetración que la información relevante a ser combinada era información sobre los componentes, o atributos, de bienes.

La discusión en el desarrollo de un paradigma para inferir fue que los investigadores pueden representar las reglas de composición que los tomadores de decisiones usan como principales efectos e interacciones de un modelo de análisis de varianza (ANOVA).

3.1 Modelo ANOVA

Los modelos de Anova parecen descriptivos intuitivamente descriptivos de muchas situaciones de juicio, sin embargo, no se han utilizado previamente para representar el proceso de juicio.

Si los estímulos de juicio se consideran como factores de tratamiento categóricos más que como variables aleatorias, y si los juicios hechos a las señales se consideran como variables dependientes, las capacidades inferenciales de la técnica de Anova pueden aplicarse al estudio del juicio. La aplicación es simple y directa: se preparan estímulos de juicio multidimensionales construyendo todas las posibles combinaciones (patrones) de los niveles de referencia en un diseño factorial completamente cruzado.

$$y_{x_1x_2\dots x_n} = \mu + \sum_{i=1}^n \alpha_{x_i}^{A_i} + \sum_{i,j=1}^n \alpha_{x_ix_j}^{A_iA_j} + \sum_{i,j,k=1}^n \alpha_{x_ix_jx_k}^{A_iA_jA_k} + \dots + \alpha_{x_1x_2\dots x_n}^{A_1A_2\dots A_n} + e_{x_1x_2\dots x_n} \quad (54)$$

Donde μ es la media general; $\alpha_{x_i}^{A_i}$ es el efecto del factor A_i en el nivel X_i ; $\alpha_{x_ix_j}^{A_iA_j}$ es el efecto de los factores A_i y A_j en X_i y X_j niveles, respectivamente y $e_{x_1x_2\dots x_n}$ es el error aleatorio Independientemente e idénticamente distribuido. En lugar de inferir la regla de composición utilizada para combinar la información, las metodologías de análisis conjuntas asumen típicamente una regla de composición e infieren los efectos del modelo de Anova. Un modelo lineal que incluya efectos de marca y atributo puede ser representado de la siguiente forma

$$y_{jx_1x_2\dots x_n} = \mu + \alpha_j^\beta + \sum_{i=1}^n \alpha_{x_i}^{A_i} + e_{x_1x_2\dots x_n} \quad (55)$$

α_j^β representa el atributo cuyos niveles son las diferentes marcas. Suponiendo que la ecuación implica que el investigador cree que las interacciones bidireccionales y de orden superior de los atributos son insignificantes. El término de error de la ecuación 55 incluye el de la ecuación 54 y la variación debido a una posible especificación errónea.

En general, si se asume un modelo factorial menor que un modelo factorial completo, se selecciona un subconjunto de todos los posibles perfiles conceptuales que permite estimar todos los parámetros del modelo reducido que se supone que toman las decisiones al combinar peso de cada nivel para llegar a las preferencias.

Si se supone que las interacciones de dos vías y de orden superior son insignificantes, puede tomarse el modelo de la ecuación 51 como la composición regla operativa y un plan experimental principal para construir perfiles. Sus marcas o niveles de atributos se codifican tradicionalmente como variables dummies utilizando codificación o polinomios ortogonales.

Los niveles de atributos dummies codificados son las variables independientes. Las estimaciones de los parámetros se interpretan como parte del valor o utilidad parte de las marcas respectivamente o atributo niveles. Una premisa clave en la interpretación de los parámetros como una indicación de algo parecido a la utilidad es que los encuestados deben hacer compensaciones entre niveles de atributos para llegar a juicios de preferencia. En el análisis de la práctica conjunta comercial, los parámetros se estiman usando mínimos cuadrados ordinarios de regresión. Modelos que incluyen interacciones son, por tanto, posibles si incluyen interacciones entre la marca y el precio podría tener sentido en algunas situaciones.

Una ventaja del enfoque de análisis conjunto es que los encuestados proporcionan suficiente información y los parámetros pueden ser estimados a nivel individual. Estas estimaciones serán posteriormente sometidas a un algoritmo de agrupamiento. Las agrupaciones resultantes se interpretan como segmentos de mercado. Utilizando estimaciones de esta manera se llama segmentación post hoc. Los miembros de cada segmento son similares con respecto a sus preferencias para productos o servicios niveles de atributos.

Los segmentos post hoc pueden ser utilizados de diferentes maneras. Por ejemplo, las preferencias de algunos segmentos pueden ser más favorables a las marcas existentes de una empresa que otros, y los vendedores pueden centrar los esfuerzos en segmentos para los que posean una ventaja diferencial. También, pueden diferenciar perfiles de los miembros de los segmentos con, estilo de vida, o variables demográficas geográficas similares para obtener información sobre cómo distribuir o promover el producto o servicio. Por último, pueden identificar los segmentos que son atractivos debido a su tamaño, la tasa de consumo, o la falta de competencia.

Una ventaja importante del análisis conjunto es que se reduce la carga de calificaciones respuesta al reducir el número de resoluciones judiciales requeridas de los encuestados: cuando no se replican, el número de calificaciones requeridas es igual al número de alternativas. Sin embargo, el uso de escalas de notas tiene consecuencias. En primer lugar, los encuestados utilizan escalas en diferentes maneras. Algunos utilizan todas las categorías; algunos no lo hacen. Algunos distribuyen sus respuestas en un extremo de la escala. Otros se inclinan por el extremo opuesto. Además, las predicciones basadas en calificación se utilizan para las preferencias.

El enfoque tradicional del análisis conjunto posee un costo de oportunidad. En los mercados competitivos, el éxito de una empresa está influenciado tanto por sus propios esfuerzos y por sus competidores. Por ejemplo, la cuota de mercado de una empresa está influenciada no sólo por su propio precio, pero también por los precios de la competencia, y el impacto sobre las cuotas de mercado de un cambio en su propio precio depende de si el cambio con los competidores coincide. Los procedimientos de análisis de conjunto tradicionales son adecuados para la captura de los efectos de los esfuerzos de las empresas. Ellos son poco adecuados para la captura de los efectos de las acciones de los competidores. Debido a que los estudios de análisis conjuntos tradicionales no requieren encuestados para realizar compensaciones entre perfiles, sólo entre los niveles de atributos para un único perfil.

3.2 Experimentación de elección discreta

Con esta técnica el experimentador organiza perfiles en k sistemáticamente construido conjuntos de opciones de k_i perfiles de marcas m , $k_i \leq m$, donde k_i es el número de perfiles en el conjunto i , $i = 1, 2, \dots, k$. En un experimento de elección, se presenta a los encuestados una serie de tales conjuntos, y para cada grupo se les pide que seleccionen el perfil que consideran el mejor. El trabajo es sólo el comienzo de cómo construir conjuntos de opciones para varios problemas. La experimentación de elección discreta ofrece dos ventajas comparada con el método tradicional. La primera ventaja es que la variable dependiente es la elección, que es similar conceptualmente al acto anterior a la compra y la segunda ventaja es que con este método el demandado implícitamente debe considerar las otras alternativas en la elección establecida al hacer una elección. Por lo tanto, este método es adecuado para detectar el efecto que tiene la disponibilidad de otras alternativas en el atractivo de una marca.

Una lista no exhaustiva de ámbitos en los que la elección discreta experimentación se ha utilizado incluye ciencias ambientales, geografía, marketing, turismo y transporte. A diferencia del análisis conjunto tradicional, la experimentación de elección discreta con los encuestados no proporciona datos suficientes de modo que no puede ser analizado a nivel individual. Los encuestados hacen una sola selección de cada conjunto de elección que se presentan. La variable dependiente de las acciones, la proporción de los N encuestados seleccionando cada alternativa en un conjunto de elección. Esta proporción típicamente se modela como el resultado de un proceso multinomial utilizando el modelo de utilidad aleatoria logit multinomial.

El modelo de utilidad aleatoria predice que los cambios en las acciones son proporcionales a los cambios en las preferencias. Como usualmente pasa con los modelos no lineales, la matriz de covarianza de las observaciones depende de los parámetros de los modelos. Esto dificulta realizar declaraciones globales acerca de los diseños experimentales óptimos. Los logits son heterocedásticos y se correlacionan, mínimos cuadrados generalizados o información completa, se debe utilizar algoritmos en la estimación de máxima verosimilitud.

3.3 Modelos de utilidad aleatorios

Los modelos de utilidad aleatorios suponen que la utilidad puede dividirse en un componente sistemático v_i y un componente aleatorio ε entonces,

$$\mu_i = v_i + \varepsilon, (56)$$

Donde $i = 1, \dots, q \in S$, con q números de alternativas para la elección. El componente v_i es común a todas las personas de la población pertinente. La suposición de independencia para los términos de error es bastante restrictiva y se han desarrollado varios modelos para acomodar la falta de independencia. Uno de estos modelos es la disponibilidad / efectos cruzados. Consideremos el conjunto de opciones S con q elementos que tienen las utilidades u_1, u_2, \dots, u_q . El modelo de utilidad aleatorio asume que, en cada ocasión, los individuos eligen la alternativa que perciben que tiene la mayor utilidad en esa ocasión; Es decir, la alternativa i se elige preferentemente a la alternativa j si y sólo si.

$$u_i > u_j, \text{ para todas las } j \neq i \in S (57)$$

Donde $i \neq j = 1, \dots, q \in S$. Las proporciones de elección observadas p_1, p_2, \dots, p_q para el set S son las proporciones de azar que se basa en una distribución q variable aleatoria de cada variable aleatoria que tiene el valor más grande, donde los sorteos son las opciones de los encuestados.

3.4 El modelo logístico

Este modelo es el más utilizado en la experimentación de elección discreta. La distribución acumulativa de las diferencias de valor extremo independiente e idénticamente distribuido varía. La distribución logit acumulativa para las probabilidades

de elección implica que los errores en los valores de utilidad son valores extremos. Esta distribución también se conoce como la distribución de valores extremos gumbel tipo 1. La función de distribución es

$$F(X) = 1 - \exp\{-\exp[\frac{x-a}{b}]\} \quad (58)$$

Números aleatorios de la variable de valor de valor extrema ε : a, b puede calcularse a partir de la relación.

$$x = a + b \ln\{\ln[\frac{1}{(1-rndu(x))}]\} \quad (59)$$

Donde el parámetro de ubicación a es la moda, el parámetro de escala b es > 0 , $rndu(x)$ es una unidad rectangular variada y la varianza de x es $b^2\pi^2/6$. La media de un valor variado extremo es -0.57721 . Sin embargo, la media no es relevante siempre y cuando las distribuciones son iid como la elección es impulsado por las diferencias entre las distribuciones.

Considere un conjunto de elección S con alternativas “ q ” elección que tienen valores de las partes de pieza “ u ” dada por la ecuación 51. Sea $\varepsilon_i: a, b; i = 1, \dots, q$ ser “ q ” valor extremo de variables aleatorias que tiene escala común un parámetro b y moda a . Con el modelo logit, la probabilidad población de elegir alternativa i de conjunto de elección S puede escribirse como

$$P_{i \in S} = e^{v_i} / \sum_{j \in S} e^{v_j} \quad (60)$$

Donde v_i se refiere a la utilidad verdadera para i , y $\sum_{j \in S} e^{v_j}$ se refiere a la suma sobre la alternativa en conjunto S . Es evidente que $\sum_{i \in S} p_i = 1.00$. Para el propósito de la estimación, se añade una alternativa de base a cada conjunto. La alternativa de base por lo general se toma como otra opción o elección retardada, en cuyo caso, asignándole un valor sistemático v_{base} de cero tiene la interpretación lógica de establecer que el valor incremental adquirido por la elección no es cero.

A veces, se utilizan los valores de base distintos de ninguna opción o elección retardada. La razón para hacer esto puede ser explícitamente para evaluar el impacto de un cambio en el producto, una cadena o una forma de hacer las cosas en cuotas de mercado actuales.

Cuando se añade una opción de base para cada conjunto de elección, el resultado es $q + 1$ alternativa en cada conjunto de elección. En este caso, la probabilidad hizo la alternativa i -ésimo será elegido de conjunto S es una función de las diferencias $\varepsilon_i - \varepsilon_{base}$.

$$P([\varepsilon_j - \varepsilon_{base}] - [\varepsilon_i - \varepsilon_{base}] < [v_i - v_{base}] - [v_j - v_{base}], \forall j \neq i). \quad (61)$$

Es importante tener en cuenta, que los valores estimados para la alternativa son en realidad contrastados, las diferencias entre el valor de la alternativa y el valor de la opción de base. La probabilidad de elegir alternativa i con una opción de base es la probabilidad que la diferencia $[v_i - v_{base}]$ es mayor que $[v_j - v_{base}]$ para $i, j \neq i$. Un cambio de base desde la otra opción a la marca dominante en un mercado afectará a las magnitudes e incluso los signos de las utilidades que fueron estimados como contrastes entre la alternativa y una base.

La variable dependiente con el modelo logit consiste en $\ln(p_{ij}/p_{i,base})$, el log de la relación de p_{ij} a $p_{i,base}$, donde p_{ij} es la proporción de participantes si se escoge la alternativa j ésima en el grupo i , y $p_{i,base}$ es la proporción que escoge la opción baja en el sistema de la opción i .

Las probabilidades de elección de probabilidad de log de cada alternativa i a la alternativa base ($i=0$) son.

$$l_{i \in t_j} = \ln\left(\frac{P_{i \in t_j}}{P_{0=base \in t_j}}\right) = \ln\left(\frac{e^{v_i} / \sum_{i \in S} e^{v_i}}{e^{v_{0=base}} / \sum_{i \in S} e^{v_i}}\right) = v_i - v_0 \quad (62)$$

La ecuación 62 implica independencia de alternativas irrelevantes; Es decir, la adición de una alternativa al conjunto no afectará las razones de posibilidades de log de las proporciones de elección de pares. Una interpretación alternativa es que los cambios en las acciones son proporcionales a los cambios en las preferencias.

La independencia de las propiedades alternativas irrelevantes es el foco de la investigación en curso porque hay muchas pruebas de que se viola tanto en los datos de elección individual como de grupo. En realidad, las alternativas no compiten entre sí.

La experimentación de elección discreta introduce muchos problemas desafiantes en el diseño de experimentos. En muchas aplicaciones, las características de las alternativas a menudo presentan costos y beneficios. Es decir, el valor de los niveles de atributos disminuye o aumenta monóticamente con el aumento del nivel del atributo. En las situaciones, los perfiles de un conjunto de opciones deben diseñarse de manera que se descarten elecciones obvias, tales como conjuntos que impliquen opciones para las cuales una alternativa domina o está dominada por otras alternativas en el conjunto de opciones. Conjuntos en los que ninguna alternativa domina o es dominada se llaman conjuntos de elección óptima de Pareto.

Además, cuando se selecciona una alternativa de un conjunto de opciones, se supone que la elección es impulsada por las características de la alternativa seleccionada y las otras alternativas disponibles. Es decir, en situaciones de elección, las características de las alternativas se comparan con las de otras alternativas disponibles. El atractivo de una alternativa en ausencia de otras alternativas se conoce como el efecto propio de la alternativa.

Los efectos de otras alternativas disponibles en una elección establecida sobre el atractivo de la alternativa se denominan efectos cruzados. Los diseños que permiten estimar los efectos propios y cruzados se llaman diseños de disponibilidad. La creación de estos conjuntos presenta problemas de diseño interesantes y desafiantes.

La migración de la investigación de encuestas a internet plantea varios problemas para el análisis conjunto. Uno de estos problemas es la pantalla limitada de verdaderas declaraciones disponibles para la presentación de los conjuntos de elección. Existen varias estrategias para abordar este problema. Se trata de utilizar diseños que tienen tamaños de conjunto de pequeñas opciones. Un segundo es crear diseños usando subconjuntos de atributos, subconjuntos de niveles o ambos. Todas estas estrategias plantean problemas de diseño experimental. También implican costos significativos en la forma del número de conjuntos de opciones que comprenden diseños completos.

Una forma de reducir el número de conjuntos de opciones necesarias para estimar los parámetros del modelo es participar en la experimentación secuencial. La experimentación secuencial implica la identificación del conjunto de perfiles necesarios para adaptarse al modelo más parsimonioso que pueda caracterizar un proceso de elección. Si este conjunto se adapta adecuadamente a los datos, el proceso experimental se detiene. De lo contrario, se recogen las opciones para un conjunto adicional mínimo de perfiles y se evalúa el ajuste a un modelo algo más complejo. Si este modelo se adapta adecuadamente a los datos, el proceso experimental se detiene. De lo contrario, se recopilan opciones adicionales para un conjunto adicional de perfiles. La tarea de diseño experimental con la experimentación secuencial es la creación de tales conjuntos secuenciales de perfil.

Hay dos consideraciones cuando se asocian costos económicos explícitos con los niveles de atributo: primero es el costo total del perfil de concepto y el segundo es la asignación de los costos totales entre los atributos. Ambos factores pueden influir en las preferencias de las personas para los perfiles. Estas cuestiones no se han separado en la literatura de experimentación de elección discreta. Los diseños de mezclas y cantidades de mezcla utilizados en experimentos industriales pueden usarse para separar los efectos de cantidad frente a mezcla en estudios de experimentos de elección discreta.

3.5 Principios del diseño experimental

Para sacar conclusiones válidas y útiles de cualquier experimento, la planificación adecuada del estudio es crucial. Esto incluiría la preparación del instrumento para recolectar datos y decisiones tales como cómo se administrará el instrumento a los sujetos, el número de temas necesarios y el tipo de análisis que se llevará a cabo. Estas y otras cuestiones se consideran en el diseño estadístico de los experimentos. Los conceptos de diseño experimental se originan a partir de experimentos agrícolas y lentamente se encuentran en otras ramas de la experimentación. La terminología de diseño, como parcelas, bloques y tratamientos, debe interpretarse en función del contexto del experimento. Una parcela es la unidad de respuesta experimental final en la que se recogen las observaciones. En la investigación social, estos son los sujetos, aunque pueden ser almacenes o regiones geográficas. Las unidades de respuesta pueden organizarse en bloques. Un bloque es una colección de unidades de respuesta

experimental homogénea y puede interpretarse como un grupo de unidades de respuesta con características similares. Las unidades de respuesta se asignan a los grupos de tratamiento. Los tratamientos son las intervenciones externas aplicadas a los elementos de los grupos de tratamiento para observar las diferencias en la respuesta.

Aleatorización, replicación y control local son tres características esenciales de cualquier experimento bien diseñado. Aleatorización es la asignación de las unidades de respuesta a los tratamientos al azar de las diferencias sistemáticas entre los grupos de tratamiento y nos permite utilizar la teoría del muestreo aleatorio como un criterio para juzgar si las diferencias observadas los grupos de tratamiento son más grandes de lo que se espera que se produzca por casualidad. La replicación nos permite controlar la magnitud de las diferencias que se espera que se produzca por casualidad. La variación de las respuestas entre los tratamientos se comparará con la variación de las respuestas dentro de los tratamientos, para lo cual la variación dentro del grupo es una indicación de la variación que se espera ocurra por azar. Más repeticiones aumentan la precisión del estudio.

El control local se utiliza en un sentido amplio para implicar los métodos que reducen las fuentes de variación en las observaciones distintas de la variabilidad debida a errores aleatorios. Una forma de control local es llevar a cabo un experimento en un ambiente controlado que elimine fuentes externas de variación externas. Las fuentes internas de variación, como las diferencias sistemáticas en las unidades de respuesta, pueden ser controladas mediante el diseño o estadísticamente realizando lo que se conoce como análisis de covarianza. Los esfuerzos de control local implican inevitablemente intercambios entre la validez interna y externa de una experiencia. La validez interna se refiere a si las diferencias observadas entre los grupos de tratamiento pueden atribuirse correctamente al tratamiento. La validez externa se refiere a si una inferencia válida de que las diferencias en los tratamientos eran la fuente de diferencias en las observaciones puede generalizarse a entornos no experimentales.

3.6 Diseño experimental versus tratamiento

La forma en que se aplican los tratamientos a los sujetos se denomina diseño experimental; La forma en que se crean los tratamientos se llama diseño del tratamiento. Los tratamientos en el análisis conjunto y la experimentación de elección discreta se

presentan en forma de perfiles consistentes en niveles declarados sobre atributos. La principal razón para presentar los tratamientos como representaciones abstractas de alternativas de elección es asegurar que los encuestados evalúen cada perfil con respecto a la misma información. Otros ejemplos del diseño del tratamiento incluyen el experimento de trabajador despedido en el que se juzga una solicitud de asistencia gubernamental basada en la raza, el género, la edad, el estado civil y el historial laboral.

Los diseños experimentales bien conocidos incluyen el diseño completamente al azar, el diseño de bloques al azar, el diseño de cuadrados latinos, el diseño de bloques incompletos y así sucesivamente. Los experimentos factoriales y las replicaciones fraccionadas son diseños de tratamiento conocidos.

3.7 Diseños de bloques incompletos equilibrados y diseños 3

Cuando las unidades experimentales o sujetos son heterogéneos, podemos agruparlos en bloques de unidades homogéneas. Las unidades dentro de un bloque son entonces homogéneas, mientras que las unidades entre bloques son generalmente heterogéneas. Si cada tratamiento se aplica una vez en cada bloque, el diseño resultante se llama diseño aleatorio de bloques. Por supuesto, tal diseño no es factible si el número de tratamientos es grande, digamos 10 o más. En tal situación, se puede utilizar un diseño de bloque incompleto. Uno de estos diseños, el diseño de bloque incompleto equilibrado es una disposición de v tratamientos en b conjuntos de tamaño k tal que cada tratamiento se produce al menos una vez en un conjunto, cada tratamiento se produce exactamente en r conjuntos y cada par de tratamientos distintos ocurre juntos en conjuntos λ .

Aquí, v , b , r , k y λ se llaman los parámetros del diseño de bloque incompleto equilibrado y satisfacen las siguientes relaciones

$$vr = bk, r(k - 1) = \lambda(v - 1), b \geq v \quad (63)$$

Se dice que un diseño de bloque incompleto y equilibrado es simétrico si $v = b$. Cualquiera dos bloques distintos de un diseño de bloque incompleto simétrico equilibrado tendrá λ tratamientos en común. Al intercambiar los roles de tratamientos y bloques en cualquier diseño de bloques incompletos equilibrados, podemos construir su

diseño dual. El diseño dual de un diseño de bloque incompleto equilibrado tendrá x tratamientos en común entre dos bloques y también se denomina diseño de bloque vinculado. Un diseño simétrico equilibrado de bloques incompletos es claramente un diseño de bloque vinculado.

Un diseño de v tratamientos dispuestos en b conjuntos de tamaño k_1, k_2, \dots , donde no todos los k son iguales y que satisface las tres condiciones de un diseño de bloque incompleto equilibrado se denomina un diseño equilibrado pairwise. Una generalización del diseño de bloque incompleto equilibrado es permitir que cada triplete de tratamientos distintos ocurra igualmente a menudo en los bloques. Dicho diseño se denomina diseño en 3, que se define como una disposición de v tratamientos en bloques b de tamaño k tal que cada tratamiento se produce al menos una vez en un bloque, cada tratamiento se produce exactamente en bloques r , cada par de tratamientos distintos ocurre juntos en λ_2 bloques, y cada triplete de tratamientos distintos ocurre juntos en λ_3 bloques.

3.8 Experimentos factoriales

La forma en que los tratamientos se aplican a los sujetos se denomina diseño experimental, mientras que la forma en que se crean los tratamientos se denomina diseño del tratamiento. Los experimentos factoriales son esencialmente diseños de tratamientos. Considere un estudio con m factores con el factor i -ésimo en los niveles s_i ($i = 1, 2, \dots, m$). Este experimento se denomina experimento $s_1 \times s_2 \times \dots \times s_m$. Si $S_1=S_2=\dots=S_m$, el experimento factorial debe de ser simétrico y se lo escribe como S^m . Si al menos dos s_i son diferentes, se dice que el experimento es asimétrico. La ventaja de los experimentos factoriales es que podemos estudiar tanto los principales efectos como las interacciones de los factores.

3.8.1 Experimentos factoriales fraccionales

A medida que aumenta el número de factores en un experimento de 2^m , el número de perfiles (o recorridos) en el experimento crece rápidamente. Por ejemplo, un experimento 2^5 tiene 32 perfiles, mientras que un experimento de 2^7 tiene 128. Por esta razón, los diseños factorial completos son factibles sólo si el número de factores es relativamente pequeño. Para un gran número de factores, el experimentador puede no tener los recursos para ejecutar el experimento factorial completo y puede optar por ejecutar sólo un subconjunto de la fracción del diseño. Tal diseño se llama un diseño

factorial fraccional. Los diseños factoriales fraccionados nos dan información sobre los principales efectos y las interacciones de bajo orden, siempre que se pueda suponer que algunas de las interacciones de orden superior son insignificantes e insignificantes. Fraccional diseños factorial se utilizan comúnmente en los experimentos de cribado en el que un gran número de factores están bajo consideración y el experimentador está interesado en la identificación de un pequeño número de factores importantes. Una vez que estos importantes factores e interacciones se identifican, se pueden estudiar con más detalle en experimentos posteriores.

Por ejemplo, en un estudio que involucraba tres atributos en dos niveles cada uno, el experimento completo requeriría ocho perfiles. Si los experimentadores tienen los recursos para obtener respuestas en sólo cuatro perfiles, pueden optar por ejecutar una mitad de la fracción del diseño completo. La pregunta, entonces, ¿cuáles son los cuatro perfiles del factorial completo que deben elegir? envés de escoger una muestra aleatoria de cuatro series de ocho, tratamos de escoger una fracción o subconjunto que permita la estimación de los tres efectos principales. Denotan los tres factores por a, b y c y los dos niveles de cada factor por los subíndices 0 y 1. $T_{a_i b_j c_k}$ será la clasificación total recibida de los clientes r por el perfil $a_i b_j c_k$ ($i, j, k = 0, 1$). Reemplazando $T_{a_i b_j c_k}$ por $a_i b_j c_k$, el experimento 2^3 posee signos positivos y negativos.

Suponga que se escogen los cuatro perfiles $a_1 b_0 c_0$, $a_0 b_1 c_0$, $a_0 b_0 c_1$, y $a_1 b_1 c_1$. Este diseño se lo llama factorial fraccional 2^{3-1} porque es mitad fracción de un experimento factorial 2^3 y posee $2^{3-1}=4$ perfiles. Para estos cuatro perfiles, la interacción de tres factores abc tiene todos los signos positivos. Por esta razón, decimos que la relación de definición de la fracción es $I = ABC$. La relación de definición para cualquier factorial fraccionario consiste en los efectos principales y las interacciones que tienen el mismo signo para los perfiles seleccionados.

El contraste utilizado para estimar el efecto principal A y la interacción de dos factores BC son idénticos. De hecho, cuando estimamos el efecto principal A, estamos en realidad estimando $A + BC$, y decimos que A es un alias con BC. El uso de una fracción de la mitad del experimento factorial completo resulta en una pérdida de información. Por lo tanto, para estimar los efectos principales A, B y C en el experimento 2^{3-1} , debemos asumir que las interacciones de dos factores son despreciables. Esta suposición se requiere porque en la fracción 2^{3-1} los efectos principales están aliados

con las interacciones de dos factores. Podemos determinar la estructura del alias del experimento 2^{3-1} examinando su relación de definición $I = ABC$. Multiplicando ambos lados de la relación de definición por el efecto principal A y ese $A^2=B^2=C^2=I$, tenemos.

$$A = A^2BC = IBC = BC \quad (64)$$

Es decir, el efecto principal A está aliado con la interacción de dos factores BC. De forma similar, el efecto principal B está aliado con AC, y el efecto principal de C está aliado con AB. Si es razonable suponer que las interacciones de dos factores tienen un efecto insignificante, el experimento 2^{3-1} se puede utilizar para estimar los efectos principales.

Los diseños factorial fraccionales pueden clasificarse en base a su resolución. Se dice que un diseño es un diseño de resolución III si los efectos principales no están aliados entre sí, pero están aliados con interacciones de dos factores y las interacciones de dos factores pueden estar aliadas con otras interacciones de dos factores. El experimento 2^{3-1} descrito es un diseño de resolución III. Se dice que un diseño es la resolución IV si los efectos principales no están aliados entre sí o las interacciones de dos factores, pero las interacciones de dos factores están aliados entre sí. Por ejemplo, considere una media fracción del diseño 2^4 con la relación definida $I = ABCD$. Esta fracción consiste en los $2^{4-1}=8$ perfiles de la columna completa. Multiplicando ambos lados de la relación de definición $I = ABCD$ por cualquier efecto principal obtenemos

$$A = A^2BCD = IBCD = BCD \quad (65)$$

Y por cualquier interacción de dos factores, obtenemos

$$AB = A^2B^2CD = ICD = CD \quad (66)$$

Podemos ver que los efectos principales no están aliados entre sí o las interacciones de dos factores, pero las interacciones de dos factores son aliadas entre sí. Finalmente, un diseño de resolución V es aquel en el que los efectos principales no están aliados entre sí o las interacciones de dos factores, y las interacciones de dos factores no están aliadas entre sí, sino que tienen un alias con tres o más interacciones de orden. La media fracción del diseño 2^5 con la relación de definición $I = ABCDE$ es un ejemplo de un diseño de resolución V. En general, se dice que un diseño es de resolución R si la

longitud mínima de palabra en la relación es R. En un diseño de resolución R, ningún efecto de factor p se confunde con ningún otro efecto que contenga menos que R-p factores. En grandes experimentos, a menudo es el caso que sólo los efectos principales y algunas interacciones de bajo orden son de interés real. Por lo tanto, cuando se utilizan diseños factorial fraccionarios, se debe elegir la fracción de resolución tal que los efectos principales y las interacciones de interés no sean alias con otros efectos de interés.

A medida que aumenta el número de factores considerados, puede ser necesario utilizar incluso una fracción más pequeña del factorial completo. Por ejemplo, en un estudio que involucra cinco atributos en dos niveles cada uno, el experimento completo 2^5 requeriría 32 perfiles. Si los experimentos tienen los recursos para obtener respuestas en sólo 8 perfiles, pueden optar por ejecutar un cuarto de fracción del diseño 2^5 , el diseño 2^{5-2} . Denotamos los cinco factores por a, b, c, d y e y asumimos que el experimentador escoge las ocho series para las que las interacciones ABD, ACE y BCDE tienen el mismo signo. La relación de definición de esta fracción es $I = ABD = ACE = BCDE$. ABD y ACE se llaman los generadores de esta fracción, mientras que BCDE es la interacción generalizada de ABD y ACE desde $(ABD)(ACE) = A^2BCDE = BCDE$. En general, la relación definitoria de cualquier cuarta fracción de un factorial completo tendrá dos generadores y su interacción generalizada. Dado que la longitud mínima de la palabra en su relación de definición es de tres, este diseño 2^{5-2} es un diseño de resolución III.

Este diseño 2^{5-2} puede obtenerse escribiendo primero un diseño completo 2^2 y luego añadiendo dos columnas para D (= AB) y E (= AC) ya que D está aliado con AB y E tiene alias con AC.

3.9 Matrices de Hadamard y matrices ortogonales

Las matrices ortogonales son útiles en el diseño de experimentos, y la discusión aquí comienza con un ejemplo simple. El siguiente es un ejemplo de una matriz ortogonal con dos niveles y una intensidad 2.

$$\begin{vmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{vmatrix} \quad (67)$$

Esta es una matriz ortogonal de fuerza 2 porque si seleccionamos dos columnas, cada combinación posible de los niveles ocurre el mismo número de veces en las dos columnas. Llamamos a esta matriz una OA ya que tiene cuatro filas, tres columnas, dos niveles y una fuerza 2, respectivamente.

En general, se dice que una matriz $n \times k$ A con entradas de un conjunto de símbolos S es una matriz ortogonal con S niveles, fuerza t e índice λ si cada subárbol $n \times t$ de A contiene cada t - tupla basada en S símbolos exactamente λ veces como una fila. Denotamos tal matriz por OA (n, k, s, f) . Nótese que no es necesario mencionar λ en esta notación porque puede determinarse mediante la ecuación $\lambda = n / s^t$. En el ejemplo, cada combinación posible de los niveles ocurre exactamente una vez en las dos columnas, así que $\lambda = 1$, y decimos que la matriz ortogonal tiene unidad de índice. El siguiente ejemplo es un OA $(8, 4, 2, 3)$, es decir, tiene ocho filas, cuatro columnas, dos niveles y una fuerza 3.

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \quad (68)$$

Es fácil verificar que $\lambda = n / s^t = 8/2^3 = 1$ en este ejemplo. Es decir, en cada tres columnas de la matriz ortogonal, cada una de las ocho posibilidades $(000,001,010,011,100,101,110,111)$ ocurre exactamente una vez. Las matrices ortogonales presentadas hasta ahora son matrices con sólo dos niveles (0 y 1). Un ejemplo de un conjunto ortogonal OA $(9, 4, 3, 2)$ con tres niveles (0, 1 y 2) es el siguiente

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 2 & 2 & 2 \\ 1 & 0 & 2 & 1 \\ 1 & 1 & 0 & 2 \\ 1 & 2 & 1 & 0 \\ 2 & 0 & 1 & 2 \\ 2 & 1 & 2 & 0 \\ 2 & 2 & 0 & 1 \end{pmatrix} \quad (69)$$

Dey and Mukherjee (1999), Stufken (1999) and Raghavarao (1971) son los trabajos más reconocidos cuando se habla de investigaciones estadísticas y matemáticas.

Las matrices Hadamard, llamadas así por el famoso algebrista Francés Jacques Hadamard, están relacionadas directamente con las matrices ortogonales. Las matrices de Hadamard son matrices cuadradas de +1s y -1s cuyas filas son ortogonales. En general, una matriz de Hadamard H de orden n es una matriz cuadrada $n \times n$ con las siguientes propiedades

i cada elemento de H es +1 o -1

$$ii \ H'H=HH'=nl \ (70)$$

En donde H' es la transpuesta de H .

De la propiedad ii queda claro que las columnas también son ortogonales. Las siguientes son matrices Hadamard de orden 2 y 4.

$$\begin{bmatrix} +1 & +1 \\ +1 & -1 \end{bmatrix} \begin{bmatrix} +1 & +1 & +1 & +1 \\ +1 & -1 & +1 & -1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & -1 & +1 \end{bmatrix} \ (71)$$

Puede demostrarse que, si existe una matriz de Hadamard, entonces n es 2 o un múltiplo de 4, es decir, $n = 2$ o $n = 4t$. El inverso, que si n es un múltiplo de 4 una matriz de Hadamard debe existir, es una conjetura. El múltiplo más pequeño de 4 para el cual ninguna matriz de Hadamard ha sido construida hasta ahora es 668. Cualquier matriz obtenida permutando las filas o columnas de H o multiplicando sus filas o columnas por -1 es también una matriz de Hadamard. Todas las matrices obtenidas de esta manera se dice que son isomorfas o equivalentes a H .

Podemos comenzar con una matriz de Hadamard y multiplicar sus filas y columnas por -1 como sea necesario para hacer que cada elemento en la primera fila y la primera columna sea igual a 1. Tal matriz de Hadamard se dice que se normaliza. Si cada elemento en la primera columna es igual a 1, decimos que la matriz es seminormalizada. Al identificar todas las columnas excepto la primera con factores $n-1 = 4t-1$, podemos ver que una matriz seminómica de Hadamard es una matriz de diseño para un plan de efectos principales ortogonales para factores $4t-1$ (a dos niveles cada uno).

Es decir, obtenemos un plan ortogonal de efectos principales de un experimento factorial fraccional de 2^{4t-1} en recorridos $4t$. Los diseños factoriales fraccionales obtenidos de las matrices de Hadamard son diseños de la resolución III. También se les llama diseños saturados ya que no se deja ningún grado de libertad para estimar el error.

3.10 Diseños Foldover

Recuerde que en un diseño de resolución III los efectos principales no están aliados entre sí, sino que son alias con interacciones de dos factores, y las interacciones de dos factores pueden ser alias con otras interacciones de dos factores. Si empezamos con un factorial fraccional de resolución III y le añadimos una segunda fracción con los signos de todos los factores invertidos, el diseño combinado se denomina diseño foldover. Este proceso, llamado plegado sobre un diseño, rompe los vínculos de alias entre los efectos principales y las interacciones de dos factores, dándonos un diseño IV de resolución. Es decir, podemos usar el diseño combinado, o el diseño foldover, para estimar los efectos principales claros de las interacciones de dos factores.

Considere el diseño de la resolución 2^{5-2} III. ABD y ACE son los generadores de esta fracción, y la relación de definición completa de esta fracción es $I = ABD = ACE = BCDE$. Resulta claro de la relación de definición que los cinco efectos principales están aliados con las interacciones de dos factores. Para separar los efectos principales y las interacciones de dos factores, podemos ejecutar una segunda fracción con los signos de todos los factores invertidos.

La relación definitoria del diseño foldover se puede obtener de la relación de definición de sus fracciones. Cuando doblamos un diseño de resolución III, de hecho, estamos cambiando los signos de los generadores que tienen un número impar de letras. En el ejemplo dado, la relación de definición de la segunda fracción es $I = -ABD = -ACE = BCDE$. Comparando las relaciones definitorias de las dos fracciones, podemos ver que tienen una palabra con el mismo signo (BCDE) y dos palabras con diferentes signos (ABD y ACE). El diseño combinado tendrá todas las palabras con el mismo signo que los generadores (aquí BCDE) más todos los productos independientes incluso de las palabras con diferentes signos. Por los productos pares, queremos decir que las palabras se toman dos a la vez, cuatro a la vez y así sucesivamente.

En este ejemplo, el producto par de (ABD) (ACE) es BCDE, por lo que la relación de definición completa del diseño foldover es simplemente $I = BCDE$. De esta relación de definición se desprende claramente que el diseño combinado es un diseño de resolución IV. El ejemplo es una ilustración de la técnica foldover para obtener diseños de resolución IV; Sin embargo, tenga en cuenta que un diseño aún mejor para cinco factores en 16 ejecuciones es la resolución V, diseño 2^{5-1} con la relación de definición $I = ABCDE$.

Recordemos que los diseños factorial fraccionarios derivados de matrices de Hadamard son diseños de resolución III. Box y Wilson (1951) describieron por primera vez el método de obtener diseños de resolución IV plegando diseños de resolución III derivados de matrices de Hadamard. También es interesante notar que, si añadimos a un diseño factorial fraccional una segunda fracción con los signos de un solo factor invertido, podemos romper los vínculos de alias entre el efecto principal de ese factor y sus interacciones de dos factores.

3.11 Experimentos de mezcla

Un experimento en el que los factores son los ingredientes o constituyentes de una mezcla se denomina experimento de mezcla. En un experimento de mezcla, la respuesta es una función de las proporciones de cada ingrediente, con la cantidad de cada ingrediente usualmente medida por peso, volumen o similar. Algunos ejemplos de pozos de experimentos de mezcla son mezclas de gasolina que combinan dos o más reservas de gasolina; Torta u otras formulaciones de alimentos que combinan ingredientes tales como polvo de hornear, harina, azúcar y agua, y hormigón formado mezclando arena, agua y cemento.

Considere una mezcla formada mezclando tres ingredientes y deje x_1 , x_2 , y x_3 representar las proporciones de cada ingrediente. Una mezcla con las proporciones $x_1 = 0,2$, $x_2 = 0,3$ y $x_3 = 0,5$ se llama mezcla completa porque está compuesta por los tres ingredientes. Una mezcla con las proporciones $x_1 = 0,7$, $x_2 = 0,3$ y $x_3 = 0,0$ se denomina mezcla binaria porque se compone de sólo dos ingredientes, mientras que una mezcla con las proporciones $x_1 = 1,0$, $x_2 = 0,0$ y $x_3 = 0,0$ es Denominada mezcla de componentes puros o de un solo componente ya que está constituida por un solo ingrediente. Tenga en cuenta que, en los tres ejemplos, $x_1 + x_2 + x_3 = 1$.

Debido a esta limitación, los niveles de los factores en un experimento de mezcla no se pueden elegir independientemente; Es decir, si sabemos que $x_1 = 0.2$ y $x_2 = 0.3$, entonces x_3 debe ser 0.5. Para generalizar, supongamos que una mezcla consiste en m componentes y x_i representa la proporción del componente *iésimo*, entonces

$$x_i \geq 0, i = 1, 2, \dots, m \quad (72)$$

$$\sum_i^m x_i = 1. \quad (73)$$

Dado que los niveles de los factores en un experimento de mezcla no son independientes, esto los hace diferentes de los experimentos que hemos descrito hasta ahora. Para una mezcla con dos componentes, el espacio factor factible incluye todos los valores de x_1, x_2 para los cuales $x_1 + x_2 = 1$. Esta restricción puede ser representada gráficamente por un segmento de línea. Para una mezcla con tres componentes, el espacio factor factible es un triángulo en el que los vértices representan mezclas puras (sólo un componente), y los bordes representan mezclas binarias (sólo dos componentes).

En general, la región experimental para una mezcla de componentes m es un simplex, es decir, una figura de lados regulares con m vértices y dimensiones $m-1$, y todos los puntos de diseño están en los vértices, en los bordes o caras, o en el interior Del simplex. Un conjunto de puntos uniformemente espaciados en un simplex se llama un diseño de celosía simplex. Considere un experimento de mezcla con m componentes y deje que cada componente tome las proporciones dadas por el $q + 1$ igualmente espaciado valorado de 0 a 1,

$$x_i = 0, \frac{1}{q}, \frac{2}{q}, \dots, 1 \quad i = 1, 2, \dots, m. \quad (74)$$

Si se utilizan todas las combinaciones posibles de las proporciones de esta ecuación, el diseño se denomina un diseño de malla simplex (m, q) para m componentes. Por ejemplo, si $m = 3$ y $q = 2$ entonces

$$x_i = 0, \frac{1}{2}, 1 \quad i = 1, 2, 3 \quad (75)$$

Y el (3, 2) diseño de red simple consiste en los puntos

$$(1, 0, 0), (0, 1, 0), (0, 0, 1), \left(\frac{1}{2}, \frac{1}{2}, 0\right), \left(\frac{1}{2}, 0, \frac{1}{2}\right), \left(0, \frac{1}{2}, \frac{1}{2}\right) \quad (76)$$

De manera similar, un diseño de red simple (3, 3) consiste en los puntos

$$(1, 0, 0), (0, 1, 0), (0, 0, 1), \left(\frac{2}{3}, \frac{1}{3}, 0\right), \left(\frac{1}{3}, \frac{2}{3}, 0\right), \left(\frac{2}{3}, 0, \frac{1}{3}\right), \left(\frac{1}{3}, 0, \frac{2}{3}\right), \left(0, \frac{2}{3}, \frac{1}{3}\right), \left(0, \frac{1}{3}, \frac{2}{3}\right), \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right). \quad (77)$$

Un modelo de primer orden está dado por

$$E(y) = \beta_0 + \sum_{i=1}^m \beta_i x_i \quad (78)$$

Donde E (y) es el valor esperado de y.

Sin embargo, debido a que $x_1 + x_2 + \dots + x_m = 1$, los parámetros B en esta ecuación no son únicos. Si multiplicamos el término B en la ecuación por $x_1 + x_2 + \dots + x_m = 1$ obtenemos

$$\begin{aligned} E(y) &= \beta_0(x_1 + x_2 + \dots + x_m) + \sum_{i=1}^m \beta_i x_i \quad (79) \\ &= \sum_{i=1}^m \beta_i^* x_i, \quad (80) \end{aligned}$$

Donde $\beta_i^* = \beta_0 + \beta_i$. Esta ecuación se denomina forma canónica del modelo de mezcla de primer orden. Al eliminar el asterisco del modelo, la forma canónica del modelo lineal es

$$E(y) = \sum_{i=1}^m \beta_i x_i \quad (81)$$

El polinomio general de segundo grado en m variables es

$$E(y) = \beta_0 + \sum_{i=1}^m \beta_i x_i + \sum_{i=1}^m \beta_{ii} x_i^2 + \sum_{i < j} \sum_{j=1}^m \beta_{ij} x_i x_j \quad (82)$$

De nuevo, aplicando el hecho de que $x_1 + x_2 + \dots + x_m = 1$, y

$$x_i^2 = x_i(1 - \sum_{j=1}^m x_j), \quad (83)$$

Obtendremos

$$E(y) = \beta_0(\sum_{i=1}^m x_i) + \sum_{i=1}^m \beta_i x_i + \sum_{i=1}^m \beta_{ii} x_i (1 - \sum_{j \neq i}^m x_j) + \sum_i \sum_{i < j}^m \beta_{ij} x_i x_j \quad (84)$$

$$= \sum_{i=1}^m (\beta_0 + \beta_i + \beta_{ii}) x_i - \sum_{i=1}^m \beta_{ii} x_i \sum_{j \neq i}^m x_j + \sum_i \sum_{i < j}^m \beta_{ij} x_i x_j \quad (85)$$

$$= \sum_{i=1}^m \beta_i^* x_i + \sum_i \sum_{i < j}^m \beta_{ij}^* x_i x_j, \quad (86)$$

Cuando los parámetros de la última ecuación son funciones simples de los parámetros originales, es decir, $\beta_i^* = \beta_0 + \beta_i + \beta_{ii}$ y $\beta_{ij}^* = \beta_{ij} - \beta_{ii} - \beta_{jj}$, $i, j = 1, 2, \dots, m$, $i < j$. Al retirar el asterisco del modelo, la forma canónica del modelo cuadrático es

$$E(y) = \sum_{i=1}^m \beta_i x_i + \sum_{i < j}^m \beta_{ij} x_i x_j. \quad (87)$$

3.12 Estimación

El problema de la estimación consiste en estimar los parámetros desconocidos de las distribuciones de probabilidad conocidas de las variables aleatorias. Los modelos comúnmente utilizados para las distribuciones de probabilidad son funciones de parámetros desconocidos.

Necesitamos estimar estos parámetros desconocidos a partir de los datos y probar las hipótesis que podemos formular sobre ellos. Dos procedimientos de estimación de uso común son el método de mínimos cuadrados y el método de máxima verosimilitud. Primero introducimos algunos modelos de probabilidad comúnmente usados y luego discutimos los métodos de estimación.

3.13 Modelos de probabilidad

Se dice que una variable aleatoria con sólo dos resultados tiene una distribución de Bernoulli. Los dos resultados se llaman éxito (denotado por 1) y el fracaso (denotado por 0). Sea la probabilidad de éxito, $P(X = 1) = p$, y la probabilidad de fracaso, $P(X = 0) = 1 - p = q$. Entonces

$$E(X) = p \text{ and } Var(X) = pq. \quad (88)$$

Supongamos que tenemos n ensayos, cada uno con sólo dos resultados, de modo que (a) el resultado de un ensayo no influye en el resultado de ningún otro ensayo, y (b) la probabilidad de éxito en n ensayos, entonces se dice X Tienen una distribución binomial con

$$P(X = x) = \binom{n}{x} p^x q^{n-x}, \quad (89)$$

Donde $x = 0, 1, \dots, n$. Aquí, $E(X) = np$ y $Var(X) = npq$.

Supongamos que tenemos n ensayos, cada uno con resultados $k + 1$ denotados por E_0, E_1, \dots, E_k , de manera que (a) el resultado de un ensayo no influye en el resultado de ningún otro ensayo, y (b) en cada uno El resultado E_i se produce con probabilidad $p_i, i = 0, 1, \dots, k$, tal que $\sum_{i=0}^k p_i = 1$. Se dice que el número de ensayos X_i , en los que $E_i (i = 0, 1, \dots, k)$, tiene una distribución multinomial con

$$P(X_i = x_i, i = 0, 1, \dots, k) = \frac{n!}{\prod_{i=0}^k x_i!} \prod_{i=0}^k p_i^{x_i}, \quad \text{Si } \sum_{i=0}^k x_i = n \quad (91)$$

$$= 0 \quad (90)$$

Se dice que una variable aleatoria X que puede tomar cualquier valor real tiene una distribución normal con media μ y varianza σ^2 si

$$P(x < X \leq x + dx) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx. \quad (92)$$

Aquí, $\frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ se denomina función de densidad de la variable aleatoria. Denotamos tal distribución por $N(\mu, \sigma^2)$.

3.14 Estimación lineal: Mínimos cuadrados y mínimos cuadrados ponderados

Consideramos que el valor esperado de las respuestas o respuestas adecuadamente transformadas son funciones lineales de parámetros desconocidos con coeficientes conocidos y estiman los parámetros como funciones lineales de respuestas o respuestas transformadas. Sean n respuestas o respuestas transformadas denotadas por y_1, y_2, \dots, y_n , asumimos

$$E(y_i) = \sum_{j=1}^k x_{ij}\beta_j, \quad (93)$$

Donde $\beta_1, \beta_2, \dots, \beta_k$ son parámetros desconocidos, y x_{ij} son constantes conocidas. Por lo general, x_{ij} son las variables demográficas de los encuestados o las características de los ítems presentados a los encuestados. $X = (x_{ij})$, $y' = (y_1, y_2, \dots, y_n)$, y $B' = (\beta_1, \beta_2, \dots, \beta_p)$. Entonces tenemos

$$E(y) = X\beta. \quad (94)$$

Aquí, X es conocida como la matriz de diseño y es de orden $n \times k$.

Se dice que una función paramétrica lineal $l'\beta$ es estimable si existe un vector a tal que

$$E(a'y) = l'\beta. \quad (95)$$

Es interesante notar que no todas las funciones lineales $l'\beta$ son estimables con cualquier matriz de diseño X , y $l'\beta$ es estimable si y sólo si l es una combinación lineal de las filas de X . Aquí, $a'y$ se llama un Estimador imparcial de $l'\beta$, y puede haber varios estimadores no sesgados de $l'\beta$. De los estimadores no sesgados, nos gusta elegir el estimador con la menor varianza, conocido como el mejor estimador lineal no sesgado de $l'\beta$. En la literatura de diseño, se demuestra que el mejor estimador lineal no sesgado de $l'\beta$ es $l'\hat{\beta}$, donde $\hat{\beta}$ es una solución de la ecuación

$$X'E^{-1}X\hat{\beta} = X'E^{-1}y, \quad (96)$$

Siendo Σ la matriz de dispersión de y . La ecuación se denomina ecuación normal ponderada y se obtiene minimizando $(y - X\beta)' \Sigma^{-1}(y - X\beta)$ con respecto a β . Puesto que $\hat{\beta}$ se obtiene a partir de mínimos cuadrados ponderados, se denomina estimador de mínimos cuadrados ponderado. Cuando $\Sigma = \sigma^2 I_n$, donde I_n es la matriz de identidad, $\hat{\beta}$ se denomina el estimador de mínimos cuadrados ordinarios.

Cuando el interés está en estimar cada uno de los k parámetros individuales B_1, B_2, \dots, B_k , necesitamos la matriz de diseño X tal que $X'\Sigma^{-1}X$ sea no singular. En ese caso,

$$\hat{\beta} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y \quad (97)$$

Son los parámetros estimados con

$$Var(\hat{\beta}) = (X' \Sigma^{-1} X)^{-1} \quad (98)$$

3.15 Estimación de máxima verosimilitud

Si X_1, X_2, \dots, X_n es una muestra aleatoria, la probabilidad conjunta de obtener los valores de la muestra expresados como una función de parámetros desconocidos se denomina probabilidad. Maximizando la probabilidad con respecto a los parámetros desconocidos, obtenemos estimadores de máxima verosimilitud de los parámetros. Maximizar la probabilidad es lo mismo que maximizar la probabilidad de log.

Consideremos la distribución normal multinomial y supongamos que observamos las frecuencias $x_0, x_1 \dots x_k$. Entonces, la probabilidad es

$$L(P_0, P_1, \dots, P_k) = \frac{n!}{\prod_{i=0}^k x_i!} \prod_{i=0}^k p_i^{x_i} \quad (99)$$

Y la probabilidad de registro es

$$l(p_0, p_1, \dots, p_k) = \ln L(p_0, p_1, \dots, p_k) = \ln n! - \sum_{i=0}^k \ln x_i! + \sum_{i=0}^k x_i \ln p_i \quad (100)$$

Diferenciando $l(p_0, p_1, \dots, p_k)$ con respecto a p_i , para $i = 1, 2, \dots, k$, anotando $\sum_{i=0}^k p_i = 1$ e igualando a cero, obtenemos

$$-\frac{x_0}{1 - \sum_{i=1}^k \hat{p}_i} + \frac{x_i}{\hat{p}_i} = 0, \quad i = 1, 2, \dots, k \quad (101)$$

Esto implica

$$\frac{x_1}{\hat{p}_1} = \frac{x_2}{\hat{p}_2} = \dots = \frac{x_k}{\hat{p}_k} = \frac{x_0}{1 - \sum_{i=1}^k \hat{p}_i} = \frac{n}{1}, \quad (102)$$

Y por lo tanto $\hat{p}_i = x_i/n$, para $i = 0, 1, 2, \dots, k$.

X_1, X_2, \dots, X_n sea una muestra aleatoria de $N(\mu, \sigma^2)$ donde μ y σ^2 son desconocidos. Entonces, la probabilidad es

$$L(\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2}} \quad (103)$$

y la probabilidad log es

$$l(\mu, \sigma^2) = \ln L(\mu, \sigma^2) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2} \quad (104)$$

$\frac{\partial l(\mu, \sigma^2)}{\partial \mu} = 0$ da $\hat{\mu} = \bar{x}$, mientras $\frac{\partial l(\mu, \sigma^2)}{\partial \sigma^2} = 0$ da como resultado $\hat{\sigma}^2 = \sum_{i=1}^n (x_i - \bar{x})^2/n$.

3.16 Pruebas de hipótesis lineales

Consideremos la transformación logit, suponemos que

$$E(L_0) = X\beta, \quad (105)$$

Donde X es una matriz k x m de rango r ($\leq m$), y β es un vector de los parámetros desconocidos de m. Queremos probar la hipótesis nula $H_0: C\beta = d$, donde C es de orden p x m y rango p ($\leq r$), y las funciones paramétricas $C\beta$ son estimables. S es la matriz de dispersión estimada.

$$S = \frac{1}{n} \left\{ D \left(\frac{1}{\hat{p}_1}, \frac{1}{\hat{p}_2}, \dots, \frac{1}{\hat{p}_k} \right) + \frac{1}{\hat{p}_0} J_k \right\} \quad (106)$$

Donde J, es una matriz k x m de todas, y $D \left(\frac{1}{\hat{p}_1}, \frac{1}{\hat{p}_2}, \dots, \frac{1}{\hat{p}_k} \right)$ es una matriz diagonal.

La falta de ajuste del modelo se prueba utilizando la estadística de prueba

$$T_1 = \widehat{L}_0' S^{-1} \widehat{L}_0 - \beta' X' S^{-1} L_0 \quad (107)$$

3.17 Diseños con atributos ordenados

Los experimentos de elección discreta a menudo, si no usualmente, tienen atributos económicos; Es decir, atributos que han ordenado niveles con beneficios asociados o costos que aumentan monótonicamente con el nivel.

Ejemplos incluyen volumen, medidas de rendimiento e incluso índices de calidad. Normalmente se supone que los beneficios aumentan a una tasa decreciente y los costos aumentan a un ritmo creciente con el aumento del nivel de atributo. Cada nivel del factor se supone que aumentan monótonicamente con beneficios. Se supone que las pérdidas asociadas a los niveles de precios disminuirán monótonamente con el aumento del nivel de precios.

En el caso de la teoría de la elección individual que subyace a la experimentación por elección discreta, se supone que hay beneficios incrementales ganados o costos incurridos al elegir un perfil sobre otro. Los tomadores de decisiones elegirán reducir los niveles de un atributo para aumentar los niveles de otro si la ganancia incremental de conmutación compensa más que la pérdida incremental. Claramente, los atributos deben ser ordenados para que las nociones de reducción y aumento de los niveles de atributos tengan sentido.

3.18 Efectos lineales, cuadráticos y cúbicos

Considere un experimento $s_1 \times s_2 \times \dots \times s_n$ con atributos A_1, A_2, \dots, A_n , donde el factor (o atributo) A_1 ocurre en s_i niveles denotados por $0, 1, 2, \dots, s_i-1$. Deje que cada ejecución (o perfil) sea replicada r veces. Con un experimento factorial completo, habrá $s_2 s_3 \dots s_n$ r observaciones con el atributo A_1 en el nivel i -ésimo. Con un diseño factorial fraccional, habrá la fracción apropiada de este número de observaciones. En general, sea $s_i^{A_1}$ el número de observaciones en el nivel i -ésimo de las observaciones del atributo A_1 , $s_{ij}^{A_1, A_2}$ con el atributo A_1 en el i -ésimo nivel y el atributo A_2 en el j -ésimo nivel y así sucesivamente.

3.18.1 Atributos a dos niveles

Hay un solo contraste lineal, que es un múltiplo de

$$A_{1L} = (T_1^{A_1} - T_0^{A_1}) \quad (108)$$

Esto representa el cambio en la respuesta del nivel más bajo al más alto del atributo A_1 . Este cambio puede ser positivo, negativo o cero.

Cuando los niveles del atributo A_1 están ordenados en escalas de intervalo o de relación, el contraste lineal A_{1L} puede interpretarse como una pendiente.

3.18.2 Atributos a tres niveles

Hay tres pares de contrastes entre los tres niveles. La suma de estos es

$$\begin{aligned} A_{1L} &= \{(T_1^{A_1} - T_0^{A_1}) + (T_2^{A_1} - T_0^{A_1}) + (T_2^{A_1} - T_1^{A_1})\} \\ &= 2(1T_2^{A_1} + 0T_1^{A_1} - 1T_0^{A_1}) \quad (109) \end{aligned}$$

Hay un solo triple de niveles, y la diferencia en el par de contrastes es

$$A_{1Q} = \{(T_2^{A_1} - T_1^{A_1}) - (T_1^{A_1} - T_0^{A_1})\} = (1T_2^{A_1} - 2T_1^{A_1} + 1T_0^{A_1}) \quad (110)$$

3.18.3 Atributos a cuatro niveles

Hay seis pares de los cuatro niveles, y cada par proporciona un parámetro de pendiente.

Combinándolos obtenemos el contorno lineal de A_1 denotado por A_{1L} .

$$\begin{aligned} A_{1L} &= \{(T_1^{A_1} - T_0^{A_1}) + (T_2^{A_1} - T_0^{A_1}) + (T_3^{A_1} - T_0^{A_1}) + (T_2^{A_1} - T_1^{A_1}) + (T_3^{A_1} - T_1^{A_1}) \\ &\quad + (T_3^{A_1} - T_2^{A_1})\} \\ &= \{-3T_0^{A_1} - T_1^{A_1} + T_2^{A_1} + 3T_3^{A_1}\} \quad (111) \end{aligned}$$

Para $s_1 = 4$, hay cuatro triples de niveles. La diferencia en las pendientes entre los tres niveles determina el efecto cuadrático. Combinamos los cuatro efectos cuadráticos para obtener el efecto cuadrático de A_1 denotado por A_{1Q} .

$$\begin{aligned} A_{1Q} &= \{(T_2^{A_1} - T_1^{A_1}) - (T_1^{A_1} - T_0^{A_1})\} + \{(T_3^{A_1} - T_1^{A_1}) - (T_1^{A_1} - T_0^{A_1})\} \\ &\quad + \{(T_3^{A_1} - T_2^{A_1}) - (T_2^{A_1} - T_0^{A_1})\} + \{(T_3^{A_1} - T_2^{A_1}) - (T_2^{A_1} - T_1^{A_1})\} \\ &= 3\{T_0^{A_1} - T_1^{A_1} - T_2^{A_1} + T_3^{A_1}\} \quad (112) \end{aligned}$$

Cuatro puntos determinan el efecto cúbico. Las diferencias entre los efectos cuadráticos son el efecto cúbico. Claramente,

$$\begin{aligned} A_{1C} &= \{T_1^{A_1} - 2T_2^{A_1} + T_3^{A_1}\} - \{T_0^{A_1} - 2T_1^{A_1} + T_2^{A_1}\} \\ &= \{-T_0^{A_1} - 3T_1^{A_1} - 3T_2^{A_1} + T_3^{A_1}\} \quad (113) \end{aligned}$$

3.18.4 Atributos en cinco niveles

Con cinco niveles de factores, hay 10 pendientes. A mediación de ellos, obtenemos

$$\begin{aligned} A_{1L} &= \{(T_1^{A_1} - T_0^{A_1}) + (T_2^{A_1} - T_0^{A_1}) + \dots + (T_4^{A_1} - T_3^{A_1})\} \\ &= \{-2T_0^{A_1} - T_1^{A_1} + 0T_2^{A_1} + T_3^{A_1} + 2T_4^{A_1}\} \quad (114) \end{aligned}$$

Hay 10 diferencias de pendientes, y obtenemos el término cuadrático como

$$\begin{aligned} A_{1Q} &= \{(T_0^{A_1} - 2T_1^{A_1} + T_2^{A_1}) + (T_0^{A_1} - 2T_1^{A_1} + T_3^{A_1}) + \dots + (T_2^{A_1} - 2T_3^{A_1} + T_4^{A_1})\} \\ &= \{2T_0^{A_1} - T_1^{A_1} - 2T_2^{A_1} - T_3^{A_1} + 2T_4^{A_1}\} \quad (115) \end{aligned}$$

Sólo hay un nivel en el que obtenemos la diferencia de cuadráticas, y el término cúbico es

$$\begin{aligned} A_{1C} &= \{T_4^{A_1} - 2T_3^{A_1} + T_2^{A_1}\} - \{T_2^{A_1} - 2T_1^{A_1} + T_0^{A_1}\} \\ &= \{-T_0^{A_1} + 2T_1^{A_1} + 0T_2^{A_1} - 2T_3^{A_1} + T_4^{A_1}\} \quad (116) \end{aligned}$$

Los efectos lineales, cuadráticos y cúbicos son múltiplos pf A, A y A. Aunque hemos discutido los resultados para el factor (atributo) A, los resultados son válidos para cualquiera de los n factores (atributos) en el experimento.

3.19 Componentes de interacción: Lineal y Cuadrática

Una combinación lineal $\sum_{i,j} d_{ij} T_{ij}^{A_1 A_2}$ se llama contraste de los efectos de A_1 y A_2 si $\sum_{i,j} d_{ij} = 0$. Hay contrastes independientes $s_1 s_2 - 1$ de los efectos de A_1 y A_2 , de los cuales $s_1 - 1$ contrastes independientes corresponden a los efectos de A_1 y $s_2 - 1$ contrastes independientes corresponden a los efectos de A_2 . Los contrastes de los efectos de A_1 y A_2 que son ortogonales a los efectos de A_1 y A_2 son los contrastes de

efecto de interacción y el número de efectos de interacción independientes de A_1 y A_2 es

$$\{(s_1 s_2 - 1) - (s_1 - 1) - (s_2 - 1)\} = (s_1 - 1)(s_2 - 1) \quad (117)$$

Considere un experimento con intervalos igualmente espaciados $s_1=3$ y $s_2=4$. El efecto lineal de A_1 es un múltiplo de $(-T_0^{A_1} + T_2^{A_1})$, y el efecto lineal de A_2 es un múltiplo de $(-3T_0^{A_2} - T_1^{A_2} + T_2^{A_2} + 3T_3^{A_2})$. El efecto lineal por interacción lineal (lineal x lineal) se obtiene hasta un término constante multiplicando los términos de la siguiente manera

$$\begin{aligned} A_{1L}A_{2L} &= (-T_0^{A_1} + T_2^{A_1})(-3T_0^{A_2} + T_1^{A_2} + T_2^{A_2} + 3T_3^{A_2}) \\ &= 3T_{00}^{A_1A_2} + T_{01}^{A_1A_2} - T_{02}^{A_1A_2} - 3T_{03}^{A_1A_2} - 3T_{20}^{A_1A_2} - T_{21}^{A_1A_2} + T_{22}^{A_1A_2} + 3T_{23}^{A_1A_2} \quad (118) \end{aligned}$$

A continuación, consideramos el efecto lineal por interacción cuadrática. El efecto lineal de A_1 es un múltiplo de $(-T_0^{A_1} + T_2^{A_1})$, y el efecto cuadrático de A_2 es un múltiplo de $(T_0^{A_2} - T_1^{A_2} - T_2^{A_2} + T_3^{A_2})$. El efecto lineal por cuadrático (lineal x cuadrático) se obtiene hasta un término constante multiplicando los términos de la siguiente manera

$$\begin{aligned} A_{1L}A_{2Q} &= (-T_0^{A_1} + T_2^{A_1})(-T_0^{A_2} - T_1^{A_2} - T_2^{A_2} + T_3^{A_2}) \\ &= -T_{00}^{A_1A_2} + T_{01}^{A_1A_2} + T_{02}^{A_1A_2} - T_{03}^{A_1A_2} + T_{20}^{A_1A_2} - T_{21}^{A_1A_2} - T_{22}^{A_1A_2} + T_{23}^{A_1A_2} \\ &= \{T_{20}^{A_1A_2} - T_{21}^{A_1A_2} - T_{22}^{A_1A_2} + T_{23}^{A_1A_2}\} - \{T_{00}^{A_1A_2} - T_{01}^{A_1A_2} - T_{02}^{A_1A_2} + T_{03}^{A_1A_2}\} \quad (119) \end{aligned}$$

Es decir, A_1A_2 es la diferencia entre el contraste cuadrático de A_1 en el nivel más alto de A_2 y el contraste cuadrático de A_1 en el nivel más bajo de A_2 . De manera similar, tenemos el efecto de interacción cuadrático por interacción lineal de A_1 y A_2 . El efecto cuadrático de A_1 es un múltiplo de t , y el efecto lineal de A_2 es un múltiplo de -3 . El efecto cuadrático por interacción lineal (cuadrático x lineal) se obtiene hasta un término constante multiplicando los términos de la siguiente manera

$$\begin{aligned} A_{1Q}A_{2L} &= (T_0^{A_1} + 2T_1^{A_1} + T_2^{A_1})(-3T_0^{A_2} - T_1^{A_2} - T_2^{A_2} + 3T_3^{A_2}) \\ &= -3T_{00}^{A_1A_2} - T_{01}^{A_1A_2} + T_{02}^{A_1A_2} - 3T_{03}^{A_1A_2} + 6T_{10}^{A_1A_2} + 2T_{11}^{A_1A_2} - 2T_{12}^{A_1A_2} - 6T_{13}^{A_1A_2} - 3T_{20}^{A_1A_2} - \\ &\quad T_{21}^{A_1A_2} + T_{22}^{A_1A_2} + 3T_{23}^{A_1A_2} \quad (120) \end{aligned}$$

Efecto cuadrático de A es múltiplo de T, y el efecto cuadrático de A es múltiplo de T. El efecto cuadrático de cuadratura (cuadrático x cuadrático) Se obtiene hasta un término constante multiplicando los términos de la siguiente manera

$$\begin{aligned}
 A_{1Q}A_{2Q} &= (T_0^{A_1} - 2T_1^{A_1} + T_2^{A_1})(T_0^{A_2} - T_1^{A_2} - T_2^{A_2} + T_3^{A_2}) \\
 &= T_{00}^{A_1A_2} - T_{01}^{A_1A_2} - T_{02}^{A_1A_2} + T_{03}^{A_1A_2} + 2T_{10}^{A_1A_2} + 2T_{11}^{A_1A_2} + 2T_{12}^{A_1A_2} + 2T_{13}^{A_1A_2} + T_{20}^{A_1A_2} - \\
 &\quad T_{21}^{A_1A_2} + T_{22}^{A_1A_2} + T_{23}^{A_1A_2} \quad (121)
 \end{aligned}$$

3.20 Diseños óptimos de Pareto

Cuando se ordenan los atributos de perfil, es posible que los perfiles sean mejores o peores que otros perfiles en todos los atributos; Es decir, un perfil dominará otros perfiles en un conjunto de opciones si ofrece más de todos los beneficios e incurre en menos de todos los costos. Por ejemplo, considere la opción de tren de tránsito de masas descritos en términos de cuatro atributos: tarifas de tren, frecuencia de trenes, rutas de trenes y seguridad en trenes. Consideramos sólo dos niveles: las tarifas actuales (0) y una reducción del 10% en las tarifas (1), la frecuencia actual del servicio (0) o la frecuencia más alta (1), las rutas actuales del tren (0) o un aumento del número de rutas (1), y nivel actual de seguridad (0) o seguridad mejorada (1).

Si presentamos los cuatro conjuntos de selección de perfiles 1111, 0111, 1011, 1101, el primer perfil está dominado y no se seleccionará. Estos ejemplos tienen en común la grasa que los encuestados no necesitan hacer concesiones a través de atributos al hacer sus elecciones. Si se considera que los parámetros estimados proporcionan información sobre la voluntad de hacer concesiones, se deben construir conjuntos de opciones para evitar las alternativas dominadas o dominantes dentro de los conjuntos. Tales conjuntos se llaman conjuntos ópticos de Pareto.

Un estudio, usando los atributos de oportunidad de empleo, ilustra las consecuencias prácticas de tener alternativas no Pareto óptimas en conjuntos de elección. Dos grupos de 55 estudiantes recibieron seis series de cuatro alternativas. Las alternativas fueron descritas en términos de cuatro atributos: salario, tiempo de vacaciones, capacitación y un índice independiente de los mejores lugares para trabajar. Los atributos fueron ordenados con números grandes indicando el nivel más deseable de atributo. El primer grupo de estudiantes recibió un conjunto mixto, consistente en dos conjuntos que eran

Pareto óptimos, dos conjuntos que tenían una alternativa dominante y dos conjuntos que tenían una alternativa dominada. El conjunto dado al segundo grupo de estudiantes fue todo óptimo para el Pareto, con los dos primeros conjuntos iguales a los conjuntos óptimos del Pareto dados al grupo 1.

Los resultados son las proporciones de los encuestados que eligen alternativas en los respectivos conjuntos de selección. Los conjuntos 1 y 2 son los mismos en ambos grupos, y las proporciones observadas son las mismas dentro del error de muestreo. Para el grupo 1, la tercera alternativa domina los tres restantes en los grupos 3 y 4. El 96% de los encuestados eligieron la alternativa dominante en el grupo 3 y el 95% eligieron el grupo dominante 4. Para el grupo 1, la tercera alternativa estuvo dominada por Cada uno de los tres restantes en el conjunto 5 y 6. Nadie escogió la alternativa dominada en ninguno de los dos conjuntos.

Observe que las alternativas dominantes en el grupo 1 aparecieron en los grupos 6 y 5 del grupo 2, respectivamente, donde recibieron sustancialmente menos del 50% de las acciones. Las alternativas dominadas en el grupo 1 aparecieron en el grupo 2 Pareto óptimo conjuntos 4 y 3, respectivamente.

El grupo 1 conjunto 5 de alternativa dominada es la alternativa más preferida en el grupo 2, conjunto 4, un conjunto óptimo de Pareto. Los resultados respaldan el argumento de que los conjuntos de opciones óptimas de Pareto deben utilizarse en estudios de experimentación de elección discreta con atributos de beneficio y costo. La pregunta es, ¿cuántos subconjuntos opuestos de Pareto se pueden construir a partir de alternativas de multi atributo con atributos ordenados?

Al igual que con el material anterior, considere un experimento de elección en m atributos, el i -ésimo atributo en s niveles denotados por $0, 1, \dots, s_i-1$ para $i = 1, 2, \dots, m$. Suponga que los niveles se ordenan de acuerdo con mayores beneficios y costos más bajos. Sea (x_1, x_2, \dots, x_m) un perfil donde x_i denota el nivel del i -ésimo atributo. Un conjunto de opciones es un subconjunto de T perfiles distintos de S , donde S es el conjunto de todos los perfiles en un diseño experimental. Note que la suma de los niveles en los perfiles respectivos va de 0 a $\sum (s_i - 1)$.

Formalmente, se dice que un subconjunto T de S es un subconjunto óptimo de Pareto si para cada dos perfiles distintos $(x_1, x_2, \dots, x_m), (y_1, y_2, \dots, y_m) \in T$ existen subscripciones i y j ($i \neq j$) tales que $x_i < y_i$ y $x_j > y_j$. Es decir, ningún perfil domina otro en los conjuntos óptimos de Pareto. Tenga en cuenta que los conjuntos de selección

$$S_l = \{(x_1, x_2, \dots, x_m) \mid \sum_{i=1}^m x_i = l\}, l = 0, 1, \dots, \sum (s_i - 1) \quad (122)$$

Son Pareto óptimos. Para el ejemplo de transporte anterior, los tres subconjuntos siguientes son Pareto óptimos con más de un perfil: $S_1 = \{1000, 0100, 0010, 0001\}$, $S_2 = \{1100, 1010, 1001, 0110, 0101, 0011\}$ y $S_3 = \{1110, 1101, 1011, 0111\}$.

Varios autores han abordado el problema de construir subconjuntos óptimos de Pareto. Wiley (1978) proporciona procedimientos para construir todos los conjuntos óptimos de Pareto (y el conjunto óptimo de Pareto más grande en particular) para m diseños, donde m es el número de atributos.

También muestra que, para esta categoría de diseño, los subconjuntos óptimos de Pareto pueden organizarse en subconjuntos equilibrados, es decir, subconjuntos en los que todos los niveles de atributo aparecen en el conjunto de opciones. Krieger y Green (1991) amplían ese trabajo. También derivan el número esperado de pares de entrada dominante en un conjunto de selección seleccionado al azar en el que todos los atributos tienen el mismo número de niveles.

Un tratamiento formal de los diseños óptimos de Pareto es proporcionado por Raghavarao y Wiley (1998). Supongamos que en un experimento de experimentación de elección discreta hay atributos m ($m \geq 2$), A_1, \dots, A_m . El atributo A_i tiene s_i niveles, designados por $0, 1, \dots, (s_i - 1)$; $i = 1, \dots, m$. Raghavarao y Wiley muestran que el conjunto $S_\alpha = \{(x_1, x_2, \dots, x_m) \mid \sum_{i=1}^m x_i = \alpha\}$ es un subconjunto óptimo de Pareto para cualquier número de atributos y sus niveles. Tenga en cuenta que este resultado puede usarse para poner perfiles en conjuntos óptimos de Pareto sin importar cómo se construyeron los perfiles. Por ejemplo, los perfiles de un diseño factorial fraccional pueden organizarse en subconjuntos usando este enfoque. El artículo proporciona tres resultados clave: $k = \sum_{i=1}^m s_i - m$,

El diseño D de los perfiles de atributos resultantes de un solo subconjunto óptimo de Pareto S_i no es un plan de efectos principal conectado. Para un experimento s^m , el diseño D basado en Pareto óptimo subconjuntos $S_{k/2}$ y $S_{(k/2)+1}$ si k es par, y $S_{(k+1)/2}$ y $S_{(k+3)/2}$ si k es impar es un plan de efectos principal conectado. En un experimento $S_1 x S_2 x \dots x S_m$, ($s_1 \leq s_2 \leq \dots \leq s_m$) el plan de efectos principal D basado en S_i , para $l = s_1 - 1, s_1, s_1 + 1, \dots, s_m - 1$, está conectado.

Un resultado importante implícito en la definición de los subconjuntos opuestos de Pareto es que los subconjuntos de los subconjuntos óptimos de Pareto son ellos mismos Pareto óptimo. En otras palabras, una vez que se tiene un subconjunto óptimo de Pareto, sus elementos pueden dividirse en subconjuntos de acuerdo con cualquier regla o algoritmo, incluso arbitrario, y los subconjuntos serán óptimos de Pareto.

3.21 Inferencias sobre efectos principales

Sea $y_{x_1 x_2 \dots x_m}$ la respuesta correspondiente al perfil (x_1, x_2, \dots, x_m) . Esta respuesta podría ser el número de puntos asignados al perfil (x_1, x_2, \dots, x_m) en el conjunto de elecciones S_i o el logaritmo de la proporción de la proporción de encuestados que eligen (x_1, x_2, \dots, x_m) a la proporción de encuestados que eligen la opción de no elección de S_i . Asumimos El modelo lineal:

$$E(y_{x_1 x_2 \dots x_m}) = u + \sum_{i=1}^m \alpha_{x_i}^{A_i} + \sum_{i,j=1}^m \alpha_{x_i x_j}^{A_i A_j} + \sum_{i,j,k=1}^m \alpha_{x_i x_j x_k}^{A_i A_j A_k} + \dots + \alpha_{x_1 x_2 \dots x_m}^{A_1 A_2 \dots A_m} \quad (123)$$

Donde u es la media general, $\alpha_{x_i}^{A_i}$ es el efecto del atributo A_i en el nivel x_i , $\alpha_{x_i x_j}^{A_i A_j}$ es el efecto de los atributos A_i y A_j en los niveles x_i y x_j , respectivamente, y así sucesivamente. Las variaciones y covarianzas de las respuestas dependen de la naturaleza de las respuestas.

3.22 Teorema 4.1

Podemos estimar todos los $m(s - 1)$ contrastes de los efectos principales usando dos conjuntos de elección consecutivos S_l y S_{l+1} donde $s - 2 \leq l \leq (m - 1)(s - 1)$.

El número de perfiles en los conjuntos de selección S_l y S_{l+1} puede ser grande; Por lo tanto, los encuestados pueden no ser capaces de elegir entre los muchos perfiles

disponibles o proporcionar una escala de calificación apropiada. En ese caso, el experimentador puede formar subconjuntos no superpuestos de S_i y S_{i+1} y recopilar datos. La elección del diseño es más complicada cuando los niveles de todos los atributos no son los mismos.

3.23 Inferencias sobre los efectos principales en experimentos de 2^m

Para un experimento de 2^m , podemos obtener resultados interesantes si asumimos que las respuestas no están correlacionadas y tienen la misma varianza común. Considere dos conjuntos opcionales de Pareto óptimos S_i y S_m con un número total de perfiles n . Nombra los perfiles 1, 2, ..., n y deja que y_1, y_2, \dots, y_n sean las respuestas medias o la respuesta transformada.

$$E(y_i) = \mu + \sum_{j=1}^m x_{ij}\beta_j, \quad (124)$$

Donde μ es la media general, β_j , es el efecto principal del atributo A_j , y X_{ij} es -1 (o 1) dependiendo del nivel bajo (o alto) del atributo A_j en el perfil i -ésimo. Asumimos además que los y_i 's no están correlacionados y la varianza de Y_i es σ^2 para cada i . Después de escribir la matriz de diseño y algún álgebra directa, se puede demostrar que la matriz de información de m efectos principales después de eliminar μ es una matriz $m \times m$

$$C = (a_0 - a_2)I_m + \left(a_2 - \frac{a_1^2}{a_0}\right)J_m \quad (125)$$

Donde I_m es una matriz de identidad $m \times m$, J_m es una matriz $m \times m$ de todos los, $a_0 = n$, y $a_1 = \sum_{i=1}^n x_{i\alpha}$, $a_2 = \sum_{i=1}^n x_{i\alpha}x_{i\beta}$ para cualquier $\alpha, \beta = 1, 2, \dots, m$;

3.24 Teorema 4.2

Para un experimento de 2^m , el diseño basado en cualquier dos subconjuntos de Pareto óptimo S_l y S_k es un plan de efectos principal conectado en el que $l \neq k$, $0 < l, k < m$.

Mientras que el teorema 4.1 establece que S_l, S_{l+1} forman un plan de efectos principal conectado para un experimento S^m , el Teorema 4.2 muestra que cualquier dos subconjuntos óptimos de Pareto distintos S_l, S_k forman un plan de efectos principal conectado para un experimento de 2^m .

Dado que tenemos varias opciones para los planes de efectos principales conectados, nos gustaría encontrar el plan que es óptimo, en cierto sentido, para un experimento de 2^m . Tenga en cuenta que todos estos paneles de efectos principales conectados utilizan diferentes números de perfiles. Por tanto, definimos información por perfil φ como el recíproco de la varianza media de los efectos principales estimados dividido por el número de perfiles utilizados en el experimento. Entonces φ es dado por

$$\varphi = \frac{m}{a_0 \text{trace}(C^{-1})} \quad (126)$$

3.25 Polinomios ortogonales

Sea $T_0^A, T_1^A, \dots, T_{s-1}^A$ las respuestas totales de un factor A en los niveles 0, 1, 2, ..., s-1. Sea $T' = (T_0^A, T_1^A, \dots, T_{s-1}^A)$. Podemos reparameterizar según sea necesario para que

$$E(T) = X\alpha, \text{ and } Var(T) = \sigma^2\epsilon \quad (127)$$

Donde X es una matriz s x (s-1) de rango s-1, y α es un vector columna de orden (s-1) de parámetros. Las ecuaciones normales ponderadas que estiman α por $\hat{\alpha}$ son

$$(X'E^{-1}X)\hat{\alpha} = X'E^{-1}T, \text{ y } Var(\hat{\alpha}) = (X'E^{-1}X)^{-1}\sigma^2. \quad (128)$$

Queremos obtener estimaciones no correlacionadas para parámetros transformados adecuadamente como lo hicimos con niveles igualmente espaciados. Sea $X'E^{-1}X = GG'$ la factorización de Cholesky, donde G es triangular inferior o se basa en valores propios y vector de $X'E^{-1}X$. Ponga $y = G'\alpha$ y $X' = X(G')^{-1}$. Entonces $E(t) = X^*y$ y las ecuaciones normales ponderadas son

$$(X'E^{-1}X^*)\hat{y} = X^*E^{-1}T \quad (129)$$

Y

$$Var(\hat{y}) = (X'E^{-1}X^*)^{-1}\sigma^2 \quad (130)$$

Tenga en cuenta que

$$X'E^{-1}X^* = G^{-1}X'E^{-1}X(G')^{-1} = G^{-1}(GG')(G')^{-1} = I_{s-1} \quad (131)$$

Por lo tanto, $\hat{y} = G^{-1}X'E^{-1}T$, y $Var(\hat{y}) = I_{s-1}\sigma^2$. Tales estimaciones ortogonales pueden obtenerse con un espaciado desigual de niveles y una replicación desigual de perfiles.

3.26 Tasa de sustitución de los atributos

Uno puede desear determinar cómo los clientes están dispuestos a sustituir un aumento incremental en un atributo para un aumento incremental en otro atributo. Por ejemplo, nos gustaría saber cómo un nuevo graduado está dispuesto a sustituir un aumento en el salario por un aumento en el número de días de vacaciones. Considere 2^3 elección discreta experimental en tres atributos en dos niveles cada uno. Los atributos y niveles son salario (0 = \$ 50,000, 1 = \$ 55,000); Tiempo de vacaciones (0 = 2 semanas, 1 = 3 semanas); y la contribución de los empleadores a la pensión (0 = 5% del salario, 1 = 7% del salario). Se forman dos conjuntos de opciones ($i = 1, 2$), y se recopilan datos sobre la elección de cuatro perfiles ($j = 1, 2, 3$ y sin opción) utilizando 100 nuevos graduados en cada uno de los dos conjuntos de selección. La matriz de dispersión estimada es

$$S = \frac{1}{100} \begin{bmatrix} S_t & o \\ o & S_2 \end{bmatrix} \quad (132)$$

Donde

$$S_1 = \begin{bmatrix} \frac{1}{.15} + \frac{1}{.05} & \frac{1}{.05} & \frac{1}{.05} \\ \frac{1}{.05} & \frac{1}{.30} + \frac{1}{.05} & \frac{1}{.05} \\ \frac{1}{.05} & \frac{1}{.05} & \frac{1}{.50} + \frac{1}{.05} \end{bmatrix} \quad (133)$$

$$S_2 = \begin{bmatrix} \frac{1}{.45} + \frac{1}{.10} & \frac{1}{.10} & \frac{1}{.10} \\ \frac{1}{.10} & \frac{1}{.40} + \frac{1}{.10} & \frac{1}{.10} \\ \frac{1}{.10} & \frac{1}{.10} & \frac{1}{.05} + \frac{1}{.10} \end{bmatrix} \quad (134)$$

Y O es una matriz 3 x 3 de ceros. Se usan los mínimos cuadrados ponderados para encontrar las estimaciones \hat{s} y \hat{v} de los parámetros desconocidos s y v como sigue.

$$\begin{bmatrix} \hat{\mu} \\ \hat{s} \\ \hat{v} \\ \hat{p} \end{bmatrix} = (X'S^{-1}X)^{-1}X'S^{-1}L \quad (135)$$

Nuestro objetivo es encontrar la tasa de sustitución c entre un aumento en el sueldo s y un aumento en el número de días de vacaciones v , es decir, determinar c para que $s = cv$. Si tanto s como v son distintos de cero, c puede estimarse como s/v , y podemos encontrar un intervalo de confianza para c usando el teorema de Fiellers. Una breve descripción del teorema de Fieller se da a continuación.

Según el teorema de Fiellers, si se piensa que las variables aleatorias a y b se distribuyen como conjuntamente normales, entonces para cualquier valor fijo r se cumple la siguiente declaración de probabilidad si z es $\alpha/2$ cuantil de la distribución normal estándar y V es la varianza Matriz de covarianza de a y b

$$\Pr((a - rb)^2 > z^2(V_{aa} - 2rV_{ab} + r^2V_{bb})) = \alpha \quad (136)$$

Por lo general, la desigualdad puede resolverse para r para producir un intervalo de confianza. Para la tasa de sustitución c , la hipótesis nula $c = c_0$ se conserva contra la alternativa bilateral

$$(\hat{s} - c_0\hat{v})^2 \leq x_{1-\alpha}^2 \{Var(\hat{s}) + c_0^2Var(\hat{v}) - 2c_0Cov(\hat{s}, \hat{v})\} \quad (137)$$

3.27 Referencias bibliográficas

- Montgomery, D. C. (2015). *Diseño y análisis de experimentos* (Segunda ed.). México: Limusa.
- Pérez López, C. (2004). *Técnicas de análisis multivariante de datos. Aplicaciones con SPSS*. Madrid, Universidad Complutense de Madrid.: Pearson Educación.
- Smith. (1960). Interpretations of adjusted treatment means and regressions in analysis of covariance. *Biometrics*, 13(3), 282-308.

CAPITULO 4: CASOS

4.1 Caso 1: Evaluación de los métodos aplicados en el análisis conjunto

4.1.1 Antecedentes

El análisis se basa en un estudio reciente de preferencias para las organizaciones de mantenimiento de la salud de una muestra de 249 jefes de hogar en una gran área metropolitana del noreste. Las OMS se describieron mediante un conjunto de atributos seleccionados mediante una revisión de la literatura, entrevistas con administradores de atención de salud y discusiones informales con posibles usuarios de organizaciones de mantenimiento de la salud durante la prueba previa de los procedimientos de entrevista. Los datos de análisis conjuntos para el estudio se recogieron en tres partes. En la primera parte (Parte A), se presentó a los encuestados 27 perfiles que describen las organizaciones de mantenimiento de la salud en términos de atributo de conveniencia.

En la segunda parte (Parte B), se mostró a los encuestados 16 perfiles de organizaciones de mantenimiento de la salud desarrollados sobre la base del alcance de la cobertura del seguro de salud. En la tercera parte (Parte C), se les mostraron 16 perfiles que describieron las organizaciones de mantenimiento de la salud en general, es decir, conveniencia, cobertura, costo, tipo de establecimiento y método de selección de médicos. En cada parte de la entrevista, se pidió a los encuestados que ordenaran los perfiles en tres pilas según su preferencia.

Este paso preliminar fue tomado para reducir la dificultad de producir el orden de prioridad estricto final. A continuación, se pidió a los encuestados que clasificaran los perfiles dentro de cada pila en orden de preferencia. Los perfiles particulares utilizados en las tres partes se generaron de acuerdo con planes de efectos principales ortogonales.

4.1.2 Diseño o método aplicado

Los datos para el estudio se analizaron mediante el algoritmo aditivo de análisis de Kruskal, MON-ANOVA (Kruskal 1965). Los rangos de preferencia de los encuestados individuales se sometieron al algoritmo y se obtuvieron estimaciones de los parámetros. Para cada encuestado, y para cada segmento del cuestionario, se seleccionó aleatoriamente un perfil y se retiró del análisis MONANOVA. Este perfil se utilizó para un tipo de prueba de validez predictiva discutida más adelante.

Se utilizaron tres métodos de evaluación. Una era una estimación interna de la maldad de ajuste del modelo aditivo a los datos clasificados, el estrés de Kruskal (Kruskal 1965). El índice mide el grado de salida de los valores ajustados del modelo aditivo de los valores que preservan los rankings de estímulos originales y, por lo tanto, es similar a la suma residual de cuadrados en la regresión. Aunque se han ofrecido reglas básicas para evaluar el estrés (Green y Wind 1973), no existen límites absolutos para distinguir los buenos de los malos resultados.

El enfoque del estudio fue comparar los valores de estrés observados con los que se obtendrían bajo condiciones de datos aleatorios. En consecuencia, se utilizó un generador de números aleatorios para producir datos de entrada para el algoritmo MONANOVA. Se realizaron tres simulaciones utilizando los diseños ortogonales de las Partes A, B y C de la entrevista, respectivamente. Para cada simulación, se generaron 100 conjuntos de datos de entrada y se realizaron 100 series de MONANOVA para determinar el estrés en condiciones de datos aleatorios. Este enfoque es similar al adoptado por Klahr (1964), Spence y Ogilvie (1973) y Stenson y Knoll (1969) para investigar el estrés en soluciones de escalamiento multidimensionales.

El segundo método para evaluar los resultados, de Green y Wind (1973), Parker y Srinivasan (1976), y Scott y Wright (1976), fue simplemente examinar las funciones de utilidad parciales para determinar si se conformaban en signo con expectativas a priori. Por ejemplo, todos los demás son iguales, uno esperaría que los encuestados prefieran un plan de salud en el que la espera para las citas es de cinco días a uno en el que la espera es de 15 días, se postularon ocho expectativas de signo para la Parte A, seis para la Parte B y cinco para la Parte C.

En la Parte C no se podían hacer expectativas a priori sobre la elección de los médicos o el tipo de práctica, Porque un encuestado podría preferir sensiblemente cualquiera de los dos niveles de estos atributos). La medida de evaluación fue simplemente el número de violaciones de las expectativas de signo a priori.

El tercer método de evaluación involucró la predictibilidad del rango de los perfiles de retención. Este método se basa en una extensión de un procedimiento propuesto por Green (1974a). Debido a que un perfil seleccionado aleatoriamente fue retenido del análisis MONANOVA para cada encuestado, el rango de este perfil podría ser predicho

usando las utilidades parciales obtenidas a través de MONANOVA, y este rango predicho podría compararse con el rango real del perfil en el original del encuestado datos.

Para proporcionar un estándar con el que comparar las predicciones que posee, se derivó una distribución para el caso de no predictibilidad. Este modelo no predecible o "nulo" puede ser construido de la siguiente manera. Suponga que se selecciona y retiene aleatoriamente un perfil. Bajo la condición nula, el rango predicho de este perfil es igualmente probable que sea cualquier número de 1 a k, donde k es el número de perfiles. Hay k^2 posibles combinaciones de rangos reales y predichos bajo el modelo nulo. Hay k maneras de obtener "hits directos", 2(k - 1) maneras de obtener una diferencia absoluta de 1 entre predicha y real, 2(k - 2) maneras de obtener una diferencia de 2, y en general 2(k - x) maneras de obtener una diferencia absoluta de x, Donde $2 \leq x \leq k - 1$. La probabilidad de un "golpe directo" es, por tanto, k / k^2 , para una diferencia absoluta de 1 es $2(k - 1) / k^2$, etc. De esta distribución, las expresiones de la media y la varianza de la diferencia absoluta entre real y Los rangos predichos, x, para cada "individuo" para el caso nulo se pueden derivar tomando los valores esperados:

$$\bar{x}_n = E(x_n) = \frac{2}{k^2} \sum_{i=1}^k i(k - i) = \frac{(k+1)(k-1)}{3k} \quad (138)$$

$$var(x_n) = \frac{2}{k^2} \sum_{i=1}^k i^2(k - i) - E(x)^2 = \frac{(k^2-1)(k^2+2)}{18k^2} \quad (139)$$

Donde N se refiere al modelo nulo. Sea \bar{x}_n y $var(x_n)$ la media observada y la varianza para los resultados conjuntos. Se requiere una prueba estadística para la diferencia entre dos medios. Debido a que n, el tamaño de la muestra, es bastante grande, el medio muestral para las distribuciones observadas y nulas será aproximadamente normal por el Teorema del Límite Central. La distribución de muestreo para X_n tendrá una media de \bar{x}_n y varianza de $var(x_n) / n$ para una muestra de tamaño n, mientras que la distribución de muestreo de X_0 tendrá una media de \bar{x}_0 y varianza de $var(x_0) / n$. Debido a que X_n y X_0 son aproximadamente distribuidos normalmente, así es su diferencia, $X_n - X_0$. La estadística de prueba, Z, es entonces aproximadamente unidad normal, y está dada por

$$Z = \frac{\bar{X}_N - \bar{X}_0}{\sqrt{var(\bar{x}_n) + var(\bar{x}_0)}} = \frac{\bar{x}_n - \bar{x}_0}{\sqrt{\frac{var(x_n)}{n} + \frac{var(x_0)}{n}}} \quad (140)$$

Esta estadística de prueba es apropiada para muestras grandes donde no se puede hacer la suposición de varianzas iguales para las dos muestras. En este caso, se espera una mayor varianza para el modelo nulo. La hipótesis nula de que $\mu_N = \mu_0$ puede probarse contra la alternativa de una cola que $\mu_0 < \mu_n$ usando Z, donde, μ_0 y μ_n se refieren a medios de población para los modelos conjunto y nulo.

4.1.3 Resultados

Las correlaciones encerradas en triángulos son las que muestran las relaciones entre los tres métodos dentro de un segmento dado de la entrevista. Todas menos una de las correlaciones en los triángulos (que entre la predicción del rango del perfil de retención y el número de infracciones de signo para los datos de la Parte B) son significativas al nivel de .01. Sin embargo, la magnitud de las correlaciones no justifica la conclusión de que los diversos métodos de evaluación son sustitutos directos.

También se muestra el estrés promedio observado con la simulación de clasificación aleatoria, el 5% menor de la cola de la distribución de los valores de estrés para los datos aleatorios, y el número de casos observados en las Partes A, B y C con valores de tensión mayor que este 5 % criterio. Los resultados indican que muchos de los valores de tensión observados son algo altos, con un número sustancial de encuestados con un estrés superior al observado en el 5% del tiempo cuando se usan datos aleatorios. La razón de la alta tensión sólo puede ser conjeturada; Es debido a "ruido" en los datos, inadecuación del modelo aditivo, o ambos. Los valores de tensión, sin embargo, no están fuera de línea con el estrés observado en otros estudios empíricos.

Las correlaciones son todas significativas ($p < .05$), pero bajas; La correlación entre el estrés en la Parte A y el estrés en la Parte B es .11, que entre el estrés en A y el estrés en C es .11, y que entre la tensión en B y la tensión en C es .21. Estos resultados sugieren una ligera tendencia a que los encuestados que tienen alto valor de estrés en una tarea de medición conjunta también tengan alto valor de estrés en otra tarea.

Los resultados muestran la distribución de las violaciones de los signos de las tres secciones del estudio. Las correlaciones entre las violaciones de los signos observadas en las tres secciones se calcularon después de la transformación de los datos utilizando $y = \ln(1 + x)$, debido a la asimetría de las distribuciones (Winer 1971).

Sólo la correlación entre las violaciones de los signos en las Partes B y C es significativa ($r = .16$; $p < .01$).

La relación entre la precisión predictiva en una parte del estudio y la precisión predictiva en otra parte se investigó mediante el cálculo de las correlaciones entre los resultados de la Parte A, B y C datos. Debido a que la variable de precisión se inclinó hacia cero, se aplicó una transformación logarítmica antes del cálculo de las correlaciones.

4.1.4 Conclusiones

La conclusión de que varias de las medidas de evaluación se correlacionaron significativamente a través de secciones del estudio sugiere que las características individuales podrían explicar algunos de los resultados observados. El análisis conjunto requiere que los encuestados realicen una serie de juicios complejos en un corto período de tiempo y requieran una buena cantidad de lectura. Por lo tanto, el historial educativo de un entrevistado podría relacionarse con la "calidad" de los resultados conjuntos. Las correlaciones se calcularon entre la educación y las medidas de evaluación.

Se examinan varios métodos de evaluación de resultados de análisis conjuntos. Aunque las evaluaciones de cada uno de los métodos están correlacionadas, las relaciones no son tan fuertes que el investigador podría utilizar los métodos de forma intercambiable. En su lugar, los métodos de evaluación deben seleccionarse de acuerdo con los propósitos de la investigación en particular. En muchos casos se utilizarían más de una, o incluso las tres técnicas.

El estrés sería útil como medida de cribado para eliminar los datos que son "ruidosos" y / o inapropiados para el modelo particular que se está considerando. Para obtener valores de tensión de referencia bajo condiciones nulas, se pueden enviar datos aleatorios simulados al algoritmo de análisis conjunto (como MONANOVA) y comparaciones de valores de tensión observados con datos originales y aleatorios. Este método es probablemente apropiado para la mayoría de los estudios conjuntos como un dispositivo de evaluación inicial.

El criterio de conformidad con el signo esperado es una medida algo problemática. El investigador desea permitir que los signos sean determinados por los datos. En muchos casos, un signo "incorrecto" sólo puede indicar que un atributo es irrelevante para el encuestado y los valores de utilidad resultantes son, por lo tanto, "no significativos" en algún sentido. Un signo "incorrecto" también podría ser correcto desde el punto de vista del entrevistado, si él o ella tiene una estructura de preferencia diferente a la esperada por el investigador.

Al menos un algoritmo conjunto (Shocker y Srinivasan 1976) permite que las restricciones sean colocadas en los signos antes del análisis. Esto probablemente debería hacerse sólo si el investigador tiene muy fuertes nociones a priori sobre los signos, tal vez derivados de otra información proporcionada por el entrevistado. Las violaciones de signos pueden ser problemáticas en estudios en los que el investigador espera reescalonar los datos a través de varios instrumentos conjuntos porque un signo "incorrecto" podría distorsionar en gran medida los resultados. En esta situación, el criterio de signo sería especialmente útil.

El tercer método para evaluar los resultados conjuntos es la predicción del rango de un perfil retenido. Este método es probablemente el más adecuado para hacer inferencias generales sobre la calidad de un conjunto de datos como un todo, en lugar de para el cribado individual de los encuestados. Es posible que las filas predichas y reales de un perfil seleccionado al azar estén muy cerca, incluso si los datos del encuestado son pobres. La situación opuesta también es posible. Sin embargo, en toda la muestra, estas casualidades no deberían ser un problema.

En el estudio presentado aquí, también se encontró evidencia de una tendencia de los encuestados a proporcionar datos deficientes a través de los diversos instrumentos conjuntos. Además, se encontró que la mayoría de las medidas de evaluación se correlacionan significativamente con el nivel educativo del encuestado. La capacidad del entrevistado y otras características necesitan ser estudiadas más para entender mejor el proceso de información que está ocurriendo durante una entrevista de análisis conjunto. Esta comprensión podría conducir a una mejor capacidad para medir las preferencias de los consumidores.

4.2 Caso 2: Optimización conjunta, un algoritmo exacto de ramificación y límite para el problema de compartir de elección

4.2.1 Antecedentes

La importancia de la fase de diseño de producto de un proceso de desarrollo de producto completo (Hauser et al. 2004) se refleja en los numerosos esfuerzos para combinar el análisis en la búsqueda de un "diseño de producto óptimo" que aparezca tanto en la investigación de marketing como de operaciones, Literaturas. En este trabajo, nos centramos en la cuota de problema de elección para el diseño de un solo producto. Por diseño de un nuevo producto se entiende el proceso de selección de un nivel para cada atributo saliente que comprende el producto.

El objetivo es encontrar la combinación de niveles de atributo que maximizan el número de encuestados para quienes la utilidad del producto construido excede una utilidad o estado de reserva. Zufryden (1977) fue el primero en usar esta idea en un modelo formal de optimización. Definió el status quoas "... la utilidad de ninguna compra en absoluto, aquella asociada con la marca favorita actual del consumidor, o ... la utilidad para la marca que sería comprada por el consumidor, dado un escenario de mercado definido".

El análisis conjunto utiliza las reacciones de un encuestado a las descripciones de los productos alternativos para estimar su participación en cada nivel de atributo asociado con una categoría de producto determinada. En su artículo de la encuesta. Greenet al. (2001) señalan que se ha utilizado un análisis conjunto para diseñar una amplia variedad de productos, incluyendo teléfonos celulares, estaciones de trabajo informáticas, fármacos, planes de mantenimiento de la salud, servicios de transporte marítimo y televisión por cable.

Se puede argumentar que la aplicación más notable y con más éxito comercial del análisis conjunto es el diseño de Courtyard by Marriott, es decir, "El hotel diseñado por viajeros de negocios". EZ-Pass, el sistema de peaje electrónico que ahora se utiliza en la mayoría de los sistemas de carreteras estatales y urbanas, también fue diseñado usando análisis conjuntos y representa un éxito notable en el sector público.

Las partes para cada nivel de atributo obtenido para cada encuestado a través del análisis conjunto se utilizan como entrada para el problema de la cuota de elección. Un diseño de producto particular proporciona una utilidad global para cada encuestado, es decir, la suma de los porcentajes de los encuestados para los niveles que componen el diseño. Los encuestados tienen un orden único de preferencia y buscamos el diseño único que con éxito equilibra estas tensiones sobre todos los encuestados.

Por ejemplo, Toubia et al. (2003) ofrecen programas de desarrollo ejecutivo que serían "impulsados por la tecnología" o "impulsados por las empresas globales". Los usuarios que reaccionan al concepto EZ-Pass enfrentan un compromiso entre las etiquetas transferibles entre vehículos a un precio más alto que las etiquetas no transferibles. En el ensayo de Dahan et al. (2002) estudio de bolsas de computadora portátil, los encuestados consideran bolsas de menor precio versus bolsas de mayor precio con soporte para teléfono celular nuevo.

Debido a que los consumidores consideran que los atributos del producto son importantes y exhiben una utilidad distinta para los niveles de un solo atributo, el problema general de la selección del diseño es complicado. Hauser et al. (2004) consideran que el problema y las extensiones a la investigación de línea de productos son una prioridad en la ciencia de marketing.

Las literaturas de investigaciones de la operación y de marketing se han centrado en diferentes aspectos del problema de la cuota de elección, lo que a su vez ha llevado al desarrollo de una variedad de enfoques. La comparación directa es difícil debido a que las soluciones de marketing tienden a ser pragmáticas y se representan usando un estilo de "ingeniería" más bien que el "teorema-prueba-prueba" estilo dominante en investigaciones de la operación. Las investigaciones publicadas en el mercadeo se han centrado principalmente en la fase de estimación parcial del problema general, enfatizando la recopilación de datos válida y la estimación estadística mejorada; La optimización ha desempeñado un papel subordinado en esta literatura. Los primeros trabajos típicamente "resolvieron" el problema usando simuladores si-entonces para evaluar la cobertura de varios perfiles preseleccionados (Green et al., 1981, Green y Krieger 1988, 1989). Su abordaje fue suplantado por heurísticas formales (Greenand Krieger 1993, Krieger et al. 2004) que consideran todo el espacio factible de soluciones, pero no garantizan la óptima globalidad.

Actualmente, las mejores soluciones representan una mezcla de contribuciones de la industria {Sawtooth Software 2003} y de la academia {por ejemplo, Ormeand Huber 2000, Huber et al. 2001, Green et al. 2003}.

En general, los algoritmos de optimización conjunta se basan en el conocimiento de las funciones individuales de utilidad del consumidor, trabajo de Allenby et al. (1995), Lenk et al. (1996), y otros (Train 2003, Allenby y Rossi 1999) utilizando la cadena de Monte Carlo (MCMC) y métodos relacionados es particularmente relevante. Estos métodos utilizaron datos de elección concreta para producir una densidad posterior completa de parciales por parte del individuo encuestado. Los resultados no sólo apoyan la optimización conjunta, sino que también proporcionan una oportunidad para explorar la robustez de cualquier algoritmo propuesto a la incertidumbre en partes estimadas.

Recientemente Hauser et al. (2004), Dahan et al. {2002}, y Toubia et al. {2003} han introducido un nuevo procedimiento de estimación utilidad parciales basado en un algoritmo poliédrico rápido y adaptable. El método utiliza una estrategia de puntos interior para adaptar dinámicamente los diseños mostrados a un encuestado en una serie de escenarios de elección discreta. Aunque se utiliza la programación matemática para resolver un problema de optimización dinámica, hasta la fecha, el conjunto poliédrico se ha utilizado exclusivamente para la estimación de partes (p. Ej. Un problema dentro del sujeto) y no en la fase de diseño del producto (un problema entre sujetos).

Con la excepción de Zufryden (1977), los investigadores de operaciones entraron en el campo de la optimización conjunta mucho más tarde. La literatura de investigaciones de operación ofrece técnicas sofisticadas para soportar el problema, ha formalizado muchos resultados y ofrece comparaciones valiosas de diferentes algoritmos (véase Kohli y Krishnamurti, 1989, Kohli y Sukumar, 1990, Balakrishnan y Jacob, 1996, y Shi y otros, 2001). Sin embargo, el trabajo publicado todavía depende exclusivamente de la heurística; El algoritmo no publicado puede garantizar la optimalidad global. Además, sin excepción, la literatura investigaciones de operación existente considera las utilidades parciales estimadas como datos sin error.

Definimos variables binarias de decisión x_{kj} . $x_{kj} = 1$ si se elige el atributo j de nivel k y 0 en caso contrario. También define $y_s = 1$ si la utilidad de un producto diseñado excede la utilidad de del estatus quo para el s -ésimo entrevistado y 0 de otro modo. Denotamos la utilidad de obstáculos para el encuestado s por $u_{kj}^s = U_{kj} - e$, donde e es un pequeño número real positivo. En todo nuestro trabajo, U_{kj} es una función del producto q de estado del usuario. Curry (2004) proporciona una discusión detallada de una variedad de obstáculos que se utilizarán en los modelos de optimización conjunta, incluyendo obstáculos basados en el status quo así como varias combinaciones basadas en reglas. Como sugirió Cochran (1997), planteamos el problema de la opción de elección como el siguiente programa lineal entero binario

$$\begin{aligned} & \text{Max} \sum_{s=1}^S y_s \\ \text{Sujeto a} & \sum_{k=1}^k \sum_{j=1}^l \mu_{kj}^s x_{kj} \geq h_s y_s \quad S = 1, 2, \dots, S \\ & \sum_{j=1}^l x_{kj} = 1 \quad k = 1, 2, \dots, k \quad (141) \end{aligned}$$

Este es un problema de cobertura de conjunto máximo (Church y Revela 1974) con coeficientes de restricción generalizados y restricciones laterales (3). La función objetivo (1) produce el número de encuestados cubiertos (es decir, el número para el cual la utilidad del status quo es excedida por la utilidad del nuevo producto). El conjunto de restricciones (2) garantiza que sólo los encuestados para los que se cumple o supera la utilidad de obstáculos se cuentan en la función objetivo. La restricción (3) asegura que se elija exactamente un nivel para cada uno de los atributos K

Como lo demostraron Kohli y Krishnamurti (1989), el problema de la cuota de elección es NP-duro, es decir, es poco probable que el problema pueda resolverse en el tiempo polinomial. Se han aplicado varios enfoques heurísticos al Problema (1) - (3), incluyendo una heurística de división y conquista (Green y Krieger 1989), heurística de búsqueda codiciosa y de programación dinámica (Kohli y Krishnamurti 1989), y Algoritmos genéticos (Balakrishnan y Jacob 1996). Shi et al. (2001) proporcionan un algoritmo de particionamiento anidado globalmente convergente basado en el muestreo aleatorio de particiones de la región factible. Debido a que su algoritmo puede incorporar los métodos heurísticos antes mencionados, lo comparan (para un tiempo de ejecución fijo) con cada heurística.

En términos de calidad de la solución, el algoritmo funciona bien. Sin embargo, los autores no revelan la duración de su tiempo de ejecución fijo y admiten que, dado el tamaño de sus problemas de prueba, las verdaderas soluciones óptimas en su estudio son desconocidas.

El trabajo presentado en este trabajo contribuye a los desarrollos en la optimización conjunta ofreciendo un algoritmo exacto y rápido que puede hacer frente a problemas muy grandes. La exactitud garantiza que se encuentre una solución globalmente óptima. Esta propiedad, que los enfoques heurísticos carecen, es crítica para la administración porque el desarrollo de nuevos productos rara vez es un ejercicio de una sola computadora. Normalmente, se deben explorar varias reformulaciones de los problemas para alcanzar los objetivos estratégicos.

Si las soluciones no son exactas, la variación entre los resultados óptimos no globales se combina con la variación de otras fuentes en estudio, lo que confunde la interpretación de la clave de información con una decisión de diseño de producto. Además, un algoritmo con giro rápido facilita la exploración que se requiere para evaluar a fondo las sensibilidades clave como el error de estimación de utilidades parciales.

En nuestra experiencia, los paquetes de optimización disponibles con el potencial de proveer soluciones exactas a (1) - (3) generalmente usan lineamientos basados en la programación lineal y son ineficaces para solucionar grandes contingencias de la parte- de- Problema de elección. El programa lineal es ineficaz porque las variables de cobertura (y ,) tienden a ser muy fraccionarias cuando se relajan las restricciones de enteros.

Mediante el uso de reglas lógicas para desarrollar y podar el árbol de búsqueda, reducimos grandemente el número de variables en las que nos ramificamos, reduciendo así el tiempo requerido para resolver problemas a la óptima verificable.

Inicializamos el algoritmo mediante la búsqueda de una solución factible utilizando un algoritmo codicioso directo. Para cada atributo $k = 1, 2, \dots, K$, los niveles iniciales se determinan de la siguiente manera:

$$x_{kt} = 1 \text{ if } \sum_{s=1}^S \mu^s = \text{MAX}\{\sum_{s=1}^S U_{kj}^s\} \quad (142)$$

El valor de la función objetivo asociado con la solución codiciosa inicial, ZQ, se establece identificando y contando los respondedores s para los cuales h^* ha sido satisfecha o superada por la elección inicial de los niveles de atributo.

Podemos definir un problema dual Lagrangiano basado en (1) - (3) en un enfoque similar al utilizado por Downs y Camm (1996) al relajar las restricciones de cobertura como sigue:

$$LD: \text{Min } L(\lambda_1, \lambda_2, \dots, \lambda_s),$$

$$s. t. \lambda_i \geq 0 \quad i = 1, 2, \dots, S, \quad (143)$$

Donde $L(\lambda_1, \lambda_2, \dots, \lambda_s)$ está dada por las siguientes expresiones:

$$\begin{aligned} \text{Max } & \sum_{s=1}^S y_s + \sum_{s=1}^S \lambda_s \left(\left(\sum_{k=1}^K \sum_{j=1}^{l_k} \mu_{kj}^s x_{kj} \right) - h_s y_s \right) \\ s. t. & \sum_{j=1}^{l_k} x_{kj} = 1 \quad k = 1, 2, \dots, K, \quad (144) \\ & x_{kj} \in (0, 1) \quad k = 1, 2, \dots, K \end{aligned}$$

λ_s Es el multiplicador de Lagrange asociado con la segunda restricción del conjunto de restricciones (3). El procedimiento de optimización del subgradiente de Downs y Camm (1996) se utiliza para obtener valores para los multiplicadores de Lagrange. El problema de irmer (6) - (8) se puede resolver mediante inspección como sigue:

$$y = \begin{cases} 1 & \text{si } 1 - \lambda_s h_s > 0, \\ 0 & \end{cases} \quad (145)$$

$$x_{kt} = \begin{cases} 1 & \text{si } \sum_{s=1}^S \mu_{kt}^s \lambda_s = \max \sum_{s=1}^S \mu_{kj}^s \lambda_s \\ 0 & \end{cases} \quad (146)$$

Nos referimos al valor óptimo del problema dual Lagrangiano como Z_{LD} . Puesto que la solución a (6) - (8) es naturalmente entera, Z_{LD} no será mejor que el límite obtenido resolviendo la relajación LP, que es lo mismo que (1) - (3) con las restricciones enteras eliminadas. En muchos algoritmos, este tipo de problema lagrangiano se utiliza para generar límites en cada nodo del árbol de búsqueda (véase, por ejemplo, Downs y Camm 1996). A través de las pruebas, hemos encontrado que esto es efectivo desde el punto de vista del tiempo de ejecución para el problema de la opción de elección. En lugar de ello, simplemente utilizamos este problema lagrangiano como un intento rápido de mejorar la solución codiciosa en la fase inicial de nuestro algoritmo. Esto es posible

porque cualquier ajuste de las variables de nivel X_{jk} . Obtenido en la solución del problema de Lagrange usando la optimización de subgradiente se puede usar para generar una solución a (1) - (3) contando el número de restricciones del tipo (2) que son viables cuando $y_s = 1$.

Nos derivamos en las restricciones (3) en orden ascendente con respecto a su cardinalidad {número de niveles dentro del atributo}. Por lo tanto, cada nivel del árbol de búsqueda corresponde a un atributo. El árbol de búsqueda se inicializa utilizando la mejor solución entre las soluciones generadas por los codiciosos y lagrangianos, y se emplea una estrategia de profundidad-primera. Las variables de cobertura (y_s) nunca aparecen explícitamente en el árbol de búsqueda. Pero se usan para evaluar la cobertura en un nodo dado y para propósitos descriptivos en las secciones siguientes.

Para demostrar la efectividad de nuestros datos de algoritmos reales, se comparan con el software de optimización de ILOG CPLEX, versión 8.0 (ILOG 2004) que funciona con el modelo de software AMPL (Fourer et al., 2002) usando el modelo (1) - (3). Las soluciones a dos problemas del mundo real se comparan con las propiedades de convergencia, el tiempo de solución y la calidad de la solución. Una empresa de investigación de mercadeo basada en el medio oeste de los Estados Unidos proporcionó el Problema A, mientras que el Profesor Paul Green (Universidad de Pensilvania) proporcionó el Problema B. Estos problemas son más grandes que la mayoría de los problemas de datos reales discutidos en la literatura (para un ejemplo reciente, . Wittink y sus colegas también proporcionan detalles sobre cientos de aplicaciones comerciales en Estados Unidos y Europa (Wittink y Cattin 1989, Wittink y otros 1994).

Se encontró su algoritmo de rama y límite exacto y verificó la solución óptima global al problema A en 1,4 segundos y el problema B en 7,8 segundos. Los resultados también ilustran las diferentes maneras en que los diversos componentes de nuestro algoritmo pueden ser eficaces. En el problema A, la solución codiciosa es mejorada por el procedimiento subgradiente, pero la solución óptima se encontró más adelante en el árbol de búsqueda. En el problema B, el procedimiento de subgradiente mejora nuevamente a los codiciosos y, de hecho, proporciona la solución probada óptima por la búsqueda.

Curiosamente, en estos dos casos el codicioso componente de nuestro algoritmo captura 95% y 99,5% de la cobertura óptima, respectivamente. Este desempeño relativamente sólido de un enfoque simple refleja el hecho de que los consumidores comparten puntos de vista de preferencia en los dos conjuntos de datos comerciales. Con respecto al tiempo transcurrido ya la cercanía a la óptima, esperamos que la calidad de la solución codiciosa se deteriore con más variación en el punto de vista (en mercados que comprenden segmentos de consumidores latentes más dispares) ya medida que se analizan más niveles. Pasamos ahora a pruebas más estrictas del algoritmo sobre datos simulados donde la variación en el punto de vista, el número de atributos, el número total de niveles y otras características del problema se controlan experimentalmente.

Los resultados sugieren la existencia de dos grupos de factores, con menor y mayor impacto, respectivamente. Tres factores tienen una influencia relativamente menor en el tiempo de solución; En el orden ascendente del impacto (con la relación correspondiente del tiempo de solución para el nivel alto frente al nivel bajo), estos son el tamaño del conjunto competitivo (1.28), el número de encuestados su (4.50) y la dificultad obstáculo relativo (6.67). Los factores con mayor impacto son el número de niveles (101,78) y el número de segmentos (418,10). Por ejemplo, se tarda aproximadamente 4,5 veces más en promedio en resolver problemas con $n = 1\ 200$ encuestados que en problemas con $n = 600$ encuestados. Por otro lado, se tarda 418 veces más en promedio para resolver problemas con cuatro segmentos que con dos segmentos.

Una consecuencia muy alentadora de estos resultados es que el uso de muestras más grandes y más representativas no ralentiza apreciablemente el algoritmo. El tamaño de la muestra sospechado en estas simulaciones ya es bastante grande según los estándares comerciales. Las soluciones a la optimalidad global requerían sólo unos pocos minutos como máximo, y los resultados sugieren que las muestras mucho más grandes funcionarán eficientemente. Los otros dos factores en el grupo de "impacto menor", el tamaño del conjunto competitivo y la dificultad relativa del obstáculo, Proceso de decisión. El tamaño de conjunto imita la cantidad de información que un individuo lleva y procesa al establecer estándares mentales de comparación entre productos, mientras que la dificultad de dificultad relativa imita el hecho de que algunas personas tienen estándares absolutamente más altos que otros cuando procesan información

independientemente de cuánta información se conserva. Juntos, estos dos factores contribuyen a la utilidad absoluta del obstáculo que un nuevo producto debe superar para satisfacer al individuo. Los resultados sugieren que las poblaciones de consumidores con estándares muy altos crearán problemas con tiempos de solución sólo marginalmente más largos que las poblaciones con estándares menos estrictos. De lejos, las dos influencias más importantes en el tiempo de solución son las variaciones en el número de niveles y el número de segmentos. El analista puede controlar el número de niveles, pero el control disponible es templado por los requisitos de la categoría de producto. Ciertas categorías son simplemente más complejas que otras y deben ser representadas por más atributos y niveles. Sin embargo, incluso para problemas con niveles de 112, promediando otros factores de diseño, los tiempos de solución son todavía razonables (697 segundos o menos de 12 minutos en un PC).

El número de segmentos de consumidores latentes en la población afecta más profundamente los tiempos de solución. Este factor es un sustituto parcial de la heterogeneidad general en las funciones de utilidad (puntos de vista de los consumidores) entre los miembros de la población consumidora. También es inobservable. Encontramos un aumento de 400 veces de la insolución para los problemas con cuatro segmentos en comparación con los dos. De hecho, La Tabla 2 revela que los problemas más difíciles provienen de una interacción de tres vías entre el número de niveles, el número de segmentos y la dificultad de obstáculos. Entre los ocho problemas con 112 niveles y cuatro segmentos, los que tienen una dificultad de obstáculos bajos tienen un promedio de 351 segundos para resolver, mientras que aquellos con una dificultad alta dificultan 2.433 segundos para resolver, una proporción de 7: 1 con los problemas más difíciles que requieren unos 40 minutos.

También se realizó análisis (i) para evaluar el grado en que el algoritmo codicioso se aproxima a la solución óptima (Modelo 3), y (ii) vincular las características del problema con la cobertura global de la solución óptima (Modelo 4). Para el Modelo 3, calculamos el porcentaje de cobertura del algoritmo codicioso en relación con la solución óptima. Para el Modelo 4, calculamos la cobertura de la solución óptima como un porcentaje de la muestra total. El logaritmo de cada variable se utiliza como variable dependiente En modelos lineales separados, con las cinco variables independientes y las interacciones de la Tabla 2 (Modelo 2) como predictores. Los resultados de los dos modelos se informan lado a lado en la Tabla 3. Obsérvese que no hubo interacciones bidireccionales

con el número de encuestados (R) ni ninguna de las interacciones de tres vías a un nivel razonable, por lo que no se informan aquí.

En el orden de varianza explicado, los tres efectos principales en ambos modelos son la dificultad relativa de obstáculos (H), el número de segmentos (S) y la interacción S x H. El impacto de la dificultad relativa del obstáculo se espera porque podemos manipular la cobertura de cero (utilizando obstáculos estrictamente imposibles) a la muestra total (usando vallas ingenuamente permisivas). Así, la dificultad relativa del obstáculo influye tanto en el rendimiento del algoritmo codicioso como en la cobertura máxima (óptima) posible en una muestra. Encontramos que, para algunos problemas, la solución codiciosa coincide con la solución óptima, mientras que el codicioso es tan bajo como la mitad del óptimo en otros. De manera similar, la solución óptima alcanza un 100% de cobertura en algunos casos, mientras que en otros alcanza una cobertura inferior al 60%.

La capacidad de la solución óptima para cubrir la muestra está directamente relacionada con la heterogeneidad de las funciones de utilidad en la muestra. Esto, a su vez, está influido por el número de segmentos en la muestra y la variación entre utilidades parciales dentro del segmento. En el presente caso, el cambio de dos a cuatro segmentos reduce la cobertura media alcanzada de cerca del 90% a alrededor del 70%. A medida que aumenta el número de segmentos, la cobertura disminuirá gradualmente, ya que ningún perfil único puede satisfacer puntos de vista cada vez más diversos. Aunque no se informa aquí, utilizamos el algoritmo para encontrar soluciones segmento por segmento para generar un portafolio de soluciones. Esta técnica es valiosa en la práctica, pero no depende de juzgar la eficiencia del algoritmo porque implica rendimientos con submuestras más pequeñas y más homogéneas.

Cada uno de los componentes principales de nuestro algoritmo (heurística codiciosa, optimización de subgradiente utilizada para intentar generar una solución mejorada y búsqueda de árbol basada en lógica) juega un papel importante en la solución de los problemas. Como era de esperar, los problemas de obstáculos más difíciles dependen más de la búsqueda de árboles, mientras que la heurística codiciosa es bastante eficaz en los problemas de obstáculos bajos. El procedimiento subgradiente es aproximadamente igual de efectivo para ambas clases.

Aunque el trabajo publicado en OR normalmente trata a las partes estimadas como datos sin errores en lugar de estimaciones estadísticas (por ejemplo, Kohli y Krishnamurti 1989, Shi y Olafsson 2000, Shi et al., 2001), una visión más realista reconoce que las partes recuperadas contienen error de estimación. Las técnicas de estimación basadas en los métodos MCMC proporcionan estimaciones de parámetros de niveles individuales, así como la capacidad de evaluar la incertidumbre en esas estimaciones. Ejemplos del uso de métodos MCMC en la comercialización pueden encontrarse en Allenby et al (1995, 1998) y Lenk et al. (1996). Estos métodos jerárquicos de análisis combinados bayesianos combinan información previa (es decir, la distribución de la heterogeneidad junto con posibles restricciones lógicas, como la utilidad decreciente monotónica en el precio) e información muestral {evaluaciones de los encuestados} para obtener estimaciones empíricas de la distribución posterior de las utilidades. El nivel individual de los encuestados (Rossi y Allenby 2003).

Debido a que nuestro algoritmo es exacto y rápido, podemos usarlo para resolver problemas de muestras repetidamente basados en distribuciones posteriores de utilidades de parciales. La exactitud garantiza que la variación en los ensayos repetidos se debe a la variación en las estimaciones de parciales en lugar de la variación atribuible a la sub-optimización. En esta sección se ilustra una forma de utilizar nuestro algoritmo para hacer frente a la incertidumbre de parciales y reportar las pruebas de la robustez de nuestro algoritmo a la variación de persona en persona en representados por estas distribuciones posteriores.

Los datos provienen de un experimento de siete atributos de elección discreta sobre las preferencias del consumidor para las características de un nuevo formato de cámara llamado Advanced Photo System (APS). El formato ha sido propuesto por un importante fabricante estadounidense de cámaras y ofrece ventajas sobre los sistemas estándar de 35 mm, incluyendo el cambio de película intermedia, la retroalimentación de la operación de la cámara y una capa magnética en la película que soporta la anotación de imágenes con títulos y fechas. Los datos se describen con más detalle en Allenby et al. (2002) y Gilbride y Allenby (2004).

Los encuestados fueron examinados y reclutados por teléfono y entrevistados en nueve localidades; N = 302 encuestados participaron. En el experimento de elección discreta, cada encuestado vio 14 escenarios de compra. Cada escenario presentado en orden

aleatorio dentro de los encuestados contenía descripciones de perfil completo de tres cámaras compactas, tres nuevas cámaras APS y una séptima opción de no comprar. (Por favor, consulte las columnas de la izquierda de la Tabla 6 para los atributos y niveles manipulados en el experimento.) El diseño completo produce $(2^2 \times 3^2 \times 4^2 \times 6) = 3,456$ perfiles.

4.2.2 Diseño o método aplicado

Parcialess se estimaron para un individuo usando un modelo probit multinomial (basado en el enfoque Bayes jerárquico descrito en Allenby et al., 2002) que produce una distribución empírica posterior de los vectores de parcialess para cada individuo en la muestra. Teníamos acceso a 100 dibujos de cada una de esas densidades. Un único sorteo para un encuestado dado es un vector 1×24 de valores de parcialess. Para probar la robustez de nuestro algoritmo a la variación en estos valores, llevamos a cabo un experimento con una fase de prueba, una fase de validación, y una fase de evaluación aleatoria.

En la fase de prueba, realizamos $N_t = 1,000$ iteraciones del siguiente proceso: dibujar aleatoriamente un vector parcial de la densidad posterior de cada encuestado para obtener una instancia de la matriz 302×24 de parcialess; Optimizar sobre esta matriz y registrar el perfil óptimo y el tiempo de cálculo; Repita este proceso hasta que todas las iteraciones estén completas. Debido a que parcialess seleccionados dentro de la persona varían en cada iteración del experimento, esperamos que nuestro algoritmo para encontrar diferentes optima durante este proceso. Si se encuentran pocas alternativas óptimas, el algoritmo de optimización demuestra robustez a la incertidumbre en las partes.

En la fase de validación presentamos cada perfil óptimo que se encuentra en la fase de prueba al modelo de utilidad de cada encuestado para registrar la cobertura de esa instancia entre los 302 encuestados. Repetimos este proceso $N_v = 1,000$ veces usando un nuevo dibujo para la función de utilidad de cada individuo en cada iteración. Nuestra esperanza es que los perfiles previamente identificados como óptimos tengan un buen rendimiento y tengan una alta cobertura durante los ensayos de validación. Como punto de referencia, realizamos este mismo experimento de validación con perfiles de diseño dibujados al azar.

Los niveles de obstáculos para la optimización se seleccionaron utilizando un modelo de profundidad de maqueta de elección como se describe en Curry (2004). Nos fijamos m para que un perfil esté en el 5% superior del rango de utilidad de un individuo sobre todos los perfiles. El nivel de obstáculos es único para cada encuestado.

4.2.3 Resultados

La fase de prueba produjo 21 perfiles únicos como óptimas más de 1.000 iteraciones. Estos 10 perfiles se encontraron en 950 de las 1.000 iteraciones. Los siguientes 11 perfiles más frecuentes representan los 50 casos restantes. Este perfil es (estilo cuerpo medio, cambio automático de midroll, entrada personalizada por Método 2, retroalimentación de operación, zoom 4x, visor regular, ajustes de retroalimentación de LCD y visor). El algoritmo encontrado y como la óptima certeza en cada caso en menos de un segundo en un ordenador portátil.

4.2.4 Conclusiones

Hemos proporcionado una rama exacta y algoritmo vinculado para la cuota de problema de elección, un problema que hasta ahora se ha resuelto heurísticamente. Hemos demostrado, en tres grandes estudios conjuntos reales y 96 problemas cuidadosamente simulados (tres repeticiones para cada uno de los 32 tratamientos experimentales) con 32 atributos y 112 niveles, que nuestro enfoque es útil para encontrar soluciones óptimamente óptimas a problemas a gran escala en una cantidad razonable de tiempo. Hemos aislado las características del problema a las que el rendimiento del algoritmo es más sensible y proporcionó resultados que sugieren que el algoritmo resuelve eficazmente problemas complejos a la optimalidad. También hemos demostrado cómo este algoritmo puede ser utilizado para abordar la incertidumbre inherente en parciales, un problema que no ha sido abordado en el trabajo algorítmico anterior sobre este problema.

El algoritmo exacto de rama-y-atado que hemos presentado es muy flexible en términos de tratar con una variedad de objetivos y restricciones. Por ejemplo, uno puede desear considerar una función objetiva que no sea la participación de elección. Un objetivo que pesa a cada encuestado por su utilidad excedente y maximiza la suma de las utilidades excedentes captaría mejor la lealtad del cliente. Esta modificación apenas altera la estructura del problema y requiere poco tiempo de cálculo adicional para calcular la

función objetivo y encontrar la solución óptima. Además, si el coste de fabricación de un producto no puede exceder una cantidad predeterminada y se conocen los costes marginales de los diversos niveles de atributo, se puede añadir a la formulación una restricción que refleje esta restricción. A diferencia de los enfoques de LP, donde la adición de una restricción de este tipo puede inducir soluciones altamente fraccionarias y conducir a tiempos de cómputo dramáticamente aumentados, el impacto en nuestro algoritmo será mínimo. La adición de tal restricción es simplemente otra prueba lógica que se implementará en cada nodo.

4.3 Caso 3: Diseño de un sistema de análisis interactivo conjunto basado en web

4.3.1 Antecedentes

Para el análisis de marketing, los diseñadores de productos quieren entender y medir la demanda de un mercado específico para diseñar el producto o servicio más atractivo. Entre muchas técnicas de análisis de marketing, el análisis conjunto es una de las herramientas de investigación de mercado más populares para diseñar productos o servicios (Gustafsson et al., 2007). Muchas empresas ya utilizan cuestionarios diseñados para recopilar información sobre productos o servicios. En estos días, la mayoría de estos cuestionarios se pueden responder con una encuesta basada en la Web (Tsuji y Takeyasu, 2009). Esto permite al cliente responder a encuestas en cualquier momento y lugar. Además, con la ayuda de la ciencia de la tecnología de la información, es una manera más sencilla de recopilar y analizar todas las respuestas de los encuestados.

En el caso de un producto complejo con muchos parámetros, el conocimiento tácito del entrevistado puede ser difícil de entender. Para obtener un análisis preciso, se debe permitir que el encuestado se corrija y reconsidere su evaluación. Con un sistema de retroalimentación, los vendedores pueden estar seguros de obtener información más precisa sobre el producto / servicio del estudio. En este artículo, explicaremos cómo este sistema ofrece la posibilidad de obtener nueva información sobre las preferencias de los consumidores.

4.3.2 Diseño o método aplicado

Al final de las tres evaluaciones, los vendedores dieron una encuesta a los encuestados con respecto a varias informaciones como su género, su edad y su grado de conocimiento en hardware portátil. Al final de las evaluaciones, pedimos a los encuestados que clasificaran los mismos atributos de sus preferencias. Con el análisis conjunto, calculamos la utilidad parcial real para cada encuestado. También clasificamos estos valores, y los vendedores pudieron observar la diferencia entre los atributos clasificados en la encuesta y las utilidades parciales clasificadas calculadas después de la evaluación.

Hemos observado que entre los 20 encuestados, nadie tiene una encuesta de clasificación completamente similar a la calculada. Sólo cuatro encuestados tienen más de dos atributos coincidentes y seis encuestados sólo tienen un atributo correspondiente. Estos resultados muestran que hay una fuerte diferencia entre lo que los encuestados piensan que les gusta y las preferencias encontradas con el análisis conjunto tradicional.

De hecho, este proceso de evaluación ayudó a los encuestados a evaluar fácilmente un producto complejo con siete atributos. Dejamos que los encuestados evalúen los perfiles en función de sus preferencias. También ayuda a los vendedores a destacar los atributos que los encuestados realmente consideran ya clasificar los atributos, de una manera similar al análisis adaptativo (Johnson, 2001).

Sin embargo, la diferencia entre las preferencias de los encuestados (expresadas en la encuesta) y las preferencias calculadas con el análisis conjunto es demasiado importante para validar la exactitud de estos resultados. Con la ausencia de retroalimentación, un vendedor no puede controlar la coherencia de la evaluación de los encuestados. Por otra parte, con el método de análisis conjunta tradicional, sólo una evaluación por conjunto de utilidades totales de perfil y simple, los encuestados no tienen la posibilidad de modificar y mejorar sus respuestas.

Con el sistema de análisis interactivo, queremos cuantificar con precisión la preferencia tácita de los encuestados en lugar de simplemente ordenarla. Con esto, los vendedores pueden determinar un grupo de atributos favoritos en lugar del atributo favorito único, o menos favorito, para los encuestados.

En el próximo capítulo, presentaremos nuestro sistema CASIMIR, diseñado para no sólo proponer una retroalimentación a los encuestados, sino también permitirles modificar y mejorar sus puntuaciones después de cada evaluación.

Nuestro análisis conjunto interactivo se propuso tener los datos más consistentes posible. Por lo tanto, para esta investigación, los encuestados se les permite interactuar durante su evaluación. De hecho, todavía evalúan varios perfiles, y también reciben un diagnóstico completo. Con este diagnóstico, pueden ser conscientes de sus preferencias de productos de forma sistemática. Además, para la próxima evaluación, pueden usar esta nueva información para reevaluar un conjunto de perfiles o evaluar nuevos perfiles. La traducción de la preferencia y del conocimiento tácito de los encuestados permite obtener respuestas de evaluación más precisas.

Después de reunir todas estas respuestas, se presentan al entrevistado los valores totales de utilidad previstos y la importancia atributiva. Con estos valores pronosticados, el encuestado puede entender su elección tácita y detectar posibles errores en las evaluaciones pasadas. Si él o ella descuida accidentalmente un atributo durante la primera evaluación, después de ver el diagnóstico, él o ella puede atribuir más importancia a este atributo durante la segunda evaluación.

Obviamente, la elección de si seguir o no este diagnóstico se da a los encuestados. Consideramos esto como una herramienta que permite a los encuestados pensar en su evaluación en lugar de una indicación que debe seguirse de forma absoluta. Los encuestados pueden utilizar el diagnóstico y los valores pronosticados para reevaluar y mejorar sus respuestas. Es importante que el análisis del producto traduzca la preferencia y el conocimiento tácito de los encuestados.

4.3.3 Resultados

La parte principal del análisis interactivo conjunto se ha presentado, y ahora, en este capítulo, presentaremos el sistema CASIMIR. Con la ayuda de este cuestionario basado en la Web, nuestra idea es predecir la utilidad total, atribuir importancia y otra información personal en un diagnóstico que se entrega después de cada evaluación.

Esta aplicación se codificó principalmente con PHP y JavaScript, permitiendo que el sistema CASIMIR se utilizara en todos los ordenadores. XAMPP, una solución de servidor web de código abierto, se utilizó para desarrollar nuestro sistema. Esta solución permite utilizar el servidor HTTP de Apache, una base de datos MySQL y un intérprete para scripts PHP. Debido a las necesidades de nuestro proyecto, esta solución fue la más interesante para nosotros.

Para calcular el análisis de regresión y todas las utilidades totales y parciales, se utiliza el freeware R. Es un potente software para la informática estadística (R Project for Statistical Computing, 2014) que utiliza su propio lenguaje de programación. R es una implementación del lenguaje S programming combinado con la semántica de alcance léxico.

Permite una programación orientada a objetos fuerte, y tiene una comunidad activa que permite a los usuarios encontrar una solución para la mayoría de los problemas. Al conectar el software R a la base de datos, los vendedores pueden acceder a todos los datos registrados. Con esto, un vendedor puede calcular la pendiente de las variables independientes, la importancia de los atributos y la utilidad total prevista y la utilidad de base, y luego todos estos nuevos datos se registran en la base de datos.

En contraste con algunas encuestas de análisis conjuntas que son anónimas, mediante el uso de este sistema, los encuestados tienen que iniciar sesión, lo que permite al comerciante rastrear y recoger cada evaluación. Para la recopilación de datos, tenemos que vincular todos los datos a los encuestados en una gran base de datos.

En primer lugar, los comercializadores deben implementar un perfil de producto o servicio en la base de datos y determinar cuántos atributos y niveles son necesarios para las encuestas (Rajaraman et al., 2010). Algunos atributos pueden estar presentes en diferentes productos, como "tamaño de pantalla" de televisores, por lo que estos atributos tienen que estar vinculados a un producto en particular. En la misma idea, algunos niveles tienen que estar vinculados a un atributo, como el "azul" y "rojo" para el atributo "color". Con estos productos y atributos detallados, los comercializadores tienen que crear perfiles.

Dependiendo del número de niveles y atributos, el número mínimo de perfiles necesarios para el análisis conjunto puede variar. El uso del software R ayuda a realizar la planificación ortogonal y determinar este mínimo. Con un segundo script R, se puede generar automáticamente un conjunto que incluya el número mínimo de perfiles requeridos. Con este método, un vendedor puede generar fácilmente algunos conjuntos de perfiles y proponerlos para su evaluación. Una vez que los perfiles de producto son creados y validados, estos perfiles están disponibles para su evaluación en la base de datos. Un mismo producto puede tener varios conjuntos de perfiles para diferentes evaluaciones.

4.3.4 Conclusiones

En primer lugar, este trabajo ha introducido y mostrado el principio y los límites del análisis conjunto tradicional. Con una experiencia simple, mostramos que puede ser difícil traducir correctamente las preferencias de los encuestados y conduce a un análisis inconsistente para los vendedores. Con nuestro sistema de cuestionarios de SSI denominado "CASIMIR", queremos diseñar una solución flexible, accesible en cualquier momento y en cualquier momento para los encuestados. También hemos demostrado la importancia de la retroalimentación en el sistema CASIMIR. Mostrar la importancia de los atributos y la coherencia personal puede ayudar a los encuestados a considerar su evaluación y preferencias. Traducir el razonamiento tácito de un encuestado es una tarea complicada. Este sistema permite al encuestado reaccionar a su propia retroalimentación. Si no validan el diagnóstico, nos damos cuenta de que tienen la oportunidad de corregir y mejorar sus evaluaciones pasadas.

Además, para tener una calificación precisa de los encuestados sin aumentar el tiempo de la evaluación, los encuestados pueden contar con los valores previstos basados en la evaluación previa. Esto permite que el sistema CASIMIR ayude a los encuestados a tomar decisiones con mayor facilidad.

4.4 Caso 4: Justificación y diseño de experimentos de elección moderada: una aplicación de precios estratégica

4.4.1 Antecedentes

Una aplicación común del análisis conjunto y técnicas relacionadas es la investigación de sensibilidad de precios (Louviere, 1988; Lazari y Anderson, 1994). En un experimento de sensibilidad a los precios, los encuestados realizan evaluaciones, generalmente opciones, en cada uno de los 20 escenarios de precios. Estos escenarios, o "conjuntos de opciones", pueden ser tan complejos como las opciones entre dos o más perfiles de producto multiatributo en los que el precio y varios otros atributos varían de un perfil a otro y el conjunto de opciones a un conjunto de opciones. También pueden ser tan despojados como un experimento de marca de precio en el que sólo el precio y la marca varían según los perfiles y los conjuntos de elección. En cualquier caso, hay una suposición de independencia entre conjuntos de opciones: la información contenida en un conjunto de opciones se supone que no afecta a las evaluaciones de los encuestados en conjuntos de elección posteriores.

Es bastante fácil imaginar situaciones en las que esta suposición es falsa. Por ejemplo, imagine un estudio donde en la marca y el posicionamiento se encuentran entre las variables. En algunas preguntas Brand X se describe como el líder de bajo costo y en otros como la marca de calidad premium. Si Brand X se describe como la marca premium en el conjunto de opciones 8, pero como el líder de bajo costo en el conjunto de opciones 7, ¿cuán creíble es su descripción como premium en el conjunto de opciones 8? Es evidente que cualquier remanente a través de conjuntos de selección puede difuminar el posicionamiento y la medición de posicionamiento para la marca X.

Una solución a esta posible contaminación por efectos de traspaso es un experimento moderado de elección (\$ MCE). En un experimento conjunto típico, todos los efectos experimentales se presentan a todos los sujetos: aunque no se modelan de manera repetida, todos los efectos se presentan como efectos dentro del sujeto. En MCE, sin embargo, algunos de los factores experimentales se presentan como efectos entre sujetos: para algunas variables, los niveles que un entrevistado ve son controlados como parte del diseño experimental.

Así que, en el ejemplo anterior, en un enfoque conjunto tradicional, el encuestado Jones recibiría todos los atributos y todos sus niveles a través de sus conjuntos de elección. En un enfoque de MCE, sin embargo, Jones vería todos los atributos y todos sus niveles excepto que en su elección establece Brand X siempre se describiría como un líder de bajo costo.

El encuestado Brown está en otra celda experimental, sin embargo. Siempre ve a Brand X con su posición de marca premium. En este MCE no hay posibilidad de un efecto de traspaso que influya en las evaluaciones de los encuestados de posicionamiento Brand X, porque el posicionamiento es un efecto entre sujetos.

Dos formatos comunes de precios utilizados por los minoristas son de alto-bajo y cada día de precio bajo. Bajo el más común de los dos, de alto a bajo, un minorista periódicamente los precios de descuento en algunos o todos los productos para que los productos individuales a veces tienen un precio más alto y en ocasiones una menor. Un minorista que adopte un posicionamiento EDLP, sin embargo, ofrece a sus clientes un precio constante que no es tan alto como el precio no descontado en alto-bajo ni tan bajo como el precio con descuento. La lógica y el éxito empírico de estas dos estrategias de los minoristas han sido el foco de atención académica (Hoetal., 1998, Hochet et al., 1994, Bell y Lattin, 1998, Lal y Rao, 1997, McGoldrick et al., 2000).

Mucho menos atención se ha centrado en las estrategias de alto-bajo y EDLP como marca. Algunas marcas de bienes de consumo en movimiento rápido han hecho cambios de alto perfil entre estas estrategias, pero la investigación publicada sobre el resultado ha sido casi inexistente, tal vez debido al entorno competitivo y la naturaleza propietaria de la investigación realizada. Incluso menos frecuentes han sido los estudios de alto-bajo versus EDLP en el contexto de productos duraderos. Esta prueba empírica, un mercado de prueba simulado del cambio de una marca a EDLP en una categoría históricamente alta-baja, se describe a continuación.

Disfrazados para propósitos competitivos, para proteger la inversión del patrocinador en investigación propia, son el cliente, Acme, y su producto, widgets. Los widgets tienen un ciclo de compra de cuatro años y están disponibles a través de múltiples canales de minoristas.

Desde el inicio de la categoría de widgets décadas atrás, cada uno de los seis fabricantes ha ofrecido una o más marcas de widgets a la vez, actualmente hay diez marcas de widgets disponibles de estos seis fabricantes. Desde principios de la vida de la categoría, cada fabricante vendió widgets durante parte del año a precio completo y durante algunos períodos promocionales, en uno de los tres niveles de descuento: un 50 por ciento de descuento en la reducción de precio temporal), un descuento del 25 por ciento O una promoción que valga el 50 por ciento del precio completo. Los períodos de descuento de los fabricantes no están sincronizados, pero sí se superponen sustancialmente.

Grandes costos financieros y administrativos asisten a la política de precios altos y bajos en Acme y sus competidores. Los reembolsos requieren atención administrativa y gastos por parte de los minoristas y los consumidores. Las reducciones temporales de los precios agregan esfuerzo administrativo a los minoristas. Ambos causan confusión entre los minoristas y los consumidores. Acme y sus competidores emplean departamentos enteros y una gran cantidad de consultores para manejar la carga administrativa de alto-bajo. La gerencia superior de Acme se pregunta por el efecto de la cuota de mercado de traducir todos esos gastos administrativos a los consumidores a través de la reducción de precios permanente menos extrema que implica la EDLP. ¿Podría Acme canalizar los fondos invertidos en alto-bajo en precios permanentemente bajos y mantener o incluso aumentar) su cuota de mercado? ¿O los competidores tomarían parte de Acme, ya sea copiando o evitando su traslado a EDL?

Una revisión de fuentes secundarias no proporcionó asesoramiento inequívoco, ya que se centra en EDLP como una estrategia de minorista, no como una estrategia de marca. En una excepción potencialmente generalizable, Bell y Lattin (1998) encontraron que los compradores de la gran cesta, compradores que compran muchas cosas en un viaje de compras, tienen menos flexibilidad en la búsqueda de la tienda de menor costo para cada artículo individual y, por lo tanto, están más inclinados Hacia las tiendas de EDLP que les ofrecen un costo de cesta más bajo.

En la medida en que el valor para los consumidores de EDLP es su capacidad de reducir el costo de su búsqueda reduciendo el costo total de cesto para tales compradores de "canastas grandes", EDLP para un bien duradero tiene menos sentido que EDLP para un minorista. Dado que la búsqueda es frecuente cada cuatro años y el tamaño de la canasta pequeña sólo el único producto duradero) la reducción total del costo de la cesta es pequeña en comparación con los ahorros ofrecidos por las compras alrededor de un descuento alto-bajo.

Se consideraron varios esfuerzos de investigación primaria: un concepto tradicional basado en pruebas de mercado de pruebas simuladas (STM) explicaría el concepto EDLP de Acme a los compradores de widgets y pedirían su probabilidad de comprar un widget de Acme. Las células experimentales separadas podrían mostrar Acme a alto-bajo y Acme en EDLP y los niveles resultantes de la intención de compra del consumidor podrían ser comparados.

Desafortunadamente, los STM tradicionales adoptan un enfoque de acción competitiva totalmente distinto. Esto en la atención a los esfuerzos competitivos realistas impide que un concepto tradicional STM basado en pruebas de abordar cuestiones importantes sobre las reacciones de la competencia y su efecto en la cuota de mercado de Acme.

La necesidad de modelar reacciones competitivas trae a la mente alguna forma de experimento conjunto (Johnson, 1974, Green y Rao, 1971, Louviere, 1988). En un experimento conjunto basado en la elección, cada encuestado hace elecciones de marca entre múltiples escenarios o "conjuntos de elección" (Louviere y Woodworth, 1983). A través de conjuntos de opciones, las estrategias de las marcas se manipularían de acuerdo con un diseño experimental, de modo que Acme sería EDLP en una pregunta, ofreciendo un TPR en el siguiente, ofreciendo el descuento en un tercero y a un precio elevado no descontado en el cuarto.

Del mismo modo, los precios de los competidores variarán de cuestión a pregunta, de nuevo variando sistemáticamente conforme al diseño experimental. Modelar la elección de la marca como una función de la marca y la fijación de precios daría lugar a una ecuación predictiva cuyos parámetros cuantificar el efecto de la EDLP, el precio completo, TPR, rebaja y TPR plus rebate.

Sin embargo, la mayor parte del valor de EDLP puede ser su capacidad para reducir los costos de búsqueda para un consumidor. Quizás sólo la constancia de EDLP le da valor. Si una marca es EDLP en una pregunta conjunta, el precio completo en el siguiente, y en TPR en un tercero, esta constancia se pierde en el encuestado. Si los encuestados recuerdan los precios de las preguntas anteriores, pueden percibir que la EDLP es sólo una TPR menos atractiva. La posible contaminación cruzada o efecto remanente de las preguntas conjuntas crea un sesgo potencial contra la EDLP.

Por estas razones, Acme eligió conducir su STM para EDLP usando un MCE. Cuando una marca está en EDLP para un encuestado, es en EDLP en cada pregunta para ese encuestado. Por lo tanto, existen diferentes versiones del cuestionario para las diferentes combinaciones experimentales de marcas en la EDLP y en niveles muy bajos.

4.4.2 Diseño o método aplicado

Acme quiere entender las implicaciones de la estrategia de precios y la táctica (EDLP, precio completo, reducción de precio temporal, descuento) para diez marcas. Para evitar la contaminación por efectos de retraso, el experimento de elección moderada se construye como un diseño de dos etapas.

El primer diseño experimental controla las estrategias de precios a través de las celdas. Determina qué marcas están en EDLP y cuáles están a precios altos-bajos. Las celdas del diseño resultan de una fracción de 12 corrientes del factorial completo de 2^{10} . En cada celda, las marcas codificadas 1 obtienen precios EDLP y las marcas codificadas 0 obtienen precios altos y bajos.

La segunda etapa del diseño experimental es el diseño dentro de la célula para las tácticas de precios. Para las marcas de alta y baja, este segundo diseño controla si la marca aparece a precio completo (0) reducción de precio temporal (1) o reembolso (2). Las marcas EDLP son constantes dentro de una celda al precio EDLP. El diseño es una fracción de ejecución de 18 de un diseño 3^5 o 3^6 , dependiendo del número de marcas de alto-bajo en la célula.

Una lista comercial de compradores de widget sirvió como marco de muestra. Un estudio por correo permitiría a los encuestados proporcionar respuestas más reflexivas que una encuesta telefónica, por lo que los encuestados fueron seleccionados para la calificación y la voluntad de participar y luego se envió una encuesta de papel y lápiz. Una encuesta basada en la Web también puede haber funcionado, pero la demografía de los compradores de widgets en el momento del estudio sugiere la penetración baja de Internet entre ellos. Un total de 2.160 encuestas fueron enviadas por correo a los encuestados calificados que declararon estar dispuestos a participar, y 1.453 (66 por ciento) fueron devueltos. Los 121 completos por célula fue mayor que el objetivo de 100 por célula.

Dos atributos (precio y marca) de cada uno de los diez productos varían en el diseño experimental y se incluyen como variables predictoras en el modelo. Con los atributos de las alternativas como predictores y la elección de marca como variable dependiente, el análisis de logit multinomial condicional es apropiado (McFadden, 1974; Ben-Akiva y Lerman, 1985). Modelado de opciones discretas a través del programa LOGIT de Salford System para MNL produjo utilidades para cada estrategia de precio / combinación de tácticas para cada marca.

Estas utilidades muestran que, para cada marca, el valor de EDLP fue mayor que el del precio completo. Para nueve de las diez marcas, EDLP era menos valiosa que la rebaja y para los diez era menos valiosa que la reducción temporal de los precios.

4.4.3 Resultados

Estas utilidades se utilizaron para simular las acciones en una variedad de escenarios y se construyeron en un sistema de apoyo a la decisión para su uso por el cliente. Las simulaciones muestran que, para cualquier marca, el paso a EDLP resulta en menor cuota. Por ejemplo, la participación de la marca insignia de Acme cae de 24,4 por ciento bajo el nivel alto-bajo al 18,1 por ciento bajo EDLP. La pérdida de acciones similares ocurre para la marca de bajo precio de Acme.

Del mismo modo, la marca de la competencia Rival Lite cae a 5,7 por ciento bajo EDLP de 9,7 por ciento bajo bajo. Incluso si Acme y sus dos competidores más grandes (Rival y Challenger) llevan a sus franquicias enteras EDLP, siguen perdiendo participación en los dos competidores más pequeños: entre ellos, Acme, Rival y Challenger caen de 81.2 por ciento bajo bajo a 71.6 por ciento bajo EDLP.

4.4.4 Conclusiones

EDLP cuesta a cualquier marca en la categoría cuota de mercado sustancial. Parece ser una estrategia de precios de marca viable. Este patrón ocurre para las marcas premium y de bajo precio de todos los fabricantes. Esto es consistente con la explicación de que la apelación de EDLP deriva del hecho de que reduce los costos de búsqueda para los clientes de la cesta grande. Dado que la actividad de búsqueda en la categoría de widgets sólo se produce cada cuatro años aproximadamente, y la cesta sólo contiene un solo producto, los ahorros de búsqueda son triviales en comparación con el precio de compra más bajo disponible la mayor parte del año bajo precios altos o bajos. A diferencia de los clientes de Bell y Lattin (1998), los compradores infrecuentes de bienes duraderos no obtienen nada de EDLP.

Cada vez que un encuestado vio una marca en EDLP, lo vio en EDLP en cada una de sus preguntas elegidas. Por lo tanto, el diseño experimental en dos etapas garantizó que no había efectos de traspaso posibles. Es evidente que esta prevención de los efectos de retraso fue fundamental para la investigación de Acme EDLP.

La prevención de efectos de traspaso puede ser importante también en otros tipos de investigación.

Estudios de precios tácticos en los que el investigador quiere evitar el "juego" por parte de los encuestados que pueden captar el hecho de que una serie de preguntas difieren sólo en los puntos de precio que muestran.

Estudios de posicionamiento en los que el investigador quiere mostrar dos o más estrategias de posicionamiento alternativas para una marca. Si la única diferencia en una marca de una pregunta de elección a la siguiente es su posicionamiento, el encuestado puede desenfocar los posicionamientos y podría ocurrir un efecto de retraso que dañaría el estudio. Prevenir estos efectos de retraso sería de nuevo crucial para el éxito del estudio.

Por lo tanto, los experimentos de elección moderada pueden ser un enfoque ampliamente aplicable, uno apropiado para experimentos de elección en los que se puede esperar que se produzcan efectos de traspaso

4.5 Caso 5: Procesamiento de información en experimentos conjuntos basados en la elección

4.5.1 Antecedentes

Desde su introducción al campo de la comercialización a principios de los años 70 por Green y sus colegas (Green y Rao, 1971, Green y Wind, 1973), el interés y la aplicación de técnicas de análisis conjuntas se ha expandido rápidamente. Ahora son el método de investigación de marketing más utilizado para analizar los compromisos de consumo (Green et al., 2001), y también han obtenido amplia aplicación en investigación en transporte, medicina, estudios ambientales y otros campos (Louviere et al., 2000). Para analizar datos conjuntos necesitamos hacer ciertas suposiciones acerca de cómo las personas procesan la información y toman decisiones.

La mayoría de las representaciones matemáticas en uso se basan en la asunción de una toma de decisiones compensatoria, lo que implica que los encuestados consideran la totalidad o la mayor parte de la información disponible y hacen concesiones entre los atributos del producto. Además, los modelos se basan en suposiciones de preferencias bien formadas y que la gente puede recuperar una respuesta apropiada a cualquier pregunta de preferencia-motivación.

En los últimos 20-30 años, la investigación en la toma de decisiones conductuales ha demostrado que las personas no siempre, si alguna vez, se comportan de acuerdo con esta perspectiva racionalista. Por ejemplo, la gente a menudo no tiene preferencias bien definidas, pero puede construirlas en el lugar cuando sea necesario (por ejemplo, Payne, Bettman y Johnson, 1992a, Payne, Bettman, Coupey y Johnson, 1992b).

Además, en lugar de utilizar una estrategia de decisión fija, los consumidores son tomadores de decisiones adaptativos que tienen un repertorio de estrategias diferentes, muchas de las cuales pueden implicar un uso selectivo de la información y la evitación de transacciones comerciales (Chatteret et al., 2003; Gigerenzer y Goldstein, 1996; Payne et al., 1993).

Estos hallazgos plantean algunas cuestiones importantes sobre la validez del análisis conjunto. Si los modelos conjuntos no son congruentes con la forma en que la gente realmente toma decisiones, pueden producir resultados erróneos y previsiones imprecisas. Trabajos recientes también han demostrado que el uso de modelos compensatorios para representar datos conjuntos puede conducir a errores graves en la estimación de modelo si, por ejemplo, los encuestados usan puntos de corte en la toma de decisiones o si hay hábito o inercia en la toma de decisiones (Cantillo y Ortu zar, 2005; Cantillo et al., 2006; Cantillo et al., 2007). Esto significa que está en juego la validez paramorfa, es decir, la correspondencia entre una estructura de modelo y el fenómeno empírico que pretende representar (Hoffman, 1960).

Los investigadores han mostrado tradicionalmente poco interés en explorar e implementar hallazgos en la investigación de decisiones conductuales dentro del contexto del análisis conjunto (Bradlow, 2005; Swait y Adamowicz, 2001). Los hallazgos iniciales de Olshavsky y Acito (1980) indicaron que los participantes conjuntos usaron estrategias de decisión no compensatorias en lugar de las reglas aditivas asumidas.

Más recientemente, Alba y Cooke (2004) han argumentado que la investigación conjunta debería incorporar la teoría de la teoría del comportamiento para desarrollar modelos y procedimientos más plausibles. Del mismo modo, McFadden (1999) reconoce que la teoría del comportamiento ha acumulado evidencia que restringe severamente la aplicabilidad de los modelos de elección racional y pidió más investigación sobre cómo minimizar los errores de respuesta causados por las irregularidades cognitivas.

4.5.2 Diseño o Método aplicado

Un total de ocho estudiantes de sexo femenino y diez varones de pregrado, con edades comprendidas entre los 19 y los 24 años de edad sirvieron como voluntarios encuestados pagados. Los estudiantes fueron escogidos para obtener un nivel de habilidades cognitivas bastante igual entre los encuestados, lo cual es importante, ya que los factores de inteligencia general pueden ser críticos para determinar el desempeño en tareas nuevas (Kraiger et al., 1993). Cada encuestado completó dos tareas conjuntas: una para un producto familiar / de alto conocimiento y otra para un producto desconocido / poco conocido.

Los teléfonos móviles fueron elegidos como el producto de alto conocimiento, ya que este es un producto para el que los estudiantes en general son muy familiares. Además, están entre los primeros en adoptar innovaciones de productos. Por ejemplo, aunque SMS inicialmente estaba dirigido hacia el mercado de negocios, fueron los jóvenes quienes impulsaron el uso. Los procesadores de alimentos fueron elegidos como el producto de bajo conocimiento.

El análisis conjunto se ha utilizado previamente con éxito en estudios de desarrollo de productos dentro de esta clase de productos (Page and Rosenbaum, 1987), y los procesadores de alimentos son un tipo de producto para el cual se esperaba que los estudiantes expresaran poco conocimiento.

Cada encuestado recibió 14 tareas de elección para la condición de alto conocimiento (teléfonos móviles) y 14 tareas de elección para la condición de bajo conocimiento (procesadores de alimentos) en un enfoque conjunto elección. Se utilizó un diseño aleatorio basado en el software CBC / Web desarrollado por Sawtooth para generar un conjunto único de tareas de elección para cada sujeto.

La selección de atributos se basó en entrevistas en profundidad con cinco estudiantes de pregrado y una revisión de informes de consumidores. Las entrevistas se centraron en cómo el individuo procedería si compra un nuevo teléfono móvil o procesador de alimentos, que atribuye (s) que consideraría y que sería el más destacado en la elección final.

También se preguntó a los sujetos sobre su familiaridad con teléfonos móviles y procesadores de alimentos. Todos se consideraban a sí mismos como poseedores de un alto conocimiento sobre los teléfonos móviles, mientras que la familiaridad con los procesadores de alimentos era inexistente o sólo menor. Los sujetos también participaron en una pre-prueba con el fin de descartar confusiones sobre el enfoque metodológico (por ejemplo, las instrucciones proporcionadas para la verbalización) y las dificultades con la respuesta a las tareas de elección. No se informaron grandes preocupaciones en estas pruebas previas.

En cada tarea se mostraron tres conceptos de producto, además de una opción de no-compra (NPO). Se recomienda incluir un NPO para imitar mejor un entorno natural, ya que los consumidores no están obligados a elegir productos que no los satisfagan (Dhar, 1997; Swait y Adamowicz, 2001). Las tareas de elección fueron presentadas en un ordenador y administradas por los sujetos. Los sujetos trabajaron a su propio ritmo y requirieron un promedio de diez minutos y cinco segundos para completar las tareas de elección para los teléfonos celulares, y nueve minutos y 40 segundos para las opciones del procesador de alimentos ($T_{17}=0.785$, $p > 0.10$). Para evitar efectos de orden, la mitad de los sujetos recibieron la tarea de alto conocimiento primero y la mitad de la tarea de bajo conocimiento.

Al pasar por las tareas conjuntas, los encuestados fueron instruidos para "pensar en voz alta", y todo lo que se dijo fue grabado en cinta de audio y luego transcrito de los simios. El experimentador (el primer autor) estaba sentado detrás del encuestado en el otro extremo de la habitación y no era visible para él / ella. Un problema general que utiliza el análisis de protocolos verbales en la investigación del procesamiento de información es la reactividad, es decir, la verbalización puede tener un impacto en el proceso de decisión y en el resultado.

En particular, dos condiciones de la situación experimental pueden amenazar la validez (Biehal y Chakravarti, 1989). En primer lugar, debido a la limitación de la capacidad cognitiva, la verbalización puede distraer al encuestado de la tarea principal de elegir. En segundo lugar, puesto que los pensamientos y las opiniones se hacen públicos, el entrevistado puede tratar de parecer más racional y de acuerdo con las normas sociales que de otra manera; Se tomaron precauciones para evitar estos efectos. Además, puesto que la verbalización toma recursos de procesamiento, mantener las demandas cognitivas para la tarea primaria a un nivel bastante bajo puede reducir los efectos de reactividad (Russoet et al., 1989).

Por lo tanto, sólo cinco atributos fueron incluidos en el diseño conjunto. Finalmente, la instrucción de "pensar en voz alta" enfatizó que los encuestados sólo deben reportar lo que pasó por sus mentes cuando respondieron a las preguntas, y se les dijo que se concentraran en completar la tarea, la instrucción de pensar en voz alta ser secundaria. Las pruebas de reactividad (ver más adelante) sugieren que la verbalización no tuvo seriamente

Al completar la tarea los entrevistados deben verbalizar continuamente sus pensamientos; Si cae en silencio puede producir lagunas y el razonamiento incoherente en el protocolo. Por lo tanto, siguiendo las recomendaciones de Ericsson y Simon (1993, pág. 256), se usó un breve recordatorio de "seguir hablando" si el entrevistado permanecía en silencio por más de cinco segundos. Esto resultó ser un problema menor; Sólo en unas pocas ocasiones fue necesariamente para recordar a los sujetos para mantener la verbalización.

Después de que se presentó la instrucción, a los encuestados se les dieron tres problemas de práctica para familiarizarse con la tarea. Los problemas de entrenamiento fueron la elección entre pisos. Después de completar la sesión de entrenamiento, se preguntó a los encuestados si habían entendido la tarea y se sentían listos para continuar. Si no, se dieron instrucciones adicionales.

Se realizó un grupo de control para probar los efectos de reactividad. Los sujetos en el grupo de control realizaron los mismos ejercicios conjuntos como el grupo de "pensar en voz alta", pero no verbalizaron sus pensamientos al completar la tarea. El grupo de control incluyó nueve mujeres y ocho estudiantes de pregrado de sexo masculino. Al igual que en el grupo de verbalización, las tareas de elección fueron presentadas en un ordenador y administradas por los sujetos. La mitad de los sujetos recibieron la tarea de alto conocimiento primero y la mitad la tarea de bajo conocimiento.

Debido a los procesos inobservables en la condición de no-verbalización, no fue posible hacer prueba astringente de reactividad aquí. En estos casos, se pueden utilizar algunas medidas de producción para comparar grupos (por ejemplo, Biehal y Chakravarti, 1989, Biggset et al., 1985, Russoet et al., 1989). Las pruebas de reactividad se hicieron de la siguiente manera: Funciones de utilidad basadas en el conjunto combinado Los datos se estimaron por separado para el alto conocimiento y las condiciones de bajo conocimiento por separado utilizando un modelo logit multinomial (MNL). Se realizaron dos modelos para cada condición.

Un modelo de efectos principales solamente; Y un modelo de interacción donde los parámetros de utilidad se expresaron como efectos principales y como una función de las condiciones experimentales.

Se utilizó una prueba de razón de verosimilitud para comparar el ajuste de los modelos. Los resultados mostraron estadísticas de prueba significativas $D(X^2 = 24.032, P < 0.001$ para teléfonos móviles y $X^2 = 27.392, P < 0.001$. Para los procesadores de alimentos), lo que indica que no se pueden descartar efectos de reactividad. Sin embargo, la inspección adicional de los parámetros de interacción revela sólo tres parámetros significativos.

Para los teléfonos móviles, el grupo de verbalización ponía menos peso en el precio y el teléfono tenía un reproductor MP-3 que los encuestados en el grupo de control. Con respecto a los procesadores de alimentos, las diferencias se encuentran para el atributo de ruido -verbalizando los encuestados hizo hincapié en el nivel de ruido más que los encuestados en el grupo de control.

Sobre la base de estos resultados, los efectos de reactividad no pueden descartarse completamente. Sin embargo, la estructura general de los atributos es bastante similar entre los grupos, lo que indica que la verbalización no afectó seriamente el procesamiento de la información.

Como cada tarea de elección fue tratada como un protocolo separado, debíamos haber tenido 14 protocolos para teléfonos móviles y 14 para procesadores de alimentos para codificar y analizar para cada encuestado. Sin embargo, debido a problemas técnicos, a tres encuestados se les dieron sólo 12 tareas de elección para teléfonos móviles. Por lo tanto, los datos incluyeron 246 protocolos para teléfonos móviles y 252 protocolos para procesadores de alimentos.

Los análisis de los protocolos siguieron un proceso de dos etapas basado en un enfoque basado en la teoría (Glaser y Strauss, 1967). En primer lugar, los datos se dividieron en frases cortas, cada una describiendo una declaración relacionada con la tarea (por ejemplo, "definir las necesidades del usuario", "definir umbrales para los niveles de atributo aceptables", "identificar la alternativa más barata").

Un problema general de definir declaraciones relacionadas con tareas es cómo establecer puntos de ruptura ya que los puntos de interrupción tienen grandes implicaciones para la interpretación de los resultados. Biehal y Chakravarti (1989) enfatizan que para los datos verbales del protocolo es necesario examinar las frases

anteriores y siguientes para asignar etiquetas adecuadas, ya que el entrevistado a menudo no verbaliza lo suficiente para obtener una imagen clara del proceso cognitivo subyacente. Siguiendo sus sugerencias, los protocolos se leyeron por primera vez y se identificaron puntos de ruptura preliminares.

Luego se examinaron las frases anteriores y siguientes a una frase dada para ver si era necesario revisar el punto de ruptura inicial se asignó un código final. Los protocolos para los cuatro primeros encuestados fueron analizados y codificados por dos de los autores. Las comparaciones de puntos de corte y categorías mostraron un nivel muy alto de acuerdo, y la tarea de codificar los restantes protocolos se dividió entre los dos.

La primera etapa de la codificación generó 112 códigos distintos que describen las operaciones cognitivas relacionadas con las tareas. A continuación, estos 112 códigos se organizaron en categorías más amplias según sus interrelaciones. Una directriz fue que las categorías deberían reflejar algunas de las principales características de las estrategias de decisión descritas en la literatura de toma de decisiones conductuales.

Por ejemplo, encontrar la mejor opción en un atributo es un elemento de una estrategia lexicográfica; un enfoque conjuntivo implica comparar los niveles de atributos con un estándar en un proceso de procesamiento de marca, etc. Estas y otras características importantes se abordaron en la definición de las categorías. El trabajo siguió un proceso iterativo donde los códigos definidos en la primera etapa fueron comparados y elaborados. En este proceso, los dos autores trabajaron juntos. Finalmente, se definieron 13 categorías, las cuales fueron utilizadas para probar las hipótesis.

Se identificaron y codificaron un total de 3.646 frases. Para evaluar la fiabilidad de nuestra codificación, un estudiante de doctorado en psicología cognitiva sirvió como codificador "ciego". Se le indicó que codificara los 498 protocolos de acuerdo con los 112 códigos definidos en la primera etapa. El porcentaje de acuerdo, definido como la coincidencia entre las observaciones a nivel de categoría, oscilaba entre el 60 por ciento (categoría 11) y el 97 por ciento (categoría 9), con un promedio de 78 por ciento. Las frases para las que los autores y el codificador "ciego" no estaban de acuerdo fueron discutidas por ellos hasta que se acordó una categoría.

4.5.3 Resultados

Las dos tareas, la elección de teléfonos móviles y procesadores de alimentos, se supone que representan tareas de elección de alto y bajo conocimiento de los encuestados, respectivamente. Para probar la manipulación se utilizó una escala de cinco elementos para el conocimiento subjetivo (Flynn y Goldsmith, 1999). El alfa de Cronbach para la escala era 0.80 para los teléfonos móviles y 0.70 para los procesadores del alimento. Los puntajes resumidos, divididos por el número de ítems proporcionaron promedios de 4.4 para teléfonos móviles y 2.5 para procesadores de alimentos ($T_{17} = 7.472$; $P < 0.000$), lo que indica que los encuestados tenían significativamente más conocimiento del producto en la primera tarea de elección que en la segunda. No se observaron diferencias significativas entre sujetos varones y mujeres para cualquier categoría de producto ($F_{1,16} = 0.037$; $P = 0.850$ para teléfonos celulares y $F_{1,16} = 0.623$; $P = 0.442$ para procesadores de comida).

La pregunta general de investigación en este estudio se refiere al procesamiento de información de los entrevistados en tareas conjuntas. Se utilizó la siguiente combinación de categorías para medir elementos relacionados con las distintas estrategias de decisión

Tomar decisiones compensatorias basadas en alternativas (categorías 3 y 11); Compensatoria y basada en los atributos (categorías 4 y 12); la toma de decisiones no compensatorias basadas en límites (categorías 5 y 8); Decisiones no compensatorias basadas en un solo atributo (categorías 6 y 10); Y la toma de decisiones basada en la experiencia previa (categorías 1 y 9).

En promedio, los encuestados usaron 7.3 operaciones cognitivas para elegir entre las tres alternativas en cada conjunto de opciones (7.2 para la categoría de alto conocimiento y 7.4 para la categoría de bajo conocimiento, $F_{1,496} = 0.392$; $P = 0.532$). El número máximo de operaciones fue 17 y el mínimo 2. Este último caso ilustra el problema de la no verbalización en los protocolos verbales, que se discute más adelante en la sección final.

Casi un tercio del razonamiento es consistente con la toma de decisiones compensatoria basada en alternativas (aditivo ponderado e igual peso), que son las estrategias más compatibles con los modelos convencionales utilizados para analizar datos conjuntos (por ejemplo, MNL y Logit anidado).

Si se incluye la estrategia compensatoria basada en atributos (la mayoría de las dimensiones de confirmación), los resultados muestran que aproximadamente la mitad de los procesos cognitivos son consistentes con el procesamiento de información compensatoria (60 por ciento si se omite las categorías 2, 7 y 13). Los procesos no compensatorios comprenden el 29 por ciento de los razonamientos y son compartidos igualmente entre los cortes y los procedimientos de atributo único.

Aunque la cantidad de procesamiento de la información compensatoria parece algo mayor en los datos actuales que los reportados en otra parte (Denstadli y Lines, 2007, Gilbride y Allenby, 2004), los encuestados mostraron un uso considerable de procesos de decisión no compensatorios. La inspección visual de los datos de nivel individual muestra que ningún encuestado utilizó un enfoque compensatorio estricto para resolver las tareas de elección.

H1 plantea que el procesamiento de la información compatible con el uso de las estrategias de corte debe ser más frecuente cuando los encuestados poseen mucho conocimiento sobre el producto, mientras que un mayor nivel de procesamiento de información compensatoria se espera cuando el conocimiento de los encuestados es bajo). Las hipótesis fueron probadas por análisis de varianza multivariante (MANOVA), incluyendo pruebas univariadas y post-hoc multiplet-tests de diferencias entre grupos. Una variable para el orden de las tareas conjuntas también se incluyó en el análisis para controlar los efectos del pedido.

Una consecuencia de estos resultados es que los experimentos conjuntos donde los encuestados tienen un conocimiento relativamente bajo sobre el producto pueden ser más adecuados para el modelado de datos que el inverso. Ambos modelos parecen razonables (sin parámetros con signo equivocado y razonables para las variables de política) y aunque no son comparables, las mediciones Log-verosimilitud y P^2 (Ortuzar, 1982) sugieren un mejor ajuste para el modelo de procesador de alimentos. Por otra parte, Price aparece como el atributo más importante en este último caso, lo cual es consistente con estudios previos que sugieren que los consumidores tienden a confiar en atributos "no funcionales" cuando el conocimiento del producto es bajo (Selnes y Troye, 1989).

H3 propuso que en experimentos conjuntos donde los encuestados tengan mucho conocimiento sobre el producto, el procesamiento de la información debería ser más consistente que cuando el conocimiento del entrevistado es bajo. Se usaron tres medidas para probar la consistencia del procesamiento de la información. En primer lugar, se calculó el rango en el número de operaciones cognitivas utilizadas para realizar una elección, donde el rango alto indica una gran variabilidad y poca consistencia en el procesamiento de la información.

En segundo lugar, se calculó una medida para la variabilidad en el tipo de operaciones cognitivas utilizadas. Para ello, primero se calculó el porcentaje de operaciones cognitivas compatibles con diferentes estrategias de decisión para cada individuo encuestado, y luego la diferencia entre los porcentajes más altos y más bajos. Un rango alto indica una estrategia de decisión dominante y, por lo tanto, un procesamiento de información más consistente.

Para las dos primeras medidas se realizó un MANOVA utilizando como variables independientes el tipo de tarea conjunta (conocimiento alto / bajo) y el orden de la tarea conjunta. Encontramos que el tipo de tarea conjunta no era significativo $F_{2,493} = 0.256$; $P = 0.774$) no proporcionando soporte para H3. La predicción de las tareas de retención de 2 x 36 mostró una tasa de éxito del 67 por ciento para el modelo de procesador de alimentos (conocimiento bajo) y una tasa de éxito del 79 por ciento para el modelo de teléfono móvil (alto conocimiento). Sin embargo, una prueba de ji cuadrado reveló diferencia significativa no estadística entre ellos ($P < 0.10$).

Inspección de los datos de nivel individual, Aunque no se encontraron diferencias en la consistencia de la elección entre las tareas de conocimiento alto y bajo, la inspección manual de los datos reveló gran variabilidad dentro del sujeto en el procesamiento de la información. En cuanto a la variabilidad en el número de operaciones cognitivas, el rango más alto en la tarea de telefonía móvil correspondía a los encuestados 1, 12 y 18 que utilizaron un mínimo de cuatro y un máximo de 14 operaciones para hacer una elección. Los encuestados 2 y 4 tuvieron el rango más pequeño con una diferencia de cuatro. En la tarea del procesador de alimentos, cinco encuestados (5,6, 15, 17 y 18) tenían rangos de 10 o más, mientras que los encuestados 7 y 10 eran tomadores de decisiones consistentes en el sentido de usar aproximadamente el mismo número de operaciones cognitivas en cada elección tarea.

El rango promedio para los 36 casos fue de ocho. En 12 casos el rango excedió la media. Una inspección más minuciosa de estos 12 casos reveló cuatro razones de alta variabilidad dentro del sujeto. En primer lugar, en siete casos la variabilidad se debió a un mayor número de operaciones cognitivas en la primera tarea de elección, donde el encuestado señaló el problema de la decisión. Si se excluyó la primera tarea, el rango quedó por debajo de la media. La segunda razón fue el uso de un proceso de decisión en dos etapas en el que las alternativas se compararon primero con las normas de rechazo o aceptación. La variabilidad se produjo porque el procesamiento de la información podía o no ser terminado después de la primera fase, dependiendo del número de alternativas que cumplieran los criterios de selección.

Por ejemplo, en las tareas del teléfono móvil, el sujeto # 18 ponderó primero las dos características más importantes para cada alternativa (capacidad de la batería y resistencia a los golpes), y rechazó cualquiera que no cumpliera con su nivel de aceptación predefinido. Una tercera razón de la variabilidad dentro del sujeto se debió a la estrategia de cambio de decisión de los encuestados, que a menudo se atribuye a las limitaciones cognitivas, la sobrecarga de la información y la falta de información. / O efectos de fatiga. Las respuestas del sujeto # 5 a las tareas del procesador de alimentos indican fuertemente un cambio en la estrategia de decisión. En las primeras diez opciones, utilizó un promedio de 8,9 operaciones cognitivas para tomar una decisión, lo que sugiere un enfoque bastante completo.

Por otra parte, sólo se utilizaron 4,7 operaciones para realizar las cuatro últimas opciones. Una observación más cercana de los protocolos revela que la estrategia compensatoria utilizada en las diez primeras selecciones fue intercambiada por un enfoque lexicográfico en el que se identificó la alternativa de precio más bajo, se evaluó brevemente en otros atributos y se eligió.

Finalmente, la experiencia de disonancia cognitiva de los encuestados (Festinger, 1957) Contribuyeron a una alta variabilidad dentro del sujeto. En estos casos, la disonancia cognitiva fue provocada por la existencia de una marca fuerte desagradable, y desencadenó un procesamiento más amplio de la información. Los signos de disonancia cognitiva se encuentran para los temas # 1 y # 10 en las tareas de telefonía móvil.

En el procesamiento de la asignación, los sujetos # 1 y # 10 descubrieron que una marca que les disgustaba (Siemens) se desempeñaba muy bien en otros atributos (por ejemplo, la capacidad de la batería). Esto aparentemente creó tensiones cognitivas y emocionales y condujo a un mayor escrutinio de la información de atributo. La mayoría de las veces, esto llevó al rechazo de la marca de aversión. Sin embargo, en la 10ª tarea de elección, el sujeto # 1 nuevamente luchó por su actitud negativa inicial de la marca al descontar la importancia de las diferencias de atributo, pero finalmente cedió y eligió la alternativa Siemens basada en el procesamiento y la integración posteriores de información de atributo

No hay criterios exactos para definir la toma de decisiones consistentes, pero con respecto a los modelos utilizados para estimar pesos de atributo conjunto, la mayoría de las operaciones cognitivas debe ser compensatoria. Por lo tanto, el 50 por ciento representaría un nivel inferior absoluto, es decir, El procesamiento de información debe definirse y codificarse como WAD, EQW y / o MCD. Aquí, 27 de los 36 casos cumplen este criterio, 13 en la tarea de telefonía móvil y 14 en la tarea de procesador de alimentos. Elevando el nivel de aceptación al 65 por ciento, se aceptan 15 casos (siete en la tarea de telefonía móvil y ocho en la tarea del procesador de alimentos), mientras que sólo el respondedor 7 en la tarea del procesador de alimentos cumplió un criterio del 80 por ciento.

Tomados en conjunto, los datos indican alta variabilidad dentro del sujeto también con respecto al tipo de procesos cognitivos utilizados en la resolución de las tareas conjuntas. Pocos encuestados son verdaderos tomadores de decisiones compensatorios, y esto viola las suposiciones hechas por métodos convencionales para analizar datos conjuntos.

4.5.4 Conclusiones

El hecho de que las personas sean tomadores de decisiones adaptativas que poseen un repertorio de estrategias diferentes en la toma de juicios y elecciones ha sido bien documentado en la literatura de toma de decisiones de comportamiento. Sin embargo, estas cuestiones han sido descritas sólo en un grado limitado dentro de un contexto conjunto, a pesar de que los modelos conjuntos se basan en supuestos de toma de decisiones compensatorias.

Los resultados del presente estudio indican que a nivel agregado sólo el 60 por ciento de los procesos cognitivos aplicados en la toma de decisiones y elecciones son compatibles con la toma de decisiones compensatorias. Aún así, con respecto a los supuestos del modelo estas cifras son más alentadoras que los resultados reportados en otra parte (Gilbride Y Allenby, 2004, Olshavsky y Acito, 1980).

No obstante, nuestros datos revelan algunas noticias inquietantes. En tareas de alto conocimiento, es probable que los entrevistados reduzcan la información sobre estímulos a características relevantes e irrelevantes y usen niveles de corte para un desempeño aceptable / inaceptable en atribuciones importantes. El uso de cut-offs se basa en una experiencia extensa de procesamiento y uso de información del producto que no está presente cuando el conocimiento es bajo. Una de las implicaciones de esto sería que las tareas de compra caracterizadas por conocimientos relativamente bajos sobre el producto son más adecuadas para el modelado conjunto que el inverso.

Esta conclusión es contraria a la sabiduría convencional, que sostiene que el modelado conjunto puede no ser adecuado para productos nuevos o altamente modificados, ya que las preferencias obtenidas en estas condiciones pueden construirse en el momento de la medición y pueden ser fácilmente modificadas (por ejemplo, Hoeffler, 2003). Si bien este efecto podría estar presente, es probable que sea parcialmente compensado por el mayor uso del procesamiento de información compensatoria cuando el conocimiento de los encuestados es bajo.

Nuestros resultados indican una alta variabilidad dentro del sujeto con respecto a cómo se procesa y se procesa la información. Sólo unos pocos encuestados usaron estrategias completas en sus elecciones, la mayoría reunió diferentes elementos de heurística y / o cambió su enfoque durante la realización de la tarea. Así, a nivel individual parece existir una clara discrepancia entre el procesamiento de la información de los encuestados y los modelos empleados.

Muchos investigadores (por ejemplo, Green y Srinivasan, 1978) han adoptado una postura apragmática con respecto a la falta de especificación del modelo, y se ha afirmado que esto es de menor importancia siempre que el rendimiento predictivo no se vea afectado. Muchas pruebas empíricas apuntan a un desempeño predictivo relativamente alto de modelos conjuntuales.

Por ejemplo, Page y Rosenbaum (1987) e informaron de un coeficiente de correlación de 0,96 entre las cuotas de mercado predichas por la simulación conjunta y los mercados reales de los procesadores de alimentos. Del mismo modo, Natter y Feurstein (2002) compararon los datos de compras reales agregadas con las predicciones realizadas por tres diferentes tipos de simuladores de mercado basados en la elección (CBC) y encontraron coeficientes de correlación entre los pronósticos CBC y las cuotas reales de mercado en torno a 0,70 para los tres modelos. Srinivasan y Park, 1997, Louviere et al., 2000).

Por otra parte, Dawes (1979) investigó el impacto de los pesos de los parámetros sobre la capacidad predictiva de los modelos lineales y concluyó que los intentos de estimar los pesos de los atributos eran infructuosos y que la parte crucial del modelado de las opciones era la identificación de los criterios de decisión utilizados en el contexto focal. El argumento de Dawes parecería socavar la utilidad del análisis conjunto como una metodología, debemos recordar que los modelos típicamente usados en este reino son altamente no lineales.

Los primeros estudios de Einhorn (1970, 1971) y Einhorn et al. (1979) parecían apoyar la idea de que los modelos lineales son robustos frente a las violaciones de las estrategias de decisión compensatoria. Específicamente, estos estudios indicaron que los modelos lineales (consistentes con las estrategias de elección compensatoria) superaron o se desempeñaron igualmente bien como modelos que intentaron representar juicios no compensatorios (modelos parabólicos e hiperbólicos), incluso en los casos en que la presencia de elementos de estrategias no compensatorias donde Demostrado (Ogilvie y Schmitt, 1979).

Más recientemente, esta posición ha sido atacada sobre la base de que los estudios comparando el desempeño predictivo de modelos compensatorios versus no compensatorios usaron representaciones bastante pobres de procesos no compensatorios, como los modelos parabólico (que representa la toma de decisiones conjuntiva) e hiperbólicos (que representan la toma de decisiones disyuntivas) Ganzachand Czaczkes, 1995).

En particular, un estudio que incluyó evaluaciones del desempeño del personal de salud encontró que un modelo de dispersión no compensatorio ajustaba mejor los datos que las especificaciones alternativas del modelo compensatorio (Brannick y Brannick, 1989). Más generalmente, los recientes intentos de modelar la elección de un amplio conjunto de alternativas como atwo Indica que las predicciones pueden ser mejoradas cuando los modelos son más fieles a los procesos de selección subyacentes (Cantillo y Ortu zar, 2005, Giledida y Allenby, 2004; Jedidi y Kohli, 2005).

Por lo tanto, la posición que los modelos lineales pueden ser utilizados sin pérdida de rendimiento predictivo independientemente del proceso de elección subyacente parece altamente debilitable. Esto implica que, al realizar experimentos conjuntos, los investigadores deben recopilar información sobre el procesamiento de información de los encuestados para saber si los procesos seguidos son consistentes con las hipótesis de la metodología y los modelos alternativos que se van a utilizar.

Sin embargo, un desafío de modelado específico surge del hallazgo de que, en ciertas condiciones, los encuestados cambian las estrategias de decisión durante la realización de la tarea conjunta. Esto implica que diferentes representaciones de modelos serían óptimas para diferentes subconjuntos del conjunto de elecciones. Es probable que las modificaciones de la técnica que aumenten la consistencia también mejoren la capacidad predictiva de los modelos derivados, ya que el procesamiento de la información a lo largo de la tarea podría ser representado por un solo modelo (compensatorio o no compensatorio).

Los cambios en las estrategias de inducción parecían ser desencadenados por la fatiga o la interacción entre las preferencias a priori de los entrevistados y la composición de un conjunto de elecciones determinado. Los efectos de fatiga pueden explicarse por el marco costo-beneficio (Payne et al., 1993), que postula que un tomador de decisiones equilibra la precisión (beneficio) de una estrategia con el esfuerzo cognitivo necesario para ejecutarlo (costo). A lo largo de la tarea, el encuestado puede poner más peso en los costos de la estrategia elegida y menos en la precisión, y cambiar a una regla de decisión que sea menos exigente. Una forma de reducir los efectos de la fatiga y aumentar la consistencia de la elección podría ser introducir roturas durante la tarea conjunta.

Esto puede hacerse dividiendo las tareas de elección en dos o más conjuntos por preguntas, ondemográficos, actitudes u otras, o incluso haciendo que los encuestados tomen las rupturas físicas. El cambio de estrategia de decisión debido a discrepancias entre las preferencias a priori y el desempeño real en la tarea conjunta es consistente con La investigación muestra que los decisores a menudo distorsionan los criterios de elección para defender sus decisiones. Por ejemplo, Svenson (1992) argumentó que los tomadores de decisiones se esfuerzan por diferenciar los alternativos aumentando las diferencias percibidas entre ellos.

Si un criterio de decisión no proporciona suficiente diferenciación, el tomador de decisiones puede cambiar su nivel. Esto sugiere que la composición de los conjuntos de elección utilizados en experimentos conjuntos debe ser probado previamente para alternativas que son altamente inconsistentes con las preferencias de las marcas de los encuestados apriori. La supresión de alternativas atractivas para marcas desagradables o alternativas no atractivas para las marcas preferidas podría ser una forma de reducir la variabilidad dentro del sujeto con respecto a la estrategia de elección

Nuestros resultados son bastante fuertes al señalar la posible existencia de estrategias no compensatorias en la toma de decisiones conjunta. Nuestro análisis también sirve para demostrar que sería totalmente imposible incorporar elementos verbales del protocolo en un estudio conjunto a gran escala. Pero dos estrategias interesantes se sugieren: (1) añadir una nueva preponderancia a la fase preparatoria comúnmente descuidada (o al menos maltratada) de un estudio conjunto, donde la buena práctica insiste en dirigir grupos de enfoque, pre-pruebas y pilotos antes de ir En el campo; Recomendamos agregar en esta etapa un análisis de protocolo verbal con una pequeña muestra, para detectar la probabilidad de estrategias no compensatorias (y en caso afirmativo, de qué tipo) que se encuentran en la encuesta propiamente dicha (2).

Modelo, basado en el paradigma del logit mixto (ML) (Train, 2003), para permitir detectar la existencia de estrategias no compensatorias a la Cantillo y colegas para el caso de un modelo EBA generalizado. Aunque esto puede hacerse en principio, sin tener "razones" para sospechar la existencia de tales estrategias, el hecho de que el ML es tan costoso de formular y estimar, y las dificultades asociadas a la interpretación de sus resultados (véase la discusión de Sillano y Ortu zar, 2005), hace que el paquete completo, por ejemplo, Estrategias (1) y (2), una proposición más viable.

El método utilizado en el presente estudio tiene algunas limitaciones. En primer lugar, los protocolos verbales concurrentes pueden no representar rastros completos de los procesos de decisión de los encuestados. Pasando por los protocolos indicó que la verbalización no siempre mantenerse al día con lo que pasó por la mente del demandado. Un examen detallado de las frases anteriores y siguientes antes de asignar un código en cierto grado fue correcto para este problema, pero ciertos matices y detalles en el procesamiento pueden haber sido omitidos en los análisis. En segundo lugar, los resultados se basan en una pequeña muestra, que puede plantear cuestiones sobre la fiabilidad de los resultados.

Se generó una gran cantidad de datos a partir de protocolos verbales que causan límites prácticos en el número de encuestados en un estudio. En el presente estudio, los 18 sujetos generaron 496 protocolos y 3.646 frases que fueron codificadas y analizadas. Los tamaños de muestra en los análisis de protocolo verbal son típicamente pequeños. Por ejemplo, Biggset (1985) utilizó 15 encuestados, Russo et al. (1989) 24, mientras que Lussier y Olshavsky (1979) reclutaron 27 estudiantes de MBA para participar en un estudio de estrategias de elección de marca.

Por lo tanto, el número de encuestados en el presente estudio está en línea con la investigación anterior. Además, se generaron pocos códigos nuevos después de analizar los primeros diez a doce encuestados, lo que sugiere que los análisis comprenden los principales procesos cognitivos involucrados en la toma de decisiones conjunta. Sin embargo, los hallazgos de estudios con tamaños de muestra pequeños deben interpretarse siempre con cierta cautela.

Finalmente, el estudio utilizó un diseño de efectos principales que no contabilizó los posibles efectos de interacción, ya que la estimación imparcial y precisa de los pesos de los atributos no estaba en el centro del problema tratado. Sin embargo, la presencia de efectos de interacción podría haber afectado los resultados ya que reflejan cómo los encuestados perciben las relaciones entre los atributos en un estímulo.

Cuando las interacciones están presentes, la evaluación de los entrevistados del nivel de un atributo depende del nivel de otro atributo. En tales casos, el procesamiento de información puede ser más exigente que en el caso de atributos independientes. Además de intercambiar los niveles de atributo, el encuestado debe ajustar la fórmula para hacer concesiones a los niveles de los atributos involucrados, creando una tensión cognitiva adicional. Esto podría hacerle adoptar otras reglas de decisión. Por lo tanto, los efectos de interacción no medidos, si están presentes, podrían crear una varianza de error en la variable dependiente (es decir, estrategia de procesamiento de información) y reducir la probabilidad de identificar el verdadero impacto de la variable independiente.

Esta investigación ha utilizado la teoría de la decisión conductual para encuadrar y comprender los mecanismos de respuesta en experimentos conjuntos. Los encuestados conjuntos han sido descritos como tomadores de decisiones adaptativos que poseen un repertorio de diferentes estrategias para tomar decisiones y elecciones, muchas de las cuales implican un uso selectivo de la información y la evitación de las compensaciones. Esto contrasta la mayoría de las representaciones matemáticas utilizadas del modelo de datos conjuntos que suponen la toma de decisiones compensatorias.

Los resultados reportados aquí claramente sugieren que los participantes conjuntos se comportan de manera incongruente con los modelos, y más aún cuando poseen un alto nivel de conocimiento sobre el producto. En estas situaciones, los encuestados hacen uso extensivo de los niveles de corte, lo que resulta en un cambio inmediato de la probabilidad de elección de una alternativa que es difícil de capturar en modelos compensatorios.

Una recomendación general para futuras investigaciones es explorar estas cuestiones más a fin de desarrollar modelos que reflejen mejor los procesos de selección subyacentes. También en este sentido es necesario investigar cómo las estrategias de decisión utilizadas en experimentos inconjuntos se relacionan con las utilizadas en las tareas de compra del mundo real.

El análisis de elección establecido, que se utilizó en el presente estudio, es quizá el enfoque más realista dentro de la familia de técnicas conjuntas, ya que elegir una alternativa preferida en el conjunto de argumentos es similar a lo que los compradores realmente hacen en el mercado. Sin embargo, las tareas de elección hipotética nunca harán una correspondencia perfecta con el mercado natural, y la situación de decisión artificial puede alterar el procesamiento de información de los encuestados.

Todavía no está claro qué relación existe entre el proceso de información adoptado en la tarea conjunta y en las decisiones de mercado. Por lo tanto, la investigación futura debe investigar el realismo en las tareas conjuntas como percibido por los encuestados y la correspondencia entre el mundo real y la toma de decisiones conjunta. Otra área para futuros estudios es desarrollar medidas que aumenten la consistencia en la toma de decisiones. Los resultados han demostrado que los encuestados pueden cambiar las estrategias de decisión durante la realización de la tarea conjunta, lo que a su vez complica la data modelación.

Se sugirieron algunas estrategias para aumentar la consistencia de la elección, pero el efecto de estas y otras medidas aún no se ha examinado. Por último, la investigación futura debe a través de la investigación cualitativa y la adaptación del diseño conjunta perseguir la posibilidad de que la estrategia de procesamiento de la información es un resultado parcial de la interacción entre el atributo de estímulo.

4.6 Caso 6: Segmentación de Mercado con Análisis Conjunto basado en la Elección

4.6.1 Antecedentes

En los últimos años, el análisis conjunto basado en la elección (Louviere y Woodworth, 1983) ha aumentado en popularidad en la comunidad de investigación de marketing (Struhl, 1994). El análisis conjunto basado en la elección tiene varias ventajas. En primer lugar, la recopilación de datos implica decisiones simuladas de compra (opciones), que es una tarea más realista y relativamente simple para los encuestados que las calificaciones o clasificaciones. En segundo lugar, las utilidades de valor parcial derivadas reflejan el impacto en la elección del producto en lugar de un cambio en la calificación o clasificación.

Por lo tanto, las estimaciones de las partes se estiman directamente en lugar de requerir un proceso de dos etapas (estimación seguida de una simulación de elección) que requiere muchas suposiciones (tales como reglas de elección). En tercer lugar, los atributos o niveles específicos del producto (características únicas de una marca, diferentes niveles de precios para cada marca) pueden ser fácilmente acomodados y las utilidades específicas de la marca pueden ser estimadas.

En cuarto lugar, en las aplicaciones de pronóstico del mercado donde se desean estimaciones más precisas de la participación y la demanda, el análisis conjunto basado en la elección tiene la capacidad de integrar una opción sin opción. En quinto lugar, a diferencia del análisis conjunto tradicional, existe una mayor flexibilidad en los diseños que abarcan una amplia gama de contextos de elección, como elegir entre tipos de productos alternativos (pedir a los encuestados que escojan entre diferentes tipos de dispositivos de comunicación portátiles) de la misma marca (diferentes planes de llamadas telefónicas), extensiones de línea y variantes de producto (como se ilustra en nuestra aplicación empírica), y así sucesivamente.

La segmentación del mercado ha sido un objetivo primordial en la realización de análisis conjuntos tradicionales (basados en calificaciones) en aplicaciones comerciales (Wittink y Cattin, 1989). Green y Krieger (1991) indican que un enfoque post hoc para la segmentación de mercado se emplea comúnmente con el análisis conjunto tradicional en el que se estima que los valores parciales estimados a nivel individual se agrupan para obtener segmentos de compradores con preferencias similares y sensibilidad a los precios.

En el caso del análisis conjunto basado en la elección, es imposible obtener estimaciones de parámetros a nivel individual debido a la falta de observaciones adecuadas por encuestado. Por lo tanto, a diferencia de las valoraciones basadas en el análisis conjunto, no es práctico formar segmentos de mercado post hoc agrupando estimaciones de parámetros individuales (Elrod, Louviere y Davey, 1992).

La práctica corriente típica es realizar una estimación de nivel agregado (Struhl, 1994). Como señala Struhl (1994, p.38): "El análisis de nivel agregado no significa ninguna utilidad a nivel individual del tipo obtenido de conjunta y la ausencia de utilidades a nivel individual significa que la segmentación no está disponible ... Segmentación basada

Sobre las utilidades conjuntas a menudo proporciona resultados tremendamente útiles, los grupos que surgen usualmente tienen deseos y necesidades muy diferentes en relación con el producto en cuestión ". Aunque todavía se puede realizar una segmentación a priori utilizando variables auxiliares (como demográficas) o basadas en las frecuencias de elección observadas, no hay garantía de que los segmentos delineados presenten diferentes preferencias y sensibilidad a los precios (Green y Krieger, 1991).

En este artículo, se discute un método para la segmentación de mercado post hoc (basado en los co-valores) con datos conjuntos basados en la elección. El método propuesto es una generalización del enfoque de segmentación probabilística de Kamakura y Russell (1989) a un contexto experimental que incluye datos de elección recogidos en conjuntos de selección conjuntos (incompletos). Utilizando este método, demostramos cómo se puede realizar simultáneamente la segmentación del mercado y calibrar los valores parciales de nivel de segmento sin recopilar datos adicionales.

El número de segmentos, los tamaños de cada segmento y los valores de los valores parciales de nivel de segmento se obtienen utilizando únicamente los datos de elección conjunta observados. Conceptualmente, el proceso es similar a DeSarbo, Wedel, Vriens y Ramaswamy (1992), quienes realizan la segmentación del mercado y la estimación de parte del valor en el nivel de segmento simultáneamente en el contexto del análisis conjunto basado en calificaciones.

Sin embargo, su enfoque no puede aplicarse al análisis conjunto basado en la elección, que implica la recopilación de datos de elección sobre diseños de conjuntos de opciones fraccionados (incompletos) (es decir, no es necesario que cada encuestado vea todos los perfiles de producto o atributo en un diseño fraccionado) Especificación de un modelo de elección discreta (como el modelo logit) para la estimación de parámetros. Dado el llamado de modelos conjuntamente basados en la elección con los profesionales, el método propuesto extiende el dominio de aplicabilidad de estos modelos para acomodar aplicaciones de segmentación de mercado.

4.6.2 Diseño o método aplicado

En un experimento conjunto basado en la elección típica, se le pide a un encuestado que elija una alternativa de un conjunto competitivo de alternativas o conjunto de opciones, cada uno de los cuales es un perfil de diferentes combinaciones de niveles o valores de un conjunto de múltiples atributos. El encuestado repite esta tarea para un número limitado de conjuntos de elección proporcionando así datos de elección repetida. Louviere y Woodworth (1983) discuten diferentes enfoques para el diseño y generación de estos conjuntos de opciones.

Estos incluyen fijo y 2^j escoge diseños. En un diseño de conjunto de elección fija, el número de alternativas (por ejemplo, marcas) en el conjunto de opciones se fija sobre la tarea del encuestado; En otras palabras, sólo hay cambios en uno o más atributos de un número fijo de alternativas. En un diseño 2^j , el tamaño del conjunto de opciones puede variar en función de la tarea del encuestado, aunque se requiere una base constante (como una opción de no elección o algún producto de referencia) para fines de estimación.

Louviere y Woodworth (1983) discuten en detalle la implementación de tales diseños 2^j . Basado en nuestra experiencia, el diseño de conjunto de opción fija es relativamente más popular entre los practicantes debido a su simplicidad. Sin embargo, en el desarrollo de nuestro enfoque, asumimos el caso más general de un diseño 2^j ; El diseño de conjunto de opción fija es un caso especial del diseño de 2^j donde en todas las alternativas forman parte de cada conjunto de opciones presentado al encuestado para su evaluación.

El procedimiento también puede ampliarse fácilmente para tratar con los diseños de conjuntos de elección aleatoria de Johnson (1992). Como se ha comentado en la introducción, los modelos conjuntos basados en la elección suelen ajustarse a nivel agregado, resultando en un solo conjunto de coeficientes de impacto (o valores parciales) que se estima para toda la muestra de encuestados. Se propone operacionalizar la noción de segmentación de mercado asumiendo la existencia de $s = 1, \dots, S$ segmentos homogéneos, latentes de los encuestados con similar preferencia / parte de valor de las estructuras.

Denotamos los tamaños relativos de estos segmentos a través de los parámetros del segmento S $\alpha_s (0 \leq \alpha_s \leq 1, \sum_{r=1}^S \alpha_r = 1)$, que representan la proporción de encuestados pertenecientes al segmento S .

Siguiendo a Kamakura y Russell (1989), expresamos la probabilidad de elección de un encuestado en términos de las probabilidades de elección de estos segmentos S . Condicionada a un encuestado perteneciente al segmento s , la probabilidad de elección está dada por

$$P_s(j \in C_n) = \frac{\exp(\beta_{0js} + \sum_{k=1}^k X_{jk} \beta_{ks})}{\sum_{a \in C_n} \exp(\beta_{0as} + \sum_{k=1}^k X_{ak} \beta_{ks})} \quad (147)$$

Donde β_{0js} es la utilidad intrínseca del brandy j al segmento s y β_{ks} es el coeficiente de impacto del atributo k en el segmento s . Por lo tanto, mientras que los parámetros del modelo conjunto basado en la elección son específicos de cada segmento, la composición de los segmentos es latente y desconocida. La probabilidad de elección incondicional de que un encuestado elija la alternativa j se puede calcular como

$$P(j \in C_n) = \sum_{s=1}^S \alpha_s P_s(j \in C_n) \quad (148)$$

Donde α_s el tamaño del segmento s , puede interpretarse como la probabilidad a priori (o inicial) de encontrar un encuestado en el segmento s . Por lo tanto, nuestro marco de modelado implica una mezcla latente de modelos logit condicionales (Kamakura y Russell, 1989, Ramaswamy y DeSarbo, 1990, Gupta y Chintagunta, 1994) para estimar los segmentos latentes usando los datos de elección conjunta observados.

En particular, modificamos la metodología de respuesta de datos de escáner de Kamakura y Russell (1989) a la escena basada en la elección donde hay repeticiones dentro de sujetos sobre presentaciones de conjuntos de opciones (incompletas). El enfoque de Kamakura y Russell (1989) puede considerarse como un caso especial del método propuesto cuando el tamaño del conjunto de opciones (número de alternativas) se fija en las elecciones consecutivas para un individuo determinado.

Para ilustrar el método propuesto, considere una aplicación comercial del análisis conjunto basado en la elección llevado a cabo para una importante empresa de bienes envasados como parte de un nuevo concepto de prueba. La categoría de productos incluye alimentos que se compran con frecuencia, ofrecidos en forma regular y dietética, excepto para marcas locales, la mayoría de las cuales tienden a ser ofrecidas sólo en forma regular. Por razones de confidencialidad, no podemos revelar el nombre específico de la categoría de producto o las identidades de las marcas. La categoría de productos está dominada por dos fabricantes nacionales, Alpha y Beta.

Fabricante Alpha tiene dos líneas de productos separadas con identidades de marca distintas (denota como Alpha 1 y Alpha 2). Bajo cada una de estas dos marcas, Alpha ofreció versiones regulares y dietéticas del producto (designadas como R y D respectivamente). Por lo tanto, los cuatro elementos ofrecidos por Alpha se denotan como Alpha 1-R, Alpha 1-D, Alpha 2-R y Alpha 2-D. El fabricante Beta, el patrocinador del estudio, tiene una marca única (Beta 1) en esta categoría de productos que se ofreció tanto en las formas regulares y de dieta. Estos se denominan Beta 1-R y Beta 1-D, respectivamente.

El fabricante Beta estaba considerando la introducción de una nueva versión del producto dietético que capitalizó en una tecnología clave del proceso, resultando en una alternativa más saludable. Management quería introducir este nuevo producto bajo una marca distinta, denominada Beta N-D. Comparado con Alpha, Beta puso relativamente más énfasis en el nombre corporativo en su empaque y su publicidad, y aparentemente los consumidores tenían asociaciones más fuertes entre el nombre de fabricante Beta y sus productos de marca.

Para entender los compromisos de los consumidores entre la marca y el nombre del fabricante, la marca nacional y local, la forma regular y dietética y el precio, se diseñó una tarea de elección conjunta utilizando las seis marcas nacionales existentes (Alpha 1-R, Alpha 1-D, Alpha 2 -R, Alfa 2-D, Beta 1-R, Beta 2-D), una alternativa de marca local (denominada Local-R) y el nuevo concepto de producto (Beta ND).

Estos productos eran todos de un tamaño de paquete estándar. Había cuatro niveles de precios probados para cada una de las dos formas de producto junto con un precio fijo (bajo) para la etiqueta privada. El precio promedio para la forma de producto de la dieta era de \$ 1.44 comparado con \$ 1.34 para la forma regular del producto, reflejando el precio de prima típico para el anterior.

La alternativa de marca local tenía un precio único fijo de \$ 1.19. En otras aplicaciones conjuntas basadas en las elecciones, en las que el conjunto competitivo no es exhaustivo o específico de una categoría de producto, esta alternativa fija podría representar una opción "sin opción u otra opción". Al igual que en todos los modelos conjuntos basados en la elección, debe haber al menos una alternativa constante a través de conjuntos de opciones para servir como base para propósitos de estimación porque los parámetros del modelo de elección logit subyacente se determinan únicamente hasta una constante aditiva.

Se generaron dos configuraciones (factoriales fraccionales incompletos que formaron una matriz ortogonal) de dieciséis conjuntos de selección, de modo que cada conjunto de opciones contenía las ocho marcas, cada una a un nivel de precios específico. Por lo tanto, el diseño del conjunto de opciones se fijó sobre la tarea de elección de un encuestado. Cada uno de las dos factoriales fraccionarias incompletos se utilizó con un grupo separado de los encuestados que fueron evaluados para asegurarse de que eran usuarios de la categoría y calificados para el estudio.

El método de recolección de datos fue a través de entrevistas personales en varios centros comerciales. Se pidió a los encuestados que imaginaran que estaban de compras en su tienda habitual con intenciones de hacer una compra en la categoría de productos. A cada uno de los encuestados se le mostraron los dieciséis conjuntos de opciones, cada uno imitando un conjunto de marcas en el estante de la tienda. Para cada conjunto de opciones, se pidió al encuestado que seleccionara una marca al precio indicado. En total, los datos de elección conjunta implican una muestra de 600 encuestados que hacen un total de 9.600 opciones.

La práctica común es realizar un análisis agregado de los datos de elección conjunta utilizando un modelo logit condicional. En este análisis, se supone que todos los consumidores tienen la misma estructura de preferencias, y sólo se estima un conjunto único de "partes-valor" para toda la muestra. Esto es equivalente a establecer $S = 1$ en el método discutido en la sección anterior.

Alternativamente, podemos realizar análisis de nivel de segmento usando el método propuesto variando el valor de S ($S > 1$). La heurística de la CAIC fue menor para la solución de cuatro segmentos. El valor de la medida basada en entropía, E_s , fue 0,941 para la solución de cuatro segmentos. Esto sugiere que las probabilidades posteriores de pertenencia estaban muy bien separadas, y el mercado puede dividirse en cuatro segmentos casi discretos.

Los valores de probabilidad de log para el análisis agregado y la solución de cuatro segmentos fueron - 17.391 .7 (CAIC = 34.864,8) y - 14.586,3 (CAIC = 29,528.5), respectivamente. Tres de los cuatro segmentos derivados son relativamente grandes, comprendiendo el 24,2 por ciento, el 34,5 por ciento y el 34,5 por ciento de la muestra. Una idea clave es que hay un segmento pequeño (6,8 por ciento) de los encuestados para quienes el nuevo concepto es relativamente poco atractivo. Además, este segmento tampoco tiene una preferencia intrínseca por las marcas nacionales.

Los encuestados en este pequeño segmento aparentemente tienen una alta utilidad intrínseca para las marcas locales y son extremadamente sensibles a los precios. Etiquetamos este segmento pequeño como un "segmento local". Por el contrario, los encuestados del tercer segmento (34,5 por ciento) parecen tener fuertes preferencias por las marcas nacionales regulares y son menos sensibles a los precios.

Este segmento es, por lo tanto, etiquetado como un segmento regular. Para las marcas nacionales (segmento regular) y las que prefieren las alternativas de marca local (segmento local). El nuevo concepto de dieta no parece atraer a los compradores de la forma regular del producto. Por último, el segmento Beta parece ser el menos sensible a los precios. Los compradores de este segmento parecen reconocer la imagen de calidad y la identidad corporativa creada por el fabricante Beta a través de su publicidad y embalaje.

Además del nuevo concepto de producto, es evidente que las marcas de la forma dietética (Beta 1-D, Alfa 2-D, Alfa 1-D) tienen una participación relativamente grande en el segmento de la dieta, con Beta 1-D Presencia dentro del segmento Beta. Además, las marcas de la forma regular (Alpha 1-R, Beta 1-R, alpha 2-R) tienen acciones relativamente grandes en el segmento regular, mientras que la marca local tiene una posición dominante en el segmento local. También es evidente de la tabla anterior que el nuevo concepto tiene una participación mucho mayor en los segmentos Beta y dieta.

En general, el mercado parece estar estructurado de acuerdo con una combinación de nombre de marca y fabricante, forma de producto (regular versus dieta) y tipo de marca (nacional versus local). En el presente estudio, algunos datos auxiliares como la evaluación de los encuestados del nuevo concepto, el comportamiento de compra pasado (preguntas de categoría y marca específica), las actitudes hacia la salud, la aptitud, la dieta y la nutrición y las características de antecedentes como la demografía fueron seleccionados de los encuestados como parte del proceso de prueba de conceptos.

En general, mediante perfiles de los segmentos derivados usando tales datos descriptivos, se pueden obtener más información sobre su composición. En las situaciones en las que las membresías de segmentos presentan alguna confusión, las probabilidades de pertenencia posteriores pueden analizarse directamente como una función de variables descriptivas relevantes (ver Ramaswamy, DeSarbo, Reibstein y Robinson, 1993) para un análisis ilustrativo de probabilidades posteriores difusas, aunque en un contexto de modelado diferente).

En el contexto actual, dada la alta separación de los segmentos derivados, se formaron particiones discretas de la muestra asignando a cada encuestado al segmento con la mayor probabilidad posterior. Las proporciones del segmento después de las asignaciones modales de los encuestados son 0.242, 0.348, 0.343 y 0.067, que son casi idénticos a los tamaños de segmento estimados. Estos segmentos discretos fueron perfilados utilizando los datos auxiliares.

Como era de esperar, los segmentos de Beta y dieta mostraron mayores puntuaciones en la parte superior-dos para la intención de compra y singularidad del nuevo concepto. Además, el segmento Beta tuvo una alta incidencia de compra de productos de Beta, lo que es consistente con la alta utilidad intrínseca de este segmento para estos productos. En contraste, los segmentos regulares y locales fueron en realidad menos favorables al concepto de nuevo producto, con el segmento local con una incidencia muy baja de compra de productos Beta.

Los encuestados pertenecientes a los segmentos Beta y dieta también parecen tener una fuerte orientación hacia la salud y la aptitud y un compromiso básico para observar su ingesta de calorías y grasa. Esto estaba en contraste directo con los encuestados en los segmentos regulares y locales, que parecían estar menos preocupados por estas cuestiones, y este último segmento tenía aún menor preocupación. Sin embargo, sólo había una débil distinción entre los segmentos con respecto a la demografía. El segmento regular tenía encuestados con ingresos relativamente más bajos de hogares algo mayores, mientras que el segmento local parecía estar formado por "nidos vacíos" con ingresos mayores a la media.

En general, los segmentos tuvieron evaluaciones distintas del nuevo concepto, las actitudes hacia la aptitud y la nutrición, las preferencias por el nuevo concepto y la conducta pasada de categoría y marca que parecía ser consistente con el esquema de segmentación derivado, otorgando cierta validez a los resultados. Dada la naturaleza del producto, no era sorprendente encontrar sólo débiles diferencias en la demografía.

En general, se ha encontrado que las características de los antecedentes, como la demografía, están débilmente asociadas con las preferencias de los atributos (Gupta y Chintagunta, 1994, Moore, 1980), y otros datos de antecedentes como la importancia percibida y la adecuación en el uso de diversas situaciones de uso pueden proporcionar mejores perspectivas para los productos relacionados con los alimentos. Sin embargo, a priori, puede no ser obvio qué conjunto de variables de fondo debería formar la base de los segmentos para permitir la estimación desagregada de modelos conjuntos basados en la elección. Si bien puede ser difícil utilizar variables de fondo para formar segmentos a priori, otra opción es agrupar a los encuestados basándose en las frecuencias de elección observada.

4.6.3 Resultados

Utilizamos un análisis de agrupamiento de K-means de las frecuencias de elección observada para formar cuatro segmentos. Después de dividir la muestra en estos cuatro segmentos a priori, se estimó un modelo conjunto separado basado en la elección para cada segmento. En primer lugar, la sensibilidad al precio no varía entre los cuatro segmentos a priori. En segundo lugar, dentro de cada segmento, las marcas dietéticas tienen utilidades de marca intrínseca más grandes que las marcas regulares; Este ordenamiento es similar al caso agregado. En tercer lugar, los cuatro tamaños de segmento del análisis de clúster de K-means son casi iguales.

En general, los resultados a priori a nivel de segmento no muestran ninguna variación llamativa a través de los segmentos y son similares a las estimaciones de nivel agregado. ¿Qué pasa con las acciones de elección dentro de cada segmento? Si bien se podría esperar una variación en las proporciones observadas dentro del segmento (dado que el análisis de conglomerados se basó en las frecuencias de elección individual observadas), no encontramos prácticamente ninguna variación en las probabilidades de elección condicional estimadas en el nivel del segmento.

Dado que el análisis de conglomerados y el análisis conjunto basado en la selección dentro del segmento no maximizan la misma unión objetiva en el caso a priori, no hay garantía de que los segmentos a priori presenten diferentes utilidades de marca intrínseca y sensibilidad a los precios. La agrupación de las frecuencias de elección da lugar a segmentos a priori con menor variación en las elecciones de marca y aumenta las posibilidades de obtener parámetros insignificantes. De hecho, la probabilidad de registro basada en los resultados de segmentación a priori fue de - 17.474,7, que no es mejor que el caso agregado.

La congruencia entre la clasificación a priori de los encuestados y la asignación post hoc de segmentos discretos (an de la permutación apropiada) es sólo el 57 por ciento. Además, como señalan DeSarbo, Wedel, Vriens y Ramaswamy (1992), diferentes procedimientos de agrupamiento pueden conducir potencialmente a diferentes esquemas de segmentación a priori y, consecuentemente, a diferentes resultados de nivel de segmento.

El objetivo fundamental de la segmentación del mercado con el análisis conjunto es capturar la heterogeneidad del mercado en las preferencias de marca y atributo y la sensibilidad a los precios. En el análisis conjunto basado en las clasificaciones tradicionales, esto se logra a través del potencial de llevar a cabo una estimación individual y luego agrupar los valores de parte individuales resultantes para formar segmentos. En una configuración conjunta basada en la elección, es ineficiente estimar parámetros individuales con un número limitado de opciones. El enfoque propuesto es una opción atractiva para llevar a cabo la segmentación del mercado con un análisis conjunto.

4.6.4 Conclusiones

Hemos discutido una aproximación para la segmentación del mercado con el análisis conjunto basado en la elección. La incapacidad para derivar segmentos de mercado ha sido hasta ahora una desventaja primaria del análisis conjunto basado en la elección (Struhl, 1994). Superamos esta desventaja al acomodar la heterogeneidad del consumidor a través de múltiples segmentos de consumidores que poseen diferentes estructuras de utilidad. Un beneficio importante del método propuesto es que los segmentos se derivan empíricamente utilizando datos de elección conjunta recogidos comúnmente sin recurrir a ninguna especificación a priori de la naturaleza de la heterogeneidad (normalmente basada en datos de antecedentes auxiliares).

El número de segmentos, sus tamaños y estructuras de utilidad se derivan empíricamente, en un marco de máxima verosimilitud. Hemos demostrado el método propuesto utilizando datos comerciales de elección conjunta recogidos en un nuevo concepto de prueba. Nuestro enfoque, sin embargo, no se limita a este tipo de aplicación y puede utilizarse en otros entornos que impliquen diferentes diseños. En general, nuestro enfoque es aplicable en cualquier caso en el que se hayan recopilado datos de elección conjunta, aunque, tal como se discutió, los investigadores pueden estar analizando los datos en el agregado o en los segmentos a priori.

Claramente, es necesario realizar más pruebas del enfoque para establecer su utilidad para los profesionales. Dado que no se recogieron perfiles de validación en la aplicación comercial, no se pudieron realizar análisis de retención para comparar las predicciones de reparto basadas en nuestro enfoque versus un enfoque de nivel agregado.

Hay algunas pruebas empíricas de la literatura de modelos de elección que demuestran que los modelos que capturan la heterogeneidad, particularmente a través de múltiples segmentos de mercado, tienen mejores resultados que los modelos de nivel agregado en términos de predicción tanto de los porcentajes de elección agregada como individual (Ramaswamy y DeSarbo, 1990). Sería fructífero realizar una evaluación de validación cruzada de nuestro enfoque y otras alternativas a la segmentación del mercado (diferentes especificaciones a priori de heterogeneidad), similar en espíritu a Green y Helsen (1989) que comparan alternativas a las calificaciones individuales, Basado en el modelo conjunto.

4.6.5 Referencias Bibliograficas

- Acito, F., & Jain, A. K. (1980). Evaluation of Conjoint Analysis Results: A Comparison of Methods. *Journal of Marketing Research*, 17(1), 106-112.
- Camm, J. D., Cochran, J. J., Curry, D. J., & Kannan, S. (2006). Conjoint Optimization: An Exact Branch-and-Bound Algorithm for the Share-of-Choice Problem. *Management Science*, 52(3), 435-447.
- Castel, D., & Tsuji, H. (2015). Design of interactive conjoint analysis Web-based system. *International Journal of Web Information Systems*, 11, 17-32.
- Chrzan, K. (2011). Rationale and design of moderated choice experiments: a strategic pricing application". *Journal of Product & Brand Management*, 10, 293-300.
- Desarbo, W. S., Ramaswamy, V., & Cohen, S. H. (1995). Market Segmentation with Choice-Based Conjoint Analysis. *Marketing Letters*, 6, 137-147.
- Ortúzar, J. M. (2012). Information processing in choice-based conjoint experiments A process-tracing study. *European Journal of Marketing*, 46, 422-446.

Diseño de experimentos y análisis de conjunto

Autor: Christian Rosero Barzola, Ph.D.

ISBN: 978-9942-960-59-7

