

# UNIVERSITAT JAUME I

Escuela Superior de Tecnología y Ciencias Experimentales.

DEPARTAMENTO DE MATEMÁTICAS

Trabajo Final de Máster:

## **DISEÑO DE ESCENARIOS PARA LA OBTENCIÓN DE PREFERENCIAS DECLARADAS**

Una trabajo final de máster, realizado por Maria Soler Martinez  
para la obtención del título de MÁSTER EN MATEMÁTICA  
COMPUTACIONAL.

---

Supervisado por:  
Irene Epifanio López



## Agradecimientos

A mi tutora Irene Epifanio López, por la paciencia y dedicación ofrecida.

A la empresa donde pude desempeñar este proyecto, Outcomes'10, en especial a Irmina Gozalbo por abrirme un mundo en el que la herramienta excel puede con todo.

Al departamento de matemáticas, por enseñarme y guiarme durante este año de máster.

*La duda es el principio de la sabiduría (Aristóteles)*

# Diseño de escenarios para la obtención de preferencias declaradas

Maria Soler Martinez

Universidad Jaume I

**Resumen** El objetivo principal del presente trabajo final de máster (TFM) es el diseño efectivo de encuestas para obtener preferencias sobre tratamientos, estas serán medidas mediante un experimento de elección discreta. Las herramientas utilizadas serán, programación en Excel (Visual Basic), método de elección discreta (discrete choice) y diseño factorial.

Se pretende estudiar las preferencias de las características que son valoradas por el consumidor a la hora de comprar un tratamiento para la cura o mejora de su enfermedad, o las características que valoran los médicos a la hora de recetar un tratamiento o medicamento (en este estudio generalizaremos la utilidad, ya que así podemos obtener una perspectiva mucho más amplia). Para ello se utilizarán encuestas, las cuales estarán diseñadas con el número óptimo de preguntas de elección, y el diseño del experimento deberá cumplir unas premisas para garantizar que se trata de un buen modelo de escenarios de elección discreta. De este modo se podrán obtener resultados fiables sobre las preferencias de los consumidores.

A lo largo del trabajo veremos cómo programar en Visual Basic una calculadora Excel (Macros) que constará de dos partes. La primera, en la cual introduciendo el número de atributos (características del tratamiento a valorar) y sus niveles (categorías de cada una de las características), nos muestra el número de escenarios óptimo. Por otro lado, la segunda parte de la calculadora, nos mostrará un posible diseño de estos (este no es único).

Para ello estudiaremos, en la sección 1, los experimentos de elección discreta, en búsqueda de indicaciones para encontrar el mejor diseño de escenarios para realizar dichos experimentos (Discrete choice), además introduciremos el modelo probabilístico que se esconde detrás de un Discrete choice. Por otra parte, veremos el diseño factorial, ya que será el método más eficiente para el estudio del efecto de cada atributo sobre la variable respuesta, en nuestro caso, las preferencias (de este modo, se puede estudiar eficientemente las preferencias de los encuestados sobre los tratamientos). Utilizaremos una variante de este, llamado diseños factoriales fraccionados, con el propósito de buscar el número mínimo de dimensiones capaces de explicar el máximo de información.

Otro problema al que nos enfrentaremos será, además de obtener las pautas para la minimización y el diseño eficiente de escenarios de calidad para estudiar las preferencias, la programación de nuestros objetivos en el lenguaje BASIC. Veremos que se trata de un lenguaje sencillo, que mediante el uso de bucles y su conjunto de herramientas nos facilitará la construcción de instrucciones necesarias para nuestro propósito.

Por lo tanto, este estudio se centrará en intentar encontrar las claves para el diseño de un buen experimento de elección discreta, en el que se tendrá en cuenta tanto la carga cognitiva como la configuración de conjuntos de elección con un buen diseño.

# Índice general

Resumen . . . . .	2
Introduction . . . . .	5
1. Experimentos de elección discreta . . . . .	9
1.1. Introducción . . . . .	9
1.2. Modelo probabilístico . . . . .	9
1.3. Modelos según la distribución de probabilidad . . . . .	12
Modelo lineal de probabilidad. . . . .	12
Modelos no lineales. . . . .	13
Modelo Probit y Logit. . . . .	14
Algunos parámetros estimados en un modelo Logit: . . . . .	18
Contraste de hipótesis: . . . . .	21
1.4. Diseño de escenarios . . . . .	22
Preferencias declaradas . . . . .	23
Diseño experimental . . . . .	24
1.5. Conclusiones . . . . .	25
2. Diseño factorial completo . . . . .	27
2.1. Introducción . . . . .	27
2.2. Modelos de un experimento factorial . . . . .	29
Análisis de la varianza . . . . .	35
2.3. Diseños $2^k$ factorial . . . . .	38
Diseño factorial $2^2$ : . . . . .	38
Diseño factorial $2^3$ : . . . . .	42
Diseño general $2^k$ : . . . . .	45
2.4. Diseño factorial $3^k$ . . . . .	47
Diseño factorial $3^2$ : . . . . .	47
Diseño factorial $3^3$ : . . . . .	48
Diseño general $3^k$ : . . . . .	49
3. Diseño factorial fraccionado . . . . .	51
3.1. Introducción . . . . .	51
3.2. Diseños factoriales fraccionados de dos niveles: . . . . .	52
Fracción un medio del diseño $2^k$ : . . . . .	52
Fracción un cuarto del diseño $2^k$ : . . . . .	55
Diseño factorial fraccionado $2^{k-p}$ : . . . . .	56
3.3. Diseño factorial fraccionado $3^{k-p}$ . . . . .	58
3.4. Conclusiones . . . . .	60
4. Programación Visual Basic . . . . .	62
4.1. Número de escenarios de elección . . . . .	63
Obtención del número de escenarios para el diseño factorial $2^k$ :	

Obtención del número de escenarios para el diseño factorial $3^k$ :	
65	
Obtención del número de escenarios para el diseño factorial general:	
65	
Programación en Visual Basic para la obtención de número de escenarios, con el mismo número de niveles para cada atributo.	
66	
Obtención del número de escenarios para diseños mixtos con 2 y 3 niveles.	
68	
Obtención del número de escenarios para diseños factoriales mixtos.	
70	
Programación en Visual Basic para la obtención del número de escenarios, con distintos niveles para los atributos.	
71	
4.2. Diseño de los escenarios de elección	
78	
4.3. Estructura de presentación de la calculadora . . . . .	97
5. Conclusiones . . . . .	100

## Introducción

El objetivo principal de este texto es el diseño de escenarios o encuestas para el estudio de preferencias. Para ello se indagará en las técnicas de análisis conjunto, en los diseños de experimentos y además se realizará un pequeño estudio de las técnicas estadísticas, para la obtención de resultados óptimos en estos métodos, de esta manera ampliar el punto de vista más allá del diseño de tales escenarios.

El análisis conjunto (AC) es una técnica de investigación utilizada tanto en el ámbito académico como empresarial para analizar las preferencias declaradas de los consumidores sobre nuevos productos y servicios que se prevé lanzar al mercado. Se trata de un método cualitativo de cuantificación de preferencias. A grandes rasgos, el análisis de conjunto tiene como objetivo averiguar qué combinación de entre un número determinado (es limitado) de atributos es el preferido por los encuestados, mediante, en este caso un experimento de elección discreta, es decir, un proceso de elección en el cual se evalúan diferentes conjuntos de perfiles de productos hipotéticos descritos por sus características, a las que llamamos atributos. Cada perfil o conjunto de elección, se configura mediante la combinación de atributos siguiendo un patrón denominado diseño experimental. Este diseño es variable, depende del número de atributos que se quieran considerar y las variaciones de cada uno de ellos, a las que llamaremos niveles de cada atributo. Por consiguiente, cuanto mayor sea el número de atributos y sus niveles, mayor será el número de escenarios o perfiles que deberán evaluar los encuestados.

Para poder aplicar las técnicas de análisis conjunto, existen diversos modelos que nos encaminan hacia el buen diseño de experimentos. En nuestro caso, optaremos por una clase de diseños experimentales de preferencias declaradas llamados modelos de elección discreta o más conocidos como Discrete choice. Se ha optado por un experimento de elección debido a que son los mejores a la hora de pronosticar el comportamiento de los usuarios, ya que ellos finalmente deciden qué bienes o servicios van a consumir, estos modelos además son capaces obtener los efectos cruzados entre los atributos, cosa muy difícil de conseguir mediante la observación directa.

La aparición de los modelos de elección discreta ha permitido el desarrollo de análisis de mercados, estimando los valores de los factores principales (atributos), y estudiando los efectos que generan las interacciones entre dos o más factores.

Para poder estimar las interacciones entre dos o más atributos se requiere de diseños con un número muy elevado de alternativas, esto implica un gran esfuerzo por parte de los encuestados, lo que genera efectos negativos en la estimación de modelo, pues esta carga cognitiva puede perjudicar a la toma de decisiones e inducir a los usuarios a no utilizar un criterio durante el proceso de selección, lo

cual implica un error en los datos y distorsiona el diagnóstico. Sin embargo, con un número amplio de escenarios se obtiene con mayor probabilidad la opción que se ajusta a las preferencias de los usuarios.

En el contexto de la obtención de preferencias, se entiende que el individuo se enfrenta a un proceso de elección donde cada una de las opciones, por sí misma, genera un grado de utilidad que la convierte en más o menos preferida (se considera que el individuo tiene la capacidad de calcular la maximización de esta utilidad). Un amplia gama de opciones puede reducir el grado de vacilación en la elección, ya que le estamos proporcionando al usuario mayor información. Por ello, excluir opciones dentro del conjunto de alternativas de elección sería equivalente a descartar datos que puedan ser relevantes en el proceso de elección.

Sin embargo, en la literatura encontramos trabajos que destacan los efectos negativos que se derivan de ofrecer un elevado número de opciones a los individuos. Por una parte, la presencia de muchos atributos reduce la cantidad de información disponible para cada atributo, y por otra parte, se considera que los individuos disponen de una racionalidad limitada, por lo que no pueden abarcar y retener toda la información disponible necesaria en un proceso de toma de decisiones conforme el número de alternativas va incrementando.

Por tanto, deducimos que para calcular el número de escenarios de elección óptimo tendremos que incluir el mayor número de atributos posible pero con un número mínimo de conjuntos de elección. Este será uno de los problemas que se nos presentará a lo largo del trabajo, y prestaremos mucha atención, ya que el tamaño de elección es determinante en la generación de error puesto que cualquier proceso de elección que realiza un encuestado es el resultado del equilibrio entre el coste generado por el esfuerzo mental de tomar la decisión y el beneficio que le aporta esta.

En la sección 1.4 veremos los tipos de errores que se cometen en los experimentos de elección discreta, pero nosotros nos centraremos mayoritariamente en dos. Por un lado, los errores derivados de las características personales de los encuestados (carga y capacidad cognitiva). Y por otro lado, los errores derivados de las características del diseño experimental (número de escenarios, combinaciones de los niveles de cada atributo).

Nos ayudaremos de dos técnicas para lograr nuestros objetivos, los diseños de experimentos de elección discreta y los diseños factoriales.

Estudiaremos los experimentos de elección discreta, además de introducirnos en el diseño de escenarios también veremos el modelo probabilístico y algunos modelos estadísticos para la interpretación de la utilidad, estos son a posteriori del diseño del experimento, utilizados para el análisis de los resultados y la obtención de las preferencias valoradas en las encuestas. Los modelos de elección

discreta son muy convenientes a la de analizar preferencias basados en una muestra de individuos a partir de encuestas, estas dan mejores resultados para evaluar las características del objeto de estudio, son más económicas en el sentido que el número de encuestas es menor al de otros modelos y permite diseñar un conjunto eficiente de situaciones de intercambio para evaluar las preferencias. El diseño de estos experimentos es una teoría muy utilizada en el campo econométrico, ya que como hemos comentado, es la clave para la obtención de resultados óptimos.

Por otro lado, para minimizar el número de escenarios o conjuntos de elección investigaremos en los diseños factoriales. Es el nombre que se le da a unos modelos que tiene como objetivo determinar la significación de los efectos principales y de las interacciones entre atributos, así de este modo estudiar el efecto de cada atributo y sus interacciones sobre la variable respuesta. Estos son una técnica de minimización de dimensiones, ya que buscan el número mínimo de dimensiones capaces de explicar el máximo de información contenida en los datos. En particular, utilizaremos un tipo de diseños derivados de estos diseños llamados diseños factoriales fraccionados. Un diseño factorial fraccionado es un diseño factorial en el cual solo se efectúa un subconjunto seleccionado, presuponiendo que los efectos de orden superior son no significativos.

Concluiremos al obtener las claves para el cálculo del número de escenarios y el diseño de los conjuntos de elección teniendo en cuenta el número de atributos y sus niveles como variables. Una vez obtenidas estas premisas veremos cómo programarlas en Visual Basic, para así construir una calculadora en excel, en la que una vez introducidos los atributos y los niveles del tratamiento del que queremos estudiar las preferencias, nos dará el número de escenarios necesarios y un posible diseño de estos.

## 1. Experimentos de elección discreta

### 1.1. Introducción

Los experimentos de elección tienen una base teórica en el modelo de Lancaster (1966) que deriva la utilidad a las características de los bienes y no a los bienes en sí como postulaba la teoría económica clásica, y tienen una base econométrica en los modelos de utilidad aleatoria (Luce (1959) y MacFadden (1974)).

El modelo de elección de Lancaster supone que el consumidor demanda bienes en virtud de sus características, y son estas las que generan utilidad. Estos estudios basados en el comportamiento de los individuos, son una buena herramienta para el análisis y predicción de la demanda.

Por su parte, la teoría de la utilidad aleatoria parte de que el individuo perfectamente racional siempre opta por la alternativa que supone de mayor utilidad. MacFadden (1974) los denominaba modelos del comportamiento, ya que, a diferencia del enfoque tradicional, no están basados en una visión descriptiva de la demanda, sino más bien tratan de representar el comportamiento de los individuos de forma explícita. A la hora de interpretar estos modelos los economistas se decantan por MacFadden (1975), este establece que la función de utilidad puede expresarse como la suma de dos componentes, la observable y la no observable, esta última de naturaleza aleatoria.

En general, los experimentos de elección presentan al usuario un conjunto de alternativas de elección, donde cada alternativa tiene diferentes combinaciones de niveles de mejora para cada atributo. Estas combinaciones óptimas se derivan de un diseño experimental, el cual estudiaremos en la sección 1.4. Estos diseños permiten jerarquizar y conocer el atributo mejor valorado.

En particular, los modelos de elección discreta permiten la modelización de variables cualitativas (haciendo uso de las técnicas de las variables discretas, donde se miden las cualidades).

Dentro de los modelos de elección discreta se distinguen los modelos de respuesta dicotómica y modelos de elección múltiple, según el número de alternativas de la variable endógena (la que se va a predecir). Y a la vez, dentro de estos, se distinguen varios modelos, según la función utilizada para la estimación de la probabilidad: el modelo de probabilidad lineal truncado, el modelo logit y el modelo probit.

### 1.2. Modelo probabilístico

En este apartado estudiaremos la interpretación de los modelos de respuesta dicotómica.

Existen varios enfoques para la interpretación de los modelos de elección discreta, nosotros nos centraremos en interpretar estos modelos de respuesta dicotómica bajo la teoría de la utilidad aleatoria, donde la alternativa que maximice la utilidad esperada en cada caso será la seleccionada.

En los modelos de utilidad aleatoria se parte de un individuo perfectamente racional que siempre opta por la alternativa que le supone de mayor utilidad frente a su alternativa complementaria (la decisión del individuo será entre dos alternativas excluyentes -la 1 o la 0- ). Si denominamos a la utilidad que aporta la alternativa  $j$  al individuo  $n$  como  $U_{nj}$ , entonces,  $U_{ni} > U_{nj}; \forall i \neq j$ , expresará que la alternativa  $i$  es más útil que la alternativa  $j$ , y esta será la escogida por el individuo  $n$ .

En el caso que se nos presenta, donde el número de alternativas de elección es 2 y además una es excluyente de la otra, tenemos que la formulación del modelo parte de la utilidad derivada de la elección  $U_{n0}$  o  $U_{n1}$ , en función de las variables explicativas de dicha elección (es decir, tanto de las características propias de cada una de las alternativas como de las características personales propias del individuo que ha de elegir). Suponiendo linealidad en las funciones se tiene:

$$U_{n0} = x'_{n0}\beta + z'_n\gamma + \varepsilon_{n0} \quad (1)$$

$$U_{n1} = x'_{n1}\beta + z'_n\gamma + \varepsilon_{n1} \quad (2)$$

Donde,  $\varepsilon_{n0}$  y  $\varepsilon_{n1}$  se refieren a las desviaciones en el comportamiento del individuo debidas a factores aleatorios. El conjunto de variables  $z_n$  recoge indicadores como género, localización, etc., que miden la heterogeneidad. Por otra parte, se encuentran los parámetros  $\beta$  y  $\gamma$  que llamaremos parámetros de preferencias (estadísticos de inferencia).

Así, si la utilidad de la alternativa 1 es mayor, la elección del individuo es 1, y observamos que se tiene:

$$U_{n1} > U_{n0} \quad (3)$$

$$x'_{n1}\beta + z'_n\gamma + \varepsilon_{n1} > x'_{n0}\beta + z'_n\gamma + \varepsilon_{n0} \quad (4)$$

$$(x'_{n1}\beta - x'_{n0}\beta) + (z'_n\gamma - z'_n\gamma) > \varepsilon_{n0} - \varepsilon_{n1} \quad (5)$$

$$(x'_{n1} - x'_{n0})\beta > \varepsilon_{n0} - \varepsilon_{n1} \quad (6)$$

(por defecto si dos alternativas son iguales en utilidad se escoge la alternativa 0,  $\gamma$  no aporta diferentes características en utilidades diferentes).

Además, tenemos que el modelo convencional para dos alternativas de elección cumple:

$$\Delta U_n = \mu_1 + (\Delta x_n)' \beta + z'_n(\gamma_1 - \gamma_0) + (\varepsilon_{n0} - \varepsilon_{n1}), \text{ ya que } \mu_0 \text{ es normalizada a } 0$$

$$E[\varepsilon_{n0} - \varepsilon_{n1} \mid X_n, z_n] = 0$$

$$\text{y } Var[\varepsilon_{n0} - \varepsilon_{n1} \mid X_n, z_n] = 1$$

La varianza igual a 1 es resultado del proceso de normalización, ya que el modelo no se altera porque se multiplique por cualquier cantidad.

Como habíamos mencionado anteriormente, si el individuo  $n$  elige la alternativa 1 su utilidad supera a la utilidad de la alternativa 0, y viceversa, de este modo, podemos modelizar el problema que nos aborda utilizando una combinación de las ideas que hemos ido mencionado hasta ahora.

Así, el problema

$$d_n = \begin{cases} 1 & \text{si } U_{n1} > U_{n0} \\ 0 & \text{si } U_{n0} > U_{n1} \end{cases} \quad (7)$$

Donde  $d_n$  es la elección del individuo  $n$ . Lo podemos expresar

$$d_n^* = \alpha + x_n' \beta + z_n' \gamma + \varepsilon_n$$

$$d_n = \begin{cases} 1 & \text{si } d_n^* > 0 \\ 0 & \text{, en otro caso} \end{cases} \quad (8)$$

Sin asumir ninguna distribución específica, la probabilidad del modelo dicotómico quedará definida por

$$\begin{aligned} Prob(d_n = 1 \mid X_n, z_n) &= Prob(d_n^* > 0 \mid x_n, z_n) \\ &= Prob(x_n' \beta + z_n' \gamma + \varepsilon_n > 0) \\ &= 1 - Prob[\varepsilon_n \leq -(x_n' \beta + z_n' \gamma)] \\ &= Prob(\varepsilon_n \leq x_n' \beta + z_n' \gamma) \\ &= F(x_n' \beta + z_n' \gamma) \end{aligned}$$

Todo esto se tiene, asumiendo una distribución asimétrica, tales como la normal y la logística.

Según sea la función distribución,  $F(x'_n\beta + z'_n\gamma)$ , asociada a  $\varepsilon_n$ , se obtienen varios modelos de experimentos de elección dicotómica. Si la función sigue una distribución uniforme, estaremos hablando del Modelo Lineal Truncado, si la función de distribución se trata de la normal tipificada, el Modelo Probit y, por otro lado, si la función de distribución es la curva logística, lo llamaremos Modelo Logit.

Del modelo de probabilidad obtenido, se sigue:

$$\begin{aligned} E[d_n | x_n, z_n] &= 0 \times Prob(d_n = 0 | x_n, z_n) + 1 \times Prob(d_n = 1 | x_n, z_n) \\ &= Prob(d_n = 1 | x_n, z_n) \end{aligned}$$

### 1.3. Modelos según la distribución de probabilidad

**Modelo lineal de probabilidad.** En cuanto al Modelo Lineal de probabilidad, se plantea como una extensión del modelo lineal general (con una simple manipulación del modelo de regresión), no entraremos a explicarlo con detalle, ya que este tiene muchas limitaciones implícitas, como problemas de heterocedasticidad (varianza no constante), la estimación por mínimos cuadrados no garantiza que sus valores estén entre 0 y 1 (este problema se puede solucionar truncando el rango de variación) y un valor limitado en el coeficiente de correlación lineal  $R^2$ .

De los resultados obtenidos en el apartado anterior, tenemos:

$$d_n = E[d_n | x_n, z_n] + a_n = Prob(d_n = 1 | X_n, z_n) + a_n$$

Donde asumimos que  $a_n$  es una variable aleatoria que se distribuye como una normal de media cero, es decir,  $a_n \sim N(0, \sigma^2)$ . De este modo el modelo lineal de probabilidad puede ser sugerido como el que sigue.

$$d_n = x'_n\beta + z'_n\gamma + a_n \quad (9)$$

Donde, la distribución que sigue  $d_n$  se trata de una Bernoulli.

$$d_n \sim Be(Prob(d_n = 1 | x_n, z_n)) \quad (10)$$

Este modelo asume que los incrementos en la probabilidad son constantes, algo poco realista, en la mayoría de los casos.

Así pues, el Modelo Lineal de Probabilidad nos será útil si nuestro objetivo es aproximar los efectos parciales de las variables sobre todo el rango de los regresores, sin embargo, de este modo no podemos observar la reacción de la variable endógena ante los cambios de las variables explicativas, en particular para aquellos valores más extremos.

**Modelos no lineales.** Un modelo paramétrico se completa especificando una distribución para  $\varepsilon_n$ .

Nosotros asumiremos una distribución simétrica, tal como la normal o la logística. De este modo, asumiendo simetría podemos decir que las probabilidades asociadas a nuestros dos casos serán:

$$Prob(d_n = 1 | x_n, z_n) = F(x'_n\beta + z'_n\gamma) \quad (11)$$

y

$$Prob(d_n = 0 | x_n, z_n) = F[-(x'_n\beta + z'_n\gamma)] \quad (12)$$

Dadas las dificultades que comprende el modelo lineal de probabilidad, se intentarán hacer transformaciones para que las predicciones caigan dentro del intervalo  $[0,1]$ .

Debería adoptarse un modelo en el que los valores de  $Prob(d_n = 1 | x_n, z_n)$  estén restringidos al intervalo  $[0,1]$ , una forma de hacerlo es tomando  $F()$  como una función distribución acumulada.

El modelo no lineal será el siguiente:

$$d_n = F(x'_n\beta + z'_n\gamma) + \varepsilon_n$$

con

$$\varepsilon_n = E[d_n | x_n, z_n] - F(x'_n\beta + z'_n\gamma)$$

Donde la función  $F(x'_n\beta + z'_n\gamma)$  debe de ser una función no lineal, en la cual las variables explicativas afecten a la variable endógena mediante un índice lineal y este luego es transformado mediante la función para que los valores de la probabilidad sean consistentes.

Como hemos comentado en varias ocasiones, tomaremos como posibles alternativas a la función distribución acumulada uniforme del modelo lineal de probabilidad, la normal y la logística, ambas acotadas entre 0 y 1. Cada una de ellas da como resultado un modelo, el modelo Probit y el Logit, respectivamente.

**Modelo Probit y Logit.** Como acabamos de ver, los problemas del modelo lineal de probabilidad se acaban sustituyendo la función de distribución uniforme por una función que toma valores estrictamente entre 0 y 1, y así se garantiza un resultado en la estimación comprendido en este rango,  $[0,1]$ . Este requisito lo cumplen las funciones de distribución, ya que son funciones continuas que toman valores comprendidos entre 0 y 1. Además, buscamos una función de distribución asimétrica.

Bajo estas condiciones nos encontramos con dos modelos que se adecuan a nuestras necesidades:

1. El Modelo de probabilidad Probit, que se asocia a la función distribución normal acumulada.

$$F(c) = \Phi(c) = \int_{-\infty}^c \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

donde

$$z \sim N(0, 1)$$

2. El modelo de probabilidad Logit, que se asocia a la función distribución logística.

$$F(c) = \Lambda(c) = \frac{1}{1 + e^{-c}} = \frac{e^c}{1 + e^c}$$

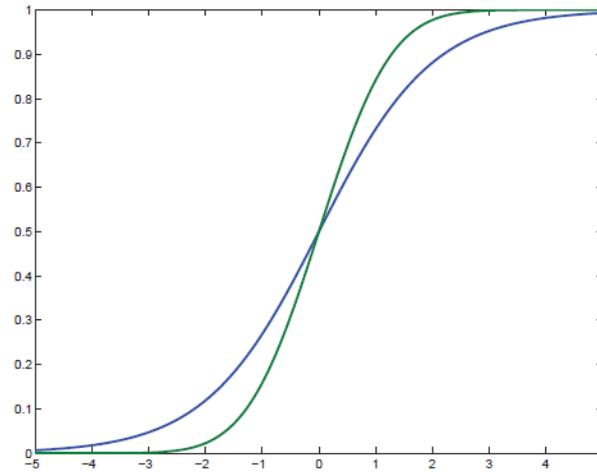
En ambos casos, tenemos que  $c$  es el polinomio que contiene el modelo de regresión lineal ( $c = x'_n \beta + z'_n \gamma$ ).

En la siguiente figura vemos las gráficas de la función de distribución logística (curva azul) y de la función de distribución normal (curva verde) con valores comprendidos entre 0 y 1. En la cual cabe destacar que hay pocas diferencias entre ambas gráficas, básicamente difieren en el comportamiento en las colas, la rapidez con la que se acercan a 0 y a 1. Donde la normal es menos achatada y alcanza más rápidamente los valores extremos, mientras, la curva logística tiene las colas más anchas, lo cual indica que la probabilidad de éxito en los extremos es mayor en el modelo logit.

Debido a la similitud de ambas curvas, podemos deducir que no habrá muchas diferencias en la estimación de ambos modelos.

Las características técnicas más importantes de ambas son muy similares:

- Ambas Funciones son positivas para cualquier valor de  $x$ .
- Asintóticas a 0 cuando la función tiende a  $-\infty$ , y asintóticas a 1 cuando tiende a  $+\infty$ .



**Figura 1.** DISTRIBUCIÓN ACUMULADA NORMAL Y LOGÍSTICA.

- Tienen máxima pendiente en el 0.
- Centradas en 0 y simétricas.

Los modelos Logit y Probit, asumen ambos que existe una variable latente no observable, pero que está conectada con la variación de los regresores. La idea es la que habíamos comentado anteriormente, en la sección 1.2, donde si llamamos a la variable latente  $d_n^*$ , solo cuando esta supera el umbral del 0 sucede el suceso  $d_n = 1$ , como ya habíamos expresado en (8). Así pues, a la hora de estimar un parámetro, estamos viendo en realidad el cambio de la variable latente ante un cambio en la variable exógena (o variable explicativa) y no en la probabilidad de que ocurra el suceso, por ello, el valor en sí de los parámetros de ambos modelos no nos aporta información estrictamente. El signo de los mismos indica la dirección en que se mueve la probabilidad (incremento o decremento) cuando aumenta la variable exógena correspondiente.

A la hora de realizar los análisis los resultados de ambas son muy similares en cuanto a significatividad, ajuste, valores de los parámetros, etc. Sin embargo, nos será mucho más fácil trabajar con los parámetros de la función logística.

Ajuste del modelo Logit:

El modelo de regresión logit será:

$$d_n = F(x_n'\beta + z_n'\gamma) + \varepsilon_n$$

Donde las probabilidades de que la variable dependiente tome valores 0 o 1 son:

$$Prob(d_n = 1 | x_n, z_n) = E[d_n | X'_n, z'_n] = \frac{e^{(x'_n\beta + z'_n\gamma)}}{1 + e^{(x'_n\beta + z'_n\gamma)}} \quad (13)$$

y

$$\begin{aligned} Prob(d_n = 0 | x_n, z_n) &= 1 - E[d_n | X'_n, z'_n] = 1 - \frac{e^{(x'_n\beta + z'_n\gamma)}}{1 + e^{(x'_n\beta + z'_n\gamma)}} \\ &= \frac{1}{1 + e^{(x'_n\beta + z'_n\gamma)}} \end{aligned}$$

(14)

Debido a la no linealidad del modelo, es decir, que la relación entre la variable dependiente y las explicativas es no lineal, no podemos aplicar el método de Mínimos Cuadrados Ordinarios o lineales para estimar los parámetros del modelo de regresión Logit.

Además, como las variables explicativas incluidas en el modelo son variables discretas combinadas de forma que se impide la obtención de observaciones repetidas de la variable endógena, entonces, la matriz de datos muestrales estará formada por  $n$  observaciones, donde la variable endógena tomará los valores 1 o 0. Esto, junto a la no linealidad comentada anteriormente, impide la utilización de métodos tradicionales, como por ejemplo, la estimación del modelo mediante el método de mínimos cuadrados generalizados.

Por tanto, se utilizará el método de máxima verosimilitud para ajustar el modelo y encontrar sus parámetros.

De este modo, la estimación de un modelo logit se obtiene maximizando su función de verosimilitud, esta se trata de una función cóncava, por tanto, los estimadores de verosimilitud existen y son únicos. Se maximizará el logaritmo de la verosimilitud.

Se consideran estimadores Máximo-verosímiles de los parámetros de una población determinada, aquellos valores que generan con mayor probabilidad la muestra observada, es decir, aquellos cuya función de densidad conjunta (función de verosimilitud) alcanza su máximo.

Suponiendo que las observaciones son independientes, tenemos la función de densidad de la variable  $d$  para cada observación  $n$ ,

$$L = \prod_{n=1}^m [F(x'_n\beta + z'_n\gamma)]^{d_n} [1 - F(x'_n\beta + z'_n\gamma)]^{1-d_n}$$

Tomando el logaritmo de la función se nos queda el logaritmo de verosimilitud:

$$\ln L = \sum_{n=1}^m d_n \ln[F(x'_n\beta + z'_n\gamma)] + \sum_{n=1}^m (d_n - 1) \ln(1 - [F(x'_n\beta + z'_n\gamma)])$$

derivando la expresión con respecto a cada coeficiente, obtenemos el gradiente del logaritmo de la verosimilitud

$$\frac{\partial \ln L}{\partial(\beta\gamma)} = \sum_{n=1}^m d_n \frac{F'[(x'_n\beta + z'_n\gamma)]}{F[(x'_n\beta + z'_n\gamma)]} (x_n z_n) + \sum_{n=1}^m (1-d_n) \left(1 - \frac{F'[(x'_n\beta + z'_n\gamma)]}{F[(x'_n\beta + z'_n\gamma)]}\right) (x_n z_n)$$

Haciendo estas ecuaciones cero, podemos encontrar los estimadores de máxima verosimilitud de los parámetros, ya que el método de estimación de máxima verosimilitud elige el estimador del parámetro que maximiza la función de verosimilitud ( $\ln L$ ).

Las derivadas de primer orden respecto a los parámetros  $\beta, \gamma$ , sabiendo que  $F(x'_n\beta + z'_n\gamma) = \frac{e^{(x'_n\hat{\beta} + z'_n\hat{\gamma})}}{1 + e^{(x'_n\hat{\beta} + z'_n\hat{\gamma})}}$ , quedan como se sigue:

$$\frac{\partial \ln L}{\partial \beta} = \sum_{n=1}^m \left( d_n - \frac{e^{(x'_n\hat{\beta} + z'_n\hat{\gamma})}}{1 + e^{(x'_n\hat{\beta} + z'_n\hat{\gamma})}} \right) x_n \quad (15)$$

$$\frac{\partial \ln L}{\partial \gamma} = \sum_{n=1}^m \left( d_n - \frac{e^{(x'_n\hat{\beta} + z'_n\hat{\gamma})}}{1 + e^{(x'_n\hat{\beta} + z'_n\hat{\gamma})}} \right) z_n \quad (16)$$

Este se trata de un sistema de ecuaciones no lineales, por eso es necesario un método iterativo o algoritmo de optimización que permita la convergencia.

Los estimadores que se obtienen bajo estas ecuaciones del modelo Logit son consistentes, si el número de observaciones tiende a infinito, los estimadores tienden en probabilidad a su valor verdadero. Además, también cumplen la propiedad de ser asintóticamente eficientes. Además, al aplicar el método de máxima verosimilitud, eliminamos el problema de la heterocedasticidad que teníamos con el modelo lineal.

**Algunos parámetros estimados en un modelo Logit:** El modelo logit, al igual que el modelo lineal de probabilidad, sirve para medir la probabilidad de que ocurra el acontecimiento objeto de estudio, en nuestro caso,  $d_n = 1$ .

En el caso de los modelos logit, como la relación entre las variables explicativas y la probabilidad de que ocurra el suceso 1 es no lineal, el aumento en una unidad de la variable explicativa (variable exógena) no siempre indican el mismo aumento en la probabilidad, ya que dependerá del nivel original de la misma.

Una interpretación más sencilla de obtención de parámetros estimados es a través de la linealización del modelo. Si llamamos  $M_n$  a la probabilidad de que ocurra la alternativa 1, de tal forma que:

$$E(d_n) = M_n = \frac{e^{(x'_n\beta + z'_n\gamma)}}{1 + e^{(x'_n\beta + z'_n\gamma)}}$$

se tiene:

$$M_n + M_n e^{(x'_n\beta + z'_n\gamma)} = e^{(x'_n\beta + z'_n\gamma)}$$

$$M_n = (1 - M_n)e^{(x'_n\beta + z'_n\gamma)}$$

$$\frac{M_n}{1 + M_n} = e^{(x'_n\beta + z'_n\gamma)}$$

De esta forma obtenemos el cociente entre la probabilidad de que la alternativa 1 sea la escogida y la probabilidad de que no se elija esta alternativa (es decir, que resulte escogida la alternativa 0), a este cociente se le denomina como *odds ratio*, esta es una forma de representar el riesgo, o por otro lado, la ventaja de la opción (o alternativa) 1 frente a la 0. Dependiendo del campo de estudio en el que nos encontremos tiene una interpretación. Es decir, si nos encontramos en el campo clínico, la *odds ratio* podría representar el riesgo de contraer una enfermedad frente al de no contraerla, sin embargo, nosotros tomaremos la interpretación que se refiere a la ventaja de escoger una alternativa frente a otra, ya que el campo que nos ocupa es el de la utilidad de tales alternativas.

A la hora de interpretar la *odds ratio* tenemos que tener en cuenta, que la *odds ratio* y la probabilidad no son equivalentes, sino que conectan con esta igualdad

$$\text{odds ratio} = \frac{M_n}{(1 - M_n)}$$

donde esta siempre será mayor o igual que 0 pero no estará limitada por un máximo, es decir, la *odds ratio* podrá variar desde 0 hasta  $+\infty$ , y su interpretación se realiza en función del valor 1 (se comparará si es mayor, menor o igual a 1). De esta forma, si *odds ratio* = 1, tendremos que  $M_n = 1 - M_n$ , y la probabilidad de que la alternativa 1 sea escogida es la misma que de que no sea escogida. Si de otro modo, *odds ratio* < 1, tenemos entonces que,  $M_n < 1 - M_n$ , con lo cual, la alternativa 1 tiene menor probabilidad de ser escogida frente a la alternativa 0.

Cuando se quiere comparar las ventajas para distintos valores de la variable explicativa (que en nuestro caso serán los niveles dentro de cada alternativa), se calcula el *cociente entre odds*, el cual nos da otra forma de cuantificar la asociación entre dos variables dicotómicas. Así, si se toma como referencia la observación (individuo) “*i*” y se quiere comparar con la del individuo “*j*”, tenemos:

$$\text{cociente entre odds} = \frac{\frac{M_{i+1}}{(1 - M_{i+1})}}{\frac{M_i}{(1 - M_i)}}$$

y de este modo, el *cociente entre odds* mide cuanto es más probable que se dé la alternativa 1 en la observación “*i*” frente que se dé la alternativa 1 en la observación “*j*”. Su interpretación es análoga a la de la *odds ratio*, es decir, si *cociente entre odds* = 1 tenemos que la probabilidad de obtener la alternativa 1 es igual para las dos observaciones, y si el valor obtenido para el *cociente entre odds* es menor que 1, la probabilidad de que la alternativa 1 sea escogida en la observación “*i*” es menor que la probabilidad de que esta sea escogida en la observación “*j*”, por el contrario, si *cociente entre odds* > 1, tendremos que la alternativa 1 será escogida en la observación “*i*” con mayor probabilidad que en la observación “*j*”.

Cuando queremos calcular la variación en la preferencia o ventaja de la observación “*i*” al verse incrementada en una unidad alguna de las variables explicativas, el cálculo del *cociente entre odds* facilita la interpretación de los parámetros estimados, obteniendo el factor de cambio del siguiente modo:

$$\begin{aligned} \text{cociente entre odds} &= \frac{\frac{M_{i+1}}{(1 - M_{i+1})}}{\frac{M_i}{(1 - M_i)}} = \\ &= \frac{e^{((x'_i+1)\beta + (z'_i+1)\gamma)}}{e^{(x'_i\beta + z'_i\gamma)}} = e^{\beta + \gamma} \end{aligned}$$

El parámetro  $e^{\beta+\gamma}$  es un factor de cambio en el *cociente entre odds* cuando las variables explicativas “ $x_i$ ” y “ $z_i$ ” aumentan en una unidad, de esta manera se estima el logaritmo de la ventaja. Por otra parte, la evaluación del valor del antilogaritmo nos dará una forma efectiva para analizar los efectos directos de los parámetros  $\beta$  y  $\gamma$  sobre la probabilidad.

Si queremos expresar el cambio que sufre la variable dependiente provocado por un cambio unitario en una de las variables independientes (variables exógenas o explicativas), calcularemos el *efecto marginal*.

El *efecto marginal* se calculará como la derivada parcial de la variable dependiente,  $d_n$ , con respecto a cada una de las variables independientes,  $x_n$ , y las variables independientes,  $z_n$ , que como ya habíamos comentado en la sección 1.2 son indicadores de género, localización, etc. Es decir, si llamamos al conjunto de todas las variables independientes  $X_{nj}$ , se interpretará el cambio producido en  $d$  cuando una de las variables explicativas aumenta en una unidad como:

$$\frac{\partial d_n}{\partial X_{jn}} = \frac{e^{-(\beta+\gamma)X_{jn}}}{(1 + e^{-(\beta+\gamma)X_{jn}})^2}(\beta + \gamma)$$

El coeficiente del *efecto marginal* será la constante dada por  $(\beta + \gamma)$ .

Por la forma en la que se define el *efecto marginal*, tenemos que en el modelo lineal de probabilidad el *efecto marginal* de cada una de las variables explicativas es constante, lo cual no es lo que esperábamos, ya que no es muy razonable que las variaciones en la probabilidad sean iguales en los valores centrales y en los valores extremos de la variable dependiente.

En el modelo logit, como acabamos de ver, el *efecto marginal* depende de los valores que toman las variables explicativas. De este modo, el efecto marginal ya no es constante como ocurría en el modelo lineal de probabilidad. Por tanto, se puede calcular los *efectos marginales* para cada observación de la muestra o, también, pueden evaluarse para el valor medio de las variables explicativas.

Puesto que la función exponencial utilizada en el modelo logit (al igual ocurre con la función utilizada en el modelo probit) se trata de una función positiva (siempre es positiva), el signo de los coeficientes siempre nos mostrará la dirección del *efecto marginal*, con esto, si el signo es positivo indicará una relación directa entre la variable explicativa y la probabilidad, mientras que si el signo es negativo la relación será inversa.

**Contraste de hipótesis:** Para ver la significatividad estadística de los parámetros estimados utilizaremos un contraste de hipótesis, donde mediante la construcción del intervalo de confianza, valoramos si se debe rechazar o aceptar la hipótesis nula.

Tenemos que la distribución del estimador del parámetro  $\beta$  es una normal:

$$\beta \sim N\left(\beta; \sqrt{\text{Var}(\hat{\beta})}\right)$$

Para realizar contrastes de hipótesis de significación individual, es decir, para testar si el valor del parámetro estimado es significativamente distinto de cero de forma individual, el contraste a realizar sería:

$$\begin{cases} H_0 : \beta = 0 \text{ El parámetro es igual a } 0 \\ H_1 : \beta \neq 0 \text{ El parámetro es distinto de } 0 \end{cases} \quad (17)$$

El intervalo de confianza proporciona un rango de posibles valores para el parámetro. Queda definido de la siguiente forma:

$$\hat{\beta} - z_{\alpha/2} \sqrt{\text{Var}(\hat{\beta})} \leq \beta \leq \hat{\beta} + z_{\alpha/2} \sqrt{\text{Var}(\hat{\beta})}$$

donde  $z \sim N(0; 1)$  y  $(1 - \alpha)$  es el nivel de confianza (o significatividad), también se puede representar en porcentaje  $(1 - \alpha)100\%$  (la forma más habitual). Una vez extraída la muestra, se confía si el valor está o no dentro del intervalo, y si repetimos el proceso con muchas muestras podríamos afirmar que el  $(1 - \alpha)100\%$  de los intervalos construidos contiene el valor verdadero.

De este modo, a partir de la expresión anterior, el contraste de hipótesis (17) seguirá la siguiente regla de decisión:

$$\text{Se rechaza } H_0 \text{ si } \left| \frac{\hat{\beta}}{\sqrt{\text{Var}(\hat{\beta})}} \right| \geq z_{\alpha/2}$$

Por otro lado, si queremos realizar contrastes de hipótesis conjunta sobre todos los coeficientes o un subconjunto de coeficientes de los estimadores, podemos utilizar el contraste de razones de verosimilitudes.

Queremos realizar el contraste definido a continuación:

$$\begin{cases} H_0 : \beta_0 = \beta_1 = \dots = 0 \\ H_1 : \beta_0 \neq \beta_1 \neq \dots \neq 0 \end{cases} \quad (18)$$

Donde cabe destacar que el rechazo de la hipótesis nula no indica un rechazo a que todos los coeficientes de los parámetros son 0, es decir, puede haber algún coeficiente 0 pero no todos lo serán.

El estadístico utilizado para este tipo de contraste (18), será:

$$-2Ln \frac{L(\hat{\beta}_j)}{L(\hat{\beta})} \sim \chi_q^2$$

Donde,  $L(\hat{\beta}_j)$  es la verosimilitud del modelo restringido, en el cual se ha impuesto la  $H_0$ ,  $L(\hat{\beta})$  es la verosimilitud del modelo completo, y  $q$  es el número de restricciones.

De este modo, la regla de decisión queda:

si  $|\chi_{1-\alpha}| \leq \chi_q^2$  se rechaza la  $H_0$ .

Valores muy próximos a cero del cociente  $\frac{L(\hat{\beta}_j)}{L(\hat{\beta})}$ , utilizado para el contraste, significa que las variables incluidas en el modelo son poco significativas. Por lo que la capacidad explicativa del modelo sería en este caso muy reducida.

Como habíamos comentado al principio de esta sección, las funciones  $\phi$  y  $\lambda$ , del modelo probit y logit, respectivamente, tienen propiedades muy similares. De este modo, los resultados obtenidos para el modelo logit como los parámetros estimados, los contrastes, etc, serán análogos para el modelo probit.

Estos modelos los habíamos introducido como medida para intentar solventar los problemas del modelo lineal de probabilidad. Estos cumplen:

- Los valores estimados de la probabilidad están entre 0 y 1.
- Estos modelos no exigen homocedasticidad y no suponen normalidad en los errores.
- La variación de la probabilidad ya no es una función lineal de los regresores.

#### 1.4. Diseño de escenarios

El tema que vamos a abordar en esta sección es el diseño de encuestas de preferencias declaradas para evaluar las características del objeto de estudio. Estas

encuestas permiten diseñar un conjunto eficiente de situaciones de intercambio para evaluar las preferencias de los usuarios.

**Preferencias declaradas** Las técnicas de preferencias declaradas son un conjunto de metodologías que se basan en juicios realizados por los individuos, se presentan situaciones hipotéticas que deben ser lo más cercanas a la realidad y estos deben emitir juicios de como actuarían ante tal situación. Estas técnicas utilizan diseños experimentales para construir estas situaciones hipotéticas que son presentadas a los encuestados en forma de alternativas de elección.

Estas técnicas de preferencias declaradas poseen las siguientes características, que son de utilidad a la hora de abordar estudios sobre valores y características de un producto:

- Las variables explicativas pueden ser extendidas, es decir, el rango de variación de los atributos puede modificarse, añadiendo factores u opciones.
- Se pueden aislar los efectos de interés de las variables, lo cual facilita su estudio posterior.
- Permite la incorporación de variables secundarias cualitativas.
- No existe error de medición de los datos (por construcción de estos).
- Facilitan la recolección de datos y su posterior análisis.

Sin embargo, hay que tener especial cuidado en determinados sesgos a la hora de utilizar estos métodos de preferencias declaradas:

- **Errores aleatorios:** Estos se deben a que los individuos seleccionan la respuesta que ellos creen que hipotéticamente elegirían en la vida real, lo cual no asegura que realmente lo harían si se les presentase la situación, además también pueden darse interpretaciones distintas a una misma pregunta, por ello, las preguntas deben ser lo más claras y concisas posible, además de acercarse a hechos reales o casi reales. También pueden darse estos errores por fatiga del entrevistado, esto se tendrá en cuenta a la hora de presentar los escenarios.
- **Errores no aleatorios:** Este tipo de errores pueden darse por percepciones cotidianas de los encuestados, debido a experiencias anteriores que pueden influir a la hora de contestar.
- **Error de afirmación:** Se pueden producir errores a la hora de contestar sobre preferencias por el afán de agradar o querer contentar al encuestador por parte del individuo encuestado.
- **Error de racionalización:** Se pueden producir errores al intentar racionalizar las respuestas, de este modo, el encuestado está proporcionando respuestas artificiales. El subconsciente del encuestado entra en conflicto, y esto

puede impactar en sus respuestas.

- **Error de política:** Este error se debe a una manipulación de las respuestas por parte del encuestado, ya que este puede que responda deliberadamente con el fin de influir en las decisiones que el cree que se basarán los resultados de la encuesta.
- **Error de no restricción:** En este caso el error viene dado por limitar las respuestas a un conjunto que no se adecua a los encuestados. Ya que al restringir las respuestas y no encontrar una práctica con su comportamiento, este puede responder de forma irreal. Por tanto, a la hora de diseñar las preguntas de un cuestionario se debe tener en cuenta que las restricciones u opciones aportadas abarquen el mayor rango posible de características.
- **Error de no respuesta:** Este tipo de errores es el más común en todo tipo de encuestas. Encontrarse con preguntas sin contestar suele ser muy habitual. Puede darse por muchos factores, ya sea por no entender la pregunta, por falta de opciones, etc.

Todos estos son errores a considerar e intentar minimizar, ya que implican errores de medición en la variable dependiente.

**Diseño experimental** El diseño experimental se trata de una técnica estadística que permite interpretar y analizar las causas de un efecto, de esta manera teniendo variables vinculadas a las causas podremos medir el efecto de otras variables relacionadas, a estas relaciones se les llama de causa-efecto. Para poder medir la causa-efecto con un grado de confianza y poder obtener resultados fiables, los diseños de experimentos nos dan unas pautas para manipular las variables.

En cuanto al diseño experimental de preferencias declaradas, se utilizan técnicas de análisis conjunto, donde cada encuestado contribuye a la muestra total en el número de preguntas contestadas. Podemos distinguir 3 tipos de diseño experimental dependiendo del tipo de respuesta:

1. **Escalamiento (Rating):** En este tipo de diseños las respuestas están organizadas en una escala numérica que indica satisfacción, atractivo o probabilidad de uso (por niveles). Esta se transforma en una escala de utilidad. Al imponer unos supuestos podremos decir que es muy probable que las medidas obtenidas estén sesgadas, aunque confiables (Louviere (2000)). Estos sirven para expresar grados de preferencia, a cada opción se le otorga una categoría de elección (category rating scale).

2. **Jerarquización (Ranking):** En la jerarquización, como su nombre indica, se jerarquiza (se compara y ordena) cada respuesta según atractivo o preferencia. Esto puede realizarse por dos caminos, por una parte pidiendo al consumidor que ordene las opciones directamente, o por otra parte, ordenar las elecciones según frecuencia de uso (para esto hace falta observaciones de las elecciones a través de un largo período de tiempo). A medida que aumenta el número de opciones aumentan las dificultades para llevarlo a cabo por parte del encuestado, además disminuirá la confiabilidad del experimento. Una vez realizada la jerarquización se realiza la estimación que permite encontrar una función de utilidad consistente con el orden de las alternativas escogidas por el encuestado.
  
3. **Elección (Choice):** En los experimentos de elección, el encuestado ha de escoger la alternativa que considere mejor, dentro de un conjunto de alternativas presentado. Estos son los más utilizados, ya que el pronóstico de la elección es lo que se busca. Estos también pueden ser utilizados en contextos de elección abstracta, es el caso de cuando se quiere realizar un estudio de un producto aún no disponible, en este caso se hacen supuestos en las características del producto para presentar las alternativas de elección. La selección de este tipo de diseño experimental se basa en la necesidad de pronosticar la probabilidad de utilización.

### 1.5. Conclusiones

Todo diseño experimental de preferencias declaradas consiste en una serie de variables independientes relacionadas con una variable dependiente (que es la variable respuesta, la variable de estudio). Cada variable independiente (a las que llamaremos factores o atributos) se caracteriza por una cantidad de niveles, que son los que representan el valor de cada atributo.

A continuación listaremos una serie de recomendaciones a considerar en el diseño experimental de preferencias declaradas, dadas por Fowkes. et Wardman (1988):

1. Los análisis de compromisos (trade-offs, en los que la idea es como una decisión que implica que esta es realizada con una comprensión total de las ventajas y desventajas, a lo mejor se da que perdemos un tipo de cualidad para poder ganar otra cualidad que a nosotros nos conviene más) presentados a los encuestados deben cubrir un amplio rango de variación en los gustos de los individuos, para intentar evitar errores de no restricción que hemos comentado en la sección 1.4.
  
2. La consideración de los efectos (atributos) debe realizarse con especial cuidado (ha de hacerse en total relación con el objeto de estudio, ofrecer situaciones de elección a los encuestados tal que pueda realizarse la estimación de

los parámetros del modelo de elección discreta especificado), ya que de estas consideraciones dependen los atributos y niveles del diseño. Además, los valores de los atributos deben seleccionarse de manera que sean interpretados como realistas por parte de los encuestados.

3. A veces será necesario igualar los valores de algunos atributos para distintas alternativas, de modo que tengan el número de niveles que nos convenga, para clarificar diferencias entre los atributos que sean imperceptibles o ignoradas por los encuestados. Esta medida en la teoría del diseño experimental es considerada como mala, sin embargo, en la práctica a veces será necesaria.
4. Para finalizar, deben realizarse pruebas de consistencia y racionalidad de las respuestas ofrecidas. Los escenarios de elección no deben ser dominantes, y si lo son algunos será para verificar la racionalidad de las respuestas.

Nuestro objetivo de estudio es analizar las preferencias sobre características de un objeto, para ello nos decantaremos por los modelos de elección discreta. Los modelos de elección discreta resultan apropiados cuando el objetivo no es predecir el comportamiento medio agregado, sino analizar los factores determinantes de la probabilidad de que un agente económico individual elija una acción dentro de un conjunto. Se presentarán al encuestado una serie de conjuntos de elección o escenarios, en función de las elecciones se analizarán y obtendrán resultados sobre las preferencias de la variable explicada.

## 2. Diseño factorial completo

### 2.1. Introducción

Los diseños factoriales son diseños de experimentos en los cuales cada nivel de cada atributo se combina con todos los niveles de los demás atributos. De este modo si un diseño factorial tiene dos atributos con dos niveles para cada uno, la cantidad de posibles combinaciones será igual a  $2^2 = 4$ , y si tenemos dos atributos con tres niveles cada uno, entonces, tendremos  $2^3 = 8$  combinaciones de posibles escenarios.

Los diseños factoriales en los que se emplean todas las posibles combinaciones de escenarios se denominan diseños factoriales completos (full factorial). Estos garantizan que todos los efectos de interés de los atributos son independientes (se dirá que los atributos son independientes por diseño), de este modo, los parámetros de interés en los modelos de elección podrán estimarse independientemente con los resultados del experimento.

Los diseños factoriales son muy eficientes por la capacidad de obtener información de los efectos entre los factores en cada observación, como habíamos comentado en la sección 1. El efecto de un factor es un cambio en la respuesta ocasionada por un cambio en el nivel de ese factor. Los efectos de interés en un diseño en general son: los efectos simples, los principales y los de interacción.

Los efectos simples son las comparaciones entre los niveles de un mismo factor, a un solo nivel del otro. Un efecto principal es la comparación de medias de cada nivel de un atributo en particular. Los efectos de interacción entre dos atributos ocurren si las preferencias de los consumidores con respecto a los niveles de un primer atributo dependen de los niveles de un segundo atributo (por ejemplo, la existencia de relación calidad-precio). De este modo podemos definir la interacción como el efecto producido por la acción de un factor, influido por la presencia de otro (también se puede hablar de interacción de más de dos factores).

La base de los diseños factoriales es determinar del total de combinaciones posibles para cada atributo con sus posibles niveles, una vez obtenido este número total de combinaciones posibles se muestra el número de situaciones hipotéticas de opciones presentadas al encuestado, cumpliendo que garanticen la independencia entre ellas (opciones). A este paso del diseño se le denomina plan maestro, un ejemplo muy utilizado de plan maestro es el diseñado por Han y Shapiro.

Por otro lado, hay veces que se sigue otra metodología para seleccionar el diseño experimental factorial adecuado. En este caso, primero se establece el número de situaciones experimentales (escenarios de elección a presentar a los encuestados). A continuación se determina el número de atributos e interacciones, y por último, el número de niveles, estos se pueden ajustar para mantener u

obtener el número máximo de escenarios dentro de los límites especificados previamente. Esta es menos común, se suele utilizar en casos donde el presupuesto lo requiere.

El diseño factorial completo permite estimar los efectos principales de todos los atributos y los efectos de interacción entre pares de atributos, que se asumen independientes de los efectos principales.

Un ejemplo de diseño factorial completo se ve reflejado en la siguiente tabla. Un diseño factorial completo de tres atributos con dos niveles cada uno (codificados con los números 0 y 1), tendría  $2^3 = 8$  escenarios de elección y sería el siguiente:

Conjunto de elección o escenario	Factor o Atributo 1	Factor o Atributo 2	Factor o Atributo 3
1	0	0	0
2	0	0	1
3	0	1	0
4	0	1	1
5	1	0	0
6	1	0	1
7	1	1	0
8	1	1	1

**Figura 2.** Ejemplo de diseño factorial completo.

En la tabla vemos que se combinan de todas las maneras posibles los dos niveles de cada uno de los tres atributos presentados. Por otro lado, notamos que la correlación entre las columnas de la tabla (las cuales representan a los atributos) es nula, esto se debe a la ortogonalidad del diseño factorial completo.

Los diseños ortogonales son más sencillos de interpretar, ya que la ortogonalidad asegura la independencia de cada uno de los efectos y de sus interacciones (en el ejemplo, al multiplicar la columna del atributo 1 con las columnas de los otros, vemos que su resultado es 0, el atributo 1 se estima de forma independiente a los demás atributos, y si repetimos de forma análoga esta operación con sus respectivos atributos, obtenemos la independencia de cada atributo con sus interacciones como queríamos).

Como cabe esperar, el número de combinaciones posibles para cada atributo con sus niveles, crece exponencialmente a medida que aumenta el número de atributos y sus posibles niveles considerados, esto implica que el diseño factorial completo requiera de la construcción de un gran número de opciones en la encuesta, lo cual puede afectar directamente a los resultados e interpretaciones de la variable respuesta, ya que se producen sesgos por fatiga del encuestado. Para evitar este tipo de errores se utilizan diseños factoriales fraccionados, que explicaremos en la siguiente sección.

A lo largo de esta sección veremos cómo describir en términos estadísticos los experimentos factoriales, nombrando los modelos más utilizados, aunque nosotros solo utilizaremos los modelos de regresión, ya que nos parecen los más aclaratorios, esto lo veremos aplicado a experimentos factoriales de dos atributos. A continuación veremos cómo realizar el análisis de la varianza mediante el cálculo de las sumas de cuadrados para la construcción de la tabla ANOVA, también para experimentos factoriales con dos atributos. Para más adelante poder centrarnos en los diseños  $2^k$  factoriales, que se trata de los experimentos factoriales de  $k$  atributos con dos niveles cada uno, para poder llegar hasta la generalización de los diseños factoriales  $2^k$ , primero estudiaremos estos diseños con dos y tres atributos, es decir, los diseños  $2^2$  y  $2^3$  factoriales, respectivamente. Procederemos de forma análoga, pero sin extendernos demasiado, para introducir los diseños factoriales  $3^k$  y los diseños factoriales mixtos.

## 2.2. Modelos de un experimento factorial

El objetivo principal de los experimentos factoriales es la interpretación de las interacciones significativas para poder reportar resultados. En esta sección veremos algunos modelos que facilitarán la representación de los experimentos factoriales. Utilizaremos como ejemplo los experimentos factoriales con dos atributos para una mejor comprensión, ya que son los más sencillos.

Las observaciones de un experimento factorial se pueden escribir como un modelo. Hay varios tipos de modelos:

- **El modelo de los efectos:**

Donde suponiendo que el diseño es balanceado, es decir, que se tiene el mismo número de observaciones en cada celda. Vease Montgomery (2004) para ver otras situaciones.

$$y_{ijk} = \mu + \tau_i + \beta_j + (\beta\tau)_{ij} + \varepsilon_{ijk}$$

con  $i = 1, 2, \dots, a$  (niveles del atributo 1) ;  $j = 1, 2, \dots, b$  (niveles del atributo 2) ;  $k = 1, 2, \dots, n$  (número de réplicas).

Donde  $\mu$  es el efecto promedio global,  $\tau_i$  es el efecto del nivel  $i$ -ésimo del atributo 1,  $\beta_j$  es el efecto del nivel  $j$ -ésimo del atributo 2,  $(\beta\tau)_{ij}$  es el efecto de la interacción entre los dos anteriores, y  $\varepsilon_{ijk}$  es la componente de error aleatorio.

■ **El modelo de las medias:**

$$y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$$

con  $i = 1, 2, \dots, a$  (niveles del atributo 1) ;  $j = 1, 2, \dots, b$  (niveles del atributo 2) ;  $k = 1, 2, \dots, n$  (número de réplicas).

Donde la media  $ij$ -ésima es:

$$\mu_{ij} = \mu + \tau_i + \beta_j + (\beta\tau)_{ij}.$$

■ **El modelo de regresión:**

Estos son utilizados cuando uno de los atributos es cuantitativo. El modelo es el descrito anteriormente como:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon$$

Si suponemos que los dos factores del diseño son cuantitativos (si no lo son siempre se puede recurrir a su transformación, mediante el uso de rangos), entonces, una forma más natural de describir el modelo del experimento factorial es utilizando el **modelo de regresión**. Una representación con un modelo de regresión del experimento factorial de dos niveles podría escribirse como:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon$$

donde  $y$  es la respuesta, las  $\beta$  son parámetros cuyos valores deben determinarse,  $x_1$  es la variable que representa al atributo 1,  $x_2$  es la variable que representa al atributo 2, y  $\varepsilon$  es el término de error aleatorio. Las estimaciones de los parámetros en este modelo de regresión resultan estar ligadas con las estimaciones de los efectos.

Veamos cómo interpretar las interacciones con el modelo de regresión mediante el ejemplo de la figura 3.

Tal como muestra el ejemplo (Montgomery (2004)), si consideramos los efectos principales como las diferencias entre el promedio de la respuesta con el nivel alto y el promedio de la respuesta con el nivel bajo de cada atributo, tenemos que los efectos principales de nuestros factores son:

$$A = \frac{40 + 52}{2} - \frac{30 + 20}{2} = 46 - 25 = 21$$

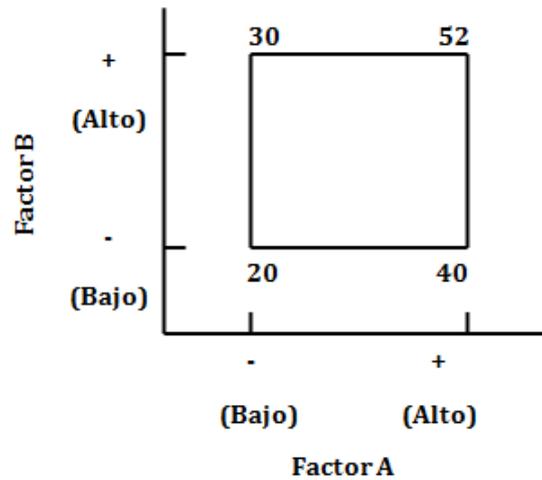
$$B = \frac{30 + 52}{2} - \frac{40 + 20}{2} = 41 - 30 = 11$$

Por otro lado, si tenemos que el efecto de A con el nivel bajo de B es  $A = 40 - 20 = 20$  y el efecto de A con el nivel alto de B es  $A = 52 - 30 = 22$ , entonces, el efecto de la interacción AB se calculará como el promedio de dichos efectos

$$AB = \frac{22 - 20}{2} = 1$$

Los valores con los que estimaremos los parámetros  $\beta_1$  y  $\beta_2$ , serán la mitad del valor del efecto principal correspondiente a cada factor, de tal manera tenemos  $\hat{\beta}_1 = 21/2 = 10,5$  y  $\hat{\beta}_2 = 11/2 = 5,5$ . Como el valor del efecto de la interacción  $AB = 1$ , tenemos que el valor del coeficiente de la interacción en el modelo de regresión es  $\hat{\beta}_{12} = 1/2 = 0,5$ . Podemos estimar el valor del parámetro  $\hat{\beta}_0$ , ya que el promedio de las cuatro respuestas,  $\hat{\beta}_0 = \frac{40+52+30+20}{4} = 35,5$ . De esta manera el modelo de regresión ajustado se queda:

$$\hat{y} = 35,5 + 10,5x_1 + 5,5x_2 + 0,5x_1x_2$$



**Figura 3.** Experimento factorial con dos atributos de dos niveles cada uno.

Las estimaciones de los parámetros obtenidas de este modo para el diseño factorial en el que todos los factores tienen dos niveles resultan ser estimaciones de mínimos cuadrados.

Vemos que el coeficiente de interacción  $\hat{\beta}_{12}$  tiene un valor muy pequeño con respecto a los coeficientes de los efectos principales,  $\hat{\beta}_{12} = 0,5$  mientras que  $\hat{\beta}_1 = 10,5$  y  $\hat{\beta}_2 = 5,5$ , por lo que podemos asumir que esta interacción es despreciable. De este modo podemos eliminar este término y el modelo de regresión se quedaría.

$$\hat{y} = 35,5 + 10,5x_1 + 5,5x_2$$

En nuestro caso, la interacción ha sido tan insignificativa que la hemos podido despreciar. Sin embargo, en los casos donde la interacción es significativa, se deberá recurrir a examinar los niveles de uno de los factores, mientras se mantienen fijos los niveles del otro, para poder sacar conclusiones acerca de los efectos principales.

La figura 4 nos muestra la gráfica de los valores de  $y$  generados por las diferentes combinaciones de  $x_1$  y  $x_2$ , esta recibe el nombre de gráfica de superficie de respuesta. Mientras que la figura 5 se muestran las líneas de contorno para las respuestas constantes en el plano  $x_1, x_2$ .

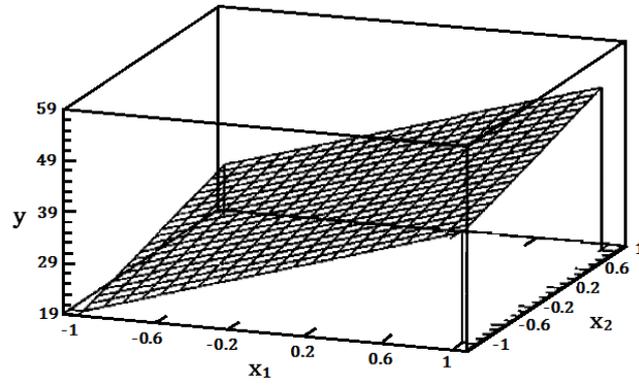


Figura 4. Superficie de respuesta (Montgomery (2004)).

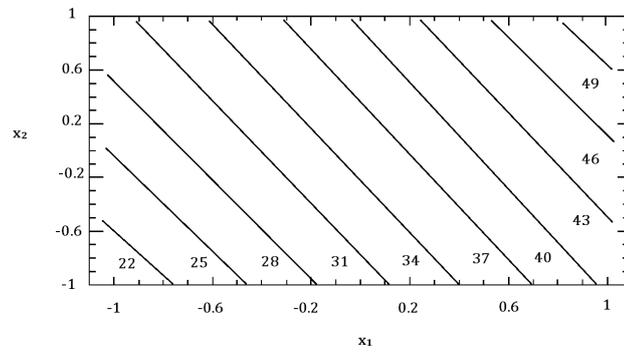


Figura 5. Gráfica de contorno (Montgomery (2004)).

En este caso tenemos que las rectas de la gráfica de contorno son paralelas, esto es debido a la ausencia de interacción entre los factores.

Por el contrario, en el caso de que sí que existiera interacción entre los factores, es decir, que el coeficiente  $\beta_{12}$  no fuera pequeño, tendríamos líneas curvas en la gráfica de contorno, y notaríamos un pequeño torcimiento del plano en la superficie respuesta.

		Factor B			
		1	2	...	b
Factor A	1	Y <sub>111</sub> , Y <sub>112</sub> , ..., Y <sub>11n</sub>	Y <sub>121</sub> , Y <sub>122</sub> , ..., Y <sub>12n</sub>		Y <sub>1b1</sub> , Y <sub>1b2</sub> , ..., Y <sub>1bn</sub>
	2	Y <sub>211</sub> , Y <sub>212</sub> , ..., Y <sub>21n</sub>	Y <sub>221</sub> , Y <sub>222</sub> , ..., Y <sub>22n</sub>		Y <sub>2b1</sub> , Y <sub>2b2</sub> , ..., Y <sub>2bn</sub>
	⋮				
	a	Y <sub>a11</sub> , Y <sub>a12</sub> , ..., Y <sub>a1n</sub>	Y <sub>a21</sub> , Y <sub>a22</sub> , ..., Y <sub>a2n</sub>		Y <sub>ab1</sub> , Y <sub>ab2</sub> , ..., Y <sub>abn</sub>

**Figura 6.** Arreglo general de un diseño factorial de dos factores.

En los diseños factoriales de dos factores, los factores de las filas y las columnas, A y B, son de igual interés. El interés se encuentra en probar hipótesis acerca de la igualdad de los efectos de los tratamientos.

Por ejemplo, un contraste de hipótesis acerca de la igualdad de los efectos de los tratamientos en las filas:

$$\begin{cases} H_0 : \tau_1 = \tau_2 = \dots = \tau_a = 0 \\ H_1 : \text{Al menos una } \tau_i \neq 0 \end{cases}$$

Otro ejemplo, un contraste de hipótesis acerca de la igualdad de los efectos de los tratamiento en las columnas:

$$\begin{cases} H_0 : \beta_1 = \beta_2 = \dots = \beta_b = 0 \\ H_1 : \text{Al menos una } \beta_j \neq 0 \end{cases}$$

También existe el interés de determinar si los tratamientos de las filas y las columnas interactúan. Por lo tanto, otro ejemplo de contraste de hipótesis será:

$$\begin{cases} H_0 : (\tau\beta)_{ij} = 0 \quad \forall i, j \\ H_1 : \text{Al menos una } (\tau\beta)_{ij} \neq 0 \end{cases}$$

Un ejemplo de un diseño factorial de dos atributos será (Montgomery (2004)): Un ingeniero está diseñando una batería que se usará en un dispositivo que se someterá a temperaturas extremas. El único parámetro del diseño que puede seleccionar en este punto es el material de la placa, y tiene tres opciones, que por experiencia sabe que valores les proporcionará cada material al someterse a ciertas temperaturas (en cuanto a duración del material). Con estos datos, mediante el uso de un diseño factorial, puede responder a preguntas como, ¿que efectos tiene el material y la temperatura sobre la vida de la batería? o ¿Existe alguna elección de material que proporcione una vida larga de la batería sin que influya la temperatura?.

**Análisis de la varianza** Para realizar el análisis de la varianza de los efectos del diseño factorial de dos atributos, y poder probar los contrastes de hipótesis antes señalados, se puede resumir en la siguiente tabla (ver tabla 1). De donde deduciremos que no podremos rechazar la hipótesis nula y tomaremos como cierto que no hay efectos en las filas, ni en las columnas, ni en las interacciones, si las estimaciones de  $MS_A$ ,  $MS_B$  y  $MS_{AB}$  son estimaciones de  $\sigma^2$ , y por lo tanto análogas a  $MS_E$ . Es decir, el estadístico  $F_0$ , que se distribuye como una  $F$  (distribución de probabilidad continua), también conocido como distribución F de Snedecor (por George Snedecor) es el que nos ayudará a contrastar las hipótesis, ya que será el que nos dará la relación entre  $MS_A$  y  $MS_E$ , por ejemplo, con  $(a - 1)$  grados de libertad en el numerador y  $ab(n - 1)$  en el denominador.

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	$F_0$
Atributo A	$SS_A$	$a - 1$	$MS_A = \frac{SS_A}{a-1}$	$F_0 = \frac{MS_A}{MS_E}$
Atributo B	$SS_B$	$b - 1$	$MS_B = \frac{SS_B}{b-1}$	$F_0 = \frac{MS_B}{MS_E}$
Interacción	$SS_{AB}$	$(a - 1)(b - 1)$	$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}$	$F_0 = \frac{MS_{AB}}{MS_E}$
Error	$SS_E$	$ab(n - 1)$	$MS_E = \frac{SS_E}{ab(n-1)}$	
Total	$SS_{Total}$	$abn - 1$		

**Cuadro 1.** Tabla del análisis de la varianza para un diseño factorial con dos atributos.

Sea  $y_{i..}$  la suma total de observaciones bajo el nivel  $i$ -ésimo del factor A,  $y_{.j}$  la suma de observaciones bajo el nivel  $j$ -ésimo del factor B,  $y_{ij}$  la suma de observaciones bajo la celda  $ij$ -ésima, y  $y_{...}$  denota el total de todas las observaciones, entonces, se tiene que las sumas de cuadrados de los efectos principales son:

$$SS_A = \frac{1}{bn} \sum_{i=1}^a y_{i..}^2 - \frac{y_{...}^2}{abn}$$

y

$$SS_B = \frac{1}{an} \sum_{j=1}^b y_{.j}^2 - \frac{y_{...}^2}{abn}$$

para obtener las sumas de cuadrados de las interacciones, se hará en dos pasos:

1. Obtendremos la suma debido a los subtotaes, que se trata de la suma de cuadrados entre los totales de las  $ab$  celdas.

$$SS_{Subtotales} = \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^b y_{ij}^2 - \frac{y_{...}^2}{abn}$$

2. Para obtener  $SS_{AB}$  tendremos que restar las sumas de cuadrados de cada uno de los efectos principales.

$$SS_{AB} = SS_{Subtotales} - SS_A - SS_B =$$

$$= \left( \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^b y_{ij}^2 - \frac{y_{...}^2}{abn} \right) - \left( \frac{1}{bn} \sum_{i=1}^a y_{i..}^2 - \frac{y_{...}^2}{abn} \right) - \left( \frac{1}{an} \sum_{j=1}^b y_{.j}^2 - \frac{y_{...}^2}{abn} \right)$$

$$= \frac{1}{n} \left( \sum_{i=1}^a \sum_{j=1}^b y_{ij}^2 - \frac{1}{b} \sum_{i=1}^a y_{i..}^2 - \frac{1}{a} \sum_{j=1}^b y_{.j}^2 \right) - \frac{y_{...}^2}{abn}$$

Además por sustracción podemos calcularnos la suma de cuadrados debida al error. De este modo sabiendo que la suma de cuadrados total es:

$$SS_{Total} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n y_{ijk}^2 - \frac{y_{...}^2}{abn}$$

tenemos que

$$SS_E = SS_{Total} - SS_A - SS_B - SS_{AB}$$

o

$$SS_E = SS_{Total} - SS_{subtotales}$$

De este modo, vemos que podemos obtener fácilmente las sumas de cuadrados de las tablas, y no será tarea difícil poder obtener el estadístico de contraste,  $F$  para cada factor. Y así poder obtener la significación tanto de los efectos principales como de las interacciones, empleando los grados de libertad correspondientes, señalados en la tabla 1.

### 2.3. Diseños $2^k$ factorial

El diseño  $2^k$  factorial es el diseño factorial más sencillo, se trata de un experimento con  $k$  factores o atributos con dos niveles cada uno. Estos niveles pueden ser cuantitativos, cualitativos o quizás pueden indicar la ausencia de un factor.

Una réplica completa de este diseño requiere  $2 \times 2 \times 2 \cdots \times 2 = 2^k$  observaciones.

Como hemos comentado en la sección 2, los diseños factoriales son cada réplica completa de un experimento, en el que se investigan todas las combinaciones posibles de los niveles de los atributos. El efecto de un factor o atributo lo podemos definir como el cambio en la respuesta producido por cambios en los niveles de dicho atributo. Estos diseños suelen ser más eficientes. Además son necesarios este tipo de diseños cuando se quieren estudiar interacciones, y permiten el estudio de los efectos de un atributo con varios niveles de otro factor.

El diseño  $2^k$  factorial es de particular utilidad, sobre todo en las fases iniciales de un experimento con muchos factores, ya que este diseño proporciona el menor número de corridas con las que pueden estudiarse  $k$  factores en un diseño factorial completo.

**Diseño factorial  $2^2$ :** El primer caso de diseño  $2^k$  factorial es el diseño de tan solo dos factores, los atributos A y B por ejemplo, con cada uno dos niveles. A este diseño se le llama diseño factorial  $2^2$ , y tendrá cuatro combinaciones de tratamientos.

El modelo de regresión para un diseño  $2^2$  es

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

donde  $x_1$  representa el atributo A y  $x_2$  representa el atributo B de forma codificada, y las  $\beta$  representan los coeficientes de regresión.

Si llamamos A al efecto del atributo A, B al efecto del atributo B, y AB al efecto de la interacción AB, y denotamos por letras minúsculas a las combinaciones de tratamientos correspondiente a los niveles altos, el nivel bajo de un factor se representará por la ausencia de letra, en su lugar, por convención se utilizará (1) para indicar que ambos factores están en el nivel bajo. Esta notación será utilizada en todos los diseños  $2^k$  factoriales.

En los casos de diseños con dos niveles, tendremos que el efecto promedio de un factor puede describirse como el cambio de respuesta producido por un cambio en el nivel de ese factor promediado para los niveles del otro factor.

Con la notación descrita, tendremos que el efecto de A en el nivel bajo de B será

$$[a - (1)]/n$$

y el efecto de A en el nivel alto del factor B

$$[ab - b]/n$$

al promediarse estas dos cantidades se tiene que el efecto principal de A es:

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

De la misma forma, podemos obtener el efecto principal de B, promediando el efecto de B con el nivel bajo y alto de A. Obtenemos que:

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

Por otro lado, el efecto de la interacción AB se define como la diferencia promedio entre el efecto de A con el nivel alto de B y el efecto de A con el nivel bajo de B. Por lo tanto,

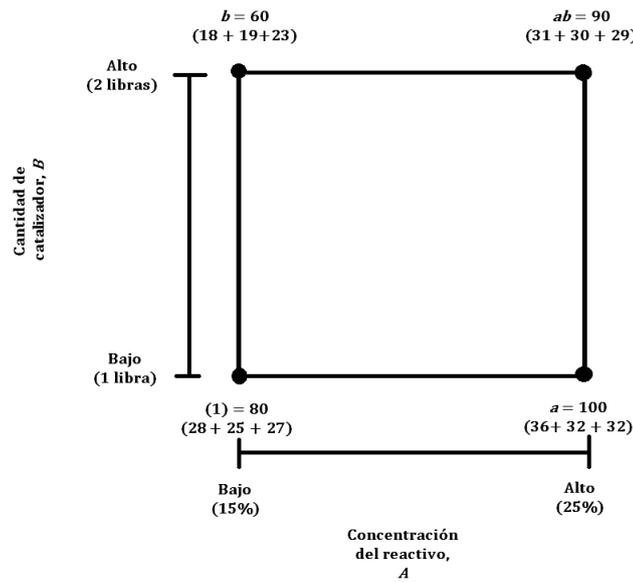
$$AB = \frac{1}{2n}[ab - b] - [a - (1)] = \frac{1}{2n}[ab + (1) - a - b]$$

Ver el ejemplo de la sección 2.2. Donde aplicamos estos cálculos para la obtención de los efectos con 1 réplica del experimento.

Como un ejemplo de diseño  $2^2$  factorial (Montgomery (2004)), la investigación del efecto de la concentración del reactivo y de la cantidad de catalizador

sobre el rendimiento de un proceso químico. Como datos se tiene que la concentración de reactivo es un factor, el factor A, cuyos niveles son 15 % y 25 %. Como factor B tendremos el catalizador, con nivel alto y bajo de 2 libras y 1 libra, respectivamente.

En la figura 7 observamos las cuatro combinaciones de los tratamientos. Vemos la notación que señalábamos antes.



**Figura 7.** Combinaciones en los tratamientos en el diseño  $2^2$  factorial.

Para este ejemplo los efectos promedios pueden estimarse como:

$$A = \frac{1}{2(3)}(90 + 100 - 60 - 80) = 8,33$$

$$B = \frac{1}{2(3)}(90 + 60 - 100 - 80) = -5,00$$

$$AB = \frac{1}{2(3)}(90 + 80 - 100 - 60) = 1,67$$

y su interpretación es la siguiente: al ser el efecto de A positivo, esto nos sugiere que al incrementar al nivel alto de A, el rendimiento aumentará. Por el contrario, como el efecto de B es negativo, tendremos que al aumentar la cantidad del catalizador se reducirá el rendimiento. En cuanto al efecto de la interacción parece ser pequeño en comparación con los dos efectos principales.

Observamos que al obtener los efectos promedios, se utiliza un contraste para estimar A, B y la interacción AB:

$$\text{Contraste}_A = ab + a - b - (1)$$

$$\text{Contraste}_B = ab + b - a - (1)$$

$$\text{Contraste}_{AB} = ab + (1) - a - b$$

a estos contrastes se les llamará el efecto principal de A, el efecto principal de B, y el efecto principal de la interacción AB, respectivamente.

Puede utilizarse el análisis de la varianza para obtener la interpretación de los efectos de los factores. Como sabemos que la suma de cuadrados del contraste es igual al cuadrado del contraste dividido por el número de observaciones en cada total, podemos obtener fácilmente la tabla para el análisis de la varianza, donde:

$$SS_A = \frac{[ab + a - b - (1)]^2}{4n}$$

$$SS_B = \frac{[ab + b - a - (1)]^2}{4n}$$

$$SS_{AB} = \frac{[ab + (1) - a - b]^2}{4n}$$

En el ejemplo de la figura 8, las sumas de cuadrados serán:

$$SS_A = \frac{(50)^2}{4(3)} = 208,33$$

$$SS_B = \frac{(-30)^2}{4(3)} = 75,00$$

$$SS_{AB} = \frac{(10)^2}{4(3)} = 8,33$$

**Diseño factorial 2<sup>3</sup>:** En este tipo de diseños tendremos tres atributos de interés, A, B, C, cada uno con sus dos niveles. En este caso tendremos 2<sup>3</sup> combinaciones posibles.

Corrida	A	B	C
1	0	0	0
2	1	0	0
3	0	1	0
4	1	1	0
5	0	0	1
6	1	0	1
7	0	1	1
8	1	1	1

**Cuadro 2.** Matriz del diseño factorial 2<sup>3</sup>.

A este cuadro se le conoce a veces como la matriz del diseño. Habrán 7 grados de libertad entre las ocho combinaciones del diseño, tres grados se asocian a los efectos principales A, B, C, y cuatro grados de libertad se asocian con las interacciones AB, AC, BC y ABC.

Las estimaciones de los efectos se realizarán de forma análoga a los diseños factoriales 2<sup>2</sup>, solo que ahora tendremos más factores para combinar los efectos. La estimación del efecto principal del atributo A por ejemplo, se realizará promediando las estimaciones del efecto de A cuando los factores B y C están en el nivel bajo, las estimaciones de A cuando B está en el nivel bajo y C en el nivel alto, las estimaciones de A cuando B está en el nivel alto y C en el nivel bajo,

y las estimaciones de A cuando tanto B como C están en el nivel alto. De este modo, tenemos:

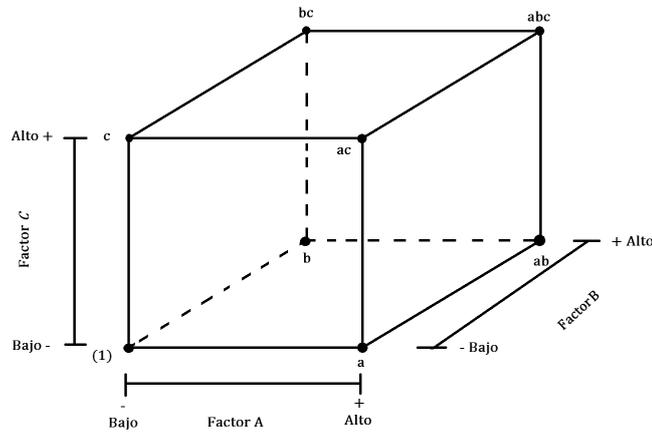


Figura 8. Diseño factorial  $2^3$ .

$$A = \frac{1}{4n} [a - (1) + ab - b + ac - c + abc - bc]$$

$$B = \frac{1}{4n} [b + ab + bc + abc - (1) - a - c - ac]$$

$$C = \frac{1}{4n} [c + ab + bc + abc - (1) - a - b - ab]$$

mientras que los efectos en las interacciones de dos factores pueden calcularse como la diferencia de los efectos promedio de un factor con los dos niveles del otro, a la mitad de esta diferencia se le llamará la interacción AB. De este modo, tenemos que la interacción AB es:

$$Diferencia = \frac{[abc - bc + ab - b - ac + c + a + (1)]}{2n}$$

y como la interacción AB es la mitad de esta diferencia, tenemos el efecto de la interacción

$$AB = \frac{[abc - bc + ab - b - ac + c + a + (1)]}{4n}$$

De forma análoga, los efectos de las interacciones de los demás factores son (viendo la figura 9, observamos que la interacción AB es la diferencia de los promedios entre las corridas de dos planos diagonales del cubo)

$$BC = \frac{[(1) + a - b - ab - c - ac + bc + abc]}{4n}$$

$$AC = \frac{[(1) - a + b - ab - c + ac - bc + abc]}{4n}$$

por otro lado, la interacción ABC se define como el promedio de los efectos entre la interacción AB y los niveles alto y bajo del factor C

$$ABC = \frac{[abc - bc] - [ac - c] - [ab - b] + [a - (1)]}{4n} = \frac{[abc - bc - ac + c - ab + b + a - (1)]}{4n}$$

Como hemos visto en la sección 2.3, al obtener los efectos promedios, se utiliza un contraste para estimar A, B, C y las interacciones. Y a partir de estos contrastes nos podemos calcular las sumas de cuadrados para así poder contrastar la significatividad de los efectos con la tabla ANOVA.

Tenemos, por ejemplo, que el contraste del factor A es

$$\text{contraste}_A = [a - (1) + ab - b + ac - c + abc - bc]$$

de este modo, tenemos

$$SS_A = \frac{[\text{contraste}_A]^2}{8n} = \frac{[a - (1) + ab - b + ac - c + abc - bc]^2}{8n}$$

De forma análoga podemos obtener las sumas de cuadrados restantes.

**Diseño general  $2^k$ :** Una vez vistas las características para los diseños factoriales  $2^2$  y  $2^3$ , estas pueden generalizarse para los diseños factoriales  $2^k$ , diseños de  $k$  factores con dos niveles cada uno.

El modelo estadístico para un diseño  $2^k$  presentará  $k$  efectos principales,  $\binom{k}{2}$  interacciones de dos factores,  $\binom{k}{3}$  interacciones de tres factores,  $\binom{k}{4}$  interacciones de cuatro factores, y así sucesivamente hasta llegar a una interacción de  $k$  factores. Es decir, un diseño factorial  $2^k$  contendrá  $2^k - 1$  efectos.

El modelo de regresión visto en los apartados anteriores, puede ser generalizado. Suponiendo que se incluyen también los términos de interacción y los efectos principales sean significativos o no, tenemos:

$$y = \beta_0 + \sum_{i=1}^k \beta_j x_j + \sum_{j < i} \beta_{ji} x_{ji} + \varepsilon$$

donde  $x_i, x_j \forall i, j = 1, \dots, k$  representan a los  $k$  atributos incluidos en el diseño.

En estos diseños se utiliza la notación descrita para los diseños factoriales  $2^2$  y  $2^3$ , el orden que se sigue a la hora de escribir las combinaciones de los tratamientos es el orden natural, o también llamado, el orden estándar. Por ejemplo, para un diseño  $2^4$  el orden estándar es (1), a, b, ab, c, ac, bc, abc, d, ad, bd, abd, cd, acd, bcd y abcd.

La obtención de la significatividad de los parámetros de las interacciones, para valorar los efectos, se realizan igualmente con la tabla ANOVA vista en la sección 2.2 ampliada para  $k$  efectos principales,  $\binom{k}{2}$  efectos de interacción de dos factores,  $\binom{k}{3}$  interacciones de tres factores,  $\dots$ , y  $\binom{k}{k} = 1$  interacción de  $k$  factores.

Para este tipo de diseños cuando  $k$  es muy grande se utilizan paquetes de software específicos, sin embargo en alguna ocasión es necesario calcular manualmente la estimación de un efecto o la suma de cuadrados de un efecto. Para ello se estimará primero el contraste de dicho efecto, estos pueden ser calculados utilizando su forma general, donde el contraste del efecto  $ABC \dots k$  se determina

$$\text{contraste}_{ABC \dots K} = (a \pm 1)(b \pm 1)(c \pm 1) \dots (k \pm 1)$$

Una vez calculados los contrastes de los efectos, pueden estimarse los efectos y calcularse las sumas de cuadrados de dichos efectos, de este modo

$$ABC \cdots k = \frac{2}{2^{kn}} (\text{contraste}_{ABC \cdots K})$$

y

$$SS_{ABC \cdots k} = \frac{1}{2^{kn}} [(\text{contraste}_{ABC \cdots K})^2]$$

donde  $n$  es el número de réplicas o corridas del experimento. En la mayoría de los casos este suele ser  $n = 1$ , solo se hace un réplica, ya que los recursos disponibles, no permiten hacer más de una sola réplica del diseño, ya que incluso para un número relativamente pequeño de atributos las combinaciones de los tratamientos en estos diseños es relativamente grande.

Una estrategia para el ahorro computacional es eliminar factores del diseño, pero esto es muy arriesgado ya que cabe la posibilidad de que alguno de los efectos de los factores que se eliminan sean significativos para el diseño, y por lo tanto estaríamos perdiendo información relevante del experimento.

Una forma de abordar el análisis del diseño factorial es suponiendo que las interacciones superiores son insignificativas para el diseño y combinar sus cuadrados medios para calcular el error. Esta técnica es conocida como **principio de efectos esparcidos**, el cual asegura que la mayoría de los sistemas están dominados por los efectos principales y algunas de las interacciones de orden inferior, con lo cual la mayoría de interacciones de orden superior son insignificantes.

Para concluir esta sección, vamos a dar un enfoque general para el análisis estadístico de los diseños factoriales  $2^k$ , dando unas pautas a seguir en la siguiente secuencia de puntos.

- **Estimar los efectos de los factores**

El investigador lo primero que tiene que obtener son los factores y estimar sus efectos, examinar sus magnitudes y signos.

- **Formar el modelo inicial**

En este paso se considera el modelo completo, con todos los atributos y niveles, para su posterior revisión y ajuste, si es necesario.

- **Realizar las pruebas estadísticas**

Utilizando el análisis de la varianza (tabla ANOVA) se comprueba formalmente los efectos de los factores y sus interacciones.

- **Refinar el modelo**

En base a los resultados obtenidos en el paso anterior se pueden eliminar las interacciones no significativas del modelo.

- **Interpretar los resultados**

Para este paso se suelen utilizar las gráficas de contorno y de superficie respuesta que hemos visto en la sección 2.2.

#### 2.4. Diseño factorial $3^k$

Durante esta sección hemos visto diseños factoriales con 2, 3, e incluso  $k$  atributos, estos diseños tenían en común que a cada factor o atributo se le asignaban 2 niveles. Ahora vamos a ver arreglos de  $k$  factores con 3 niveles cada uno, se hará referencia a los tres niveles como bajo, intermedio y alto (en lugar de alto y bajo), y se representarán con los dígitos 0 (bajo), 1 (intermedio) y 2 (alto), aunque en algunas ocasiones se representarán con -1 (bajo), 0 (intermedio) y 1 (alto), esta notación es común cuando los factores son cuantitativos.

**Diseño factorial  $3^2$ :** Es el diseño  $3^k$  más simple, cuyas combinaciones de los tratamientos se muestran en la figura 9.

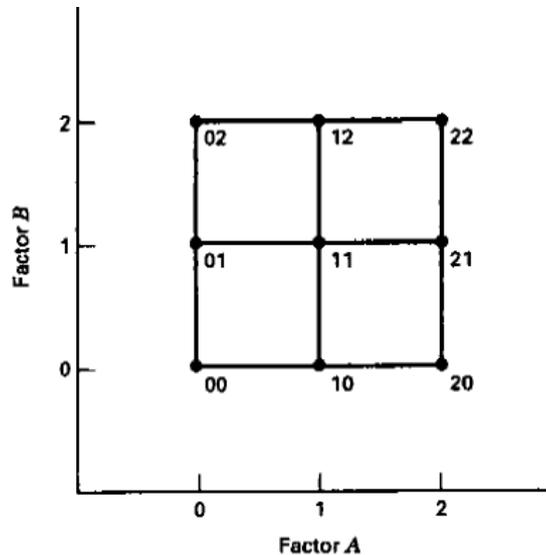


Figura 9. Combinaciones de tratamientos.

El modelo de regresión asociado a este tipo de diseños viene dado por la siguiente ecuación

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_{12}x_1x_2 + \beta_{11}x_1^2 + \beta_{22}x_2^2 + \varepsilon$$

Observamos que el modelo de regresión que relaciona la respuesta  $y$  con  $x_1$  y  $x_2$  de un diseño factorial con dos atributos de 3 niveles cada uno, necesita de la adición de un término cuadrático.

En los diseños factoriales  $3^2$  están presentes  $3^2 = 9$  combinaciones de tratamientos, hay ocho grados de libertad entre estas combinaciones. Los efectos principales de A y B tienen dos grados de libertad cada uno, y la interacción AB tiene cuatro grados de libertad. De este modo, calculando las sumas de cuadrados podremos obtener la tabla ANOVA. Las sumas de cuadrados de A, B y AB pueden calcularse mediante los métodos utilizados para los diseños factoriales  $2^k$ , sin embargo en este tipo de diseños cada efecto principal puede representarse con un componente lineal y uno cuadrático. Para los efectos de las interacciones se separarán en cuatro partes, en cada una se ajustan los términos  $\beta_{12}x_1x_2$ ,  $\beta_{122}x_1x_2^2$ ,  $\beta_{112}x_1^2x_2$ , y  $\beta_{1122}x_1^2x_2^2$ .

**Diseño factorial  $3^3$ :** Suponemos ahora que hay tres factores (A, B y C) con cada uno tres niveles dispuestos en un experimento factorial. Las combinaciones de los tratamientos de este tipo de diseños se muestran en la figura10.

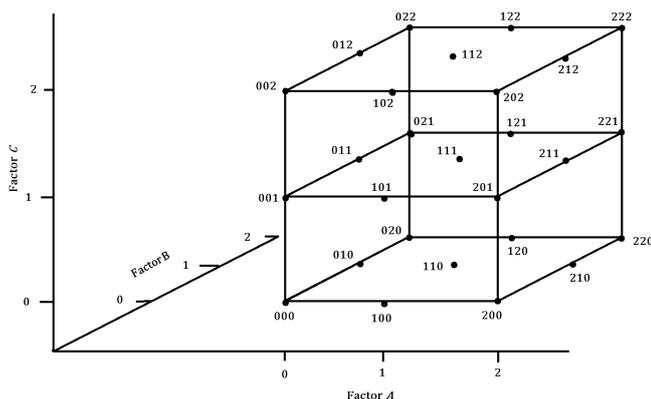


Figura 10. Combinaciones de tratamientos.

El modelo de regresión que representa la relación entre la variable respuesta y los factores para el diseño factorial  $3^3$  vendrá dado por

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 + \varepsilon$$

En este tipo de diseños tendremos  $3^3 = 27$  combinaciones de tratamientos, que tienen 26 grados de libertad. Cada efecto principal tiene dos grados de libertad, cada interacción de dos factores tiene 4 grados de libertad y la interacción de tres factores tiene 8 grados de libertad.

Las sumas de cuadrados podrán calcularse utilizando los métodos de la sección anterior. Además nos puede ser de ayuda la descomposición de los efectos en un componente cuadrático y otro lineal. La descomposición de las interacciones de dos factores podrán hacerse en cuatro partes según los efectos (un grado de libertad para cada una): lineal x lineal, lineal x cuadrático, cuadrático x lineal y cuadrático x cuadrático. Y por último, la interacción de tres factores se hará en ocho componentes (las posibles combinaciones de los tres factores según si son lineales o cuadráticos).

**Diseño general  $3^k$ :** Los conceptos utilizados en los diseños  $3^2$  y  $3^3$ , pueden ser generalizados para los diseños  $3^k$  (compuesto por  $k$  factores de tres niveles cada uno).

El modelo de regresión generalizado para los diseños factoriales  $3^k$  será:

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i < j} \beta_{ij} x_i x_j + \sum_{i=1}^k \beta_{ii} x_i^2 + \varepsilon$$

Tendremos  $3^k$  combinaciones de tratamientos con  $3^k - 1$  grados de libertad entre ellas. Estas combinaciones de tratamientos permiten  $\binom{k}{2}$  interacciones de dos factores (cada uno con cuatro grados de libertad),  $\binom{k}{3}$  interacciones de tres factores,  $\dots$ , y una interacción de  $k$  factores (con  $2^k$  grados de libertad).

Las sumas de cuadrados de los efectos y las interacciones se calcularán con los métodos usuales. No hará falta hacer ninguna descomposición adicional de las interacciones de ordenes superiores.

Enseguida nos damos cuenta, que en este tipo de diseños se hace más notable el incremento de combinaciones posibles de tratamientos, el tamaño del diseño

se incrementará rápidamente con  $k$  pequeña. Por ejemplo, un diseño  $3^3$  hemos visto que tiene 27 combinaciones de tratamientos por réplica, un diseño  $3^4$  tendrá 81, un diseño  $3^5$  tendrá 243 combinaciones de tratamientos. Por este motivo muchas veces sólo se considerará una sola réplica del diseño  $3^k$ .

### 3. Diseño factorial fraccionado

#### 3.1. Introducción

En la sección anterior (sección 3) vemos que el problema principal de los diseños factoriales es el gran tamaño que toma el diseño, ya que se quieren estudiar todas las combinaciones posibles de tratamientos. Comentamos que una técnica para mejorar este problema sería fraccionar el diseño, mediante el uso de los diseños factoriales fraccionados.

Estos diseños desprecian las interacciones de algunos o todos los atributos, suponiendo que estas tienen una influencia despreciable en la respuesta (estos efectos quedan confundidos con 1), en otras palabras, los diseños factoriales fraccionados son formas de seleccionar subconjuntos de tratamientos combinados del diseño factorial completo sistemáticamente.

En general, los diseños factoriales fraccionados implican una pérdida de información estadística, requiriendo realizar supuestos sobre la no significatividad de efectos de orden superior (por ejemplo, interacciones entre dos o más atributos).

Una estrategia para reducir el número de escenarios de un diseño experimental adecuado sería utilizar una fracción ortogonal del diseño factorial completo. Por ejemplo, del diseño factorial completo de  $2^6$ , tomaríamos la fracción de  $2^{6-3} = 2^3 = 8$  conjuntos de elección. Este diseño conservaría la ortogonalidad entre los distintos atributos de toda la matriz donde se combinan los niveles de los atributos y permite estimar el modelo de utilidad sin multicolinealidad en las variables (lo que implicará que no existe una fuerte correlación entre variables explicativas del modelo, por lo tanto, todas las variables aportarán información al modelo) y haciendo que los atributos estén independientemente distribuidos.

En esta sección también veremos, cómo el uso de bloques nos será muy útil. El uso de bloques es un medio para reducir el error experimental con el fin de lograr una mayor precisión. Cualquier factor que afecta a la variable respuesta y tiene un rango de variación entre las unidades experimentales, aumenta la varianza del error y disminuye la precisión (cuanto más factores o atributos experimentales influyen en la variable respuesta, puede llevar a más error). La utilización de bloques agrupa las unidades experimentales, hace grupos homogéneos de forma que la variación de las unidades experimentales disminuye. Por ejemplo, un grupo formado por parcelas adyacentes formaría un grupo en un experimento de agronomía; en el campo de la salud, un experimento para contrastar los efectos de un tratamiento, un bloque sería agrupar a los pacientes por peso y estaturas similares.

### 3.2. Diseños factoriales fraccionados de dos niveles:

Veamos primero la técnica de reducción de combinaciones aplicada a los diseños factoriales de dos niveles, donde tengamos los atributos que tengamos, todos ellos tendrán dos niveles asociados.

Por ejemplo, en una réplica completa de un diseño  $2^6$  se requieren 64 combinaciones con 63 grados de libertad, de los cuales 6 corresponden a los efectos principales, y 15 de ellos a las interacciones de dos factores, los demás 42 grados de libertad se asocian con las interacciones de tres o más factores. Si suponemos que hay interacciones de orden superior que pueden despreciarse, este diseño se puede disminuir de tamaño, quedándonos con una fracción del mismo.

El uso de los diseños factoriales fraccionados lleva con frecuencia a una gran economía y eficiencia en la experimentación, en particular si las corridas pueden hacerse secuencialmente.

**Fracción un medio del diseño  $2^k$ :** Se denomina fracción  $1/2$ , cuando el diseño factorial fraccionado es el resultante de quitarle  $1/2$  a las  $2^k$  combinaciones posibles, es decir, cuando nos quedamos con un diseño  $2^{k-1}$  y reducimos el número de combinaciones de tratamientos a la mitad. Por ejemplo, cuando tenemos un diseño de tres factores con dos niveles cada uno, lo cual nos da  $2^3 = 8$  combinaciones de tratamientos, sin embargo, por algún motivo solo estamos en condiciones de realizar 4 corridas del experimento, en este caso utilizaremos una fracción un medio del diseño. De este modo tendremos  $2^{3-1} = 2^2 = 4$  (3 grados de libertad) combinaciones de los tratamientos, como queríamos.

Al utilizar la fracción  $1/2$ , hemos separado el diseño en dos fracciones, como vemos en la figura 11 y la figura 12.

La fracción principal que vemos en la figura 11 es la que nos servirá para obtener los efectos principales, junto a la tabla de signos que se muestra a continuación.

De este modo obtenemos los efectos de los factores principales

$$A = \frac{1}{2}(a - b - c + abc)$$

$$B = \frac{1}{2}(-a + b - c + abc)$$

$$C = \frac{1}{2}(-a - b + c + abc)$$

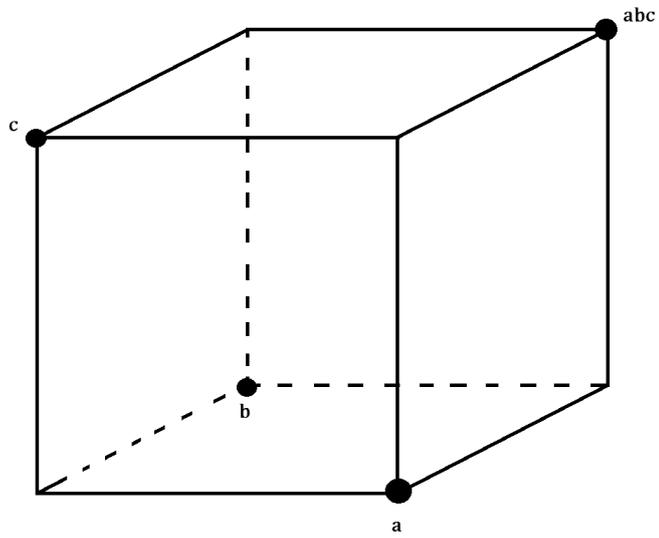


Figura 11. Fracción principal.

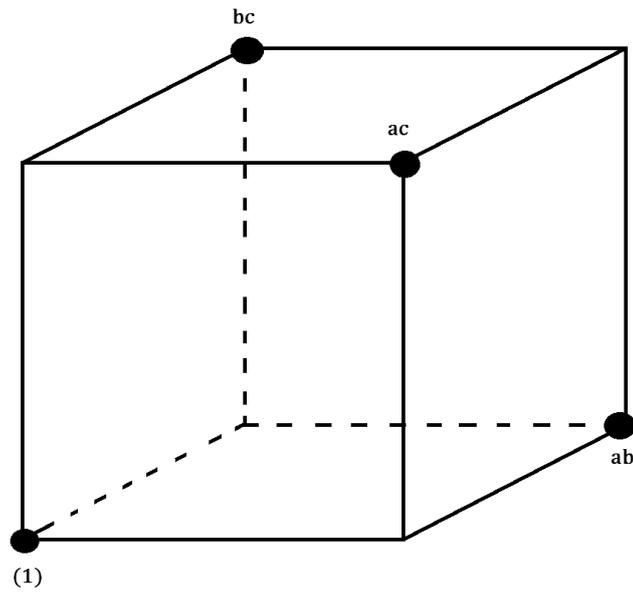


Figura 12. Fracción alterna.

Combinaciones de tratamientos	I	A	B	C	AB	AC	BC	ABC
a	+	+	-	-	-	-	+	+
b	+	-	+	-	-	+	-	+
c	+	-	-	+	+	-	-	+
abc	+	+	+	+	+	+	+	+
ab	+	+	+	-	+	-	-	-
ac	+	+	-	+	-	+	-	-
bc	+	-	+	+	-	-	+	-
(1)	+	-	-	-	+	+	+	-

**Cuadro 3.** Tabla de signos del diseño factorial  $2^3$ .

y los efectos para estimar las interacciones de dos factores son

$$AB = \frac{1}{2}(-a - b + c + abc)$$

$$AC = \frac{1}{2}(-a + b - c + abc)$$

$$BC = \frac{1}{2}(a - b - c + abc)$$

donde podemos observar que los efectos de A son iguales a los de BC, los efectos de B son iguales a los efectos de AC, y los efectos de C son iguales a los efectos de AB. A los efectos que poseen esta propiedad se les llama alias. En este caso, por ejemplo, se puede decir que el efecto de A es alias del efecto de la interacción BC.

Este tipo de diseños que acabamos de describir se les llamará diseños de resolución III, ya que los efectos principales son alias de los efectos de interacción. Según estén repartidos los alias, los diseños se pueden separar en:

- **Diseños de resolución III:** Se trata de diseños en los que ninguno de los efectos principales es alias de otro efecto principal, pero los efectos principales sí pueden ser alias de las interacciones de dos factores, y algunas de las

interacciones de dos factores pueden ser alias entre sí. Como el ejemplo del diseño  $2^{3-1}$  que acabamos de ver.

- **Diseños de resolución IV:** Se trata de diseños en los que ninguno de los efectos principales es alias de ningún otro efecto principal, ni lo es de ninguna de las interacciones de dos factores, pero las interacciones de dos factores son alias entre sí. Un ejemplo de este tipo de diseños es el diseño  $2^{4-1}$ .
- **Diseños de resolución V:** Se trata de diseños en los que ninguno de los efectos principales ni de las interacciones de dos factores son alias de otro efecto principal ni son alias de las interacciones de dos factores, pero las interacciones de dos factores pueden ser alias de las interacciones de tres factores.

El uso de los diseños factoriales fraccionados lleva con frecuencia a una gran economía y eficiencia en la experimentación, en particular si las corridas pueden hacerse secuencialmente. En el caso de los diseños en los que se aplica una fracción de  $1/2$ , cada una de las dos fracciones, la principal y la alterna, representan bloques del diseño completo con las interacciones de orden superior confundidas con los bloques (en el ejemplo que hemos visto del diseño  $2^{3-1}$ , ABC estará confundida).

**Fracción un cuarto del diseño  $2^k$ :** Acabamos de ver la utilidad de los diseños de fracción un medio, en los que reducimos el número de combinaciones de tratamientos a la mitad, sin embargo, hay ocasiones en las que será necesario reducir aun más el diseño, por lo que serán necesarias fracciones menores del diseño  $2^k$ .

La fracción  $1/4$  del diseño nos permitirá reducir el número de combinaciones de tratamientos en la cuarta parte. Este diseño tendrá  $2^{k-2}$  corridas y es común llamarlos diseño factorial fraccionado  $2^{k-2}$ .

Este tipo de diseños pueden ser construidos primero anotando el diseño básico compuesto por las corridas asociadas a un diseño factorial completo de  $(K-2)$  factores y asociando después las dos columnas con las interacciones elegidas que incluyan los  $k-2$  factores.

Veámoslo con un ejemplo, si tenemos un diseño factorial  $2^6$ , este necesita 64 corridas o combinaciones de tratamientos, sin embargo, suponemos que solo podemos realizar 16 corridas. En este caso utilizaremos una fracción un cuarto del diseño, y tendremos  $2^{6-2} = 2^4 = 16$  combinaciones posibles de tratamientos. Para construir este diseño se anota primero el diseño básico de 16 corridas de un diseño completo de 4 factores, A, B, C, D. Después se añaden los factores E, F, asociando sus niveles a los niveles de las interacciones de tres factores ABC y BCD.

Otra forma de construir este diseño es haciendo cuatro bloques del diseño  $2^6$  con ABCE y BCDF confundidas y eligiendo el bloque donde ABCE y BCDF son positivas (la fracción principal).

Veamos una tabla de signos que nos será de mucha ayuda, ya que se ve la estructura de la construcción del diseño.

Corrida	A	B	C	D	E= ABC	F= BCD
1	-	-	-	-	-	-
2	+	-	-	-	+	-
3	-	+	-	-	+	+
4	+	+	-	-	-	+
5	-	-	+	-	+	+
6	+	-	+	-	-	+
7	-	+	+	-	-	-
8	+	+	+	-	+	-
9	-	-	-	+	-	+
10	+	-	-	+	+	+
11	-	+	-	+	+	-
12	+	+	-	+	-	-
13	-	-	+	+	+	-
14	+	-	+	+	-	-
15	-	+	+	+	-	+
16	+	+	+	+	+	+

**Cuadro 4.** Construcción del diseño  $2^{6-2}$  de resolución IV.

Vemos cómo las cuatro primeras columnas (sin contar la primera que especifica el número de corridas) son los factores principales del diseño básico de un diseño  $2^4$ , y las dos últimas columnas son los efectos confundidos.

**Diseño factorial fraccionado  $2^{k-p}$ :** Al planear un experimento en forma fraccionada en los diseños  $2^k$ , como la finalidad es tratar de que cada nivel de los factores aparezca el mismo número de veces en el total de combinaciones de tratamientos (así se podrá medir sus efectos con igual precisión), se debe dividir el total de combinaciones de tratamientos por una potencia de dos, y utilizar una de esas fracciones para el experimento. A las nombradas fracciones en las que dividimos el total de combinaciones se les denominará bloques. Por ejemplo, si queremos dividir en dos bloques un experimento  $2^3$  (experimento que constará de 3 atributos con dos niveles para cada uno), el resultado será  $2^3/2 = 2^2 = 4$  combinaciones de tratamientos para cada bloque.

A un diseño factorial fraccionado  $2^k$  que contiene  $2^{k-p}$  combinaciones de tratamientos se le llama fracción  $1/2^p$  del diseño  $2^k$  o diseño factorial fraccionado

$2^{k-p}$ .

Cuando hablamos de diseños factoriales fraccionados  $2^{k-p}$ , se usa el símbolo  $2^{k-p}$  para designar un experimento fraccionado a  $1/2^p$ . Donde se eligen inicialmente  $p$  generadores independientes y sus  $2^p - p - 1$  interacciones generalizadas.

Deberá prestarse atención a la hora de elegir los generadores, para que los efectos de interés no sean alias entre sí. Como habíamos comentado en la sección 2, se debe tener especial cuidado en que los niveles de todos los factores aparezcan el mismo número de veces para que no haya dominancia de unos atributos o factores sobre otros, para evitar sesgos al analizar los efectos que tienen dichos atributos sobre la variable respuesta, lo cual puede llevar a errores a la hora de interpretar los resultados del experimento. Esto se podrá evitar haciendo una buena elección de los generados de modo que se obtengan las mejores relaciones de los alias para no confundir ningún efecto de interés, se suele suponer que las interacciones de ordenes superiores son no significativas, con lo cual simplifica la estructura de los alias (se suelen elegir los alias de forma que tengamos la resolución más alta, ya que lo que esto implica es que las interacciones superiores son las que son confundidas).

Las estimaciones de los efectos de los diseños factoriales fraccionados  $2^{k-p}$  suelen realizarse en computadora, sin embargo, este diseño también puede analizarse recurriendo a los métodos básicos utilizados en las secciones anteriores, de este modo el efecto  $i$ -ésimo se estima con la combinación lineal

$$l_i = \frac{2(\text{contrast}_i)}{N}$$

donde  $N = 2^{k-p}$  es el número de observaciones total. Además el  $\text{contrast}_i$  se puede calcular fácil utilizando la tabla de los signos en la columna correspondiente al factor  $i$ .

En ocasiones, los diseños factoriales después de aplicarles la fracción  $1/2^p$  siguen requiriendo igualmente de muchas corridas, tantas que no es posible realizarlas todas de forma homogénea. En estas ocasiones suele ser útil el uso de bloques.

Si por ejemplo, consideramos un diseño factorial fraccionado  $2^{6-2}$ , cuyo diseño se muestra en la tabla 4. Este diseño fraccionado contiene 16 combinaciones, se podrán dividir en dos bloques de 8 corridas cada uno.

Acabamos de ver la utilidad de los diseños factoriales fraccionados como técnica de reducción de combinaciones de tratamientos, concluimos resumiendo las decisiones que hace falta tomar en este tipo de diseños para mayor efectividad.

- Decidir la fracción del diseño que se va a utilizar, es decir, el valor de  $p$ . Teniendo en cuenta las corridas que se pueda permitir el investigador se elegirá un valor de  $p$ , para que el número de combinaciones sea el buscado.
- Decidir los efectos independientes que se van a confundir con 1, esto quiere decir saber los efectos que no se van a poder cuantificar (que en nuestro caso van a ser cuantificados los más posibles). Es deseable que no sean efectos principales. Se prefiere confundir interacciones de alto orden. Es decir, los alias serán entre las interacciones de ordenes superiores (los diseños serán de la más alta resolución posible).

### 3.3. Diseño factorial fraccionado $3^{k-p}$

Al igual que en los diseños factoriales fraccionados  $2^k$  las fracciones del diseño eran bases de 2, en este caso las fracciones serán base 3, es decir, se partirá el diseño en  $1/3, 1/3^2, \dots, 1/3^p$ .

La fracción más grande del diseño  $3^k$  es la fracción un tercio, la cual al aplicarla se dice que estamos frente a un diseño factorial fraccionado  $3^{k-1}$ . Esta fracción reduce el número de corridas en la tercera parte que las del diseño factorial completo.

Para construir un diseño factorial fraccionado  $3^{k-1}$  se selecciona una componente de interacción, normalmente la de orden más alto, y se hace una partición del diseño  $3^k$  completo en tres bloques con  $3^{k-1}$  combinaciones posibles de tratamientos en cada bloque, donde cada efecto principal o componente de interacción estimado a partir del diseño  $3^{k-1}$  tiene dos alias. Por ejemplo, si consideramos una fracción de un tercio del diseño  $3^3$ , puede seleccionarse cualquier componente de la interacción ABC para construir el diseño (ABC,  $AB^2C$ ,  $ABC^2$  o  $AB^2C^2$ ). Supongamos que elegimos la componente  $AB^2C^2$ . Cada fracción del diseño  $3^{3-1}$  resultante contendrá  $3^2 = 9$  combinaciones de tratamientos que deben satisfacer

$$x_1 + 2x_2 + 2x_3 = u \pmod{3}$$

donde  $u = 0, 1$  ó  $2$ .

Así que la estructura de los alias resultante nos da  $A = ABC = BC$ ,  $B = AC^2 = ABC^2$ ,  $C = AB^2 = AB^2C$ ,  $AB = AC = BC^2$ , por consiguiente solo se estimarán cuatro efectos a partir de los ocho grados de libertad del diseño. Además vemos que se trata de un diseño de resolución III, ya que los efectos principales son alias de efectos de interacción.

En general, las combinaciones de tratamientos en un diseño  $3^{k-1}$  pueden construirse utilizando un método similar al que se empleó en las series  $2^{k-p}$ . Primero se escriben las  $3^{k-1}$  corridas del diseño completo para  $k-1$  factores de tres niveles, y después se introduce el factor que falta con el componente apropiado de la interacción de orden más alto que cumpla  $AB^{\alpha_2}C^{\alpha_3}, \dots, (k-1)^{\alpha_{k-1}}$ , mediante la relación

$$x_k = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{k-1} x_{k-1}$$

donde  $\beta_i = (3 - \alpha_k)\alpha_i \pmod{3}$  para  $1 \leq i \leq k-1$ . De este modo se obtiene el diseño de resolución más alta.

Veámoslo con un ejemplo, si consideramos una fracción de un tercio para el diseño  $3^4$ , tenemos el diseño factorial fraccionado  $3^{4-1}$ , queremos construir este diseño. Los tres primeros de cada combinación de tratamientos serán las 27 combinaciones dadas por el diseño  $3^3$  completo. Ahora nos falta añadir la cuarta y última columna. Seleccionamos la componente de interacción  $AB^2CD$ , para la cual se tiene  $\alpha_1 = \alpha_3 = \alpha_4 = 1$  y  $\alpha_2 = 2$ , con lo cual se tiene

$$\begin{aligned}\beta_1 &= (3-1)\alpha_1 \pmod{3} = 3-1 = 2 \\ \beta_2 &= (3-1)\alpha_2 \pmod{3} = (3-1)2 = 4 \pmod{3} = 1 \\ \beta_3 &= (3-1)\alpha_3 \pmod{3} = 3-1 = 2\end{aligned}$$

por lo tanto,

$$x_4 = 2x_1 + x_2 + 2x_3$$

y con esta combinación se puede obtener el resto de combinaciones del diseño (sin más que sustituir los valores de  $x_1, x_2, x_3$  de las columnas anteriores).

Para un mayor fraccionamiento de los diseños factoriales  $3^k$  se recurre a un fraccionamiento mayor. Veamos como se tratan en general los fraccionamientos de estos diseños, que recibirán el nombre de diseños factoriales fraccionados  $3^{k-p}$ , donde se construye una fracción  $1/3^p$  del diseño  $3^k$  para  $p < k$ .

Al aplicar este tipo de diseños nos quedamos con  $3^{k-p}$  corridas del diseño. El procedimiento para construir estos diseños consiste en seleccionar  $p$  componentes en  $3^p$  bloques, donde cada bloque es un diseño factorial fraccionado  $3^{k-p}$ . Las corridas que definen un diseño factorial fraccionado  $3^{k-p}$  pueden generarse seleccionando primero las combinaciones de tratamientos de del diseño completo  $3^{k-p}$ , y después introduciendo los  $p$  factores que faltan mediante la selección de componentes de interacción, e igualándolos entre ellos, como hemos visto para el ejemplo de diseño factorial fraccionado  $3^{4-1}$ .

### 3.4. Conclusiones

A lo largo de esta sección hemos visto cómo los diseños factoriales son diseños eficientes para la obtención de los efectos entre los atributos. De este modo, estos diseños serán de gran ayuda para la obtención de preferencias. Haciendo uso de un buen diseño factorial, podemos ver qué atributos influyen o son más significativos en la variable respuesta.

Aunque los ejemplos que hemos ido dando a lo largo de la sección son ejemplos de diseños aplicados a procesos industriales, los diseños factoriales también pueden aplicarse al estudio de preferencias, ya que los diseños factoriales al mostrarnos los atributos que más influyen en la variable respuesta, nos indican las características que más influyen en la elección, es decir, en las preferencias de los individuos.

También hemos visto que este tipo de diseños tienen un inconveniente, ya que para hacer un estudio completo de los efectos tanto de los factores como de las interacciones, se necesita estudiar todas las posibles combinaciones de tratamientos. Esto hace que por pocos atributos que tengamos, como tenemos que combinar cada uno con los niveles del resto de atributos, el número de combinaciones o corridas para el experimento aumenta considerablemente rápido, tanto que a veces el experimentador no se puede hacer cargo de su coste. Este inconveniente puede llevar incluso a generar errores en los resultados del experimento, ya que si se trata de un estudio de preferencias, el encuestado al encontrarse con tantos escenarios de elección (tantos como combinaciones posibles de tratamientos en el diseño factorial completo) puede llegar a contestar sin seguir una lógica debido a la gran carga cognitiva que le está ocasionando.

Este inconveniente se soluciona con una técnica que hemos comentado en esta sección, el uso de diseños factoriales fraccionados. Estos diseños disminuyen el número de combinaciones o escenarios posibles, sin intervenir demasiado en la buena eficiencia de los diseños factoriales a la hora de estudiar los efectos de los atributos y sus interacciones, ya que supone que las interacciones de orden superior son poco significativas en el modelo, y por tanto despreciables.

Las conclusiones que obtenemos del diseño factorial fraccionado, es que mediante su uso podemos reducir el número de escenarios de elección a la cantidad deseada. De este modo, será el investigador el que decida sobre esta medida, una vez se tengan el número de atributos y los niveles para cada uno de ellos, el investigador decidirá qué fracción aplicar al diseño, según los objetivos de este. Lo que sí que intentará siempre el investigador es elaborar el diseño factorial fraccionado para al menos calcular los efectos principales.

Además, también comentamos durante la sección el uso de bloques, esta técnica puede ser una forma de que el investigador se asegure de que no se está obviando o pasando por alto ningún efecto de algún factor principal, ya que en

cada bloque se estudian los factores principales del objeto de estudio.

En el caso que nos ocupa, el investigador además de decidir el número de escenarios de elección utilizados en el diseño, tendrá que asegurarse de que no haya dominancia ni efectos principales confundidos. Es decir, en el diseño de este tipo de encuestas para obtener las preferencias, se tendrá que cumplir, por una parte que cada nivel aparezca el mismo número de veces, para evitar que haya niveles dominantes y por lo tanto tengan más probabilidad de ser elegidos. Por otra parte, para asegurarse de que los efectos principales no estén confundidos (correlacionados), cada nivel de un atributo se debe cruzar con cada nivel del resto de atributos el mismo número de veces. De esta forma, cumpliendo estas premisas, podemos asegurarnos de que el diseño del experimento se trata de un diseño eficiente.

## 4. Programación Visual Basic

Como comentamos en la sección 1, la finalidad del presente trabajo es el diseño de escenarios para el estudio de preferencias de pacientes y otras partes interesadas sobre tratamientos en cuidados de salud.

Para dar un enfoque en el ámbito de la salud del estudio de preferencias, encontramos en la literatura, que los últimos años se está incrementando el uso de las técnicas de análisis conjunto, en especial los experimentos de elección discreta tienen un enfoque aplicado al cuidado de la salud.

El ISPOR (International Society for Pharmacoeconomics and Outcomes Research) nos da unas pautas a seguir sobre las buenas prácticas para la realización de un análisis conjunto:

### 1. Identificación de los atributos y niveles:

El primer paso para la realización de un análisis conjunto será la elección de los atributos y distribuir sus niveles. En este paso se seleccionan todas las características que se quieran obtener sobre el tratamiento objeto de estudio (realizando revisiones en la literatura es la forma más común).

### 2. Selección de los atributos y niveles:

Una vez se tienen todas estas características, en el paso dos, se confirma si estos atributos seleccionados son las características más valoradas de dicho tratamiento. Solo serán de interés aquellos atributos mejor valorados o que cualifican de una mejor forma los tratamientos objeto de estudio. Además, una amplia variedad de atributos hace que el estudio de sus efectos no sea tan eficiente.

### 3. Diseño experimental y elaboración de los escenarios a presentar:

Este paso es el de nuestro cometido, en el que se diseña el formato y el número de elecciones a presentar al participante. Un diseño estadísticamente eficiente hace referencia a una minimización del rango de variación de los intervalos de confianza para los parámetros estimados.

Cuando hablamos de diseño experimental o diseño de experimentos, nos estamos refiriendo al proceso de generar combinaciones específicas de atributos y niveles que los encuestados deben evaluar en las preguntas o escenarios de elección. Para ello utilizaremos las técnicas de diseños factoriales fraccionados, ya que nos dan una técnica para la minimización de escenarios y unas premisas para la construcción de estas combinaciones.

Además de una minimización del número de escenarios, se va a querer minimizar la amplitud del intervalo de confianza para los atributos que se van a

estimar, de esta manera, tendremos un diseño estadísticamente eficiente.

Los diseños estadísticamente eficientes son aquellos que están balanceados, cada nivel aparece el mismo número de veces para un atributo, y se trata de un diseño ortogonal, es decir, que cada uno de los pares de niveles (en el escenario) aparece una misma frecuencia dentro de los pares de atributos en el diseño (Johson F.R. (2013)).

En la presente sección vamos a ver cómo se ha seguido el proceso de la construcción y programación en excel de estos escenarios. Para ello dividiremos la sección en dos partes, estas serán las partes que conforman la calculadora (excel). En la primera parte de la calculadora, calcularemos el número de escenarios de elección para un diseño óptimo, teniendo en cuenta el número de atributos y los niveles de cada uno de ellos. En la segunda parte, se mostrará un posible diseño (este no es único) de las combinaciones de escenarios.

#### 4.1. Número de escenarios de elección

Para esta primera parte del diseño de escenarios para la obtención de preferencias, nos centraremos en obtener el número de escenarios óptimo, es decir, el número mínimo de escenarios o combinaciones de niveles posibles para asegurar un buen diseño.

Utilizaremos las técnicas del diseño factorial fraccionado para nuestro propósito. En la sección 3 hemos visto cómo los diseños factoriales completos son diseños eficientes para el estudio de preferencias, debido a que son una técnica que facilita el estudio de los efectos entre los factores o atributos e incluso de las interacciones entre ellos, sin embargo, son diseños que a medida se introducen factores van creciendo de tamaño, ya que los diseños factoriales completos utilizan todas las posibles combinaciones de niveles para el estudio de sus efectos. Por este motivo, se convierte en una necesidad el uso de una variante de estos diseños, el diseño factorial fraccionado, el cual reduce el número de combinaciones o escenarios tanto como se desee. Pudiendo así ofrecer un diseño que es eficiente en la obtención de los efectos de los atributos y además con el tamaño o número de escenarios que más se adecue a las circunstancias.

En nuestro caso, para el estudio de preferencias declaradas, se tiene en cuenta que los escenarios de elección se tienen que presentar al encuestado de forma que este siempre utilice la lógica a la hora de hacer las elecciones (se intentará hacer todo lo que esté en nuestra mano). Para ello, se deberá evitar presentar un gran número de escenarios al encuestado.

Tras revisar la literatura, llegamos a la conclusión de que uno de los errores de racionalización más comunes se debe a la carga cognitiva que genera tener que responder un gran número de preguntas (en nuestro caso, las preguntas serían

los escenarios de elección). Ya que cuando se llega a las últimas preguntas, las respuestas ya no siguen la lógica que se precisa.

Por ello, nuestra prioridad a la hora de elegir el número de escenarios será obtener un número de escenarios, tal que, no se llegue a ocasionar esta carga cognitiva que influirá negativamente a la calidad de las respuestas.

Se han hecho estudios relacionados con la sobrecarga informativa que le supone al entrevistado el tener frente a él un número excesivo de escenarios de elección a los que responder. La mayoría de investigadores utilizan entre 10 y 25 escenarios o combinaciones de niveles como rango adecuado para presentar un diseño experimental en el que interviene la elección de los encuestados. En nuestro caso, por petición de los investigadores, fijaremos el número de escenarios máximo a presentar en 16 combinaciones de niveles de los atributos, y de este modo poder obtener resultados eficientes, sin sesgos por la carga cognitiva del encuestado.

En base a las investigaciones presentadas en este trabajo, obtenemos que la forma de llegar al número óptimo de escenarios presentados al encuestado para la obtención de preferencias declaradas, consiste en encontrar la fracción adecuada del diseño factorial de modo que el número de escenarios sea el deseado. Es decir, la tarea consistirá en encontrar la  $p$  del diseño factorial fraccionado  $2^{k-p}$ , si se trata de atributos con dos niveles cada uno, o la  $p$  del diseño factorial fraccionado  $3^{k-p}$ , si se tiene que el diseño consta de atributos con 3 niveles cada uno.

#### **Obtención del número de escenarios para el diseño factorial $2^k$ :**

En los casos en los que tenemos  $k$  atributos con dos niveles cada uno, para la obtención del número óptimo de escenarios, tendremos que encontrar la fracción  $1/2^p$  adecuada.

Por lo tanto, primero tendremos que obtenernos la  $p$  que nos dará la fracción del diseño imponiéndole que como máximo el número de escenarios presentados al encuestado será 16, de este modo, obtenemos

$$\frac{2^k}{2^p} \leq 16$$

entonces,

$$p = \frac{\log\left(\frac{2^k}{16}\right)}{\log 2}$$

donde  $k$  es conocida, ya que es el número de atributos que se quieren asociar al diseño.

Vemos que este cálculo de  $p$  no siempre nos dará un número entero, así que tomaremos  $p'$  como el entero más cercano al valor de  $p$ , para asegurarnos que no sobrepasamos los 16 escenarios de elección.

De este modo, el número óptimo de escenarios sería :

$$n^{\circ} \text{escenarios} = \frac{2^k}{2^{p'}}$$

### **Obtención del número de escenarios para el diseño factorial $3^k$ :**

Del mismo modo, en los casos en los que queremos calcular el número de escenarios óptimo para un diseño factorial con  $k$  factores de tres niveles cada uno. Obtendremos la  $p$  que nos dará el diseño factorial fraccionado  $3^{k-p}$  con la condición de que el número de escenarios no debe ser mayor a 16, y luego aplicaremos la fracción del diseño  $1/3^p$ . Así

$$p = \frac{\log(\frac{3^k}{16})}{\log 3}$$

Igual que antes redondeamos el valor de  $p$  al entero más cercano, y lo llamamos  $p'$ , de este modo, aplicándole la fracción  $1/3^{p'}$  al diseño obtenemos el número de escenarios.

$$n^{\circ} \text{escenarios} = \frac{3^k}{3^{p'}}$$

donde  $k$  es conocida, ya que es el número de atributos asociados al diseño, y  $p'$  es la fracción calculada.

### **Obtención del número de escenarios para el diseño factorial general:**

Acabamos de ver cómo obtener el número óptimo de escenarios de elección para los diseños factoriales  $2^k$  y  $3^k$ , donde no obtenemos diferencias significativas, con lo cual podemos generalizar este cálculo para atributos con cualquier nivel, es

decir, para diseños a los que se asocien  $k$  atributos con el mismo número de niveles  $h$ .

Del mismo modo que antes tenemos

$$p = \frac{\log(\frac{h^k}{16})}{\log h}$$

donde  $k$  y  $h$  son conocidos, ya que serán el número de atributos y sus niveles asociados al diseño.

Quedándonos entero más cercano asociado al valor de la  $p$  obtenida, tenemos  $p'$ , y de este modo, el número óptimo de escenarios viene dado por

$$n^{\circ}escenarios = \frac{h^k}{h^{p'}}$$

**Programación en Visual Basic para la obtención de número de escenarios, con el mismo número de niveles para cada atributo.**

Viendo la fórmula general, acabamos de ver que nos da el número de escenarios óptimo para el diseño factorial fraccionado, el cual nos dará un diseño eficiente para un experimento de elección discreta, en el que se busca como resultado el estudio de preferencias declaradas.

Su programación en Visual Basic será muy sencilla como se muestra en el siguiente código, donde  $nA$  es el número de niveles de cada atributo y  $a$  el número de atributos.

```
Public Function escenarios1(nA, a)
Dim p As Long
If ((nA ^ a) < 16) Then
    escenarios1 = nA ^ a
Else
p = Log((nA ^ a / 16)) / Log(nA)
If ((p - Int(p)) <= 0.5) Then
    p = Int(p)
Else
    p = Int(p) + 1
End If
escenarios1 = (nA ^ a) / (nA ^ p)

End If

End Function
```

Esta fórmula general solo se puede aplicar en el caso de que todos los atributos del diseño tengan el mismo número de niveles.

Por otra parte, para asegurar que el número de escenarios obtenido da lugar a una combinación de niveles de atributos cuyo diseño sea estadísticamente eficiente, encontramos en el ISPOR TASK FORCE REPORT sobre buenas practicas en el análisis de conjunto, para que un diseño sea estadísticamente eficiente tendremos que ratificar que:

1. Para cada atributo, el número de veces que aparece cada uno de sus niveles es el mismo, a los diseños con esta propiedad se les llamará diseños balanceados.
2. Cada par de niveles emparejado (se refiere a la combinación de dos niveles por atributo en cada escenario) aparece el mismo número de veces para cada atributo. A esta propiedad se le llama ortogonalidad.

De esta manera lo que tenemos que buscar es un diseño ortogonal y balanceado, esto requiere que el número total de alternativas debe ser divisible por el número de niveles de cada atributo. Una forma más eficiente de hacerlo es separando los escenarios por bloques (de forma que en cada bloque no haya más de 16 escenarios), es decir, de lo que tenemos que asegurarnos es de que el número de escenarios en cada bloque sea divisible por el número de niveles de cada atributo.

Por tanto, en el caso que nos ocupa donde tenemos  $k$  atributos con  $h$  niveles (los atributos asociados al diseño tienen todos el mismo número de niveles), solo tendremos que asegurarnos de que el número de escenarios obtenido sea divisible por  $h$ . Lo cual no hará falta, ya que al tener todos los atributos el mismo número de niveles, para obtener el número de escenarios estamos dividiendo una potencia del número de niveles entre otra potencia (se entiende que menor que la del numerador) del número de niveles, de este modo el resultado obtenido también será una potencia de niveles.

Es decir, tendremos que  $escenarios = h^{k-p'}$ , por consiguiente, el número de escenarios es divisible por  $h$ , ya que  $escenarios \bmod h = 0$ .

De igual modo, creamos una función Visual Basic, donde nos aseguraremos de que el número de escenarios por bloques sea divisible por el número de niveles de los atributos. Esta la utilizaremos, por un lado, a modo de test para comprobar y asegurarnos de que hemos obtenido correctamente el número de escenarios, y por otro lado, es una forma de clasificar los escenarios en bloques, ya que como hemos visto durante el trabajo, la partición de los escenarios en bloques es una buena técnica para minimizar el número de escenarios.

Haciendo uso de las congruencias, una forma de asegurarse que un número es divisible por otro, es si su módulo es 0.

Utilizamos la siguiente función en Visual Basic, donde *es* se refiere al número de escenarios que será obtenido en la función anterior, *nA* como ya habíamos dicho es el número de niveles y *a* el número total de atributos.

```
Public Function bloquesok1(es, nA, a)
  If (es > 16) Then
    If (es / 2 Mod nA = 0) Then
      bloquesok1 = 2
    ElseIf (es / 3 Mod nA = 0) Then
      bloquesok1 = 3
    ElseIf (es / 4 Mod nA = 0) Then
      bloquesok1 = 4
    Else
      bloquesok1 = 5
    End If
  Else
    bloquesok1 = 1
  End If
End Function
```

Podemos observar que hacemos la comprobación de los bloques para un máximo de 5, ya que tener más de 5 bloques sería un caso extremo. Supondría un número de escenarios totales (después de aplicar la fracción al diseño completo) muy elevado, cosa que solo se obtendría con una cantidad de atributos y niveles alto. Como hemos comentado a lo largo del trabajo, esto empeora el diseño.

### **Obtención del número de escenarios para diseños mixtos con 2 y 3 niveles.**

Hasta ahora hemos visto la programación del número de escenarios para el caso en el que el número de niveles es el mismo para todos los atributos. En esta sección veremos el caso en el que no todos los atributos tienen el mismo número de niveles, en particular, habrán dos grupos de atributos, cada uno con un número de niveles, 2 y 3 niveles en concreto.

Sea un diseño factorial comprendido por *A* atributos con 2 niveles y *B* atributos con 3 niveles, tendremos que el número de escenarios del diseño factorial completo será  $2^A \times 3^B$ . Del mismo modo que en la sección anterior, tendremos que buscar una fracción adecuada del diseño para que el número de escenarios no exceda de 16.

En este caso, las fracciones buscadas serán potencia de 2 y 3, el número de niveles asociado al diseño. Por tanto, habrá que dividir el número de escenarios totales del diseño factorial completo entre  $(1/2^{p_1}) \times (1/3^{p_1})$ .

El número de escenarios óptimo obtenido será de la forma

$$\frac{2^A}{2^{p_1}} \times \frac{3^B}{3^{p_1}}$$

Tendremos que encontrar el valor de  $p_1$  que nos darán un número de escenarios con menos de 16 escenarios por bloque y con el número de escenarios en cada bloque divisible entre 2 y 3.

Primero nos centraremos en encontrar un número de escenarios que cumpla la restricción por carga cognitiva, de forma que

$$\frac{2^A \times 3^B}{2^{p_1} \times 3^{p_1}} \leq 16$$

es decir,

$$\frac{2^A \times 3^B}{(2 \times 3)^{p_1}} \leq 16$$

de este modo,

$$p_1 = \frac{\log\left(\frac{2^A \times 3^B}{16}\right)}{\log(2 \times 3)}$$

donde  $A$  es conocido ya que se trata del número de atributos de dos niveles y  $B$  también, ya que es el número de atributos de tres niveles. Del mismo modo que antes, redondeamos el valor de  $p_1$  al entero más próximo,  $p'_1$ .

En el siguiente paso, el cálculo de escenarios aplicando la fracción del diseño obtenida, tenemos que tener especial cuidado cuando el diseño tiene asociados atributos con distintos niveles, ya que el valor de  $p'_1$  obtenido puede resultar mayor que alguna de las potencias, es decir, se puede dar el caso, por ejemplo, de que  $A \leq p_1$ , y esto nos lleva a que el número de escenarios resultante pueda no ser divisible entre el número de los niveles de atributos, e incluso se puede dar que el número de escenarios obtenidos al aplicar la fórmula de fraccionamiento no sea una cantidad entera, lo cual es un resultado erróneo.

Esto nos lleva a distinguir casos a la hora de calcular el número de escenarios, dependiendo de si el valor obtenido de  $p'_1$  es mayor o no que  $A$  o que  $B$ . El modo de solucionar esta circunstancia es, para el nivel asociado a los atributos cuyo valor es superado por  $p'_1$  ( $A$  ó  $B$ ), tomaremos la medida de aplicarle la mayor fracción posible  $q$  de modo que  $A - q = 1$  (así tenemos asegurado que el número de escenarios es divisible entre el número de niveles), y al otro nivel se le aplicará la fracción  $p_1 + 1$  como sigue.

Si suponemos que el valor obtenido  $p'_1$  es mayor que  $A$  tendremos,

$$escenarios = \frac{2^A \times 3^B}{2^{A-1} \times 3^{p_1+1}}$$

es decir,

$$escenarios = \frac{2 \times 3^B}{3^{p_1+1}}$$

### **Obtención del número de escenarios para diseños factoriales mixtos.**

En este apartado vamos a generalizar las pautas vistas en el apartado anterior para la obtención del número de escenarios, discutiendo el caso en el que tenemos un diseño factorial completo con atributos con dos clases de niveles asociados (las clases de niveles hace referencia a distintas cantidades de niveles).

Del mismo modo, llamando  $A$  a los atributos de una clase y  $nA$  al número de niveles asociados a dichos atributos, y llamando  $B$  a los atributos de la otra clase y  $nB$  al número de niveles asociado a estos atributos, se obtiene primero la fracción, valor de  $p_1$  utilizando la siguiente fórmula,

$$p_1 = \frac{\log\left(\frac{nA^A \times nB^B}{16}\right)}{\log(nA \times nB)}$$

donde, del mismo modo que en los demás casos, para calcular el número de escenarios nos quedaremos con  $p'_1$ , el entero más próximo a  $p_1$ .

Mientras el cálculo de los escenarios se queda como sigue:

$$escenarios = \frac{nA^A \times nB^B}{(nA \times nB)_{p_1}^{p_1}}$$

en el caso de que  $p_1' > A$ , tenemos

$$escenarios = \frac{nA \times nB^B}{nB^{p_1'+1}}$$

y en el caso de que  $p_1' > B$ , tenemos

$$escenarios = \frac{nA^A \times nB}{nA^{p_1'+1}}.$$

El caso en el que  $p_1' > A$  y  $p_1' > B$ , no lo contemplamos ya que no se dará, por la forma en la que está obtenida  $p_1'$ .

### **Programación en Visual Basic para la obtención del número de escenarios, con distintos niveles para los atributos.**

En esta sección vamos a ver las funciones que hemos utilizado en Visual Basic para el caso general de un diseño con atributos con un número de niveles asociados y otros atributos con otro número de niveles asociado, de esta forma, tendremos dos clases de atributos, los atributos  $A$  con  $nA$  número de niveles asociado y los atributos  $B$  con  $nB$  número de niveles asociado. Con lo que estamos trabajando con un diseño factorial completo de  $nA^A \times nB^B$  combinaciones de niveles posibles para la construcción de los escenarios.

Tal como hemos visto en la sección anterior, el primer paso para calcular el número de escenarios óptimo, es encontrar la  $p_1$  que nos dará la fracción del diseño que vamos a aplicar, como vemos en el siguiente código, esta computación es muy similar a la realizada para la obtención del número de escenarios en el caso de tener todos los atributos de un mismo nivel.

```

Public Function op2(a, B, nA, nB)
Dim p As Long
If (nA = 0) Then
p = (Log((1 ^ a) * (nB ^ B) / 16)) / Log(nB)
  If (p < 0) Then
    op2 = 0
  ElseIf (p - Int(p) < 0.5) Then
    op2 = Int(p) 'esta me devuelve el menor entero
  Else
    op2 = Int(p) + 1
  End If

ElseIf (nB = 0) Then
p = (Log((nA ^ a) * (1 ^ B) / 16)) / Log(nA)
  If (p < 0) Then
    op2 = 0
  ElseIf (p - Int(p) < 0.5) Then
    op2 = Int(p) 'esta me devuelve el menor entero
  Else
    op2 = Int(p) + 1
  End If

Else
p = (Log((nA ^ a) * (nB ^ B) / 16)) / Log(nA * nB)
  If (p < 0) Then
    op2 = 0
  ElseIf (p - Int(p) < 0.5) Then
    op2 = Int(p) 'esta me devuelve el menor entero
  Else
    op2 = Int(p) + 1
  End If
End If

End function

```

Podemos observar que en este caso también estamos teniendo en cuenta el caso en que el número de niveles es el mismo para todos los atributos, es decir, el caso en el que o bien  $nA = 0$  o bien  $nB = 0$ , este paso quizás no haría falta pero lo añadimos a modo de asegurarnos e intentar salvar errores.

El siguiente paso para el cálculo de escenarios, sería obtener el número de estos, aplicando la fracción con el valor de  $p'_1$  obtenido que nos devuelve la función anterior.

```

Public Function escenarios2(p, a, B, nA, nB)
If (nA = 0) Then
    escenarios2 = (nB ^ B) / ((nB) ^ p)
ElseIf (nB = 0) Then
    escenarios2 = (nA ^ a) / ((nA) ^ p)
Else
    'tengo que distinguir los casos, para siempre tener un numero
    de escenarios m ltiple del numero de niveles
    If (B - p <= 0) Then
        escenarios2 = ((nA ^ a) * nB) / (nA ^ (p + 1))
    ElseIf (a - p <= 0) Then
        escenarios2 = (nA * (nB ^ B)) / (nB ^ (p + 1))
    Else
        escenarios2 = (nA ^ a) * (nB ^ B) / ((nA * nB) ^ p)
    End If
End If
End Function

```

La función que acabamos de ver nos devolverá el número total de escenarios, a modo de test nos creamos una función para distribuir estos escenarios en bloques, tal como habíamos hecho en el caso anterior.

```

Public Function bloques(esc)
If (esc <= 16) Then
    bloques = 1
Else
    If ((Int(esc / 2) - esc / 2) = 0 And (esc / 2) <= 16)
        Then
            bloques = 2
        ElseIf ((Int(esc / 3) - esc / 3) = 0) Then
            bloques = 3
        Else
            bloques = 4
        End If
End If
End Function

```

```

Public Function bloquesok2(es, B, nA, nB)

```

```

If (nA = 0 Or nB = 0) Then
    bloquesok2 = B
Else
    If ((es / B) Mod nA = 0 And (es / B) Mod nB = 0) Then
        bloquesok2 = B
    Else
        bloquesok2 = B + 1
    End If
End If

End Function

```

Esta función la hemos dividido en dos, en la primera dividimos los escenarios obtenido en bloques y en la segunda pasamos el test, para comprobar que el número de escenarios en cada bloque sea divisible por el número de niveles de cada atributo.

Acabamos de ver cómo programar las funciones que nos han llevado a la obtención del número de escenarios óptimo, en el caso de estar frente a un diseño factorial con dos clases de atributos según el número de niveles asociado a cada uno de ellos. Es decir, hemos programado el cálculo de los escenarios de un diseño factorial fraccionado, donde su diseño factorial completo tiene atributos asignados a dos números de niveles.

A continuación veremos que para obtener el número de escenarios óptimo en el caso de tener un diseño factorial con tres clases de atributos según sus niveles se procede de forma análoga.

Sea  $A$  el número de atributos con  $nA$  niveles (donde  $nA$  es la cantidad de niveles asociados al atributo  $A$ ), sea  $B$  el número de atributos con  $nB$  niveles y sea  $C$  el número de atributos con  $nC$  cantidad de niveles, se obtiene el número de escenarios óptimo cuyo diseño factorial completo tendrá  $nA^A \times nB^B \times nC^C$  combinaciones de niveles posibles, como sigue.

Primero obtendremos el valor de la fracción que se le va aplicar al diseño factorial.

```

Public Function op(A, B, nA, nB, C, nC)

    p = (Log((nA ^ A) * (nB ^ B) * (nC ^ C) / 16)) / Log(nA *
        nB * nC)
    If (p < 0) Then
        op = 0
    ElseIf (p - Int(p) < 0.5) Then

```

```

op = Int(p) 'esta me devuelve el menor entero
Else
op = Int(p) + 1
End If

End Function

```

Una vez obtenida la fracción, para calcular el número de escenarios observamos que se procede de forma análoga al caso anterior pero con más casos.

```

Public Function Escenarios(p, a, B, c, nA, nB, nC)

If (a - p <= 0 And B - p <= 0 And c - p > 0) Then
Escenarios = ((nC ^ c) * nB * nA) / (nC ^ p)
ElseIf (a - p <= 0 And B - p > 0 And B - p <= 0) Then
Escenarios = (nC * nA * (nB ^ B)) / (nB ^ p)
ElseIf (a - p <= 0 And B - p > 0 And c - p > 0) Then
Escenarios = (nA * (nB ^ B) * (nC ^ c)) / ((nB * nC) ^
p)
ElseIf (a - p > 0 And B - p <= 0 And c - p <= 0) Then
Escenarios = ((nA ^ a) * nB * nC) / (nA ^ p)
ElseIf (a - p > 0 And B - p > 0 And c - p <= 0) Then
Escenarios = ((nA ^ a) * (nB ^ B) * nC) / ((nA * nB) ^
p)
ElseIf (a - p > 0 And B - p <= 0 And c - p > 0) Then
Escenarios = ((nA ^ a) * (nC ^ c) * nB) / ((nA * nC) ^
p)
Else
Escenarios = (nA ^ a) * (nB ^ B) * (nC ^ c) / ((nA *
nB * nC) ^ p)
End If

End Function

```

Y del mismo modo que antes, dividimos los escenarios en bloques.

```

Public Function bloques(esc)
If (esc <= 16) Then
bloques = 1
Else
If ((Int(esc / 2) - esc / 2) = 0 And (esc / 2) <= 16)
Then
bloques = 2
ElseIf ((Int(esc / 3) - esc / 3) = 0) Then

```

```

        bloques = 3
    Else
        bloques = 4
    End If

End If
End Function

```

Le pasamos una función a modo de test, que comprueba si el número de escenarios es divisible por el número de niveles de cada atributo.

```

Public Function bloquesok(es , B, nA, nB, nC)
If (nA = 0 And nB = 0) Then
    bloquesok = B
ElseIf (nA = 0 And nB <> 0 And nC <> 0) Then
    If ((es / B) Mod nC = 0 And (es / B) Mod nB = 0) Then
        bloquesok = B
    Else
        bloquesok = B + 1
    End If
ElseIf (nA <> 0 And nB = 0 And nC <> 0) Then
    If ((es / B) Mod nA = 0 And (es / B) Mod nC = 0) Then
        bloquesok = B
    Else
        bloquesok = B + 1
    End If
Else
    If ((es / B) Mod nA = 0 And (es / B) Mod nC = 0 And (es / B
    ) Mod nB = 0) Then
        bloquesok = B
    Else
        bloquesok = B + 1
    End If
End If

End Function

```

Podemos observar que la computación para obtener el número de escenarios en el caso que acabamos de ver, donde el diseño está comprendido por tres clases de niveles asociados a los atributos, es análoga a la computación del caso que hemos visto anteriormente, en el que el diseño tiene dos clases de atributos según sus niveles. La única diferencia significativa es que se tiene que tener en cuenta más casos a la hora de programar, ya que estamos ante un nivel más que en el

anterior.

Se podría pensar frente a un diseño con  $A$  atributos de  $nA$  niveles asociados,  $B$  atributos de  $nB$  niveles asociados,  $C$  atributos con  $nC$  niveles asociados y  $D$  atributos con  $nD$  niveles, que la manera de proceder para el cálculo óptimo del número de escenarios sería la misma o muy similar, sin embargo, si tenemos en cuenta que al añadir otra cantidad de niveles al diseño estamos aumentando considerablemente el número de escenarios (aunque el número de atributos asociados a este nivel sea tan solo 1), lo cual hace tarea casi imposible obtener un número de escenarios que cumpla la restricción por carga cognitiva de 16 escenarios de elección, cumpliendo a la vez que el número de escenarios por bloque sea divisible entre el número de niveles para cada atributo.

Si consideramos el menor diseño posible con estas características, es decir, el menor diseño que tiene cuatro niveles asociados a cuatro clases de atributos asociados a tales niveles. Suponemos cuatro atributos con cada uno un número de niveles distintos, tomamos las menores cantidades de niveles posibles, de este modo, estamos ante un diseño factorial con 1 atributo de dos niveles (ya que no tendría sentido tener 1 atributo de 1 nivel), 1 atributo de 3 niveles, un atributo de 4 niveles y por último 1 atributo con 4 niveles asociados. En este caso estamos ante un diseño factorial completo con las siguientes combinaciones de niveles que conforman el escenario de elección.

$$nA^A \times nB^B \times nC^C \times nD^D = 2 \times 3 \times 4 \times 5 = 5! = 120$$

Ya que a la hora de minimizar el número de escenarios se tiene que tener en cuenta que la cantidad de escenarios por bloque tiene que ser divisible entre el número de los distintos niveles del diseño, no podremos aplicar una fracción posible al diseño, ya que esto nos quitaría la divisibilidad del número de escenarios por los niveles.

Con el ejemplo que hemos considerado, obtenemos que el número como mínimo de escenarios por bloque será  $3 \times 4 \times 5 = 60$ , ya que 4 es divisible por 2, el 2 lo podemos excluir. Sin embargo, 60 no es un número de escenarios por bloque aceptable, ya que se estaría incumpliendo la restricción de 16 escenarios por bloque, y este diseño no obtendría resultados eficientes a la hora de analizar las preferencias de los encuestados, se obtendrían sesgos debido a la carga cognitiva que supondría responder a 60 preguntas de elección.

Vemos que utilizando las mínimas cantidades posibles para el caso en el que un diseño factorial consta de 4 clases de atributos, una para cada cantidad de niveles, se tiene que no se puede obtener un diseño factorial fraccionado (no podemos obtener la fracción que nos dé el número de escenarios óptimo) que cumpla las restricciones para un buen diseño.

#### 4.2. Diseño de los escenarios de elección

En esta sección vamos a ver la programación en Visual Basic de la parte de la calculadora en la que se muestra un posible diseño de los escenarios obtenidos.

Como hemos ido mencionando a lo largo del trabajo, nuestra finalidad es obtener un diseño eficiente (estadísticamente hablando) de los escenarios de elección o encuestas de elección presentadas a una serie de encuestados, para poder realizar un estudio de preferencias declaradas mediante un experimento de elección discreta.

La técnica utilizada para obtener los diseños eficientes de los escenarios de elección es el diseño factorial, este nos da un método sencillo para la reducción del número de escenarios de elección (combinaciones de niveles del diseño), aplicando una fracción al diseño factorial, como hemos visto hasta ahora (diseño factorial fraccionado).

Por otra parte, una vez se tiene el número de combinaciones de niveles de atributos que van a formar el diseño, se tendrán que elaborar los pares de elección, hablamos de pares de elección ya que el método utilizado para el análisis de las preferencias serán experimentos de elección binaria (dos opciones a elegir una). Para elaborar dichos pares de elección (escenarios de elección) se procurará que el diseño sea estadísticamente eficiente. Como ya habíamos comentado, según las recomendaciones del ISPOR, un diseño estadísticamente eficiente, tiene que ser un diseño balanceado, es decir, que el número de niveles por atributo aparezca el mismo número de veces, y además tendrá que tratarse de un diseño ortogonal, con lo cual, los pares de niveles tienen que aparecer el mismo número de veces para todos los atributos.

El diseño ortogonal balanceado que acabamos de describir asegura que las estimaciones de los parámetros no estén correlacionadas y sean independientes de los otros atributos, proporcionando resultados robustos para todos los niveles evaluados, sin embargo, este puede generar alternativas dominadas. Las alternativas dominadas son aquellas donde todos los niveles de una opción son sin duda mejores que los de la otra alternativa de elección (Tribaldos (2016)).

Para generar este tipo de diseños, vamos a estudiar antes el algoritmo Mix-and-Match. En términos generales, la expresión "Mix-and-Match" se utiliza para definir que las partes o elementos a los que se refiere se pueden combinar entre sí sin problemas, y esta definición es justo lo que nosotros queremos hacer, queremos combinar los niveles de atributos del diseño.

El algoritmo Mix-and-Match es un algoritmo utilizado en criptografía para la computación de texto cifrado, ya que la codificación trata de encontrar combinaciones lógicas que cifren el texto de forma que sea ininteligible para la

persona que no conozca la clave. Este algoritmo está comprendido por una serie de operaciones lógicas.

Este algoritmo aplicado al diseño de escenarios procede de la siguiente forma, escoge unos escenarios generados como referencia y modifica cada nivel de cada atributo sumándole una constante, así se genera otros escenarios.

Otras ideas de computación que nos serán de ayuda para el diseño de los escenarios de elección serán las funciones de soporte e implementación de experimentos de elección básicas utilizadas en el paquete estadístico de R, **support.CEs** (H.Aizaki (2012)). Este paquete está comprendido por funciones que para crear el diseño de experimentos de elección se basan en matrices de efectos principales ortogonales, tratan el diseño de los experimentos de elección como matrices ortogonales. No entraremos a estudiar estas funciones de R, pero sí que nos quedaremos con la idea principal que manejan para el diseño de los escenarios de experimentos de elección. Esta idea en general será generarse matrices ortogonales e ir comparando dos a dos, y quedarse con la de menor dominancia.

Formalmente, diremos que una matriz cuadrada  $A$  es dominante por columnas si se tiene:

$$a_{i,j} > \sum_{i=1}^n a_{i,j}, \forall j \in 1, 2, \dots, n$$

La dominancia en un diseño de elección hace que el diseño no aporte información para el estudio de las preferencias, ya que si hay un escenario de elección dominante para una alternativa siempre se escogerá este.

Como hemos comentado, se tratará el diseño de los escenarios de elección como una matriz. Estas matrices tendrán como filas los niveles asociados a cada atributo, y como columnas los escenarios de elección, que se tratarán como pares, ya que cada escenarios contiene dos alternativas de elección.

Veámoslo mejor con un ejemplo. Suponemos que tenemos que mostrar el diseño en forma de matriz de un experimento de elección, comprendido por dos atributos con dos niveles cada uno. Se tratará de un diseño factorial  $2^2$ , al que no hará falta aplicarle una fracción, ya que el número de escenarios o combinaciones de niveles posibles será  $2^2 = 4$ . En la siguiente tabla (tabla 5) se muestra un posible diseño de los escenarios de elección para este diseño.

Donde, A111 y A121 serán las alternativas 1 y 2 del escenario 1, respectivamente. Del mismo modo, A112 será la alternativa 1 del escenario 2 y A122 la alternativa 2 del escenario 2, el resto sigue de forma análoga, el primer índice indica la alternativa o tratamiento y el segundo índice se refiere al escenario de elección en el que se encuentra. Por otra parte, a1 se refiere al nivel 1 del atributo A, a2 al nivel 2 del atributo A, mientras que b1 y b2 indican el nivel 1 y 2 del

	A111	A121	A112	A122	A113	A123	A114	A124
Atributo A	a1	a2	a1	a1	a2	a2	a2	a1
Atributo B	b1	b2	b2	b1	b1	b2	b2	b1

**Cuadro 5.** Diseño de escenarios matricial.

atributo B, respectivamente.

En la tabla 5 podemos observar cómo este diseño cumple las premisas de un diseño estadísticamente eficiente, ya que se trata de un diseño ortogonal y balanceado. Vemos que tanto para el atributo A como para el atributo B, se tiene que cada nivel aparece el mismo número de veces, es decir, el nivel a1 aparece 4 veces a lo largo del diseño igualmente el nivel a2 aparece 4 veces, lo mismo ocurre con los niveles del atributo B, con lo cual esto indica que el diseño está equilibrado o balanceado como habíamos denominado anteriormente a esta propiedad. En cuanto a la ortogonalidad, la definíamos como la característica de que en el diseño los pares de niveles asociados para cada atributo tengan la misma frecuencia, es decir, que estas parejas aparezcan el mismo número de veces en los escenarios. Si nos fijamos en la tabla 5 vemos que cada pareja asociada a cada alternativa de cada escenario aparece dos veces, es decir, tanto el par (a1,b1), (a1,b2), (a2,b2), (a2,b1) aparece el mismo número de veces, como queríamos.

Una vez clara la estructura matricial que tiene que seguir el diseño, vamos a ver cómo programar la construcción de esta matriz. La idea que se va a seguir será construir una matriz de números enteros aleatorios positivos, definiremos el rango de estos como el número de niveles asociados a cada atributo, de este modo, si por ejemplo tenemos que el atributo A tiene 3 niveles, asignaremos 3 como valor de  $n_A$ , por lo tanto nos tendremos que generar aleatoriamente un número entero entre 1 y 3. Esta matriz será construida de forma que el diseño sea ortogonal y balanceado.

Para ello, computaremos una función en Visual Basic donde nos asignaremos estos números enteros aleatorios primero a las columnas impares, y luego para las columnas pares asignaremos números aleatorios también, pero teniendo en cuenta que el número asignado no sea igual al de la columna anterior, ya que este par de columnas será un escenario de elección (no queremos elegir entre dos alternativas de elección iguales), para a continuación recorrer la matriz ya construida del diseño y hacer las comprobaciones pertinentes de diseño ortogonal y balanceado.

A continuación vamos a ver un ejemplo para el caso más sencillo, en el que tenemos dos atributos, cuyos niveles son  $n_A$  y  $n_B$ , en este caso nos dará igual que sean iguales o distintos los niveles para cada atributo, ya que tendremos que fijar el rango para el que se genera el número entero aleatorio que asignaremos

a cada atributo en cada columna. Ya que se trata de un código largo vamos a separarlo en tres partes, aunque tenemos que tener en cuenta que se trata de la misma función dentro de Visul Basic.

En la primera parte mostrada a continuación, vemos cómo asignamos a las columnas impares números aleatorios, imponiendo la condición de que en cada fila se le asigne un número aleatorio dentro de un rango especificado por el número de niveles que corresponde al atributo de dicha fila.

```

Private Sub Atri2_Click()

Dim nA As Integer , nB As Integer
nA = Range("N27").Value
nB = Range("N29").Value

Atri = Range("S27").Value
esc = Range("S30").Value
Dim i As Integer , j As Integer , k As Integer , n As Integer , n1
    As Integer

ReDim Aux(Atri , (2 * esc)) As Integer
Dim m1 As Integer

'Por ahora solo le voy a asignar a las columnas impares
    n meros aleatorios asignando niveles para cada atributo

' por separado para cada fila , es decir , atributo
For k = 1 To esc
    For j = 1 To 2 * esc
        If (j = (2 * k) - 1) Then

            Randomize
            n = (nA - 1) * Rnd + 1 'primera fila , atributo 1

            Aux(1 , j) = n

            Randomize
            n = (nB - 1) * Rnd + 1 'segunda fila , atributo
                2
            n = Int(n)
            Aux(2 , j) = n

        End If
    Next j

```

```

Next k
' Aqui ya tengo asignado para cada escenario de tratamiento 1
  un nivel de cada atributo

```

Como habíamos comentado, vemos que para la primera fila le asignamos un número entero aleatoriamente entre 1 y nA (número de niveles que tendrá el atributo A), mientras que en la segunda fila asignamos el número entero aleatoriamente entre 1 y nB (número de niveles para el atributo B).

A continuación vemos cómo asignar el entero aleatorio correspondiente a las columnas pares, en este caso, solo vamos a exigir que no sea el mismo número que el asignado a la columna anterior.

```

'ahora voy a asignar niveles a las columnas pares

For k = 1 To esc
For j = 1 To 2 * esc
  If (j = 2 * k) Then

    Randomize
    n = (nA - 1) * Rnd + 1
    m1 = Int(n) 'Atributo1
                ' si ese nivel i ya ha salido , no
                puede salir
    Do While m1 = Aux(1, j - 1)

      Randomize
      n = (nA - 1) * Rnd + 1
      m1 = Int(n)

    Loop
    Aux(1, j) = m1

    n1 = Aux(2, j - 1) 'Atributo 2
    Do While m1 = n1
      Randomize
      n = (nB - 1) * Rnd + 1
      m1 = Int(n)
    Loop
    Aux(2, j) = m1

  Else
  End If
Next j
Next k

```

La idea de este bucle es asignar a las columnas pares números distintos a los de las columnas impares para cada fila, para ello comparamos con la fila anterior, si ha aparecido el nivel asignado para el atributo entonces en la siguiente le asigno aleatoriamente otro nivel. Esto lo conseguimos utilizando la función *Do While*, esta se trata de un bucle, cuyo objetivo es hacer lo asignado, mientras se cumpla una condición impuesta, en nuestro caso será que el número aleatorio obtenido en la columna que ocupamos sea igual al de la columna anterior. De este modo, se irá obteniendo números aleatorios mientras sean iguales, en cuanto se obtenga uno diferente la función parará y nos devolverá el último número obtenido (que será distinto al asignado a la columna anterior).

Por último vamos a ver cómo hemos programado en Visual Basic la función que me pasa los test y me verifica que el diseño obtenido de la matriz es el diseño eficiente que buscamos.

```

Dim sumni As Integer
Dim sum1 As Integer
Dim sum2 As Integer

Dim m As Integer

For i = 1 To nA      'Atributo 1 '
  sumni = 0
  For j = 1 To 2 * esc 'columnas impares

    If (Aux(1, j) = i) Then
      sumni = sumni + 1
    Else
      sumni = sumni
    End If

  Next j

  If (sumni < ((2 * esc) / nA)) Then
    For j = 1 To esc Step 2
      If (Aux(1, j) <> i Or Aux(1, 2 * j) <> i) Then '
        si no aparece en este par ya puede aparecer
        n = Aux(1, j)
        m = Aux(1, j + 1)
        sum1 = 0
        sum2 = 0
        'aqui le asignar el nivel a la columna donde
        est el nivel que mas veces aparece
      End If
    Next j
  End If
Next i

```



```

                End If
            Next k
            If (sum1 > sum2) Then
                Aux(2, j) = i
            Else
                Aux(2, j + 1) = i
            End If
            sumni = sumni + 1
        Else
        End If
    Next j
Else
End If

```

En este código se puede observar que se está comprobando primero para el atributo A y luego para el atributo B, que el número de niveles obtenido en la matriz del diseño aparezca el mismo número de veces. De la forma que hemos obtenido el número de escenarios de elección que forman parte del diseño (en el apartado anterior), sabemos las veces que tiene que aparecer cada nivel, ya que el número de escenarios obtenido es divisible por todos los niveles de los atributos y del hecho de que cada nivel tiene que tener la misma frecuencia para cada atributo. Por ejemplo el nivel 1 del atributo A tendrá que aparecer  $(2 * \text{escenarios}) / nA$ , donde  $nA$  será el número de niveles asociados al atributo A.

Lo que hacemos es guardarnos las veces que aparece cada nivel, lo hacemos recorriendo para cada fila las columnas, y guardando para cada nivel la vez que aparece y si vuelve a aparecer lo sumamos. Una vez tenemos las veces que aparece el nivel 1 del atributo A, por ejemplo, mediante un condicional dado por la función *If() Then*, en el caso que el número de veces que aparece sea menor que el número de veces que debería aparecer, tendremos que añadir ese nivel las veces que haga falta y lo añadiremos en la columna donde esté el nivel que más veces aparece, pero siempre teniendo en cuenta que no las sobrepase.

De la forma que hemos obtenido la matriz del diseño y comprobando que el número de niveles aparece el mismo número de veces para cada atributo, se deriva directamente que el diseño es ortogonal, es decir, cada par de elección aparecerá con la misma frecuencia.

Teniendo que la matriz del diseño obtenida cumple las premisas de un diseño estadísticamente eficiente, ahora solo faltará comprobar que no haya dominancia, es decir, que no haya alternativas que dominen y que de este modo sean siempre elegidas sin dar opción a poder comparar los efectos de todos los niveles de los atributos.

Para poder comparar la dominancia, supondremos que los niveles están enumerados de forma ascendente, es decir, se le asignará el número 1 al nivel más bajo de cada atributo y así irá aumentando hasta llegar al nivel más alto que se

le denominará con el mayor número, el máximo número de niveles que tiene el atributo. Por ejemplo, si tenemos que el atributo A tiene tres niveles asociados, se le asignará 1 al nivel más bajo (en el caso de ser niveles cualitativos, se asignará la característica menos valorada) y 3 al nivel más alto.

```

Dim Max1 As Integer , Mayor As Integer , Max2 As Integer
Dim Min1 As Integer , Min2 As Integer

'Primero tendr que saber el m ximo nivel

Mayor = 0
If (nA >= nB) Then
    Mayor = nA
Else
    Mayor = nB
End If

'a continuaci n los comparo

For j = 1 To esc Step 2 'columnas impares
    For k = 2 To 2 * esc 'columnas pares

        Max1 = 0
        Min1 = 0
        For i = 1 To Atri
            Min1 = Min1 + 1 * Aux(i, j) 'minimo de la
                columna impar
            Max1 = Max1 + Mayor * Aux(i, k) 'maximo de la
                columna par
        Next i

        'Ahora las comparo
        Do While (Min1 = 1 And Max1 = 2 * Mayor) 'si en la
            columna 1 siempre tengo asignado el peor nivel y
            en la segunda el mejor
            Randomize
            n = (nA - 1) * Rnd + 1
            m1 = Int(n)

            Do While m1 = Aux(1, j)

                Randomize
                n = (nA - 1) * Rnd + 1
                m1 = Int(n)
            
```

```

    Loop
    Aux(1, k) = m1

    'Atributo 2

    Do While m1 = Aux(2, j)
    Randomize
    n = (nB - 1) * Rnd + 1
    m1 = Int(n)
    Loop
    Aux(2, k) = m1

    Loop
    ' lo he hecho para el caso de que en las impares se
    asignen los niveles minimos y en las pares los
    maximos
    ' ahora tendr que ver que al rev s tampoco pasa

    Next k
Next j

For j = 1 To esc Step 2 'columnas impares
    For k = 2 To 2 * esc 'columnas pares

        Max2 = 0
        Min2 = 0
        For i = 1 To Atri
            Min2 = Min2 + 1 * Aux(i, k) 'minimo de la
            columna par
            Max2 = Max2 + Mayor * Aux(i, j) 'maximo de la
            columna impar
        Next i

        'Ahora las comparo
        Do While (Min2 = 1 And Max2 = 2 * Mayor) 'si en la
            columna 1 siempre tengo asignado el peor nivel y
            en la segunda el mejor
            Randomize
            n = (nA - 1) * Rnd + 1
            m1 = Int(n)

            Do While m1 = Aux(1, j)

                Randomize
                n = (nA - 1) * Rnd + 1
                m1 = Int(n)

            Loop
            Aux(1, k) = m1

```

```

        ' Atributo 2

        Do While m1 = Aux(2, j)
        Randomize
        n = (nB - 1) * Rnd + 1
        m1 = Int(n)
        Loop
        Aux(2, k) = m1

    Loop
    ' lo he hecho para el caso de que en las impares se
    asignen los niveles minimos y en las pares los
    maximos

Next k
Next j

```

Para evitar dominancia, en el código vemos que primero tendré que comprobar cuál es el nivel máximo, como en este caso solo tengo dos opciones de niveles, los comparo entre ellos, es decir, comparo nA con nB, y tendremos como máximo el mayor de los dos. A continuación comparo cada columna impar con las columnas pares. Lo que quiero evitar es que, por ejemplo, que en la columna par esté asignado el peor nivel y en la columna impar el mejor. Para ello, si se da este le asigno otro nivel a la columna par.

```

For i = 1 To 10
    For j = 1 To 2 * esc
        ActiveSheet.Cells(i + 20, j + 25).Value = Aux(i, j)
    Next j
Next i
End Sub

```

Para finalizar la función tendremos que mostrar la matriz obtenida, para ello la recorreremos y le asignamos el valor obtenido para cada fila y cada columna en una hoja de excel.

Acabamos de ver la función que devuelve la matriz que nos mostrará el diseño de los escenarios de elección, dado el número de niveles asociado a cada atributo. En este caso el número de atributos es dos, es el caso más simple, aunque para los demás casos se procederá de forma análoga, pero se tendrá que tener en cuenta para cada caso el número de atributos, ya que según el número de atributos habrán más casos y el código se hará más grande.

Veamos cómo sería la función para el caso de tener tres atributos, a modo de comparativa y con el fin de entender como evoluciona el diseño a medida que aumenta el número de atributos.

En este caso tenemos tres atributos, como acabamos de comentar, los datos iniciales que necesita la función es el número de niveles de cada atributo, nA, nB, nC para los atributos A, B y C, respectivamente. También le pasaremos el número de escenarios óptimo obtenido en la sección anterior.

```

Private Sub Atri3_Click()

Dim nA As Integer, nB As Integer, nC As Integer, nD As Integer
, nE As Integer, nF As Integer, nG As Integer, nH As
Integer, nI As Integer, nJ As Integer, Atri As Integer,
esc As Integer
nA = Range("N27").Value
nB = Range("N29").Value
nC = Range("N31").Value

Atri = Range("S27").Value
esc = Range("S30").Value
Dim i As Integer, j As Integer, k As Integer, n As Integer, n1
As Integer

ReDim Aux(Atri, (2 * esc)) As Integer
Dim m1 As Integer

' por separado para cada fila , es decir , atributo
For k = 1 To esc
For j = 1 To 2 * esc
If (j = (2 * k) - 1) Then

Randomize
n = (nA - 1) * Rnd + 1

n = Int(n)
Aux(1, j) = n

Randomize
n = (nB - 1) * Rnd + 1
n = Int(n)
Aux(2, j) = n

Randomize
n = (nC - 1) * Rnd + 1

```

```

        n = Int(n)
        Aux(3, j) = n

    Else
    End If
Next j
Next k

For k = 1 To esc
For j = 1 To 2 * esc

    If (j = 2 * k) Then

        Randomize
        n = (nA - 1) * Rnd + 1
        m1 = Int(n)

        Do While m1 = Aux(1, j - 1) 'mientras sean iguales
            que me genere otro, si no son iguales que pare y
            me lo devuelva

            Randomize
            n = (nA - 1) * Rnd + 1
            m1 = Int(n)

        Loop
        Aux(1, j) = m1

        Do While m1 = Aux(2, j - 1) 'Atributo 2
            Randomize
            n = (nB - 1) * Rnd + 1
            m1 = Int(n)
        Loop
        Aux(2, j) = m1

        'Atributo 3

        Do While m1 = Aux(3, j - 1)
            Randomize
            n = (nC - 1) * Rnd + 1
            m1 = Int(n)
        Loop
        Aux(3, j) = m1

    Else

```

```

    End If
Next j
Next k

' Hasta aqu solo asignare niveles. Faltar pasar los test



---



Dim sumni As Integer
Dim sum1 As Integer
Dim sum2 As Integer

Dim m As Integer

For i = 1 To nA 'Atributo 1 '
    sumni = 0
    For j = 1 To 2 * esc

        If (Aux(1, j) = i) Then
            sumni = sumni + 1
        Else
            sumni = sumni
        End If

    Next j

    If (sumni < ((2 * esc) / nA)) Then
        For j = 1 To esc Step 2 ' asi columnas impares
            If (Aux(1, j) < i Or Aux(1, 2 * j) < i) Then ' si
                no aparece en este par ya puede aparecer
                n = Aux(1, j)
                m = Aux(1, j + 1)
                sum1 = 0
                sum2 = 0 'aqui le asignar el nivel a la
                    columna donde est el nivel que mas veces
                    aparece
                'me tendr que volver a recorrer y que me
                    guarde el numero de veces que aparece n y m
                    y los comparo
                For k = 1 To 2 * esc
                    If (Aux(1, k) = n) Then
                        sum1 = sum1 + 1
                    ElseIf (Aux(1, k) = m) Then
                        sum2 = sum2 + 1
                    Else
                    End If
                
```

```

        Next k
        If (sum1 > sum2) Then
            Aux(1, j) = i
        Else
            Aux(1, j + 1) = i
        End If
        sumni = sumni + 1
    Else
    End If
Next j
Else
End If
Next i
,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,
For i = 1 To nB 'Atributo 2
    sumni = 0
    For j = 1 To 2 * esc
        If (Aux(2, j) = i) Then
            sumni = sumni + 1
        Else
            sumni = sumni
        End If
    Next j

    If (sumni < ((2 * esc) / nB)) Then
        For j = 1 To esc Step 2 ' asi columnas impares
            If (Aux(2, j) < i Or Aux(2, 2 * j) < i) Then ' si
                no aparece en este par ya puede aparecer
                n = Aux(2, j)
                m = Aux(2, j + 1)
                sum1 = 0
                sum2 = 0 'aqui le asignar el nivel a la
                    columna donde est el nivel que mas veces
                    aparece
                'me tendr que volver a recorrer y que me
                    guarde el numero de veces que aparece n y m
                    y los comparo
                For k = 1 To 2 * esc
                    If (Aux(2, k) = n) Then
                        sum1 = sum1 + 1
                    ElseIf (Aux(2, k) = m) Then
                        sum2 = sum2 + 1
                    Else
                    End If
                Next k
                If (sum1 > sum2) Then
                    Aux(2, j) = i

```

```

        Else
            Aux(2, j + 1) = i
        End If
        sumni = sumni + 1
    Else
        End If
    Next j
Else
End If

Next i

,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,
For i = 1 To nC 'Atributo 3
    sumni = 0
    For j = 1 To 2 * esc
        If (Aux(3, j) = i) Then ' contador
            sumni = sumni + 1
        Else
            sumni = sumni
        End If
    Next j

    If (sumni < ((2 * esc) / nC)) Then
        For j = 1 To esc Step 2 ' asi columnas impares
            If (Aux(3, j) <> i Or Aux(3, 2 * j) <> i) Then ' si
                no aparece en este par ya puede aparecer
                n = Aux(3, j)
                m = Aux(3, j + 1)

                sum1 = 0
                sum2 = 0 'aqui le asignar el nivel a la
                    columna donde est el nivel que mas veces
                    aparece
                'me tendr que volver a recorrer y que me
                    guarde el numero de veces que aparece n y m
                    y los comparo
                For k = 1 To 2 * esc
                    If (Aux(3, k) = n) Then
                        sum1 = sum1 + 1
                    ElseIf (Aux(3, k) = m) Then
                        sum2 = sum2 + 1
                    Else
                        End If
                Next k
                If (sum1 > sum2) Then
                    Aux(3, j) = i
                End If
            End If
        Next j
    End If
End For

```

```

                Else
                Aux(3, j + 1) = i
                End If
                sumni = sumni + 1
            Else
            End If
        Next j
    Else
    End If
Next i

,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,
Dim Max1 As Integer, Mayor As Integer, Max2 As Integer
Dim Min1 As Integer, Min2 As Integer

'Primero tendr que saber el m ximo nivel

Mayor = 0
If (nA >= nB) Then
    Mayor = nA
Else
    Mayor = nB
End If
If (nC >= Mayor) Then
    Mayor = nC
Else
End If

For j = 1 To esc Step 2 'columnas impares
    For k = 2 To 2 * esc 'columnas pares

        Max1 = 0
        Min1 = 0
        For i = 1 To Atri
            Min1 = Min1 + 1 * Aux(i, j) 'minimo de la
            columna impar
            Max1 = Max1 + Mayor * Aux(i, k) 'maximo de la
            columna par
        Next i

        'Ahora las comparo
        Do While (Min1 = 1 And Max1 = 3 * Mayor) 'si en la
            columna 1 siempre tengo asignado el peor nivel y
            en la segunda el mejor
    
```

```

Randomize
n = (nA - 1) * Rnd + 1
m1 = Int(n)

Do While m1 = Aux(1, j)

Randomize
n = (nA - 1) * Rnd + 1
m1 = Int(n)

Loop
Aux(1, k) = m1

'Atributo 2

Do While m1 = Aux(2, j)
Randomize
n = (nB - 1) * Rnd + 1
m1 = Int(n)
Loop
Aux(2, k) = m1

'Atributo 3

Do While m1 = Aux(3, j)
Randomize
n = (nC - 1) * Rnd + 1
m1 = Int(n)
Loop
Aux(3, k) = m1

Loop
' lo he hecho para el caso de que en las impares se
asignen los niveles minimos y en las pares los
maximos
' ahora tendr que ver que al rev s tampoco pasa
Max2 = 0
Min2 = 0
For i = 1 To Atri
Min2 = Min2 + 1 * Aux(i, k) 'minimo de la
columna par
Max2 = Max2 + Mayor * Aux(i, j) 'maximo de la
columna impar
Next i

'Ahora las comparo

```

```

Do While (Min2 = 1 And Max2 = 3 * Mayor) 'si en la
    columna 2 siempre tengo asignado el peor nivel y
    en la primera el mejor
Randomize
n = (nA - 1) * Rnd + 1
m1 = Int(n)

Do While m1 = Aux(1, j)

Randomize
n = (nA - 1) * Rnd + 1
m1 = Int(n)

Loop
Aux(1, k) = m1

'Atributo 2

Do While m1 = Aux(2, j)
Randomize
n = (nB - 1) * Rnd + 1
m1 = Int(n)
Loop
Aux(2, k) = m1

'Atributo 3

Do While m1 = Aux(3, j)
Randomize
n = (nC - 1) * Rnd + 1
m1 = Int(n)
Loop
Aux(3, k) = m1

Loop
Next k
Next j

,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,
For i = 1 To 10
    For j = 1 To 2 * esc
        ActiveSheet.Cells(i + 20, j + 25).Value = Aux(i, j)
    Next j
Next i
End Sub

```

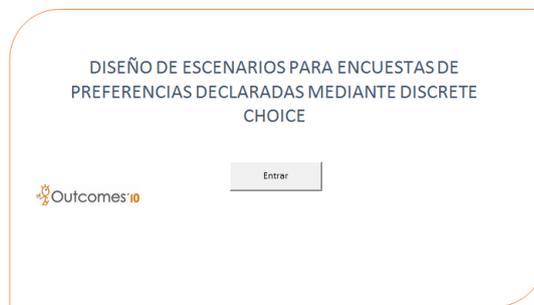
Como habíamos comentado vemos que la función abarca más casos y va creciendo a medida que aumenta el número de niveles. En la calculadora ex-

cel hemos supuesto que el número de atributos máximo es 10. En la literatura no encontramos ningún ejemplo con más de 10 atributos, lo cual nos indica que no habrá muchos casos en los que se dé un diseño factorial con más de 10 niveles.

### 4.3. Estructura de presentación de la calculadora

Una vez explicado cómo se ha obtenido el número de escenarios y el diseño de estos, vamos a ver cómo se presenta la calculadora, y su funcionamiento.

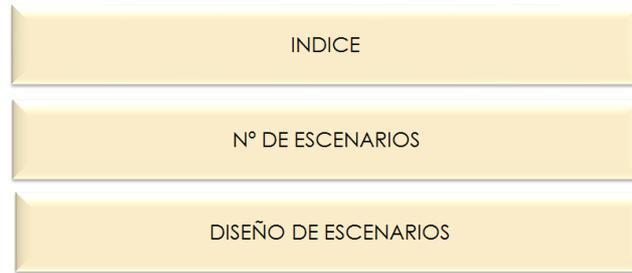
En la figura 13 vemos la portada, en la cual mediante el botón *Entrar*, accedemos al índice.



**Figura 13.** Portada de la calculadora

En el índice tendremos tres posibilidades, acceder a los objetivos donde se explica brevemente la funcionalidad de la misma, ir directamente a la parte de calcular el número de escenarios o por otra parte ir directamente al diseño de escenarios (esto será útil en el caso que solo queramos obtener el diseño de los escenarios con la cantidad de estos ya conocida), cómo vemos en la figura 14.

Si entramos a la opción del cálculo del número de escenarios óptimo, se preguntará, cuántas clases de niveles distintas tendremos, recordemos que a la hora de obtener el número de escenarios separábamos por casos según la cantidad de atributos con distintos niveles. De este modo, una vez seleccionado el número de niveles se mostrará por pantalla un cuadro de forma de excel como se muestra en la imagen 15, en el que introduciremos el número de niveles y la cantidad de atributos que tienen ese nivel, al pulsar el botón *Calcular*. Dentro del mismo nos mostrará el número de escenarios obtenido, el número de bloques y el número de escenarios por bloque.



**Figura 14.** Índice de la calculadora

En la imagen 15 podemos ver un ejemplo del caso de estar frente a un diseño con 3 atributos de 2 niveles y 2 atributos de 3 niveles.

Por último, en la parte de la calculadora en la que se muestra el diseño de los escenarios de elección, tendremos que introducir los niveles para cada atributo, como se muestra en la imagen 16.

Al continuar, se mostrará la matriz que dará un diseño eficiente para los escenarios de elección (figura 17).

¿ CANTIDADES DISTINTAS DE NIVELES? Dos

**calculo2** ×

Niveles	Atributo
<input style="width: 80%;" type="text" value="2"/>	<input style="width: 80%;" type="text" value="3"/>
<input style="width: 80%;" type="text" value="3"/>	<input style="width: 80%;" type="text" value="2"/>

Calcular

Número de escenarios

Número de escenarios por bloque

Número de bloques

Distintos Nivele  
¿Cuantos niveles distintos?

Figura 15. Introducción de los datos en la calculadora

Niveles del Atributo 1 <input style="width: 50%;" type="text" value="2"/>	Número de Atributos <input style="width: 50%;" type="text" value="2"/>
Niveles del Atributo 2 <input style="width: 50%;" type="text" value="2"/>	Número de escenarios <input style="width: 50%;" type="text" value="4"/>
Niveles del Atributo 3 <input style="width: 50%;" type="text"/>	
Niveles del Atributo 4 <input style="width: 50%;" type="text"/>	
Niveles del Atributo 5 <input style="width: 50%;" type="text"/>	
Niveles del Atributo 6 <input style="width: 50%;" type="text"/>	
Niveles del Atributo 7 <input style="width: 50%;" type="text"/>	

Continuar ➔

Figura 16. Introducción de los datos para obtener el diseño

Atributos	Escenario 1		Escenario 2		Escenario 3		Escenario 4	
	Tratamiento A	Tratamiento B						
Atributo 1	1	2	1	1	2	2	2	1
Atributo 2	1	2	2	1	1	2	2	1

Cuantos atributos hay? DosA

Figura 17. Salida por pantalla del diseño de escenarios

## 5. Conclusiones

A la hora de realizar este trabajo se han ido obteniendo las pautas o premisas para obtener un buen diseño de escenarios de elección. Este diseño es muy importante ya que será el que se le presentará al encuestado a la hora de realizar un estudio de preferencias. El encuestado deberá elegir entre varias alternativas (dos en nuestro caso), y mediante las técnicas de los experimentos de elección discreta se obtendrán los resultados. El trabajo se basa en la recolección de información sobre estos métodos y los métodos que nos han sido de ayuda durante el proceso.

Por otra parte, a la hora de programar en Visual Basic, hemos podido observar que se trata de un lenguaje simple y amplio, lo cual nos ha facilitado la tarea. Y se ha descubierto que excel es una herramienta para el análisis de datos que tiene un amplio abanico de posibilidades.

Aunque se haya generalizado a lo largo del trabajo, la calculadora presentada se utilizará en Outcomes'10, una consultoría farmacéutica dedicada a la farmacoeconomía y la obtención de resultados en salud. Uno de los trabajos que se desempeñan en la empresa es estudiar las preferencias tanto de pacientes como de médicos e incluso de farmacéuticos o personas relacionadas con el campo de la salud, sobre los tratamientos utilizados para la cura o mejora de una enfermedad específica. Para ello, el investigador encargado del proyecto debe, identificar los atributos y sus niveles que caracterizan al tratamiento mediante revisiones sistemáticas en la literatura, seleccionar los atributos y niveles finales llegando a un consenso de expertos en esta enfermedad, después se tienen que diseñar los escenarios a presentar y ahí es donde la calculadora será de gran ayuda, ya que este paso se reducirá a la introducción del número de atributos y niveles. Antes se tenía que recurrir al diseño de estos mediante el uso de funciones en R, y se tenía que encargar el trabajo a una persona especializada en este software, ahora con la calculadora, cualquier usuario puede obtener el diseño de escenarios, es de fácil manejo, ya que todo el código está por detrás. Una vez obtenido el diseño de los escenarios en la calculadora, se plantean las encuestas que van a ser presentadas a los encuestados y de donde se obtendrán resultados de preferencias mediante el uso de uno de los modelos estadísticos presentados en la sección 3, cuando estudiábamos los experimentos de elección discreta.

## Bibliografía

- FOWKES., A. et WARDMAN, M. (1988). The design of stated preference travel choice experiments. *Journal of Transport economics and policy*, Volume XXII, nº 1.
- H.AIZAKI (2012). Basic functions for supporting an implementation of choice experiments in R. *Journal of stadistical software*, Vol. 50.
- JOHSON F.R., Lancsar E., M. D. e. a. (2013). Constructing experimental designs for discrete choice experiments: Report of the ISPOR conjoint analysis experimental design good research practices task force. *Value in Health*, 16:3–13.
- LANCASTER, K. J. (1966). A new approach to consumer theory. *The Journal of Political Economy*, 74(2):132–157.
- LOUVIERE, J. (2000). Stated choice methods: Analysis and application. *Cambridge university press*.
- LUCE, R. (1959). Individual choice behavior. *Wiley, New York*.
- MACFADDEN, D. (1974). Conditional logit analysis of qualitative choice behavior. *FRONTIERS IN ECONOMETRICS*, Academic Press:105–142.
- MACFADDEN, D. (1975). The measurement of urban travel demand. *Journal of Public Economics*, 3:303–328.
- MONTGOMERY, D. C. (2004). *Diseño y análisis de experimentos*. Universidad estatal de Arizona.
- TRIBALDOS, M. (2016). *Preferencias de pacientes y médicos por los tratamientos de la diabetes mellitus tipo 2 en España y Portugal: disponibilidad a pagar por obtener beneficios adicionales y evitar efectos secundarios*. Thèse de doctorat, Universitat Jaume I.