



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

DESARROLLO DE MODELO LEXICOGRÁFICO DE ORDENAMIENTO
PARCIAL DE ATRIBUTOS

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN GESTIÓN DE
OPERACIONES
MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

MARÍA PILAR HENRÍQUEZ ARRATIA

PROFESOR GUÍA:
RICARDO MONTOYA MOREIRA

MIEMBROS DE LA COMISIÓN:
ANDRÉS MUSALEM SAID
DENIS SAURE VALENZUELA

SANTIAGO DE CHILE
2014

RESUMEN DE MEMORIA PARA OPTAR AL TÍTULO DE: Ingeniero Civil Industrial y grado de Magíster en Gestión de Operaciones
POR: María Pilar Henríquez Arratia
FECHA: 15/09/2014
PROFESOR GUÍA: Ricardo Montoya

DESARROLLO DE MODELO LEXICOGRÁFICO DE ORDENAMIENTO PARCIAL DE ATRIBUTOS

Los modelos de elección compensatorios han sido los modelos más utilizados para representar la toma de decisiones de compra de los consumidores. Sin embargo, diferentes escenarios puede motivar la toma de decisiones evitando compensaciones.

Para representar este tipo de situaciones, los modelos lexicográficos se convierten en una alternativa que puede reportar beneficios por ahorro de tiempo y esfuerzo.

Los modelos lexicográficos desarrollados asumen que los individuos tienen total conocimiento respecto a los atributos con los que aparecen descritas las alternativas y son capaces de valorizarlos y utilizarlos para decidir la alternativa preferida. En esta tesis se propone un modelo de elección lexicográfico que elimina el supuesto de capacidad de ordenamiento completa de atributos.

El modelo propuesto admite incertidumbre en el orden de atributos y también errores de elección por parte de los consumidores a la hora de declarar la alternativa preferida. Se proponen dos enfoques de estimación: homogéneo y heterogéneo en la capacidad individual de ordenamiento de atributos. En cada uno de ellos, se utiliza el procedimiento Bayesiano para estimar las probabilidades lexicográficas individuales de elección.

Mediante simulación, se confirma la robustez del modelo propuesto en términos de conseguir estimar la cantidad de atributos que los individuos están realmente considerando para tomar sus decisiones, recuperar los parámetros del modelo y predecir las elecciones de los consumidores.

Este trabajo finaliza con una aplicación a una muestra real encuestada respecto al mercado de *tablets*. En esta parte, se compara el modelo heterogéneo con un modelo lineal. Si bien el segundo logra mejores tasas de acierto, el modelo heterogéneo es a su vez mejor que cualquier variante al modelo lineal del tipo q-compensatorio.

Se concluye que con el modelo propuesto es posible obtener más información respecto a cómo los consumidores toman decisiones, transformándose en un modelo que logra explicar de mejor forma los procesos de elección sin perder con ello capacidad predictiva.

AGRADECIMIENTOS

Esta tesis marca el término de una etapa hermosa, llena de esfuerzo y aprendizaje. Muchas son las personas que están detrás de este logro y a quienes debo agradecer por ser parte importante en mi formación y por su incondicional apoyo en el proceso de estudio que culmina con este trabajo.

A mi profesor guía, Ricardo Montoya, por darme los lineamientos necesarios para esta tesis, por su apoyo constante, por estar atento a mis inquietudes, por ser paciente y por continuar depositando su confianza en mí cuando mi trabajo profesional comenzó a consumir parte importante del tiempo dedicado a esta tesis.

A mis padres, Isabel y Hernán, por ser un ejemplo de esfuerzo, por inculcarme el valor del estudio y la perseverancia, y por darme todas las herramientas, los tiempos y los espacios que necesité.

Gracias a mis hermanas, María Isabel y Ana Lía, por sus consejos personales y académicos, por estar siempre atentas y dispuestas a brindar ayuda. Agradezco a Cristóbal Castro por su paciencia y comprensión, por celebrar mis triunfos, por sacrificar su tiempo enseñándome materias donde soy más débil, por sus ideas y visión crítica. También estoy agradecida de Iván Elgueta, Carlos Landaeta y Ernesto Aldunate por motivarme a terminar, por alentarme y recordarme la importancia de confiar.

A todas estas personas, gracias por sus enseñanzas, por su ánimo, motivación y por todos los buenos momentos que me regalaron en este proceso.

TABLA DE CONTENIDO

	Página
INTRODUCCIÓN	1
CAPÍTULO I: MARCO TEÓRICO	4
1.1. Modelos compensatorios	4
1.2. Modelos no compensatorios	5
1.3. Reglas de decisión lexicográficas	6
1.4. Modelo Probabilístico Lexicográfico	8
CAPÍTULO II: CONSTRUCCIÓN DEL MODELO DE ELECCIÓN LEXICOGRÁFICO DE ORDENAMIENTO PARCIAL DE ATRIBUTOS	13
2.1 Enfoque de estimación homogéneo	13
2.2 Enfoque de estimación heterogéneo	16
CAPÍTULO III: VALIDACIÓN DEL MODELO CON SIMULACIÓN DE MONTE CARLO	17
3.1. Procedimiento	18
3.2. Diseño 1: Ortogonal	19
3.2.1. Resultados diseño 1	20
3.2.1.1. Simulación homogénea - Estimación homogénea	20
3.2.1.2. Simulación heterogénea - Estimación heterogénea	21
3.2.2. Análisis de resultados del diseño 1	21
3.3. Diseño 2: Con coincidencias en niveles	23
3.3.1. Resultados diseño 2	24
3.3.1.1. Simulación homogénea (20 preguntas, 125 individuos) – Estimación homogénea	24
3.3.1.2. Simulación heterogénea (20 preguntas, 125 individuos) – Estimación heterogénea	25
3.3.1.3. Simulación homogénea (40 preguntas, 125 individuos) – Estimación homogénea	26
3.3.1.4. Simulación heterogénea (40 preguntas, 125 individuos) – Estimación heterogénea	26
3.3.1.5. Simulación homogénea (20 preguntas, 250 individuos) – Estimación homogénea	27
3.3.1.6. Simulación heterogénea (20 preguntas, 250 individuos) – Estimación heterogénea	27
3.3.1.7. Simulación homogénea (40 preguntas, 250 individuos) – Estimación homogénea	28
3.3.1.8. Simulación heterogénea (40 preguntas, 250 individuos) – Estimación heterogénea	28
3.3.2. Análisis de resultados del diseño 2	29

CAPÍTULO IV: APLICACIÓN EMPÍRICA	40
4.1. Resultados con enfoque de estimación homogéneo	41
4.2. Resultados con enfoque de estimación heterogéneo	43
4.3. Análisis de resultados aplicación empírica	44
CONCLUSIONES	49
BIBLIOGRAFÍA	50
ANEXOS	52
ANEXO A: Resultado simulación con Diseño 1	52
1. Simulación homogénea (20 preguntas, 125 individuos) y enfoque de estimación homogéneo	52
2. Simulación heterogénea (20 preguntas, 125 individuos) y enfoque de estimación heterogéneo	56
ANEXO B: Resultado simulación con Diseño 2	57
1. Simulación homogénea (20 preguntas, 125 individuos) y enfoque de estimación homogéneo	57
2. Simulación heterogénea (20 preguntas, 125 individuos) y enfoque de estimación heterogéneo	61
3. Simulación homogénea (40 preguntas, 125 individuos) y enfoque de estimación homogéneo	62
4. Simulación heterogénea (40 preguntas, 125 individuos) y enfoque de estimación heterogéneo	66
5. Simulación homogénea (20 preguntas, 250 individuos) y enfoque de estimación homogéneo	67
6. Simulación heterogénea (20 preguntas, 250 individuos) y enfoque de Estimación heterogéneo	71
7. Simulación homogénea (40 preguntas, 250 individuos) y enfoque de estimación homogéneo	72
8. Simulación heterogénea (40 preguntas, 250 individuos) y enfoque de estimación heterogéneo	76
ANEXO C: Tasa de éxito post estimación <i>out of sample</i>	77

INDICE DE TABLAS

	Página
Tabla 1.1. Reglas de Decisión Lexicográficas	7
Tabla 3.1. Ejemplo atributos/niveles encuesta <i>notebooks</i>	17
Tabla 3.2. Ejemplo pregunta encuesta <i>notebooks</i> con diseño ortogonal	17
Tabla 3.3. Condiciones experimentales en análisis simulación	19
Tabla 3.4. MLL en diseño 1 con simulación homogénea (20 preguntas, 125 individuos) y estimación homogénea	20
Tabla 3.5. Matriz de confusión porcentual en diseño 1 con simulación heterogénea (20 preguntas, 125 individuos) y estimación heterogénea	21
Tabla 3.6. Ejemplo pregunta encuesta <i>notebooks</i> con diseño con coincidencias en niveles	23
Tabla 3.7. MLL en diseño 2 con simulación homogénea (20 preguntas, 125 individuos) y estimación homogénea	25
Tabla 3.8. Matriz de confusión porcentual en diseño 2 con simulación heterogénea (20 preguntas, 125 individuos) y estimación heterogénea	25
Tabla 3.9. MLL en diseño 2 con simulación homogénea (40 preguntas, 125 individuos) y estimación homogénea	26
Tabla 3.10. Matriz de confusión porcentual en diseño 2 con simulación heterogénea (40 preguntas, 125 individuos) y estimación heterogénea	26
Tabla 3.11. MLL en diseño 2 con simulación homogénea (20 preguntas, 250 individuos) y estimación homogénea	27
Tabla 3.12. Matriz de confusión porcentual en diseño 2 con simulación heterogénea (20 preguntas, 250 individuos) y estimación heterogénea	27
Tabla 3.13. MLL en diseño 2 con simulación homogénea (40 preguntas, 250 individuos) y estimación homogénea	28

Tabla 3.14.	Matriz de confusión porcentual en diseño 2 con simulación heterogénea (40 preguntas, 250 individuos) y estimación heterogénea	28
Tabla 3.15.	Tasas globales de acierto en la estimación del parámetro q bajo el enfoque de estimación heterogéneo	30
Tabla 3.16.	RMSE promedio en estimación del parámetro \bar{v}_i bajo el enfoque de estimación homogéneo	31
Tabla 3.17.	Tasa de éxito <i>in sample</i> post estimación bajo el enfoque de estimación homogéneo	35
Tabla 3.18.	Tasa de éxito <i>in sample</i> post estimación bajo el enfoque de estimación heterogéneo	36
Tabla 4.1.	Atributos y niveles en aplicación empírica <i>tablets</i>	40
Tabla 4.2.	MLL en estimación homogénea de aplicación empírica	42
Tabla 4.3.	Estimación de parámetro \bar{v}_i e intervalo de confianza en estimación homogénea de aplicación empírica	42
Tabla 4.4.	Estimación de parámetro $\bar{\alpha}_{i1}$ e intervalo de confianza en estimación homogénea de aplicación empírica	42
Tabla 4.5.	Medidas de desempeño en estimación homogénea de aplicación empírica	43
Tabla 4.6.	Porcentaje de individuos con capacidad de ordenamiento de 1, 2, 3 y 4 atributos en estimación heterogénea de aplicación empírica	43
Tabla 4.7.	Estimación de parámetro \bar{v}_i e intervalo de confianza en estimación heterogénea de aplicación empírica	43
Tabla 4.8.	Estimación de parámetro $\bar{\alpha}_{i1}$ e intervalo de confianza en estimación heterogénea de aplicación empírica	43
Tabla 4.9.	Medidas de desempeño en estimación heterogénea de aplicación empírica	44
Tabla 4.10.	Tasas de acierto en predicción de elecciones de modelos lineal q -compensatorio y lexicográfico de ordenamiento parcial de atributos	46
Tabla 4.11.	Tasas de acierto en predicción de elecciones de modelo lexicográfico de ordenamiento parcial de atributos	46

Tabla A.1.	Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 1 con simulación homogénea (20 preguntas, 125 individuos) con $q=1$ y estimación homogénea	52
Tabla A.2.	Medidas de desempeño en diseño 1 con simulación homogénea (20 preguntas, 125 individuos) con $q=1$ y estimación homogénea	52
Tabla A.3.	Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 1 con simulación homogénea (20 preguntas, 125 individuos) con $q=2$ y estimación homogénea	53
Tabla A.4.	Medidas de desempeño en diseño 1 con simulación homogénea (20 preguntas, 125 individuos) con $q=2$ y estimación homogénea	53
Tabla A.5.	Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 1 con simulación homogénea (20 preguntas, 125 individuos) con $q=3$ y estimación homogénea	54
Tabla A.6.	Medidas de desempeño en diseño 1 con simulación homogénea (20 preguntas, 125 individuos) con $q=3$ y estimación homogénea	54
Tabla A.7.	Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 1 con simulación homogénea (20 preguntas, 125 individuos) con $q=4$ y estimación homogénea	55
Tabla A.8.	Medidas de desempeño en diseño 1 con simulación homogénea (20 preguntas, 125 individuos) con $q=4$ y estimación homogénea	55
Tabla A.9.	Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 1 con simulación heterogénea (20 preguntas, 125 individuos) y estimación heterogénea	56
Tabla A.10.	Medidas de desempeño en diseño 1 con simulación heterogénea (20 preguntas, 125 individuos) y estimación heterogénea	56
Tabla B.1.	Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación homogénea (20 preguntas, 125 individuos) con $q=1$ y estimación homogénea	57
Tabla B.2.	Medidas de desempeño en diseño 2 con simulación homogénea (20 preguntas, 125 individuos) con $q=1$ y estimación homogénea	57
Tabla B.3.	Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación homogénea (20 preguntas, 125 individuos) con $q=2$ y estimación homogénea	58
Tabla B.4.	Medidas de desempeño en diseño 2 con simulación homogénea (20 preguntas, 125 individuos) con $q=2$ y estimación homogénea	58

Tabla B.5.	Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación homogénea (20 preguntas, 125 individuos) con $q=3$ y estimación homogénea	59
Tabla B.6.	Medidas de desempeño en diseño 2 con simulación homogénea (20 preguntas, 125 individuos) con $q=3$ y estimación homogénea	59
Tabla B.7.	Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación homogénea (20 preguntas, 125 individuos) con $q=4$ y estimación homogénea	60
Tabla B.8.	Medidas de desempeño en diseño 2 con simulación homogénea (20 preguntas, 125 individuos) con $q=4$ y estimación homogénea	60
Tabla B.9.	Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación heterogénea (20 preguntas, 125 individuos) y estimación heterogénea	61
Tabla B.10.	Medidas de desempeño en diseño 2 con simulación heterogénea (20 preguntas, 125 individuos) y estimación heterogénea	61
Tabla B.11.	Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación homogénea (40 preguntas, 125 individuos) con $q=1$ y estimación homogénea	62
Tabla B.12.	Medidas de desempeño en diseño 2 con simulación homogénea (40 preguntas, 125 individuos) con $q=1$ y estimación homogénea	62
Tabla B.13.	Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación homogénea (40 preguntas, 125 individuos) con $q=2$ y estimación homogénea	63
Tabla B.14.	Medidas de desempeño en diseño 2 con simulación homogénea (40 preguntas, 125 individuos) con $q=2$ y estimación homogénea	63
Tabla B.15.	Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación homogénea (40 preguntas, 125 individuos) con $q=3$ y estimación homogénea	64
Tabla B.16.	Medidas de desempeño en diseño 2 con simulación homogénea (40 preguntas, 125 individuos) con $q=3$ y estimación homogénea	64
Tabla B.17.	Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación homogénea (40 preguntas, 125 individuos) con $q=4$ y estimación homogénea	65
Tabla B.18.	Medidas de desempeño en diseño 2 con simulación homogénea (40 preguntas, 125 individuos) con $q=4$ y estimación homogénea	65

Tabla B.19.	Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación heterogénea (40 preguntas, 125 individuos) y estimación heterogénea	66
Tabla B.20.	Medidas de desempeño en diseño 2 con simulación heterogénea (40 preguntas, 125 individuos) y estimación heterogénea	66
Tabla B.21.	Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación homogénea (20 preguntas, 250 individuos) con $q=1$ y estimación homogénea	67
Tabla B.22.	Medidas de desempeño en diseño 2 con simulación homogénea (20 preguntas, 250 individuos) con $q=1$ y estimación homogénea	67
Tabla B.23.	Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación homogénea (20 preguntas, 250 individuos) con $q=2$ y estimación homogénea	68
Tabla B.24.	Medidas de desempeño en diseño 2 con simulación homogénea (20 preguntas, 250 individuos) con $q=2$ y estimación homogénea	68
Tabla B.25.	Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación homogénea (20 preguntas, 250 individuos) con $q=3$ y estimación homogénea	69
Tabla B.26.	Medidas de desempeño en diseño 2 con simulación homogénea (20 preguntas, 250 individuos) con $q=3$ y estimación homogénea	69
Tabla B.27.	Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación homogénea (20 preguntas, 250 individuos) con $q=4$ y estimación homogénea	70
Tabla B.28.	Medidas de desempeño en diseño 2 con simulación homogénea (20 preguntas, 250 individuos) con $q=4$ y estimación homogénea	70
Tabla B.29.	Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación heterogénea (20 preguntas, 250 individuos) y estimación heterogénea	71
Tabla B.30.	Medidas de desempeño en diseño 2 con simulación heterogénea (20 preguntas, 250 individuos) y estimación heterogénea	71
Tabla B.31.	Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación homogénea (40 preguntas, 250 individuos) con $q=1$ y estimación homogénea	72
Tabla B.32.	Medidas de desempeño en diseño 2 con simulación homogénea (40 preguntas, 250 individuos) con $q=1$ y estimación homogénea	72

Tabla B.33.	Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación homogénea (40 preguntas, 250 individuos) con $q=2$ y estimación homogénea	73
Tabla B.34.	Medidas de desempeño en diseño 2 con simulación homogénea (40 preguntas, 250 individuos) con $q=2$ y estimación homogénea	73
Tabla B.35.	Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación homogénea (40 preguntas, 250 individuos) con $q=3$ y estimación homogénea	74
Tabla B.36.	Medidas de desempeño en diseño 2 con simulación homogénea (40 preguntas, 250 individuos) con $q=3$ y estimación homogénea	74
Tabla B.37.	Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación homogénea (40 preguntas, 250 individuos) con $q=4$ y estimación homogénea	75
Tabla B.38.	Medidas de desempeño en diseño 2 con simulación homogénea (40 preguntas, 250 individuos) con $q=4$ y estimación homogénea	75
Tabla B.39.	Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación heterogénea (40 preguntas, 250 individuos) y estimación heterogénea	76
Tabla B.40.	Medidas de desempeño en diseño 2 con simulación heterogénea (40 preguntas, 250 individuos) y estimación heterogénea	76
Tabla C.1.	Tasa de éxito <i>out of sample</i> post estimación bajo el enfoque de estimación homogéneo	77
Tabla C.2.	Tasa de éxito <i>out of sample</i> post estimación bajo el enfoque de estimación heterogéneo	77

INDICE DE ILUSTRACIONES

	Página
Figura 1.1. Selección web de <i>notebooks</i>	4
Figura 4.1. Pregunta 1 encuesta aplicación empírica <i>tablets</i>	40
Figura 4.2.: Heterogeneidad en la capacidad de ordenamiento y preferencia de atributos utilizando la Regla Lexicográfica por Atributos (LBA) en aplicación empírica	47

ÍNDICE DE GRÁFICOS

	Página
Gráfico 3.1. Distribución en estimación individual de q en simulación heterogénea (20 preguntas, 125 individuos) y estimación heterogénea	37
Gráfico 3.2. Distribución en estimación de q en simulación heterogénea y estimación heterogénea	37
Gráfico 3.3. Tasa de acierto en predicción de q en simulación heterogénea y estimación heterogénea	38
Gráfico 3.4. Tasa de acierto post estimación <i>in sample</i> en simulación heterogénea y estimación heterogénea	39
Gráfico 4.1. Nivel de conocimiento de atributos aplicación empírica <i>tablets</i>	41
Gráfico 4.2. Distribución en estimación de q en estimación heterogénea de aplicación empírica	45

INTRODUCCIÓN

En la actualidad, mercados cada vez más competitivos y globalizados, han hecho posible el acceso a una mayor variedad de productos, situación que viene acompañada de mayores dificultades para tomar decisiones de compra.

Los consumidores enfrentan decisiones aún más complejas toda vez que deben lidiar con la creciente escasez de tiempo y donde la comparación exhaustiva entre alternativas es, en ocasiones, impracticable.

Entender cómo los consumidores resuelven esta dificultad tiene para la empresa un carácter estratégico y representa una ventaja competitiva si logra identificar y modelar los patrones que dirigen las elecciones de los consumidores. En efecto, el análisis de los procesos de elección es de vital importancia para la toma de decisiones respecto al diseño de productos, diseño de sitios *web*, fijación de precios e incluso de segmentación de clientes, permitiendo una mejor dirección de los esfuerzos de marketing.

Para estudiar cómo los consumidores toman decisiones de compra, los modelos más utilizados han sido los “modelos de elección compensatorios”, los cuales se basan en la idea de que el consumidor es capaz de comparar las alternativas y “compensar” la ausencia de una característica preferida, por la presencia de otra(s) que reporten en conjunto una utilidad o satisfacción superior.

Si bien los modelos compensatorios han sido la mejor opción computacionalmente abordable, muchos autores cuestionan la utilización de este tipo de modelos argumentando que bajo ciertas circunstancias pueden no ser suficientemente realistas.

En efecto, Payne, Bettman y Johnson (1993) sostienen que los procesos de elección dependen factores tales como la información respecto a las características del producto, relevancia de la decisión, dificultad de comparación, entre otras. Drolet y Luce (2004) incorporan los “factores emocionales” como elementos determinantes de las decisiones. Estudios experimentales como los de Dieckmann, Dippold y Dietrich (2009) plantean que los procesos de elección dependen de la dificultad de la decisión, dada por la cantidad opciones de compra y sus atributos.

Así, los consumidores podrían no detenerse a analizar los pros y contras de cada alternativa bajo ciertas circunstancias. Existen individuos que sólo considerarán automóviles de procedencia japonesa, o habrá padres dispuestos a pagar cualquier precio por una buena cobertura de salud para sus hijos.

Procesos como los anteriores, son del tipo “no compensatorio”. En ellos, los consumidores hacen uso de “reglas lexicográficas” para establecer rápidamente un *ranking* de preferencia de alternativas en función de los atributos que posean, procediendo de forma similar a como se ordenan las palabras en un diccionario.

La regla lexicográfica más común, y la que será abordada en este trabajo, es la Regla Lexicográfica por Atributos. Esta regla es utilizada toda vez que el consumidor

identifica el atributo más importante y ordena las alternativas en consecuencia (ejemplo: prefiere los vehículos japoneses a cualquier otro), luego ordena cada subconjunto de alternativas en función del segundo atributo más importante (ejemplo: ordena los vehículos japoneses de acuerdo al tipo de combustible que utilizan y hace lo mismo con aquellos de otra procedencia) y así sucesivamente. El beneficio en el uso de esta regla se explica por la mayor facilidad en identificar la alternativa preferida, siendo necesario únicamente un *ranking* de preferencia de atributos.

Es posible encontrar evidencias, como las presentadas por Bröder (2000) y Bröder y Schiffer (2003), que sugieren que un modelo más realista será uno del tipo lexicográfico que, si bien puede incorporar cierta pérdida de precisión, puede también reportar importantes beneficios por ahorro de tiempo y esfuerzo.

Investigaciones recientes realizadas por Kohli y Jedidi (2007), Yee, Dahan, Hauser y Orlin (2007) y Jedidi, Kohli y Montoya (2011), han abordado el problema de elección de los consumidores desde un punto de vista “no compensatorio”. Estos modelos, se basan en el supuesto de que los individuos tienen la capacidad de establecer un *ranking* u orden de preferencia de atributos completo, es decir, donde los individuos asignan importancia relativa a todos los atributos con que aparecen descritas las alternativas.

Lo anterior no necesariamente es válido para todos los consumidores. Puede ocurrir que un consumidor sea indiferente entre los atributos menos preferidos o que se encuentre poco familiarizado con las características del producto de tal manera que sólo le sea posible establecer un *ranking* u ordenamiento compuesto por un subconjunto de atributos, que es lo que en este trabajo se denominará “ordenamiento parcial de atributos”. Surge así la pregunta: ¿cómo determinar por cuántos atributos está compuesto este orden parcial?

Hasta la fecha no se han desarrollado modelos lexicográficos que busquen precisar cuántos (y cuáles) atributos son capaces de considerar realmente los consumidores a la hora de tomar su decisión de compra.

La respuesta a ésta pregunta contribuirá a mejorar el análisis de un proceso de compra. Utilizando esta información, diseñadores, fabricantes y minoristas podrán enfatizar (o relegar) aquellas características de los productos que resulten determinantes en las elecciones, tener nociones de qué tan rápidamente están tomando decisiones y qué atributos no están siendo considerados. Esta nueva información permitirá ser más eficaz y efectivos a la hora de definir segmentos, ofertas, canales, etc.

Es así como, este trabajo tiene como objetivo general la construcción de un modelo de elección no compensatorio llamado Modelo de Elección Lexicográfico de Ordenamiento Parcial de Atributos, que elimina el supuesto de capacidad de ordenamiento completa de atributos.

Los objetivos específicos de este trabajo son:

- Desarrollar un modelo de elección lexicográfico que permita obtener la capacidad de ordenamiento de los atributos de los consumidores.

- Desarrollar un método para estimar los parámetros asociados al modelo desarrollado.
- Estudiar los alcances y limitaciones del modelo en una aplicación real.

La metodología conducente a lograr los objetivos propuestos comienza con la recopilación y lectura de literatura para proceder a continuación a desarrollar del modelo utilizando como base principal el método Bayesiano descrito por Jedidi, Kohli y Montoya (2011). A continuación, se estudiarán los alcances tanto del modelo como de la metodología de estimación mediante simulación.

Finalmente, se aplicará el modelo a una muestra poblacional real. En la última parte de este trabajo, se presentarán los resultados obtenidos junto con el análisis correspondiente a esa muestra particular.

CAPÍTULO I: MARCO TEÓRICO

En este trabajo se considerarán procesos de elección en los cuales los productos están representados por sus atributos y cada atributo por sus niveles. El consumidor debe decidir qué producto comprar o consumir dentro de un conjunto de alternativas posibles.

La Figura 1.1. representa una situación donde aparecen distintas alternativas de computadores portátiles o *notebooks* descritos por sus atributos o características. Por ejemplo, la memoria RAM es uno de los atributos y 2GB, 4GB, 6 GB, 8GB son sus niveles.

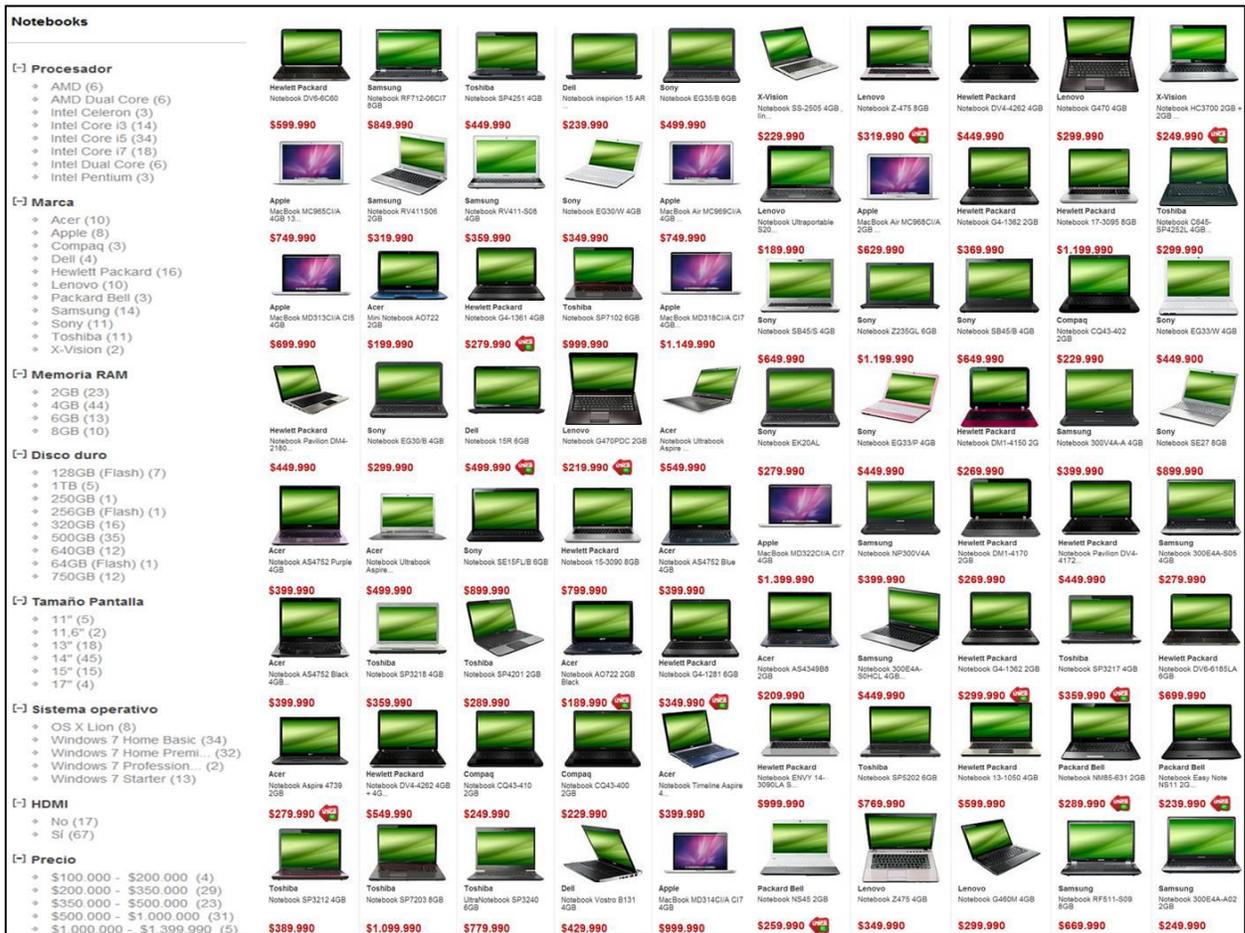


Figura 1.1. Selección web de *notebooks*
Fuente: www.falabella.com

1.1. Modelos compensatorios

Para describir los procesos de elección, típicamente se utilizan los modelos de elección compensatorios, los cuales asumen que el consumidor es capaz de "compensar" la ausencia de algún nivel deseado, por la presencia de otros que, en conjunto, reporten una mayor satisfacción.

De esta forma, utilizando la descripción dada por Green y Rao.(1971), el individuo maximizará la utilidad que reporta el bien, calculada como la suma ponderada de la utilidad otorgada por los atributos que lo describen. Es decir:

$$U_k = \beta_0 + \sum_{j=1}^J \sum_{n=1}^{N_j} \beta_{jn} x_{jn}^k + \varepsilon_k \quad (1)$$

Donde: U_k : Utilidad de la alternativa k.
 β_0 : Intercepto.
 β_{jn} : Coeficiente asociado al nivel n del atributo j.
 x_{jn}^k : Variable *dummy* cuyo valor es 1 si la alternativa k posee en nivel n-ésimo del atributo j (y 0 en caso contrario).
 ε_k : Error asociado a la alternativa k.

1.2. Modelos no compensatorios

Un modelo no compensatorio refleja situaciones donde hay niveles preferidos cuya ausencia no se puede compensar con la presencia de otros.

Por ejemplo, algunos consumidores no comprarán un *notebook* con menos de 4 GB de memoria RAM y no se detendrán a analizar alternativas que pueden tener otro conjunto de atributos muy convenientes pero con otro tipo de memoria. Esta situación, refleja el típico caso donde la valoración de un aspecto es tan alta que no se está dispuesto a hacer compensaciones.

Sin embargo, muchas otras situaciones revelan ausencia de compensación. En particular, Payne, Bettman y Johnson (1988) plantean que los consumidores usualmente no se detienen a analizar los *trade-offs* de cada alternativa, sino que toman decisiones rápidas, adoptando un proceso de decisión heurístico e incorporando “pérdidas de precisión” en su elección pero también ganancias en tiempo y esfuerzo.

Muchas aplicaciones son consistentes con esta idea, donde el consumidor es primero incitado a eliminar del conjunto de alternativas aquellas que poseen niveles inaceptables y reducir así la complejidad del problema a enfrentar.

En este contexto, la utilización de estrategias o reglas de decisión lexicográficas aparece como una de las heurísticas más comunes a la hora de llevar a cabo un proceso no compensatorio. Es posible encontrar ejemplos de lo anterior en Bettman, Luce y Payne (1998), Bröder (2000), Bröder y Schiffer (2003), Gigerenzer y Goldstein (1996), Martignon y Hoffrage (2002), Payne, Bettman y Johnson (1993) y Urban y Hauser (2004).

1.3. Reglas de decisión lexicográficas

Las reglas de decisión lexicográficas son utilizadas por el consumidor para ordenar alternativas de forma similar a como se ordenan las palabras en un diccionario. Con ellas, el individuo acepta o descarta alternativas siguiendo un proceso del tipo no compensatorio.

Por ejemplo, una persona que se dispone a comprar un *notebook*, utiliza una regla lexicográfica si prefiere estrictamente Windows® a cualquier otro sistema operativo. Luego, si entre de los *notebooks* con sistema operativo Windows® siempre prefiere un procesador Intel® y luego utiliza, de forma secuencial, la marca, la memoria y el espacio en disco para tomar su decisión.

Este tipo de reglas han sido reportadas por Slovic (1975), quien informó de su uso para decidir entre dos alternativas igualmente valoradas, en un estudio donde la mayoría de los individuos opta por consumir la alternativa que posee características poco deseables excepto en el más importante atributo, donde es superior a la otra opción, apoyando así la idea de que las decisiones están determinadas por mecanismos que son fáciles de explicar.

Investigaciones realizadas por Bröder (2000) y Bröder y Schiffer (2003) sugieren que los consumidores, muchas veces de manera inconsciente, simplifican elecciones complejas (con muchas alternativas o atributos) utilizando estrategias o reglas lexicográficas.

Drolet y Luce (2004) señalan que los consumidores a menudo utilizan estas reglas cuando tienen razones emocionales para evitar compensaciones.

Yee, Dahan, Hauser y Orlin (2007) concluyó que aproximadamente dos tercios de los individuos utiliza este tipo de reglas para evaluar *Smart-Phones*.

Así, si consideramos la Figura 1.1., el sitio *web* de venta online de *notebooks* alienta al consumidor a filtrar su búsqueda basándose en características tales como marca, procesador, memoria RAM, precio, de tal forma que sea posible eliminar rápidamente aquellos productos con atributos no deseados o menos preferidos.

Existen diversas reglas lexicográficas que pueden ser utilizadas para establecer un orden de preferencia de alternativas. En la primera columna de la Tabla 1, se ilustra la regla Lexicográfica por Atributos (LBA, del inglés, *Lexicographic by Attributes*) que será la regla a abordar en este trabajo. En el ejemplo, se observa el *ranking* construido por un individuo que establece como primer criterio la preferencia de marca (HP® es preferido a Sony®, Sony® es preferido a Toshiba® y Toshiba® es preferido a Dell®) y luego, se ordena las alternativas de una misma marca utilizando como segundo criterio la preferencia de memoria RAM (6 GB es preferido a 4 GB).

Otra posibilidad es que un consumidor prefiera un *notebook* HP®, en segundo lugar e independiente de la marca, prefiera una memoria RAM de 6 GB, luego, para un mismo tipo de memoria RAM, prefiera los *notebooks* marca Sony® y a continuación los Toshiba®. Este tipo de regla se denomina Lexicográfica de Aceptación por Niveles (LAL, del inglés, *Lexicographic by Acceptance Levels*).

El tercer caso ilustrado en la Tabla 1.1. es aquel en que el individuo sitúa en la última posición a los *notebooks* marca Dell®, luego, como segundo criterio, la memoria RAM de 4GB es menos preferida que la de 6 GB y, finalmente, reconoce que un *notebook* marca Toshiba® le agrada menos que uno Sony® y este, a su vez, es menos preferido que uno HP®. Ante esta situación, el individuo estará utilizando una regla Lexicográfica de Eliminación de Niveles (LEL, del inglés, *Lexicographic by Elimination Levels*).

Regla	Preferencia de atributos/niveles	Ranking de preferencia de alternativas
LBA	HP>Sony>Toshiba>Dell, 6 GB > 4 GB	
LAL	HP, 6 GB, Sony, Toshiba	
LEL	Dell, 4 GB, Toshiba, Sony	

Tabla 1.1. Reglas de Decisión Lexicográficas

Kohli y Jedidi (1999) proponen una función de utilidad que puede ser usada para modelar preferencias lexicográficas:

$$U_k = \frac{x_{r_1}^k}{n_{r_1}} + \frac{x_{r_2}^k}{n_{r_1} n_{r_2}} + \dots + \frac{x_{r_j}^k}{n_{r_1} n_{r_2} \dots n_{r_j}} \quad (2)$$

Donde: $U(x_k) = U_k$: Utilidad de la alternativa k.
 r_j : j-ésimo atributo más importante.
 n_{r_j} : Cantidad de niveles del j-ésimo atributo más importante.
 $x_{r_j}^k \in \{0, 1, \dots, n_{r_j}-1\}$: Nivel del j-ésimo atributo más importante de la alternativa k, donde el nivel 1 es preferido al 0 y, así sucesivamente, $n_{r_j}-1$ es el mejor nivel para el atributo r_j .

Así, el consumidor elegirá la alternativa s si:

$$U(x_s) > \max_{k=1, \dots, K, k \neq s} \{U(x_k)\} \quad (3)$$

A modo de ejemplo, si se considera:

Alternativa 1, que sólo posee el nivel preferido en el atributo más importante, siendo altamente desfavorable en todos los demás atributos, es decir, $x_{r_1}^1 = n_{r_1}-1$ y $x_{r_j}^1 = 0$ para $j=2, 3, \dots, J$.

Alternativa 2, con presencia del nivel preferido en todos los atributos excepto el más importante, donde posee el nivel que le sigue inmediatamente al preferido, es decir, $x_{r_1}^2 = n_{r_1}-2$ y $x_{r_j}^2 = n_{r_j}-1$ para $j=2, 3, \dots, J$.

Un individuo revelará sus preferencias lexicográficas si prefiere la alternativa 1 a la 2, pese a que esta última domina en todos menos el más importante atributo.

La función de utilidad anterior captura en efecto este hecho pues:

$$\begin{aligned}
 U(x_1) &= \frac{n_{r_1}-1}{n_{r_1}} = 1 - \frac{1}{n_{r_1}} \\
 U(x_2) &= \frac{n_{r_1}-2}{n_{r_1}} + \frac{n_{r_2}-1}{n_{r_1}n_{r_2}} + \frac{n_{r_3}-1}{n_{r_1}n_{r_2}n_{r_3}} + \frac{n_{r_4}-1}{n_{r_1}n_{r_2}n_{r_3}n_{r_4}} + \dots + \frac{n_{r_j}-1}{n_{r_1}n_{r_2}n_{r_3}n_{r_4}\dots n_{r_j}} \\
 &= 1 - \frac{2}{n_{r_1}} + \frac{1}{n_{r_1}} - \frac{1}{n_{r_1}n_{r_2}} + \frac{1}{n_{r_1}n_{r_2}} - \frac{1}{n_{r_1}n_{r_2}n_{r_3}} + \dots + \frac{1}{n_{r_1}n_{r_2}n_{r_3}\dots n_{r_{j-1}}} - \frac{1}{n_{r_1}n_{r_2}n_{r_3}\dots n_{r_j}} \\
 U(x_1) - U(x_2) &= \frac{1}{n_{r_1}n_{r_2}n_{r_3}\dots n_{r_j}} > 0 \quad \forall j \in \mathbb{N}
 \end{aligned}$$

Luego, la utilidad otorgada por la alternativa 1, será siempre mayor a la utilidad proporcionada por la alternativa 2.

Es escasa la literatura que aborda los modelos lexicográficos con una representación matemática. Sin embargo, es posible encontrar estudios donde modelos lexicográficos proporcionan predicciones que son más precisas que modelos compensatorios (Brighton y Gigerenzer, 2008) y otros que arrojan resultados casi igualmente buenos (Bröder y Gaissmaier, 2007).

Otras aplicaciones donde se pone a prueba la capacidad predictiva del modelo compensatorio *versus* un modelo lexicográfico (Dieckmann, Dippold y Dietrich, 2009, Yee, Dahan, Hauser y Orlin, 2007) confirman la idea intuitiva de que al aumentar las alternativas de elección, el modelo lexicográfico es más preciso que el modelo compensatorio para predecir el comportamiento del consumidor.

Queda abierta la interrogante si los modelos lexicográficos logran también un buen ajuste cuando el conjunto de alternativas es más reducido.

1.4. Modelo Probabilístico Lexicográfico

Jedidi, Kohli y Montoya (2011) proponen un modelo probabilístico de elecciones lexicográficas que incorpora incertidumbre en el orden de preferencia de los atributos.

La introducción de incertidumbre responde a la posibilidad de que una persona no esté segura de la valoración que otorgará a los atributos. También puede ocurrir, como lo reporta Tversky en 1972, que se valore de manera diferente un mismo atributo en diferentes situaciones. Por ejemplo, qué tan saludable es la comida puede ser más importante a la hora de cocinar en el hogar que si se trata de ordenar en un restaurant, o de forma similar, el consumo energético de los artículos de iluminación puede ser más relevante a la hora de equipar un domicilio permanente que una casa de veraneo.

Así, comenzando por los postulados del modelo MROL (del inglés, *Rank-Ordered Logit Model*, introducido por Beggs, Cardell y Hausman (1981), para determinar las preferencias de individuos, se representa la importancia de los atributos como la suma de una parte sistemática y una aleatoria, es decir:

$$w_j = v_j + \varepsilon_j \quad (4)$$

Donde: w_j : Importancia del atributo j .
 v_j : Importancia sistemática del atributo j .
 ε_j : Error.¹

Si ahora se considera el vector $r = (r_1, r_2, \dots, r_J)$, donde r_j es el j -ésimo atributo más importante, entonces $P(r_1)$ será la probabilidad de que el atributo r_1 sea el más importante, $P(r_2|r_1)$ será la probabilidad de que el atributo r_2 sea el segundo más importante dado que r_1 es el primero, y así sucesivamente hasta que, de manera más general, $P(r_s|r_1, \dots, r_{s-1})$, con $s = 1, 2, \dots, J$, denotará la probabilidad de que el atributo r_s sea el s -ésimo más importante dado que r_1, r_2, \dots, r_{s-1} es la secuencia de atributos de mayor relevancia.

Luego, si se considera el vector $v = (v_1, v_2, \dots, v_J)$ de importancia sistemática, la probabilidad de observar un *ranking* de atributos r dado v , será:

$$P(r | v) = P(r_1) P(r_2|r_1) \dots P(r_J | r_1, \dots, r_{J-1}) \quad (5)$$

De esta manera, se obtiene una expresión simplificada de $P(r|v)$:

$$P(r_1) = P(w_{r_1} > \max\{w_{r_2}, \dots, w_{r_J}\}) = \frac{e^{v_{r_1}}}{e^{v_{r_1}} + \dots + e^{v_{r_J}}} \quad (6)$$

$$P(r_s | r_1, \dots, r_{s-1}) = P(w_{r_s} > \max\{w_{r_{s+1}}, \dots, w_{r_J}\}) = \frac{e^{v_{r_s}}}{e^{v_{r_s}} + \dots + e^{v_{r_J}}}; \forall s = 2, \dots, J-1 \quad (7)$$

$$P(r | v) = P(w_{r_1} > \dots > w_{r_J}) = \prod_{s=1}^{J-1} \frac{e^{v_{r_s}}}{e^{v_{r_s}} + \dots + e^{v_{r_J}}} \quad (8)$$

Finalmente, incluso cuando la incertidumbre en el orden de atributos se encuentra resuelta, un consumidor puede incurrir en errores de juicio o reacción, cometiendo errores al declarar sus preferencias. Así, si un individuo se enfrenta a K alternativas de elección, con una alta probabilidad el individuo elegirá la alternativa que es lexicográficamente preferida pero no es seguro que esto ocurra.

Jedidi, Kohli y Montoya (2011) incorporan este hecho estimando el vector de probabilidad de elección $a = (a_1, a_2, \dots, a_K)$, donde K es la cantidad de alternativas y a_k es la probabilidad de que se elija la k -ésima alternativa más preferida, es decir, existiendo $k-1$ alternativas con mayor utilidad que U_k .²

¹ Se asume que ε_j sigue una distribución independiente e idénticamente distribuida tipo doble exponencial.

² En general, un buen modelo debiera tener un valor de a_1 superior y valores a_2, a_3, \dots, a_K sucesivamente decrecientes.

Así, se pueden distinguir tres situaciones:

- Si $a_k = \frac{1}{K}$ para todo k , entonces la elección es aleatoria.
- Si $a_1 = 1$ y $a_k = 0$ para todo $k > 1$, entonces el orden de preferencia de atributos (r) captura perfectamente el orden de preferencia de alternativas y el modelo probabilístico se convierte en determinístico.
- Si $a_k = \frac{1-a_1}{K-1}$ para todo $k = 2, \dots, K$, entonces todas las alternativas, con excepción de aquella con mayor utilidad, puede ser escogida con igual probabilidad.

De manera directa se individualiza todo lo anterior definiendo:

- $w_i = (w_{i1}, w_{i2}, \dots, w_{iJ})$: Vector de importancia de atributos para el individuo i .
 $v_i = (v_{i1}, v_{i2}, \dots, v_{iJ})$: Vector de importancia sistemática de atributos para el individuo i .³
 $r_i = (r_{i1}, r_{i2}, \dots, r_{iJ})$: Vector de *ranking* de atributos para el individuo i .
 $a_i = (a_{i1}, a_{i2}, \dots, a_{iK})$: Vector de probabilidad de elección para el individuo i .
 $\alpha_i = (\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{iK})$: Transformación de a_i , donde $a_{ik} = \frac{e^{\alpha_{ik}}}{1+e^{\alpha_{ik}}}$.

El modelo probabilístico lexicográfico estima v_i y a_i utilizando como datos las respuestas entregadas por un grupo de consumidores a una encuesta donde se les presenta T conjuntos de elección (preguntas) y, en cada uno de ellos, cada consumidor debe declarar qué opción prefiere de un conjunto de K alternativas de elección. En lo que sigue, esto se denota como:

$$y_{ikt} = \begin{cases} 1 & \text{; si el consumidor } i \text{ elige la alternativa } k \text{ en la ocasión } t \\ 0 & \text{; en otro caso} \end{cases}$$

Como primer paso en el procedimiento de inferencia, el modelo captura la heterogeneidad en la importancia latente de atributos y en las probabilidades de elección de los consumidores asumiendo que:

$$v_i \sim N(\bar{v}, \Sigma) \text{ y } \alpha_i \sim N(\bar{\alpha}, \psi) \quad (9)$$

Donde: $\bar{v} = (\bar{v}_1, \dots, \bar{v}_J) \sim N(0, 10^3 I_{(J-1) \times (J-1)})$ es la media de importancia latente de los atributos en la población y $\Sigma \sim IW(J+4, J+4) I_{(J-1) \times (J-1)}$ es la matriz de covarianza que captura cómo varía \bar{v} dentro de la población.

$\bar{\alpha} = (\bar{\alpha}_1, \dots, \bar{\alpha}_K) \sim N(0, 10^3 I_{(K-1) \times (K-1)})$ es la media de probabilidades de elección de las alternativas y $\psi \sim IW(K+4, (K+4) I_{(K-1) \times (K-1)})$ es la matriz de covarianza que captura cómo varía $\bar{\alpha}$ en la población.⁴

³ Se fija $v_{iJ} = 0$ para todo i pues el modelo se define sólo si una de las componentes de v_i es cero. Los demás valores son determinados en referencia al atributo base, J .

A continuación, se utiliza el método de estimación bayesiana MCMC (del inglés *Markov Chain Monte Carlo*) que procede como sigue:

- i. Comenzar con valores arbitrarios para los vectores de orden de atributos (r_i), importancia latente (v_i) y probabilidades de elección (a_i).
- ii. Dado r_i , actualizar v_i con un vector propuesto o “candidato” v_i^c que es aceptado con probabilidad:⁵

$$\min \left\{ \frac{P(v_i^c | r_i)}{P(v_i | r_i)}, 1 \right\} \quad (10)$$

Utilizando el Teorema de Bayes se puede reescribir la probabilidad anterior como:

$$\min \left\{ \frac{P(r_i | v_i^c) \frac{P(v_i^c)}{P(r_i)}}{P(r_i | v_i) \frac{P(v_i)}{P(r_i)}}, 1 \right\} = \min \left\{ \frac{P(r_i | v_i^c) P(v_i^c)}{P(r_i | v_i) P(v_i)}, 1 \right\} \quad (11)$$

Donde $P(r_i | v_i)$ está definida en la ecuación (8).

- iii. Actualizar r_i por un vector de orden de atributos candidato, r_i^c , que se construye intercambiando un par de componentes adyacentes de r_i extraídas mediante una distribución multinomial.

Utilizando los valores actualizados de v_i y de manera análoga al paso anterior, se acepta el candidato con probabilidad:

$$\min \left\{ \frac{L_i(Y_i | X, a_i, r_i^c) P(r_i^c | v_i^c) P(v_i)}{L_i(Y_i | X, a_i, r_i) P(r_i | v_i) P(v_i)}, 1 \right\} = \min \left\{ \frac{L_i(Y_i | X, a_i, r_i^c) P(r_i^c | v_i^c)}{L_i(Y_i | X, a_i, r_i) P(r_i | v_i)}, 1 \right\} \quad (12)$$

- Donde:
- $Y_i = ((y_{ikt}))$: Matriz de $K \times T$ que indica si el individuo i eligió o no la alternativa k en la ocasión (o pregunta) t .
 - $X = ((x_{jt}^k))$: Arreglo de $K \times J \times T$ que indica el nivel del atributo j que posee la alternativa k de la ocasión t .
 - $L_i(Y_i | X, a_i, r_i)$: Verosimilitud de datos condicional del individuo i , que se obtiene calculando previamente:
 - $U_i = ((u_{ikt}))$: Matriz de $K \times T$ donde u_{ikt} es la utilidad que la alternativa k de la ocasión t le reporta al individuo i , calculada con la ecuación (2).

⁴ Se utiliza la notación I para la matriz identidad y N e IW para hacer referencia a las distribuciones Normal y Wishart Invertida respectivamente.

⁵ Método “Random walk Metropolis-Hastings”.

$$z_{ikt}^h = \begin{cases} 1; & \text{si la alternativa } k \text{ de la ocasi3n } t \text{ tiene la } h\text{-3sima m3s alta} \\ & \text{utilidad entre las alternativas } 1, \dots, K. \\ 0; & \text{en otro caso} \end{cases}$$

Luego,

$$L_{it}(y_{it}|X, z_{it}, a_i, r_i) = \prod_{k=1}^K \left(a_{i1}^{z_{ikt}^1} a_{i2}^{z_{ikt}^2} \dots a_{iK}^{z_{ikt}^K} \right)^{y_{ikt}} \quad (13)$$

$$L_i(Y_i|X, a_i, r_i) = \prod_{t=1}^T L_{it} \quad (14)$$

iv. Utilizando los nuevos v_i y r_i , actualizar a_i aceptando a_i^c con probabilidad:

$$\min \left\{ \frac{L_i(Y_i|X, a_i^c, r_i) P(a_i^c|\bar{\alpha}, \psi)}{L_i(Y_i|X, a_i, r_i) P(a_i|\bar{\alpha}, \psi)}, 1 \right\} \quad (15)$$

Donde $P(a_i|\bar{\alpha}, \psi) = P(\alpha_i|\bar{\alpha}, \psi)$ al ser a_i una transformaci3n logar3tmica de α_i .

v. Volver al paso ii iterando con los nuevos valores de r_i , v_i y a_i .

Finalmente, se obtienen estimaciones de los par3metros poblacionales, \hat{v} , $\hat{\Sigma}$, $\hat{\alpha}$ y $\hat{\psi}$, calculando los valores medios de \bar{v} , $\bar{\Sigma}$, $\bar{\alpha}$ y $\bar{\psi}$ respectivamente.

CAPÍTULO II: CONSTRUCCIÓN DEL MODELO DE ELECCIÓN LEXICOGRÁFICO DE ORDENAMIENTO PARCIAL DE ATRIBUTOS

El Modelo Probabilístico Lexicográfico asume que los individuos son capaces de llevar a cabo el ejercicio racional de establecer un *ranking* compuesto por todos los atributos de acuerdo a su valoración o importancia. Sin embargo, este supuesto puede no ser siempre cierto. Podría ocurrir que, por desconocimiento o falta de experiencia, el consumidor no se encuentre familiarizado con parte de los atributos. También puede ocurrir que el consumidor sea indiferente entre los atributos menos preferidos.

En este capítulo se eliminará el supuesto de que todos los individuos son capaces de establecer dicho *ranking* completo de atributos (r), para construir así un modelo de elección lexicográfico de ordenamiento parcial de atributos utilizando como base el modelo probabilístico lexicográfico.

El modelo que se propone en este trabajo se plantea bajo enfoques homogéneo y heterogéneo en la capacidad de ordenamiento. En ambos casos, contempla modificaciones respecto al modelo probabilístico lexicográfico y algunos elementos nuevos a objeto de poder incorporar la capacidad parcial de ordenamiento de atributos.

En primer lugar, en ambos enfoques de este modelo se asume que todas las alternativas, a excepción de aquella con mayor utilidad lexicográfica, se eligen con igual probabilidad, es decir:

$$a_{ik} = \frac{1-a_{i1}}{K-1} \quad \forall 2 \leq k \leq K \quad (16)$$

Con ello, la estimación de la probabilidad de elección se reduce a la probabilidad de elección de la alternativa con mayor utilidad por individuo (a_{i1})

La transformación que vinculará a_{i1} con α_{i1} será:

$$a_{i1} = \frac{1}{1+e^{-\alpha_{i1}}} \quad (17)$$

Otros elementos, junto con el detalle de cada enfoque de estimación se explican a continuación.

2.1. Enfoque de estimación homogéneo

Como una primera aproximación en la búsqueda de la cantidad de atributos con que los individuos construyen su *ranking* u ordenamiento, el modelo considerará un ordenamiento parcial homogéneo de atributos, es decir, se asumirá que todos los individuos de la muestra establecen un *ranking* compuesto por igual cantidad de atributos (tienen igual capacidad de ordenamiento).

Bajo este supuesto, se asumirá también que si un individuo sólo es capaz de establecer un *ranking* entre los q atributos más valorados ($q \in \{1, \dots, J-1\}$) entonces, los restantes $J-q$ atributos, igualmente necesarios para construir el *ranking* r , son ordenados aleatoriamente en las posiciones finales. Dicho orden aleatorio puede ser diferente ante diferentes decisiones.

Luego, utilizando la ecuación (6), en lo que sigue, la probabilidad de observar un *ranking* de atributos r_i dado v_i y dado q , será:

$$P(r_i|v_i, q) = P(w_{ir_{i1}} > w_{ir_{i2}} > \dots > w_{ir_{iq}} > \max\{w_{ir_{iq+1}}, \dots, w_{ir_{iJ}}\}) \quad (18)$$

$$= \frac{1}{(J-q)!} \prod_{s=1}^q \frac{e^{v_{ir_{is}}}}{e^{v_{ir_{is}} + \dots + e^{v_{ir_{iJ}}}}} \quad (19)$$

Donde el factor $\frac{1}{(J-q)!}$ rescata la probabilidad de observar un ordenamiento particular de los últimos $J-q$ atributos.⁶

Con una cantidad total de N individuos, la verosimilitud de orden de atributos será:

$$L(r_i|v_i, q) = \prod_{i=1}^N \frac{1}{(J-q)!} \prod_{s=1}^q \frac{e^{v_{ir_{is}}}}{e^{v_{ir_{is}} + \dots + e^{v_{ir_{iJ}}}}} \quad (20)$$

La log-verosimilitud, en tanto, queda establecida como:

$$\ln(L(r_i|v_i, q)) = \sum_{i=1}^N \ln\left(\frac{1}{(J-q)!} \prod_{s=1}^q \frac{e^{v_{ir_{is}}}}{e^{v_{ir_{is}} + \dots + e^{v_{ir_{iJ}}}}}\right) \quad (21)$$

$$= \sum_{i=1}^N \left[\sum_{s=1}^q \ln\left(\frac{e^{v_{ir_{is}}}}{e^{v_{ir_{is}} + \dots + e^{v_{ir_{iJ}}}}}\right) - \ln((J-q)!) \right] \quad (22)$$

Por otra parte, la verosimilitud de datos condicional del individuo i ahora será $L_i(Y_i|X, a_i, \bar{r}_i, q)$, donde $\bar{r}_i = (r_{i1}, \dots, r_{iq}, \bar{r}_{iq+1}, \dots, \bar{r}_{iJ})$ es un vector de orden de atributos que comparte las primeras q componentes con r_i pero las restantes componentes corresponden a los $J-q$ atributos de menor valoración ordenados al azar.

Las correcciones incorporadas tienen incidencia directa en la probabilidad de aceptar los candidatos v_i^c , r_i^c y a_i^c , ya que ahora serán también dependientes de la cantidad de atributos q que puede ordenar un individuo y que, bajo este enfoque, es igual para toda la muestra.

⁶ Notar que si $q=J-1$, entonces estamos en el caso anterior (capacidad de ordenamiento completa), puesto que al establecer un *ranking* de $J-1$ atributos, el atributo J -ésimo sólo puede ocupar la última posición y, por lo tanto, el *ranking* está completo.

Así, incorporando los nuevos elementos a las ecuaciones (11), (12) y (15), las probabilidades de aceptar los candidatos v_i^c , r_i^c y a_i^c ahora serán, respectivamente:

$$\min \left\{ \frac{P(r_i|v_i^c, q) P(v_i^c)}{P(r_i|v_i, q) P(v_i)}, 1 \right\} \quad (23)$$

$$\min \left\{ \frac{L_i(Y_i|X, a_i, \bar{r}_i^c, q) P(r_i^c|v_i^c, q)}{L_i(Y_i|X, a_i, \bar{r}_i, q) P(r_i|v_i, q)}, 1 \right\} \quad (24)$$

$$\min \left\{ \frac{L_i(Y_i|X, a_i^c, \bar{r}_i, q) P(a_i^c|\bar{a}, \psi)}{L_i(Y_i|X, a_i, \bar{r}_i, q) P(a_i|\bar{a}, \psi)}, 1 \right\} \quad (25)$$

La verosimilitud de datos condicional de la población se calculará como:

$$L(Y_i|X, a_i, \bar{r}_i, q) = \prod_{i=1}^N \prod_{t=1}^T \prod_{k=1}^K (a_{i1}^{z_{ikt}^1} a_{i2}^{z_{ikt}^2} \dots a_{iK}^{z_{ikt}^K})^{y_{ikt}} \quad (26)$$

$$\ln(L(Y_i|X, a_i, \bar{r}_i, q)) = \sum_{i=1}^N \ln \left(\prod_{t=1}^T \prod_{k=1}^K (a_{i1}^{z_{ikt}^1} a_{i2}^{z_{ikt}^2} \dots a_{iK}^{z_{ikt}^K})^{y_{ikt}} \right) \quad (27)$$

$$= \sum_{i=1}^N \sum_{t=1}^T \sum_{k=1}^K \ln(a_{i1}^{z_{ikt}^1} a_{i2}^{z_{ikt}^2} \dots a_{iK}^{z_{ikt}^K})^{y_{ikt}} \quad (28)$$

Esta última expresión es de especial relevancia ya que para efectos de este trabajo, el criterio de estimación bajo el enfoque homogéneo será el de máxima MLL (del inglés *Marginal Log-likelihood*). Para calcular el valor máximo, se utilizará el método de clase latente, con J-1 clases que conformarán distintas versiones del modelo.

De esta manera, se estimarán los parámetros J-1 veces, primero asumiendo que el *ranking* de todos los individuos de la muestra está compuesto sólo por 1 ($q=1$) de un total de J atributos, luego asumiendo que está compuesto por 2 atributos ($q=2$), y así sucesivamente, hasta estimar el modelo que asume que los individuos son capaces de generar un *ranking* de J-1 atributos (lo cual es equivalente a un ordenamiento completo pues el atributo no considerado sólo puede ocupar la posición final).

Luego, se obtienen los valores de \hat{q} , \hat{v} , y \hat{a}_1 utilizando el modelo que posee el máximo valor de MLL.

El valor de q estimado se puede interpretar como la cantidad de atributos que el conjunto mayoritario de individuos es capaz de considerar en su *ranking* u ordenamiento de atributos. Es decir, contemplando la posibilidad de que los individuos tengan diferentes capacidades de ordenamiento, se espera que lo más frecuente sea encontrar individuos que consideren la cantidad q de atributos recién estimada.

Este procedimiento se puede entender como una “estimación externa” y tiene la dificultad de que es necesario realizar la estimación J-1 veces, sin embargo, es un camino para acercarse al resultado buscado y una forma preliminar para validar el modelo propuesto mediante el trabajo de simulación que se detallará más adelante.

2.2. Enfoque de estimación heterogéneo

En esta parte se quiere permitir que, dentro de la muestra de individuos, existan diferentes capacidades de ordenamiento de atributos. Así, ahora cada individuo podrá tener su propio q_i representando la cantidad de atributos sobre los cuales construye su *ranking* de atributos y ya no existirá una cantidad única, válida para todos los individuos.

De esta manera, como adaptación inmediata a la ecuación (19), la probabilidad de observar un *ranking* de atributos r_i dado v_i y dado q_i , será:

$$P(r_i|v_i, q_i) = \frac{1}{(J-q_i)!} \prod_{s=1}^{q_i} \frac{e^{v_{ir_{is}}}}{e^{v_{ir_{is}}} + \dots + e^{v_{ir_{iJ}}}} \quad (29)$$

Y el vector de orden de atributos adaptado para el caso homogéneo, \bar{r}_i , bajo este enfoque será $\bar{r}_i = (r_{i1}, \dots, r_{iq_i}, \bar{r}_{iq_i+1}, \dots, \bar{r}_{iJ})$.

Con el enfoque heterogéneo, ya no es necesario utilizar clase latente para decidir el mejor q de acuerdo al modelo con mejor ajuste. Esta vez, la estimación de q_i es interna, siendo el propio modelo el que se encarga de actualizar y decidir cada q_i .

Para lograrlo, se modifica la parte final del método MCMC, incorporando un paso intermedio, es decir, el paso v. ahora será:

- v. Considerando los valores actualizados de v_i , r_i y a_i , se actualiza q_i aceptando el candidato q_i^c , que se obtuvo de una distribución uniforme discreta con valores hasta $J-1$, con probabilidad:

$$\min \left\{ \frac{L_i(Y_i|X, a_i, \bar{r}_i, q_i^c) P(r_i|v_i, q_i^c)}{L_i(Y_i|X, a_i, \bar{r}_i, q_i) P(r_i|v_i, q_i)}, 1 \right\} \quad (30)$$

- vi. Volver al paso ii iterando con los nuevos valores de r_i , v_i , a_i y q_i .

Finalmente, para cada individuo, se obtienen estimaciones \hat{q}_i calculando el valor modal de q_i a lo largo de todas las iteraciones.

CAPÍTULO III: VALIDACIÓN DEL MODELO CON SIMULACIÓN DE MONTE CARLO

Con el objetivo de validar el modelo propuesto y evaluar la precisión de la recuperación de parámetros, se realizaron sucesivas simulaciones de Monte Carlo para dos diseños de encuestas diferentes.

Ambos diseños coinciden en su estructura donde, en cada pregunta, se presentan cuatro alternativas de elección para un mismo tipo de producto (que se conserva a lo largo de toda la encuesta) y el encuestado debe declarar qué alternativa prefiere. Para ello, sólo se provee el “perfil” de cada alternativa, el cual corresponde a una serie de cinco características, cada una de las cuales equivale a uno de cuatro niveles de un atributo particular.

Por ejemplo, un estudio de *notebooks* podría describir sus alternativas utilizando los siguientes atributos y niveles:

	Atributo 1: Precio	Atributo 2: Procesador	Atributo 3: Memoria RAM	Atributo 4: Espacio en disco	Atributo 5: Tamaño pantalla
Nivel 1	\$ 750.000	Intel Dual Core	2 GB	128 GB	11"
Nivel 2	\$ 475.000	Intel i3	4 GB	320 GB	13"
Nivel 3	\$ 225.000	Intel i5	6 GB	500 GB	14"
Nivel 4	\$ 150.000	Intel i7	8 GB	750 GB	15"

Tabla 3.1. Ejemplo atributos/niveles encuesta *notebooks*

Así, el perfil de una alternativa podría ser:

\$ 275.000	Intel Dual Core	8 GB	500 GB	11"
-------------------	------------------------	-------------	---------------	------------

Y una pregunta podría ser la siguiente:

Si éstas fueran sus únicas opciones de compra, ¿cuál escogería?					
	Precio	Procesador	Memoria RAM	Espacio en disco	Tamaño pantalla
Notebook 1	\$ 750.000	Intel i3	4 GB	500 GB	15"
Notebook 2	\$ 475.000	Intel i5	6 GB	750 GB	11"
Notebook 3	\$ 225.000	Intel i7	8 GB	128 GB	13"
Notebook 4	\$150.000	Intel Dual Core	2 GB	320 GB	14"

Tabla 3.2. Ejemplo pregunta encuesta *notebooks* con diseño ortogonal

Para efectos de la simulación, no es relevante el tipo de producto a evaluar ni cuáles son específicamente sus atributos y niveles pues se trata de un diseño genérico,

con cinco atributos, cada uno con cuatro niveles (ordenados de manera creciente en cuanto a su atractivo) y donde, en cada pregunta, el encuestado (ficticio) debe elegir una de cuatro alternativas descritas por su perfil.⁷

A continuación, se explica en detalle el procedimiento a seguir, las características de cada diseño y los principales resultados obtenidos.

3.1. Procedimiento

- i. Se establece la cantidad de atributos que cada individuo puede considerar en el *ranking* de atributos (q_i ⁸) y la media de importancia latente de los atributos de la población (\bar{v}). Luego se genera, para cada individuo, el vector v_i a partir de una distribución normal de media \bar{v} y matriz de covarianza Σ . A continuación, se obtiene la importancia del atributo j para el individuo i (vector w_i) agregando un error de distribución tipo I independiente e idénticamente distribuido.
- ii. Con los valores anteriores, se obtiene el vector individual de ordenamiento de atributos para cada conjunto de elección t , r_{it} . Para ello, si el individuo i es capaz de ordenar sólo q_i de los $J=5$ atributos, entonces los $J-q_i$ atributos restantes ocupan una posición aleatoria dentro de las últimas $J-q_i$ posiciones del *ranking*. Dicho ordenamiento no necesariamente es el mismo en los $J-q_i$ atributos restantes para diferentes conjuntos de elección (preguntas). De ahí la necesidad de establecer un ordenamiento por individuo y por conjunto de elección.
- iii. A continuación, de forma análoga a la generación del vector v_i , se construye el vector α_i ⁹:
- iv. Finalmente, por cada pregunta, se calcula la utilidad lexicográfica de cada alternativa y se generan las elecciones de los individuos.

En este trabajo se simularon muestras poblacionales que responden dos tipos de encuesta diferentes. El primero de ellos corresponde a una encuesta con diseño ortogonal. El segundo tipo de encuesta está hecho en base a un diseño que se denominará “diseño con coincidencias en niveles”.

Así, con el objeto de evaluar la capacidad predictiva de los enfoques de estimación homogéneo y heterogéneo, se aplicó el modelo a las siguientes diez condiciones experimentales:

⁷ Más adelante, se detalla una aplicación empírica donde sí es necesario asignar atributos y niveles conformes a la realidad del producto en estudio.

⁸ Con $q_i=q$ para todo i cuando el enfoque de estimación es homogéneo.

⁹ A lo largo de este capítulo de simulación se asume un bajo error de juicio por parte de los consumidores, lo cual se logra conservando el valor $\bar{\alpha}_1 = 3,5$ que da origen a una precisión media $\bar{a}_1 = 97\%$.

Diseño 1 u ortogonal			
Simulación	Estimación	Cantidad de Preguntas	Tamaño de la muestra
Homogénea	Homogénea	20	125
Heterogénea	Heterogénea	20	125
Diseño 2 o con coincidencias en niveles			
Simulación	Estimación	Cantidad de Preguntas	Tamaño de la muestra
Homogénea	Homogénea	20	125
Heterogénea	Heterogénea	20	125
Homogénea	Homogénea	40	125
Heterogénea	Heterogénea	40	125
Homogénea	Homogénea	20	250
Heterogénea	Heterogénea	20	250
Homogénea	Heterogénea	40	250
Heterogénea	Heterogénea	40	250

Tabla 3.3. Condiciones experimentales en análisis simulación

A los dos enfoques de estimación se les exigirá ser capaces de recuperar los parámetros poblacionales q , \bar{v}_i y $\bar{\alpha}_{i1}$, donde \bar{v}_i es la media poblacional del vector v_i y $\bar{\alpha}_{i1}$ es la media poblacional de α_{i1} . El criterio general será que los intervalos de confianza al 95% de las estimaciones \hat{v}_i y $\hat{\alpha}_{i1}$, contengan al parámetro poblacional real. El criterio para evaluar la recuperación de q se detallará conforme se expongan los resultados.

Debido a la variedad de condiciones experimentales que se abordarán, sólo se expondrán en este capítulo los resultados vinculados a la recuperación del parámetro q y una noción respecto a la recuperación de los demás parámetros. Es posible ver el detalle de estimaciones de parámetros y otras medidas de desempeño en los anexos.

La principal limitante en este trabajo de simulación es el tiempo computacional requerido para alcanzar la convergencia del modelo. Por ello se acotó la cantidad de simulaciones a 170 ensayos experimentales en total. En cada estimación, el procesamiento se detuvo al concluir la iteración número 100.000, cuyo tiempo computacional requerido osciló entre las 4 y 15 horas¹⁰.

3.2. Diseño 1: Ortogonal

Utilizando este diseño, se crea una encuesta de 20 preguntas, construida de manera ortogonal. En cada simulación, se tomará una muestra ficticia de 125 personas.

¹⁰ El tiempo computacional requerido depende principalmente del tamaño de la muestra simulada y cantidad de preguntas de la encuesta

Este tipo de diseño es el más eficiente por cuanto permite examinar simultáneamente muchos factores con una cantidad reducida de preguntas. Es por esta razón que lo encontramos frecuentemente en aplicaciones prácticas pues es una buena forma de reducir los esfuerzos de diseño estadístico.

En un diseño ortogonal, cada nivel de cada atributo debe aparecer la misma cantidad de veces en cada conjunto de elección (diseño “balanceado”) y lo mismo debe ocurrir con cada combinación de par de niveles de distintos atributos.

3.2.1. Resultados diseño 1

3.2.1.1. Simulación homogénea - Estimación homogénea

La Tabla 3.4. muestra la *marginal log-likelihood*, en adelante, MLL, para cuatro simulaciones con capacidades homogéneas de ordenamiento ($q=1, 2, 3$ y 4 en las filas), cada una de ellas, estimada cuatro veces (una vez bajo el supuesto de una capacidad homogénea de ordenamiento diferente).

El enfoque de estimación homogéneo quedará validado, en términos de la recuperación del parámetro q si, por fila, el máximo valor de MLL se obtiene con una estimación que asume el mismo valor de q que posee la población en la simulación, es decir, si los máximos valores se encuentran en la diagonal.

Capacidad de ordenamiento de la población en simulación	Capacidad de ordenamiento de la población en estimación			
	$q = 1$	$q = 2$	$q = 3$	$q = 4$
$q = 1$	-375,90	-375,91	-377,14	<u>-374,25</u>
$q = 2$	-464,50	-461,54	<u>-459,93</u>	-460,96
$q = 3$	-381,29	-381,87	<u>-376,79</u>	-381,36
$q = 4$	-422,64	-423,84	<u>-421,98</u>	-422,40

Tabla 3.4. MLL en diseño 1 con simulación homogénea (20 preguntas, 125 individuos) y estimación homogénea

Los parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en general se recuperan, tanto en los modelos que asumen una capacidad de ordenamiento igual a aquella con la que fue simulada la muestra (modelos de la diagonal de la Tabla 3.4.), como en los modelos que arrojaron un mayor valor de MLL (modelos subrayados en la Tabla 3.4.). El parámetro \bar{v}_{i4} es el que presenta mayores dificultades en su estimación (Anexo A1).

3.2.1.2. Simulación heterogénea - Estimación heterogénea

A diferencia del caso anterior, ahora sólo se simula una vez una población con capacidad heterogénea de ordenamiento de atributos, es decir, parte de los individuos puede establecer un *ranking* de 1 atributo, otros de 2, otros de 3 y otros de 4 atributos.

Se estima el modelo una vez y se obtiene la matriz de confusión de la Tabla 3.5. En ella se indica, en la fila con $q=1$, de la totalidad de individuos que sólo pueden establecer un *ranking* compuesto por 1 atributo, qué porcentaje estima el modelo que efectivamente tiene una capacidad de ordenamiento $q=1$, qué porcentaje estima que posee $q=2$, $q=3$ y $q=4$ (errando en éstos tres últimos casos). Lo mismo para las filas con $q=2$, $q=3$ y $q=4$ cambiando respectivamente el universo de individuos considerado.

Bajo esta condición experimental, el enfoque de estimación heterogéneo se validará en caso de que los mayores valores de cada fila se encuentren en la diagonal.

Capacidad de ordenamiento en simulación	Capacidad de ordenamiento estimada			
	$q = 1$	$q = 2$	$q = 3$	$q = 4$
$q = 1$	30,00	16,67	<u>33,33</u>	20,00
$q = 2$	<u>30,00</u>	26,67	23,33	20,00
$q = 3$	<u>36,36</u>	27,27	30,30	6,06
$q = 4$	43,75	<u>31,25</u>	9,38	15,63

Tabla 3.5. Matriz de confusión porcentual en diseño 1 con simulación heterogénea (20 preguntas, 125 individuos) y estimación heterogénea

En este caso, la media de importancia latente de los atributos para la población (vector \bar{v}) es diferente dependiendo de cuál sea la capacidad de ordenamiento de los individuos. Así, el vector \bar{v} del subgrupo que posee $q=1$ difiere del subgrupo con $q=2$ y del con $q=3$ y $q=4$.

Luego, para analizar la recuperación de los parámetros \bar{v}_i y $\bar{\alpha}_{i1}$, se divide la muestra en cuatro subgrupos de acuerdo a la capacidad de ordenamiento de atributos estimada y se analizan los valores medios estimados por subgrupo.

Procediendo de dicha forma, se observa que todos los subgrupos son capaces de recuperar todos los parámetros a excepción de \bar{v}_{i4} (Anexo A2).

3.2.2. Análisis de resultados del diseño 1

Observando los resultados anteriores es posible notar que ambos enfoques fallan en la estimación de la capacidad de ordenamiento real de atributos.

En efecto, en el caso simulación homogénea y estimación homogénea, los máximos valores de MLL (aquellos que aparecen subrayados en la Tabla 3.4.), sólo coinciden con la capacidad de ordenamiento real de la muestra para el caso $q=3$.

Queda de manifiesto la débil capacidad para discriminar o decidir cuál es el modelo que generó los datos. A modo de ejemplo, para la muestra simulada con capacidad de ordenamiento de un atributo, si bien el modelo estima erróneamente una capacidad de ordenamiento de cuatro atributos, la diferencia absoluta entre el máximo MLL y el mínimo es sólo de 2,89.

Los resultados expuestos en la Tabla 3.5., correspondientes al caso simulación heterogénea y estimación heterogénea, confirman que no se puede recuperar el parámetro q pues considerando el porcentaje de acierto en las distintas capacidades de ordenamiento se obtiene un acierto global de 25,6%, que es similar a lo que se podría obtener si, sin mediar estudio alguno, se estimara que todos los individuos de la muestra poseen la misma capacidad de ordenamiento.

Por otra parte, es posible notar que en la mayoría de los casos se recuperan los parámetros \bar{v}_i y $\bar{\alpha}_{i1}$. En particular, en el caso de simulación homogénea y estimación homogénea, cuando se estima bajo el supuesto de capacidad de ordenamiento igual a la capacidad de ordenamiento con que fue simulada la muestra, la recuperación de \bar{v}_i y $\bar{\alpha}_{i1}$ es igual o mejor que cuando se estima bajo el supuesto de $q=4$ (supuesto que se relaja en esta investigación).

Además, al simular una muestra heterogénea en la capacidad de ordenamiento de atributos, la recuperación de los parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ es igualmente buena cuando se estima de manera heterogénea que cuando se estima bajo el supuesto de $q=4$.

En la búsqueda de la razón por la cual no es posible recuperar el parámetro q (principal objetivo de este trabajo), surge la interrogante de si esto se debe a errores en el modelo, error en el criterio de estimación (máxima MLL) o características del diseño que impiden la correcta estimación de la capacidad de ordenamiento de atributos.

Una posible explicación a lo ocurrido es que, dado que se asumen preferencias lexicográficas, si se considera un diseño donde las cuatro alternativas de cada pregunta son muy diferentes entre sí, entonces los individuos no utilizarán más que un atributo para tomar su decisión y, en consecuencia, no se les da la oportunidad de revelar realmente con cuántos atributos son capaces de construir su *ranking*.

En efecto, este tipo de diseño, en la búsqueda de una mayor eficiencia estadística, en cada pregunta le presenta a los encuestados conjuntos de elección con alternativas de perfiles muy diferentes (con pocas o ninguna coincidencia en sus niveles).

Por ejemplo, si para un individuo el atributo precio es lexicográficamente preferido a los demás, entonces le bastará con escoger la alternativa que tenga mejor precio. Lo mismo ocurrirá con todas las preguntas. Así, difícilmente entregará información respecto a cuál era su capacidad real de ordenamiento de atributos pues en la gran mayoría de los casos le bastó con utilizar sólo uno.

Sin embargo, si dentro del set de alternativas existe más de una con el mismo precio conveniente, entonces el individuo tendrá la necesidad de considerar el segundo atributo más relevante en su *ranking* para tomar una decisión. Si con eso no consigue llegar a una decisión, entonces pasará a evaluar el tercer atributo, y así sucesivamente.

Para verificar la hipótesis anterior, se procedió a evaluar el modelo bajo un diseño diferente, que es el que se describe a continuación.

3.3. Diseño 2: Con coincidencias en niveles

El segundo diseño, responde a la necesidad de averiguar cuál es la real capacidad de ordenamiento de atributos de los individuos.

Este diseño difiere del anterior por cuanto, dentro de un mismo conjunto de elección, los niveles que definen el perfil de una alternativa pueden coincidir con los de otra(s) alternativa(s) de la misma pregunta.

Por ejemplo, en una pregunta cualquiera con 4 alternativas, últimas tres pueden tener exactamente el mismo nivel para el atributo precio, luego, las alternativas 2 y 3 pueden coincidir en el procesador, y así sucesivamente. No puede suceder que existan dos alternativas iguales por conjunto de elección.

Para la construcción de este diseño se utilizó como base el diseño ortogonal y se modificaron luego algunos de los niveles de ciertas alternativas de modo que exista mayor similitud entre las alternativas de un mismo conjunto de elección, aun cuando este procedimiento sacrifique eficiencia estadística.

Así, en este diseño es posible encontrar preguntas como:

Si éstas fueran sus únicas opciones de compra, ¿cuál escogería?					
	Precio	Procesador	Memoria RAM	Espacio en disco	Tamaño pantalla
Notebook 1	\$ 750.000	Intel i3	4 GB	500 GB	15"
Notebook 2	\$ 475.000	Intel i3	6 GB	750 GB	15"
Notebook 3	\$ 475.000	Intel i3	2 GB	128 GB	13"
Notebook 4	\$475.000	Intel Dual Core	2 GB	320 GB	13"

Tabla 3.6. Ejemplo pregunta encuesta *notebooks* con diseño con coincidencias en niveles

Esta vez, en caso que los individuos puedan establecer un *ranking* de, por ejemplo, tres de los cinco atributos, este diseño promueve que utilice dichos tres atributos para tomar su decisión por lo menos en una parte de la totalidad de preguntas de la encuesta, pese al supuesto de preferencias lexicográficas.

En este tipo de casos, si un individuo tiene el *ranking* de preferencia de atributos Precio – Procesador – Espacio en disco (y los otros dos atributos los ordena al azar

dentro de las últimas dos posiciones en cada decisión) en orden decreciente de importancia, entonces se verá forzado a revelar sus preferencias puesto que sólo considerando el atributo más importante (en este caso el precio) no le bastará pues no será capaz de decidir entre las alternativas 2, 3 y 4. Luego, utilizando los primeros dos atributos (en este caso precio y procesador) tampoco es capaz de tomar una decisión pues las alternativas 2 y 3 son exactamente iguales en dichas características. Así, sólo cuando utiliza tres atributos puede establecer qué alternativa le brinda mayor utilidad lexicográfica (en este caso la alternativa 2).

Este modelo supone que los individuos cuyo *ranking* contemple un atributo (como por ejemplo, el precio) o dos atributos (como por ejemplo, Precio - Procesador) entonces, en esta pregunta deberán escoger de manera aleatoria entre las alternativas 2, 3 y 4 en el primer caso y 2 y 3 en el segundo.

Los individuos que sean capaces de construir un *ranking* de 4 atributos, como Precio – Procesador – Espacio en disco – Memoria RAM, podrán no utilizar todo el *ranking* para tomar la decisión, de forma similar a lo que ocurría con el diseño ortogonal.

Sin embargo, en este diseño existen preguntas con diferentes intensidades de coincidencias, desde algunas con alternativas muy diferentes (diseño ortogonal) hasta otras con alternativas prácticamente iguales, todo esto, con el objeto de poner a prueba diferentes capacidades de ordenamiento.

Con este diseño se construyeron dos encuestas. La primera, compuesta por 20 conjuntos de elección y la segunda por 40. Para ambos casos, se realizaron simulaciones de conjuntos poblacionales de 125 y 250 individuos.

3.3.1. Resultados diseño 2

Mayores detalles de resultados se encuentran en el Anexos B.

3.3.1.1. Simulación homogénea (20 preguntas, 125 individuos) - Estimación homogénea

Al igual que en el diseño 1, bajo esta condición experimental y todas las que siguen a continuación, se validará el enfoque de estimación homogéneo en términos de la recuperación del parámetro q si los máximos valores de MLL de cada fila se ubicaran en la diagonal de la siguiente tabla:

Capacidad de ordenamiento de la población en simulación	Capacidad de ordenamiento de la población en estimación			
	q = 1	q = 2	q = 3	q = 4
q = 1	<u>-478,94</u>	-951,45	-1107,90	-1143,60
q = 2	-569,66	<u>-471,63</u>	-671,13	-799,60
q = 3	-468,59	-402,07	<u>-388,25</u>	-569,95
q = 4	-501,25	-450,20	-455,68	<u>-448,27</u>

Tabla 3.7. MLL en diseño 2 con simulación homogénea (20 preguntas, 125 individuos) y estimación homogénea

3.3.1.2. Simulación heterogénea (20 preguntas, 125 individuos) - Estimación heterogénea

El enfoque de estimación heterogéneo quedará validado si los máximos valores de cada fila se ubican en la diagonal de la siguiente matriz de confusión:

Capacidad de ordenamiento en simulación	Capacidad de ordenamiento estimada			
	q = 1	q = 2	q = 3	q = 4
q = 1	<u>86,67</u>	10,00	3,33	0,00
q = 2	6,67	<u>46,67</u>	30,00	16,67
q = 3	0,00	6,06	<u>57,58</u>	36,36
q = 4	6,25	12,50	12,50	<u>68,75</u>

Tabla 3.8. Matriz de confusión porcentual en diseño 2 con simulación heterogénea (20 preguntas, 125 individuos) y estimación heterogénea

3.3.1.3. Simulación homogénea (40 preguntas, 125 individuos) - Estimación homogénea

Capacidad de ordenamiento de la población en simulación	Capacidad de ordenamiento de la población en estimación			
	q = 1	q = 2	q = 3	q = 4
q = 1	<u>-1022,60</u>	-1864,00	-2235,50	-2278,40
q = 2	-1051,00	<u>-888,09</u>	-1447,40	-1578,90
q = 3	-1072,30	-847,97	<u>-784,75</u>	-1099,90
q = 4	-1168,50	-893,63	-863,44	<u>-855,13</u>

Tabla 3.9. MLL en diseño 2 con simulación homogénea (40 preguntas, 125 individuos) y estimación homogénea

3.3.1.4. Simulación heterogénea (40 pregunta, 125 individuos) - Estimación heterogénea

Capacidad de ordenamiento en simulación	Capacidad de ordenamiento estimada			
	q = 1	q = 2	q = 3	q = 4
q = 1	<u>86,67</u>	6,67	3,33	3,33
q = 2	0,00	<u>90,00</u>	6,67	3,33
q = 3	3,03	12,12	<u>54,55</u>	30,30
q = 4	0,00	6,25	12,50	<u>81,25</u>

Tabla 3.10. Matriz de confusión porcentual en diseño 2 con simulación heterogénea (40 preguntas, 125 individuos) y estimación heterogénea

3.3.1.5. Simulación homogénea (20 preguntas, 250 individuos) - Estimación homogénea

Capacidad de ordenamiento de la población en simulación	Capacidad de ordenamiento de la población en estimación			
	q = 1	q = 2	q = 3	q = 4
q = 1	<u>-947,53</u>	-1902,90	-2178,10	-2247,00
q = 2	-961,88	<u>-806,54</u>	-1214,20	-1490,40
q = 3	-927,36	<u>-809,58</u>	-824,66	-1159,40
q = 4	-1068,10	-922,01	<u>-921,06</u>	-923,20

Tabla 3.11. MLL en diseño 2 con simulación homogénea (20 preguntas, 250 individuos) y estimación homogénea

3.3.1.6. Simulación heterogénea (20 preguntas, 250 individuos) - Estimación heterogénea

Capacidad de ordenamiento en simulación	Capacidad de ordenamiento estimada			
	q = 1	q = 2	q = 3	q = 4
q = 1	<u>87,72</u>	8,77	1,75	1,75
q = 2	0,00	<u>57,14</u>	19,05	23,81
q = 3	4,62	6,15	<u>44,62</u>	<u>44,62</u>
q = 4	7,69	10,77	6,15	<u>75,38</u>

Tabla 3.12. Matriz de confusión porcentual en diseño 2 con simulación heterogénea (20 preguntas, 250 individuos) y estimación heterogénea

3.3.1.7. Simulación homogénea (40 preguntas, 250 individuos) - Estimación homogénea

Capacidad de ordenamiento de la población en simulación	Capacidad de ordenamiento de la población en estimación			
	q = 1	q = 2	q = 3	q = 4
q = 1	<u>-2096,10</u>	-3818,30	-4516,20	-4629,30
q = 2	-2051,20	<u>-1690,00</u>	-2789,70	-3038,30
q = 3	-2038,00	-1677,10	<u>-1597,30</u>	-2209,50
q = 4	-2062,70	-1639,20	<u>-1520,00</u>	-1535,40

Tabla 3.13. MLL en diseño 2 con simulación homogénea (40 preguntas, 250 individuos) y estimación homogénea

3.3.1.8. Simulación heterogénea (40 pregunta, 250 individuos) - Estimación heterogénea

Capacidad de ordenamiento en simulación	Capacidad de ordenamiento estimada			
	q = 1	q = 2	q = 3	q = 4
q = 1	<u>84,21</u>	15,79	0,00	0,00
q = 2	4,76	<u>77,78</u>	9,52	7,94
q = 3	0,00	4,62	<u>72,31</u>	23,08
q = 4	1,54	6,15	7,69	<u>84,62</u>

Tabla 3.14. Matriz de confusión porcentual en diseño 2 con simulación heterogénea (40 preguntas, 250 individuos) y estimación heterogénea

En cuanto a la recuperación de los parámetros \bar{v}_i y $\bar{\alpha}_{i1}$, cuando se trata del enfoque de estimación homogéneo, los modelos situados en la diagonal de las Tablas 3.7., 3.9., 3.11. y 3.13, (aquellos que asumen una capacidad de ordenamiento igual a aquella con la que fue simulada la muestra), son capaces de recuperar la mayoría de las componentes del vector \bar{v}_i y, sólo en algunos casos, recupera también $\bar{\alpha}_{i1}$ (ver Anexos B 1, 3, 5 y 7).

Cuando el enfoque de estimación es heterogéneo, separando las estimaciones de acuerdo a los subgrupos compuestos por individuos con igual estimación del parámetro q, el modelo recupera las componentes relevantes del vector \bar{v}_i pero en general no se recupera $\bar{\alpha}_{i1}$ (ver Anexos B 2, 4, 6 y 8).

3.3.2. Análisis de resultados del diseño 2

Estimación del parámetro q

A diferencia de los resultados obtenidos al estimar el modelo utilizando el diseño 1, esta vez se obtienen resultados que son favorables en términos de la recuperación del parámetro q .

Los resultados anteriores revelan que, cuando se trata de muestras poblacionales de 125 individuos y del enfoque de estimación homogéneo, siempre fue posible estimar la cantidad de atributos que los individuos son capaces de incorporar en un *ranking* de orden de atributos, ya sea ellos revelen sus preferencias en una encuesta de 20 o 40 preguntas.

En efecto, las Tablas 3.7. y 3.9. muestran que de acuerdo al criterio de máxima MLL, ya sea se simule una muestra poblacional donde todos tengan una capacidad de ordenamiento de atributos de 1, 2, 3 o 4 atributos, el modelo es siempre capaz de estimar cuál es el valor del parámetro q que posee la población.

Pese a los buenos resultados obtenidos para muestras de 125 individuos, cuando se trata de 250 individuos ocurre algo inesperado: al estimar utilizando el enfoque homogéneo, el modelo falla en tres de los ocho escenarios. Los primeros dos errores son visibles en las dos últimas filas de la Tabla 3.11. donde, utilizando como datos las respuestas a una encuesta de 20 preguntas, al simular poblaciones con capacidad homogénea de ordenamiento de 3 o 4 atributos, se estima erróneamente $q=2$ y $q=3$ respectivamente.

El tercer error se observa en la última fila de la Tabla 3.13. donde, al simular una muestra con capacidad de ordenamiento de 4 atributos respondiendo a una encuesta con 40 preguntas, el modelo estima que el parámetro es $q=3$.

Una posible explicación a lo ocurrido radica en la necesidad de mayor tiempo computacional para llevar a cabo más iteraciones.

Para probar esta hipótesis, se realizaron nuevas estimaciones de estos casos particulares pero, esta vez, comenzando a iterar desde los valores simulados. Como resultado, uno de los tres errores se corrige y los otros reportan un MLL superior, pero no lo suficiente como para convertirse en el máximo.

Por otro lado, las tasas globales de acierto en la estimación del parámetro q del enfoque de estimación heterogéneo son:

Cantidad de preguntas de la encuesta	Tamaño de la muestra simulada	
	125	250
20	65,6%	65,6%
40	77,6%	79,6%

Tabla 3.15. Tasas globales de acierto en la estimación del parámetro q bajo el enfoque de estimación heterogéneo

Estos resultados revelan la capacidad del enfoque heterogéneo para estimar la capacidad de ordenamiento de atributos a nivel individual y son consistentes con la idea de que al aumentar la cantidad de preguntas, cada individuo entrega más información respecto a sus preferencias personales y, por lo tanto, el modelo realiza una mejor estimación.

Respecto a la precisión en la estimación de la capacidad de ordenamiento de atributos, cabe mencionar que todas condiciones experimentales analizadas bajo el enfoque de estimación heterogéneo comparten un patrón: en términos netos, la cantidad de individuos cuya capacidad de ordenamiento estimada es $\hat{q}=1$, se encuentra subestimada y, por otro lado, la cantidad de individuos cuya capacidad de ordenamiento estimada es $\hat{q}=4$, está sobre estimada.

En efecto, la cantidad de individuos cuya capacidad de ordenamiento fue estimada como $\hat{q}=2, 3$ o 4 teniendo en realidad $q=1$ es superior a la cantidad de individuos cuya capacidad de ordenamiento fue estimada como $\hat{q}=1$ teniendo en realidad $q=2, 3$ o 4 . Por otro lado, la cantidad de individuos cuya capacidad de ordenamiento fue estimada como $\hat{q}=1, 2$ o 3 teniendo en realidad $q=4$ es inferior a la cantidad de individuos cuya capacidad de ordenamiento fue estimada como $\hat{q}=4$ teniendo en realidad $q=1, 2$ o 3 .

Estimación del parámetro \bar{v}_i

Utilizando el enfoque de estimación homogéneo, es posible observar que, para las cuatro condiciones estudiadas (muestras de 125 o 250 personas declarando preferencias en encuestas de 20 o 40 preguntas), el modelo que utiliza el supuesto correcto de capacidad de ordenamiento (mismo valor de q que el utilizado para generar la muestra) es capaz de recuperar la gran mayoría de las componentes de \bar{v}_i , con dificultades para estimar \bar{v}_{i4} (ver Anexos B 1, 3, 5 y 7) cuando la muestra simulada posee una capacidad de ordenamiento de atributos $q=1$.

Esta menor capacidad para estimar \bar{v}_{i4} puede explicarse por la gran aleatoriedad involucrada en la toma de decisiones. En efecto, si $q=1$, cuando un atributo no sea suficiente para tomar una decisión (lo cual sucede a menudo dado el tipo de diseño utilizado), entonces los individuos eligen aleatoriamente entre las alternativas que posean el nivel más favorable del atributo más importante.

Sumado a lo anterior, los parámetros poblacionales son tales que para la gran mayoría de la muestra el atributo más importante es el primero. Por efectos de la distribución de probabilidad del error involucrado, para algunos es más importante el

segundo pero nunca es el cuarto atributo el más importante y, por lo tanto, el modelo es incapaz de estimar el parámetro asociado.

Por otro lado, independiente del tamaño de la muestra o de la cantidad de preguntas de la encuesta, la recuperación del parámetro \bar{v}_i es mejor (o a lo sumo igual) cuando se estima asumiendo la capacidad de ordenamiento que realmente tiene la muestra que cuando se asume $q=4$.

Este resultado es importante por cuanto revela que, pese a que con el supuesto antiguo de capacidad de ordenamiento completo de atributos (todos los individuos poseen $q=4$), se obtienen buenas estimaciones, éstas son aún mejores cuando el modelo asume la capacidad de ordenamiento real de la muestra, cantidad que en una aplicación real puede ser estimada de acuerdo al criterio de máxima MLL expuesto en este trabajo.

Observando el detalle de recuperación de las componentes del vector \bar{v}_i , se puede apreciar que cuando la cantidad de preguntas de la encuesta es mayor, la recuperación de \bar{v}_i es mejor. La medida de desempeño que cuantifica esta apreciación es el error cuadrático medio (RMSE), que corresponde a una medida de distancia media entre las componentes del vector \bar{v}_i y la estimación \hat{v}_i .

Así, si se calcula un RMSE promedio de las estimaciones donde se asume la capacidad de ordenamiento real de la muestra se obtiene:

Cantidad de preguntas de la encuesta	Tamaño de la muestra simulada	
	125	250
20	0,120	0,10
40	0,10	0,09

Tabla 3.16. RMSE promedio en estimación del parámetro \bar{v}_i bajo el enfoque de estimación homogéneo

Para analizar los resultados provenientes del enfoque de estimación heterogéneo, al igual que en el Diseño, 1, ellos se diferencian de acuerdo a la estimación del parámetro q (ver Anexos B 2, 4, 6 y 8), es decir, se construyen cuatro subgrupos de individuos donde, los integrantes del primer subgrupo son aquellos que poseen $\hat{q}=1$, los miembros del segundo poseen $\hat{q}=2$ y así sucesivamente. Esto porque el parámetro poblacional \bar{v}_i con que se construyó la muestra no es único sino que depende de la capacidad de ordenamiento que posea el individuo. Luego, se cree conveniente analizar los resultados también por separado.

De esta forma, se puede apreciar que de los diferentes subgrupos se obtienen estimaciones de las componentes de \bar{v}_i que son muy similares entre sí pero, a la hora de analizar si recuperan los parámetros poblacionales reales se detecta el siguiente patrón:

- Los subgrupos con capacidad de ordenamiento estimada $\hat{q}=1$ en general son capaces de recuperar únicamente la componente \bar{v}_{i1} del vector \bar{v}_i .

- Los subgrupos con capacidad de ordenamiento estimada $\hat{q}=2$ en general son capaces de recuperar únicamente las componentes \bar{v}_{i1} y \bar{v}_{i2} del vector \bar{v}_i .
- Los subgrupos con capacidad de ordenamiento estimada $\hat{q}=3$ en general son capaces de recuperar \bar{v}_{i1} , \bar{v}_{i2} y \bar{v}_{i3} del vector \bar{v}_i .
- Los subgrupos con capacidad de ordenamiento estimada $\hat{q}=4$ en general son capaces de recuperar todas las componentes del vector \bar{v}_i .

Si se calculan los valores medios globales de la estimación de \bar{v}_i (sin separación por subgrupo), se detecta, en términos de la recuperación de la media global \bar{v}_i , que los resultados son igual de buenos que cuando se estima bajo el antiguo supuesto de capacidad homogénea de ordenamiento $q=4$.

Así, pese a que con el enfoque de estimación heterogéneo se obtienen buenas cifras de acierto en la estimación de la capacidad de ordenamiento de atributos a nivel individual, es necesario tener conciencia de que, en términos individuales, los individuos cuya estimación \hat{q} sea menor a $J-1$ (cuatro en este caso) sólo recuperarán parte de las componentes de \bar{v}_i . Este hecho se ve reflejado en valores de RMSE superiores a los obtenidos en el caso homogéneo.

Estimación del parámetro $\bar{\alpha}_{i1}$

El parámetro que más dificultades presenta para su correcta estimación es $\bar{\alpha}_{i1}$. Sin embargo, cuando se simula una muestra que posee capacidad de ordenamiento $q=4$, es posible recuperar el valor real. Cuando la muestra posee una menor capacidad de ordenamiento, en general la estimación del parámetro se encuentra subestimada, es decir, el modelo interpreta que la probabilidad de elegir la alternativa con mayor utilidad lexicográfica es menor que la real.

La menor capacidad para estimar $\bar{\alpha}_{i1}$ es una pérdida que no se atribuye al modelo utilizado sino a las características del diseño de la encuesta.

En efecto, cuando se utiliza el diseño ortogonal, es posible recuperar $\bar{\alpha}_{i1}$ independiente de la capacidad de ordenamiento de la muestra. Este hecho es consistente con el camino que se esperaba siguieran los resultados ya que con un diseño ortogonal la decisión nunca será aleatoria pues, como se explicó con anterioridad, el atributo más importante es suficiente para tomar una decisión, razón por la cual, quienes puedan construir un *ranking* compuesto por 1, 2, 3 o 4 atributos estarán en iguales condiciones para elegir la alternativa preferida.

Cuando el diseño posee coincidencias en sus niveles, quienes sólo puedan construir un *ranking* parcial de atributos, en parte de las preguntas tomarán decisiones aleatorias, lo cual reduce la probabilidad de elegir la alternativa con mayor utilidad lexicográfica, situación que se refleja en una estimación de $\bar{\alpha}_{i1}$ inferior a la real.

Verosimilitud de orden y tasa de éxito

En el detalle de resultados que de los Anexos A y B. se mencionan otras dos medidas de desempeño: El logaritmo de la verosimilitud de orden y la tasa de éxito.

El logaritmo de la verosimilitud de orden se calcula directamente utilizando la ecuación (22), expresión que es creciente en q , aspecto que es consistente con los resultados obtenidos ya que, independiente de la condición experimental, las estimaciones homogéneas bajo supuestos de capacidad de ordenamiento superiores arrojan mayores valores de esta medida de desempeño.

La tasa de éxito corresponde al porcentaje medio de iteraciones en que la alternativa que es lexicográficamente preferida por cada individuo de acuerdo a la estimación coincide con la alternativa escogida.

Para calcularla, en cada iteración, en cada pregunta y para cada individuo, se detecta la alternativa con mayor utilidad lexicográfica dado el candidato de orden de atributos (r_i^c). Si coincide con la alternativa escogida en la simulación entonces se tiene un acierto. La tasa de éxito registrada se determina calculando el acierto porcentual medio.

Entendiendo este procedimiento y los supuestos bajo los cuales se construyó el modelo, es posible comprender por qué la tasa de éxito es no decreciente en q , es decir, se obtienen tasas mejores (o a lo sumo iguales) cuando se estima bajo el supuesto de $q=4$ (capacidad completa de ordenamiento de atributos).

En efecto, en el proceso de estimación, la tendencia es proponer cada vez mejores candidatos r_i^c . Cuando en la estimación se asume capacidad completa de ordenamiento ($q=4$), el candidato r_i^c es utilizado íntegramente para escoger la alternativa lexicográficamente preferida, obteniendo así mejores predicciones.

Cuando en la estimación se asume capacidad parcial de ordenamiento ($q=1, 2$ o 3 en este caso), para el cálculo de la tasa de acierto sólo se conservan las q primeras componentes del candidato r_i^c y las demás se ordenan aleatoriamente. Así, en cada nueva iteración sólo se aprovecha de manera parcial el vector de orden de atributos.

Capacidad de predicción de elecciones

En relación a la capacidad de predicción de elecciones que posee el modelo, queda por resolver:

- En el caso del enfoque de estimación homogéneo, ante el comportamiento de la tasa de acierto, surge la interrogante de si, para efectos de predicción de las elecciones de los individuos, conviene simplemente aplicar el modelo bajo supuesto de capacidad completa de ordenamiento ($q=4$) incluso cuando, utilizando el criterio de máxima MLL, se haya estimado que la capacidad de ordenamiento mayoritaria de la muestra es en realidad de 1, 2 o 3 atributos.

- Por otro lado, pese a que las tasas de acierto de predicción del parámetro q del enfoque de estimación heterogéneo revelan que el modelo propuesto es eficiente en la estimación de la capacidad de ordenamiento de atributos, surge la interrogante de si la falta de precisión en la estimación de parte de las componentes de los parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ de los individuos cuya capacidad de ordenamiento estimada es parcial ($\hat{q} < 4$), traerá consecuencias en la predicción de las preferencias del consumidor.

Para dar respuesta a lo anterior se calcula una tasa de acierto que se denominará en lo que sigue “tasa de éxito post estimación” y que difiere de la anterior por cuanto sólo se calcula una vez finalizada la estimación de los parámetros \bar{v}_i y $\bar{\alpha}_{i1}$, procediendo como se indica a continuación:

- Se aplica el modelo y se obtienen las estimaciones individuales \hat{v}_i , $\hat{\alpha}_{i1}$ y \hat{q}_i con $i=1, \dots, 125$ o $i=1, \dots, 250$ según corresponda¹¹.
- Utilizando la ecuación (29), se calcula, para cada individuo y para cada uno de los G ordenamientos (*rankings*) que se pueden construir con 5 atributos, la probabilidad de haber tomado decisiones utilizando el *ranking* de orden de atributos r_g , $P_{ig}(r_g | \hat{v}_i, \hat{q}_i)$, con $g=1, \dots, 120$.
- Utilizando la ecuación (2), se determina la alternativa que es lexicográficamente preferida (aquella con mayor utilidad lexicográfica), para cada individuo, en cada pregunta y bajo el supuesto de toma de decisiones utilizando cada posible *ranking* de atributos. Así, se define:

$$z_{ikt}^g = \begin{cases} 1; & \text{si la alternativa } k \text{ de la pregunta } t \text{ es lexicográficamente preferida} \\ & \text{por el individuo } i \text{ cuando utiliza el } \textit{ranking} \text{ de orden de atributos } g, \\ & \text{con } k=1, \dots, 4 \text{ y } t=1, \dots, 20 \text{ o } t=1, \dots, 40 \text{ según corresponda.} \\ 0; & \text{en otro caso} \end{cases}$$

- Se estima la probabilidad de que el individuo i elija la alternativa k en la pregunta t como:

$$p_{ikt} = \sum_{g=1}^{120} (\hat{a}_{i1})^{z_{ikt}^g} (1 - \hat{a}_{i1})^{1 - z_{ikt}^g} [P_{ig}(r_g | \hat{v}_i, \hat{q}_i)] \quad (31)$$

Donde,

$$\hat{a}_{i1} = \frac{1}{1 + e^{-\hat{\alpha}_{i1}}} \quad (32)$$

¹¹ Si se trata del enfoque homogéneo, se utilizan \hat{v}_i y $\hat{\alpha}_{i1}$ provenientes de la estimación que arrojó el mayor valor de MLL y $\hat{q}_i = \hat{q}_j = \hat{q} \forall i, j$, donde \hat{q} es la capacidad de ordenamiento de atributos utilizada como supuesto en dicha estimación y que, de acuerdo a lo visto anteriormente coincide con la capacidad de ordenamiento de atributos real de la muestra.

- v. Así, para predecir qué alternativa escoge el individuo i en la pregunta t , lo más adecuado es apostar por la alternativa que, posea la mayor probabilidad de elección, es decir, se escoge la alternativa k' , donde $p_{ik't} = \max_k \{p_{ikt}\}$.
- vi. Finalmente, se compara la predicción de elección con la alternativa que fue escogida en la simulación y se contabiliza el porcentaje de aciertos (considerando todas las preguntas y todos los individuos).

La siguiente Tabla comparativa, muestra las tasas de acierto post estimación¹² cuando la estimación es homogénea y se realiza bajo el supuesto correcto de capacidad de ordenamiento, es decir, asumiendo que todos los individuos poseen la capacidad de ordenamiento que efectivamente tiene la muestra simulada, versus las tasas obtenidas cuando se utiliza el supuesto de capacidad completa de ordenamiento de atributos (todos los individuos poseen $q=4$):

Cantidad de preguntas de la encuesta	Capacidad de ordenamiento de la muestra simulada	Capacidad de ordenamiento en estimación	Tamaño de la muestra simulada	
			125	250
20	q=1	q=1 q=4	76,60% 71,60%	70,13% 69,23%
	q=2	q=2 q=4	69,75% 64,90%	68,73% 65,30%
	q=3	q=3 q=4	69,45% 63,65%	72,43% 63,10%
	q=4	q=4	64,00%	63,95%
40	q=1	q=1 q=4	79,98% 79,00%	79,00% 78,95%
	q=2	q=2 q=4	75,50% 74,10%	73,95% 72,48%
	q=3	q=3 q=4	75,78% 69,50%	76,69% 72,06%
	q=4	q=4	64,98%	64,69%

Tabla 3.17. Tasa de éxito *in sample* post estimación bajo el enfoque de estimación homogéneo

Cuando se trata del enfoque de estimación heterogéneo *versus* la estimación bajo el supuesto $q=4$ se obtiene:

¹² Tasas de acierto *in sample*. Ver en Anexo C las tasas de acierto *out of sample*.

Cantidad de preguntas de la encuesta	Capacidad de ordenamiento de la muestra simulada	Capacidad de ordenamiento en estimación	Tamaño de la muestra simulada	
			125	250
20	Heterogénea	Heterogénea q=4	68,70% 59,20%	67,28% 62,78%
40	Heterogénea	Heterogénea q=4	73,00% 66,00%	71,00% 65,75%

Tabla 3.18. Tasa de éxito *in sample* post estimación bajo el enfoque de estimación heterogéneo

De las Tablas 3.17. y 3.18. se desprende que el modelo propuesto en este trabajo es eficiente en términos de la predicción de elecciones de los individuos.

En efecto, cuando se elimina el supuesto de capacidad completa de ordenamiento de atributos ($q=4$ para todos los individuos), se obtiene mejores tasas de acierto en la predicción de elecciones. Este hecho reafirma el beneficio de estimar correctamente la capacidad de ordenamiento de atributos utilizando alguno de los enfoques del modelo de ordenamiento parcial en vez de asumir que los individuos poseen capacidad completa de ordenamiento de atributos.

Finalmente, es posible observar que se obtienen mejores tasas de acierto a mayor cantidad de preguntas pero aumentando el tamaño de la muestra no necesariamente se consiguen mejores predicciones lo cual es consistente con los demás resultados expuestos en éste capítulo.

Relación entre estimación de q_i y la tasa de éxito post estimación en enfoque heterogéneo

Tal y como se mencionó con anterioridad, bajo el enfoque de estimación heterogéneo se obtiene estimaciones de q_i con el valor modal de \hat{q}_i a lo largo de todas las iteraciones.

Es decir, si de un total de 100 iteraciones, en sesenta de ellas $\hat{q}_i=1$, en veinte $\hat{q}_i=2$, en diez $\hat{q}_i=3$ y en diez $\hat{q}_i=4$, entonces el método arrojará $\hat{q}_i=1$ y se declarará que con un 60% de probabilidad, el individuo i utiliza un atributo para tomar su decisión de compra.

Así, además de la estimación \hat{q}_i , es posible obtener distribuciones de probabilidad en la estimación de q a nivel individual.

El siguiente gráfico representa la situación de los primeros quince individuos para el caso de encuesta con veinte preguntas y muestra poblacional de 125 individuos:

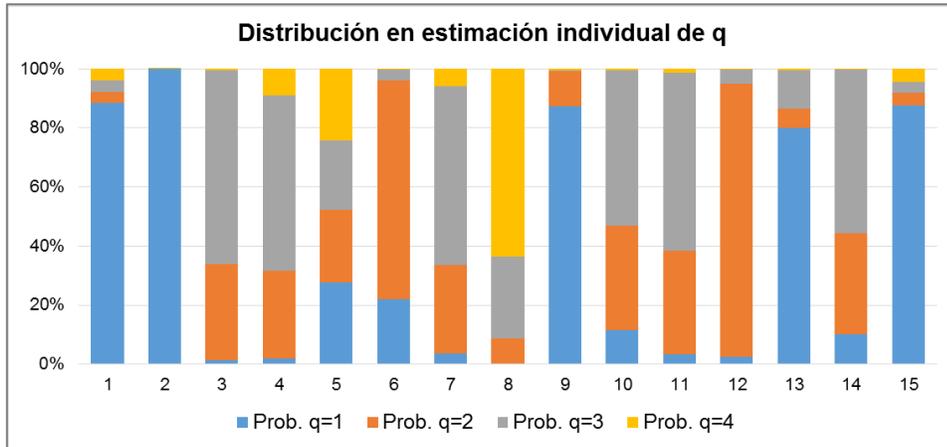


Gráfico 3.1. Distribución en estimación individual de q en simulación heterogénea (20 preguntas, 125 individuos) y estimación heterogénea

El gráfico anterior revela que existe individuos con una marcada prevalencia de algún valor de q particular, como son los casos de los individuos 1, 2 (con $\hat{q}_i=1$), 6 (con $\hat{q}_i=2$), etc. Pero, para otros individuos no es tan claro cuál es el valor de q predominante. Casos como ese se observan en el individuo 5, donde cada valor de q tiene una frecuencia similar, o como el individuo 14 donde, pese a que muestra prevalencia por $\hat{q}_i=3$, también posee una cantidad importante de iteraciones donde $\hat{q}_i=2$.

Luego, si se agrupan los individuos de cada muestra de acuerdo al valor de q estimado se obtiene:

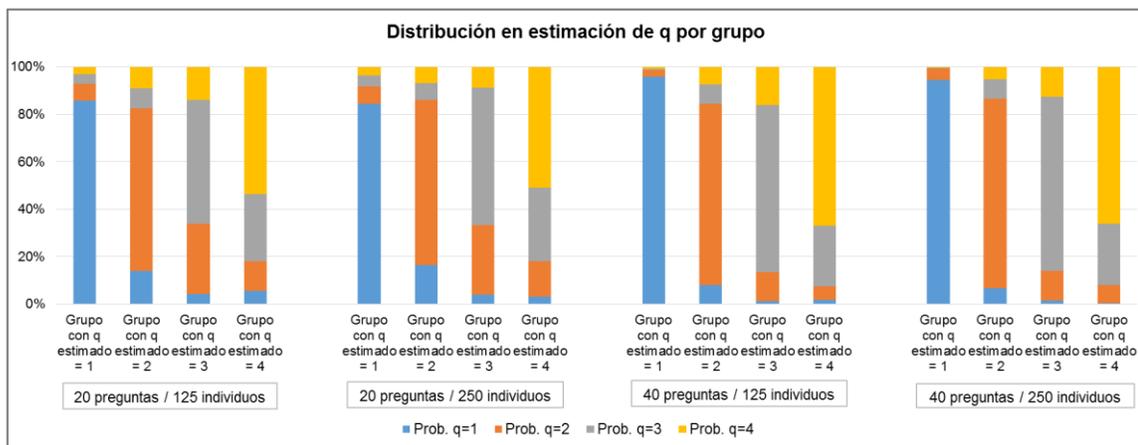


Gráfico 3.2. Distribución en estimación de q en simulación heterogénea y estimación heterogénea

Para resolver si los comportamientos de las distribuciones de probabilidad tienen relación con la tasa de acierto en la estimación de q, se define:

P_{iq} : Probabilidad estimada de que el individuo i posea $\hat{q}_i=q$, $q = 1, 2, 3, 4$.

PQ_i : Diferencia entre las dos más altas probabilidades P_{iq} , $q = 1, 2, 3, 4$.

Valores mayores de PQ_i implican mayor prevalencia por un \hat{q}_i particular.

En el siguiente gráfico se presenta la tasa de acierto en la estimación de q para diferentes grupos de individuos. Así, por ejemplo, en el eje vertical aparece la tasa de acierto en la estimación de q cuando se considera la muestra completa (125 o 250 individuos según corresponda) y, sobre la línea punteada, la tasa de acierto cuando solo se considera el grupo de individuos que posee $PQ_i \geq 0,5$.

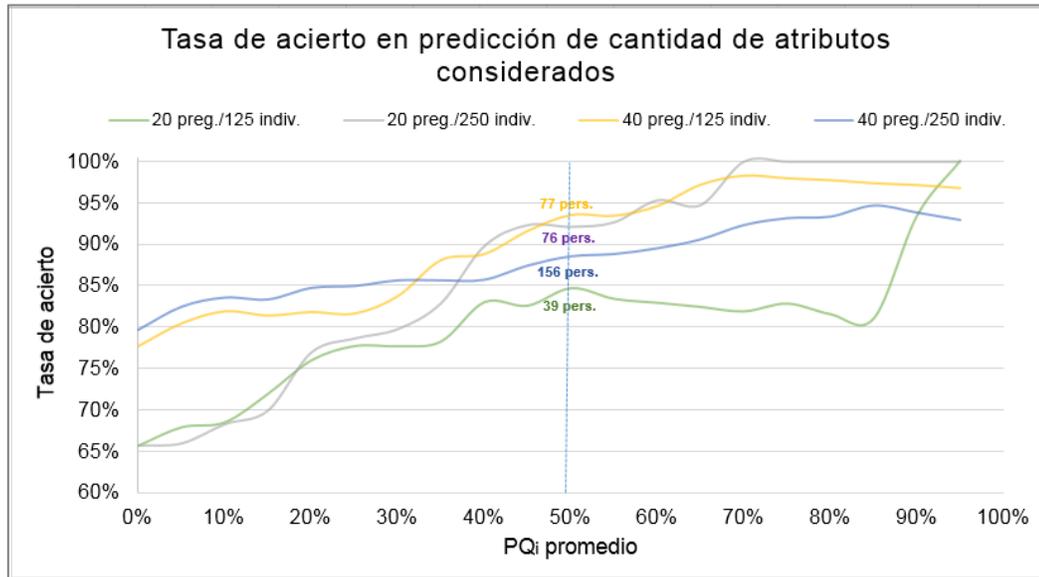


Gráfico 3.3. Tasa de acierto en predicción de q en simulación heterogénea y estimación heterogénea

Se confirma así la idea de que a mayor prevalencia de un valor de q particular a lo largo de las iteraciones, mayor es la probabilidad de acertar en la estimación de q_i .

Finalmente, si la distribución de \hat{q}_i da nociones de la precisión en la estimación de q_i , ¿es también un indicador de la tasa de éxito post estimación (porcentaje de acierto en la predicción de elecciones)?

Se realiza entonces un ejercicio similar pero esta vez considerando la tasa de éxito post estimación promedio que posee cada grupo considerado.

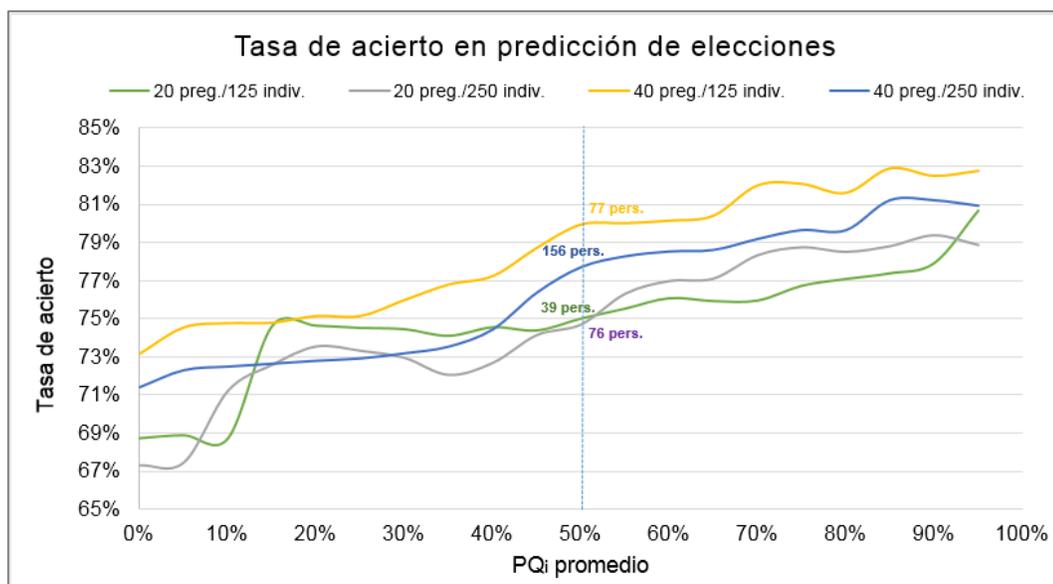


Gráfico 3.4. Tasa de acierto post estimación *in sample* en simulación heterogénea y estimación heterogénea

En los dos gráficos anteriores se puede apreciar tendencias crecientes, es decir, en general, aumentan las tasas de acierto en la medida que sólo se consideran segmentos de individuos donde su estimación \hat{q}_i aparece repetido mayor cantidad de veces a lo largo de todas las iteraciones.

Por otra parte, en ambos casos las tasas de acierto son mejores cuando la cantidad de preguntas por individuo es superior, lo cual es consistente con lo esperado. Más aún, cuando se utilizan 40 preguntas, la cantidad de individuos que conforman el segmento considerado en el corte 50% (77 y 156 personas cuando las muestras originales son de 125 y 250 individuos respectivamente) es, en ambos casos, cercano al 62% de la muestra total, que es muy superior a los otros casos donde sólo el 31% de la muestra total está contenido en corte 50%.

CAPÍTULO IV: APLICACIÓN EMPÍRICA

En este capítulo, se utilizará el modelo de elección lexicográfico de ordenamiento parcial de atributos para analizar las preferencias de elección reales respecto al consumo de *tablets* de un grupo compuesto por 125 estudiantes universitarios de entre 18 y 26 años.

De manera individual y supervisada, a cada uno de ellos se les solicitó contestar una encuesta *online* y cronometrada¹³, compuesta por dos partes principales.

En la primera parte se presentan 20 preguntas de elección construidas de acuerdo al diseño con coincidencia en niveles. En cada una de ellas, tal y como fue expuesto en el capítulo de simulación, el encuestado debió escoger una de cuatro alternativas. Cada alternativa se expuso descrita por cinco atributos, cada uno de ellos compuesto por los dos niveles que se detallan en la tabla que se presenta a continuación:

	Atributo 1: Sistema Operativo	Atributo 2: Capacidad	Atributo 3: Conectividad	Atributo 4: Tamaño de Pantalla	Atributo 5: Precio
Nivel 1	Android	16 GB	3G	7"	\$ 400.000
Nivel 2	IOS (Apple)	32 GB	4G	9"	\$ 250.000

Tabla 4.1. Atributos y niveles en aplicación empírica *tablets*

Así, los encuestados se vieron enfrentados a 20 preguntas sucesivas como la siguiente:

1. Si estas fueran sus únicas opciones de compra, descritas por su sistema operativo, capacidad, conectividad, tamaño de pantalla y precio... ¿cuál escogería?

Por favor haga click sobre su elección:

Android 16 GB 4G 7" \$ 400 mil <input type="radio"/>	IOS (Apple) 32 GB 4G 7" \$ 250 mil <input type="radio"/>	Android 32 GB 3G 9" \$ 400 mil <input type="radio"/>	Android 16 GB 3G 9" \$ 400 mil <input type="radio"/>
--	--	--	--

Figura 4.1. Pregunta 1 encuesta aplicación empírica *tablets*

¹³ Los encuestados no tuvieron conocimiento de que se midió su tiempo de respuesta.

Con el objeto de no influir en el nivel de información que poseen los encuestados respecto al producto y, por lo tanto, en su capacidad de ordenar o establecer un *ranking* de atributos, cada encuestado responde sin previa indicación o descripción del producto y sus características.¹⁴

En la parte final de la encuesta se incorporaron preguntas demográficas y respecto al nivel de información que se tiene del producto objeto la encuesta. De esta manera fue posible obtener mayor información respecto a la muestra encuestada. En particular, el 80% pertenece al género masculino y un 31% ha comprado alguna vez un *Tablet*.

El siguiente gráfico muestra el porcentaje de individuos que revela tener conocimiento práctico o conceptual de 1, 2, 3, 4 o la totalidad de atributos con que aparecen descritas las opciones o alternativas.

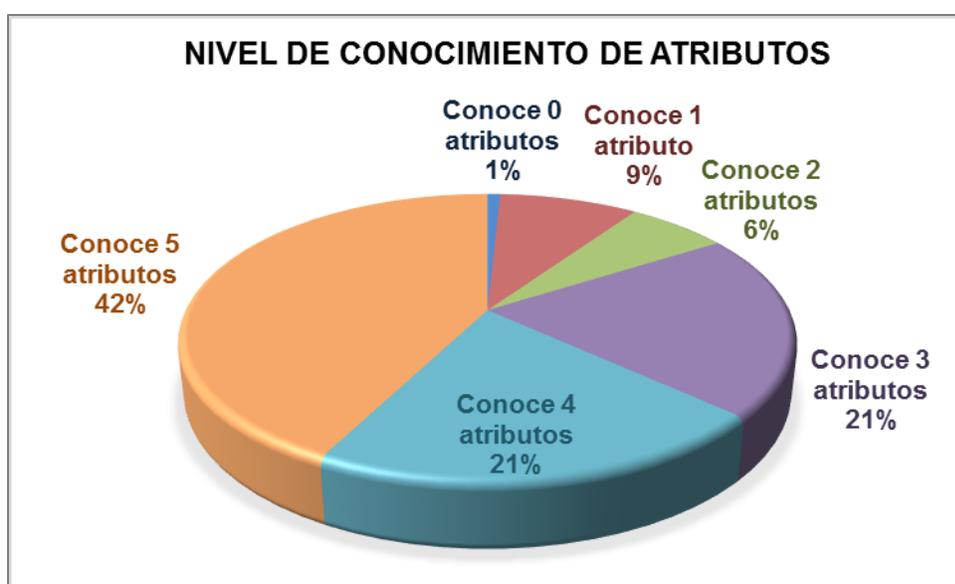


Gráfico 4.1. Nivel de conocimiento de atributos aplicación empírica *tablets*

A continuación, se muestran los resultados obtenidos luego de aplicar el Modelo de Elección Lexicográfico de Ordenamiento Parcial de Atributos para luego finalizar éste capítulo con un análisis que busca dar respuesta a interrogantes como: ¿Qué tan heterogéneos son los consumidores en su capacidad para establecer *rankings* de orden de atributos?, ¿existe algún patrón común en la importancia asignada a los atributos?

4.1. Resultados con enfoque de estimación homogéneo

Siguiendo la metodología explicada en capítulos anteriores, dado que las alternativas aparecen descritas por cinco atributos, se estima el modelo cuatro veces, la

¹⁴ Al dar información se provee la posibilidad de que todos establezcan un *ranking* completo de atributos y, por lo tanto, pierde sentido parte importante del estudio.

primera bajo el supuesto de capacidad homogénea de ordenamiento de un atributo, es decir, todos los individuos pueden establecer un *ranking* de sólo un atributo (sólo reconocen el más importante), la segunda bajo el supuesto de $q=2$, la tercera con $q=3$ y finalmente bajo el supuesto $q=4$.

La siguiente Tabla muestra el logaritmo de la verosimilitud de datos marginal (MLL) que se obtuvo en cada caso, considerando 100 mil iteraciones de las cuales se descartan las primeras 50 mil por razones de convergencia:

	Capacidad de ordenamiento de la población en estimación			
	q = 1	q = 2	q = 3	q = 4
MLL	<u>-1481,80</u>	-1509,60	-1543,70	-1561,50

Tabla 4.2. MLL en estimación homogénea de aplicación empírica

Así, $\hat{q}=1$ es, bajo el enfoque de estimación homogéneo, la estimación de la capacidad de ordenamiento de atributos que posee la porción mayoritaria de individuos encuestados.

Luego, considerando el modelo que asume capacidad homogénea de ordenamiento con $q=1$, los restantes parámetros estimados son:

Estimación \hat{v}_i	Intervalo de confianza al 95%
$\hat{v}_{i1} = -1,98$	[-2,80 , -1,24]
$\hat{v}_{i2} = -2,78$	[-4,10 , -1,85]
$\hat{v}_{i3} = -2,87$	[-4,31 , -1,88]
$\hat{v}_{i4} = -5,14$	[-8,64 , -3,35]

Tabla 4.3. Estimación de parámetro \hat{v}_i e intervalo de confianza en estimación homogénea de aplicación empírica

Estimación $\hat{\alpha}_{i1}$	Intervalo de confianza al 95%
$\hat{\alpha}_{i1} = 1,39$	[1,22 , 1,56]

Tabla 4.4. Estimación de parámetro $\hat{\alpha}_{i1}$ e intervalo de confianza en estimación homogénea de aplicación empírica

Otras medidas de desempeño obtenidas bajo el enfoque de estimación homogéneo, como el logaritmo de la verosimilitud de orden (calculada utilizando la ecuación (22)) y la tasa de éxito, son las siguientes:

Medida de desempeño	Resultados obtenidos en estimación con capacidad homogénea de ordenamiento
Log-verosimilitud de orden	-502,96
Tasa de éxito post estimación <i>in sample</i> (%)	55,75
Tasa de éxito post estimación <i>out of sample</i> (%)	58,80

Tabla 4.5. Medidas de desempeño en estimación homogénea de aplicación empírica

4.2. Resultados con enfoque de estimación heterogéneo

La siguiente Tabla muestra el porcentaje de individuos que se estima posean cada capacidad de ordenamiento q .

Capacidad de ordenamiento	$q = 1$	$q = 2$	$q = 3$	$q = 4$
Porcentaje de individuos (%)	31,20	24,80	13,60	30,40

Tabla 4.6. Porcentaje de individuos con capacidad de ordenamiento de 1, 2, 3 y 4 atributos en estimación heterogénea de aplicación empírica

Los parámetros poblacionales estimados (sin segmentación de acuerdo a la capacidad de ordenamiento estimada) y medidas de desempeño son:

Estimación \widehat{v}_i	Intervalo de confianza al 95%
$\widehat{v}_{i1} = -3,93$	[-5,86 , -2,55]
$\widehat{v}_{i2} = -2,06$	[-2,74 , -1,45]
$\widehat{v}_{i3} = -2,33$	[-3,13 , -1,65]
$\widehat{v}_{i4} = -4,00$	[-5,38 , -2,85]

Tabla 4.7. Estimación de parámetro \widehat{v}_i e intervalo de confianza en estimación heterogénea de aplicación empírica

Estimación $\widehat{\alpha}_{i1}$	Intervalo de confianza al 95%
$\widehat{\alpha}_{i1} = 1,70$	[1,43 , 1,96]

Tabla 4.8. Estimación de parámetro $\widehat{\alpha}_{i1}$ e intervalo de confianza en estimación heterogénea de aplicación empírica

Medida de desempeño	Resultados obtenidos en estimación con capacidad homogénea de ordenamiento
MLL	-1419,70
Log-verosimilitud de orden	-398,38
Tasa de éxito post estimación <i>in sample</i> (%)	65,45
Tasa de éxito post estimación <i>out of sample</i> (%)	65,60

Tabla 4.9. Medidas de desempeño en estimación heterogénea de aplicación empírica

4.3. Análisis de resultados aplicación empírica

Los resultados expuestos en la Tabla 4.2. revelan que, cuando se utiliza el enfoque homogéneo de estimación, la capacidad de ordenamiento que posee la mayoría de la muestra es de un atributo (por ser el supuesto bajo el cual se obtiene el máximo valor de MLL), es decir, reconocen el atributo más importante y, en caso de no ser suficiente para tomar una decisión, eligen aleatoriamente entre las alternativas que poseen el nivel más favorable de dicho atributo.

Observando las estimaciones del vector \bar{v}_i y recordando que la última componente (correspondiente al precio), tiene el valor $\widehat{v}_{i5}=0$, se concluye que el atributo más importante para este grupo de individuos es el precio.

Estos resultados son consistentes con los obtenidos utilizando el enfoque de heterogéneo donde, de acuerdo a la estimación del parámetro q , la mayoría de la muestra construye un *ranking* compuesto únicamente por un atributo.

Este hecho es relevante por verificar que, en términos de la estimación de la capacidad de ordenamiento de atributos, se obtiene el mismo resultado ya sea por la vía del enfoque homogéneo o del enfoque heterogéneo.

Si bien la cantidad de individuos con $\hat{q}=1$ es muy similar a la cantidad de individuos con $\hat{q}=4$, de acuerdo a lo visto en el capítulo anterior, es posible prever que la cantidad de individuos cuya capacidad de ordenamiento estimada es $\hat{q}=1$ está subestimada, versus la cantidad de individuos cuya capacidad de ordenamiento estimada es $\hat{q}=4$, que se encuentra sobre estimada.

De forma análoga al análisis realizado en el capítulo de simulación, si bien la determinación del valor \hat{q} para cada individuo se realiza con el valor modal obtenido en las iteraciones, es posible rescatar una distribución individual de \hat{q}_i . Así, si se agrupa la muestra poblacional de acuerdo a la capacidad de ordenamiento de atributos estimada, se obtiene:

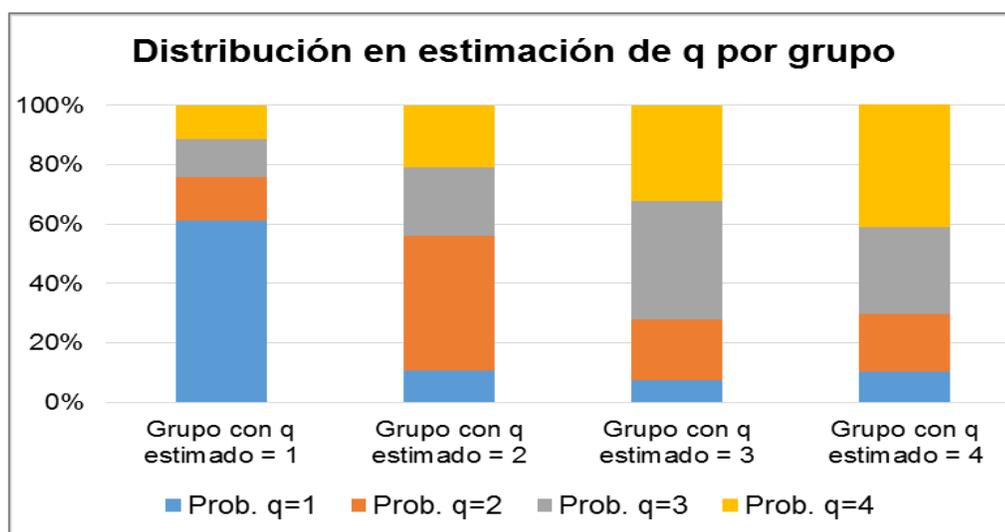


Gráfico 4.2. Distribución en estimación de q en estimación heterogénea de aplicación empírica

El gráfico anterior ratifica la idea de que existe una mayor probabilidad de errar al declarar que un individuo posee $\hat{q}=4$, que cuando se declara $\hat{q}=1$.

Por otro lado, de acuerdo a los valores estimados de $\bar{\alpha}_{i1}$ y utilizando la ecuación (17), es posible prever que los individuos escojan la alternativa que es lexicográficamente preferida con una probabilidad de 80,06% o 84,55% dependiendo de si la estimación se realiza bajo el enfoque homogéneo o heterogéneo respectivamente aunque, de acuerdo a lo analizado en el capítulo anterior, es posible que dichas cantidades estén levemente subestimadas.

Los valores de MLL y tasa de éxito post estimación registrados en las Tablas 4.2., 4.5. y 4.9. permiten definir que la estimación realizada bajo el enfoque heterogéneo es superior a aquella realizada bajo el enfoque homogéneo.

Al comparar el modelo heterogéneo con un modelo lineal (ecuación (1)), se espera que la flexibilidad de este último le provea la facultad de obtener mejores tasas de acierto en la predicción de preferencias. Esto porque el modelo lineal puede igualmente representar preferencias lexicográficas al otorgar un valor suficientemente alto a alguno de sus coeficientes (part worths). Una comparación más justa puede darse con modelos q-compensatorios¹⁵, introducidos por Yee *et al.* (2007), donde se restringe los valores que pueden tomar cualquier par de parámetros de tal forma que su cociente sea menor o igual a un valor fijo q.

¹⁵ Aquí, q no tiene relación con la capacidad de orden de atributos denotada igualmente por q en este trabajo.

Modelo	<i>In sample</i>	<i>Out of sample</i>
Lineal	78,28%	72,38%
q-compensatorio con q=4	62,61%	61,99%
q-compensatorio con q=3	59,39%	60,17%
q-compensatorio con q=2	55,24%	56,33%
q-compensatorio con q=1	47,72%	49,53%
Lexicográfico de ordenamiento parcial	65,45%	65,60%

Tabla 4.10. Tasas de acierto en predicción de elecciones de modelos lineal, q-compensatorio y lexicográfico de ordenamiento parcial de atributos

Luego, se confirma que la tasa de éxito del modelo lineal es superior, sin embargo, el modelo lexicográfico es superior a cualquier modelo q-compensatorio.

Separando las tasas de acierto del modelo lexicográfico de ordenamiento parcial de acuerdo a la capacidad de ordenamiento estimada que poseen los individuos se obtiene:

Capacidad de ordenamiento	<i>In sample</i>	<i>Out of sample</i>
1 atributo	61,06%	60,90%
2 atributos	71,77%	70,97%
3 atributos	72,43%	64,71%
4 atributos	61,68%	66,45%

Tabla 4.11. Tasas de acierto en predicción de elecciones de modelo lexicográfico de ordenamiento parcial de atributos

Finalmente, utilizando los resultados para caracterizar diferentes segmentos de la muestra de acuerdo a la capacidad individual de ordenamiento de atributos bajo el enfoque heterogéneo, es posible construir la siguiente Figura:

Algunas de las conclusiones que se desprenden de la figura anterior son las siguientes:

- Un 86,4%¹⁶ de los encuestados sitúa al precio en el primer lugar del *ranking* de orden de atributos. Para el 13,6% restante, el sistema operativo ocupa la primera posición y, cuando ellos pueden construir un *ranking* con dos o más atributos, el precio ocupa la siguiente posición del *ranking*.
- El tamaño de la pantalla no forma parte del *ranking* de quienes pueden ordenar menos de cuatro atributos. Sólo ocupa la cuarta o quinta posición del *ranking* de quienes tienen capacidad completa de orden de atributos.
- La capacidad y conectividad ocupan mayoritariamente posiciones intermedias de los *rankings* de quienes pueden ordenar 2, 3 o 4 atributos.

¹⁶ $0,31 \times 0,85 + 0,25 \times 0,77 + 0,14 \times 0,88 + 0,3 \times 0,95 = 86,4\%$

CONCLUSIONES

Se da cumplimiento al objetivo principal de este trabajo, por cuanto el modelo propuesto es capaz de reflejar decisiones de los consumidores eliminando el supuesto de capacidad completa de ordenamiento de atributos.

Así mismo, se ratifica que este modelo consigue avanzar en la estimación de la capacidad de ordenamiento de atributos (parámetro q), lo cual no debe ser confundido con la cantidad de atributos de los cuales los individuos tienen algún nivel de conocimiento ya que, por ejemplo, una persona puede tener conocimiento respecto de todos los atributos y, sin embargo, sólo ser capaz de construir un *ranking* de preferencias con parte de ellos.

Para tener éxito en la estimación de la cantidad de atributos que los individuos consideran para tomar sus decisiones, es necesario utilizar una encuesta que promueva que se revele en parte de las elecciones el *ranking* individual de orden de atributos.

Al mismo tiempo, el modelo es efectivo a la hora de estimar los parámetros relacionados con la importancia que se le asigna a los atributos pero se pierde capacidad de estimar con precisión el parámetro a .

Cuando se utiliza el enfoque de estimación homogéneo, el criterio de máxima *Marginal Log-likelihood* es efectivo como herramienta para determinar la capacidad de ordenamiento de atributos que posee la parte mayoritaria de la muestra y, con ello, los demás parámetros asociados, pero, si se quiere obtener la capacidad de ordenamiento individual, se debe utilizar el enfoque de estimación heterogéneo.

Bajo el enfoque heterogéneo, es posible caracterizar de mejor manera a los individuos que son objeto de estudio, permitiendo construir diferentes segmentos de acuerdo a la estimación individual de la capacidad de ordenamiento de atributos estimada.

Finalmente, se concluye que con el modelo propuesto es posible obtener más información respecto a cómo los consumidores toman decisiones, transformándose en un modelo que logra explicar de mejor forma los procesos de elección sin perder con ello capacidad predictiva.

BIBLIOGRAFÍA

- [1] BEGGS, S., CARDELL, S. y HAUSMAN, J. 1981. Assessing the potential demand for electric cars. *Journal of Econometrics*, 17 (1): 1 – 19.
- [2] BETTMAN, J., JOHNSON, E. y PAINE, J. 1988. Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 14: 534 – 552.
- [3] BETTMAN, J., JOHNSON, E. y PAINE, J. 1993. *The adaptive decision maker*. Cambridge University Press, Cambridge, England.
- [4] BETTMAN, J., LUCE, M. y PAINE, J. 1998. Constructive choice processes. *Journal of Consumer Research*, 25 (3): 187 – 217.
- [5] BRIGHTON, H. y GIGERENZER, G. 2008. Bayesian brains and cognitive mechanisms: Harmony or dissonance?. *The Probabilistic Mind: Prospects for Bayesian Cognitive Science*, 9: 189 – 208.
- [6] BRÖDER, A. 2000. Assessing the empirical validity of the ‘take the best’ heuristic as a model of human probabilistic inference. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 26 (5): 1332 – 1346.
- [7] BRÖDER, A. y GAISSMAIER, W. 2007. Sequential processing of cues in memory-based multi-attribute decisions. *Psychonomic Bulletin and Review*, 14: 895 – 900.
- [8] BRÖDER, A. y SCHIFFER, S. 2003. Take the best versus simultaneous feature matching: probabilistic inferences from memory and effects of representation format. *Journal of Experimental Psychology: General*, 132 (2): 277 – 293.
- [9] DAHAN, E., HAUSER, J. ORLIN, J. y YEE, M. 2007. Greedoid-based non compensatory consideration-then-choice inference. *Marketing Science*, 26 (4): 532 – 549.
- [10] DIECKMANN, A., DIETRICH, H. y DIPPOLD, K. 2009. Compensatory versus non compensatory models for predicting consumer preferences. *Judgment and Decision Making*, 4 (3): 200 – 213.
- [11] DROLET, A., y LUCE, M. 2004. The rationalizing effects of cognitive load on emotion-based trade off avoidance. *Statistical Science*, 7 (4): 457 – 472.
- [12] GIGERENZER, G. y GOLDSTEIN, D. 1996. Reasoning the fast and frugal way: Model of bounded rationality. *Psychological Review*, 103 (4): 650 – 669.
- [13] GREEN, P. y RAO, V. 1971. Conjoint measurement for quantifying judgemental data. *Journal of Marketing Research*, 8: 355 – 363.

- [14] HAUSER, J. y URBAN, G. 2004. Listening-in to find and explore new combinations of customer needs. *Journal of Marketing Research*, 68: 72 – 87.
- [15] HOFFRAGE, U. y MARTIGNON, L. 2002. Fast, frugal and fit: simple heuristics for paired comparison. *Theory and Decision*, 52 (1): 29 – 71.
- [16] JEDIDI, K. y KOHLI, R. 2007. Representation and inference of lexicographic preference models and their variants. *Marketing Science*, 26 (3): 380 – 399.
- [17] JEDIDI, K. y KOHLI, R. 2008. Inferring latent class lexicographic rules from choice data. *Journal of Mathematical Psychology*, 52: 241 – 249.
- [18] JEDIDI, K., KOHLI, R y MONTOYA, R. 2011. Probabilistic lexicographic models, Working Paper, Columbia University.
- [19] SLOVIC, P. 1975. Choice between equally valued alternatives. *Journal of Experimental Psychology: Human Perception Performance*, 1: 280 – 287.
- [20] TVERSKY, A. 1972. Elimination by aspects: a theory of choice. *Psychological Review*, 79 (4): 281 – 299.

ANEXOS¹⁷

ANEXO A: Resultado simulación con Diseño 1

1. Simulación homogénea (20 preguntas, 125 individuos) y enfoque de estimación homogéneo

Simulación con capacidad de ordenamiento de 1 atributo (q=1)

Media poblacional \bar{v}_i	Estimación \widehat{v}_i en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
7,85	6,06 *	6,14	6,04	6,31 *
0,16	-0,37 *	-0,35 *	-0,39 *	-0,95 *
0,08	-0,80 *	-0,76 *	-0,61 *	0,09 *
0,20	-6,31	-6,41	-3,64	-3,44
Media poblacional $\bar{\alpha}_{i1}$	Estimación $\widehat{\alpha}_{i1}$ en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
3,46	3,67 *	3,67 *	3,66 *	3,67 *

Tabla A.1. Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 1 con simulación homogénea (20 preguntas, 125 individuos) con q=1 y estimación homogénea

Medida de desempeño	Resultados obtenidos en estimación con capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
MLL	-375,90	-375,91	-377,14	-374,25
Log-verosimilitud de orden	-412,23	-334,11	-292,93	-239,50
Tasa de éxito (%)	96,50	96,50	96,50	96,50
RMSE	0,20	0,21	0,15	0,14

Tabla A.2. Medidas de desempeño en diseño 1 con simulación homogénea (20 preguntas, 125 individuos) con q=1 y estimación homogénea

¹⁷ (*) Indica que el parámetro pertenece al intervalo de confianza al 95%.

Simulación con capacidad de ordenamiento de 2 atributos (q=2)

Media poblacional \bar{v}_i	Estimación \widehat{v}_i en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
7,85	8,51 *	9,09 *	11,60	8,86 *
6,16	7,02 *	7,57 *	10,11	7,40 *
0,08	-8,67	-3,07 *	-6,17	-10,93
0,20	-1,07 *	-5,62 *	-1,38 *	-5,22
Media poblacional $\bar{\alpha}_{i1}$	Estimación $\widehat{\alpha}_{i1}$ en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
3,46	3,36 *	3,36 *	3,35 *	3,35 *

Tabla A.3. Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 1 con simulación homogénea (20 preguntas, 125 individuos) con q=2 y estimación homogénea

Medida de desempeño	Resultados obtenidos en estimación con capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
MLL	-464,50	-461,54	-459,93	-460,96
Log-verosimilitud de orden	-449,84	-282,79	-183,98	-78,34
Tasa de éxito (%)	95,50	95,50	95,50	95,50
RMSE	0,13	0,09	0,10	0,10

Tabla A.4. Medidas de desempeño en diseño 1 con simulación homogénea (20 preguntas, 125 individuos) con q=2 y estimación homogénea

Simulación con capacidad de ordenamiento de 3 atributos (q=3)

Media poblacional \bar{v}_i	Estimación \widehat{v}_i en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
7,85	9,37 *	6,51 *	7,44 *	9,03 *
6,16	7,55 *	4,63 *	5,59 *	7,22 *
4,08	5,89 *	2,99 *	3,92 *	5,48 *
0,20	3,38 *	0,28 *	1,70 *	2,67
Media poblacional $\bar{\alpha}_{i1}$	Estimación $\widehat{\alpha}_{i1}$ en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
3,46	3,69 *	3,70 *	3,70 *	3,69 *

Tabla A.5. Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 1 con simulación homogénea (20 preguntas, 125 individuos) con q=3 y estimación homogénea

Medida de desempeño	Resultados obtenidos en estimación con capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
MLL	-381,29	-381,87	-376,79	-381,36
Log-verosimilitud de orden	-468,94	-363,14	-278,28	-206,13
Tasa de éxito (%)	96,40	96,40	96,40	96,40
RMSE	0,10	0,10	0,11	0,10

Tabla A.6. Medidas de desempeño en diseño 1 con simulación homogénea (20 preguntas, 125 individuos) con q=3 y estimación homogénea

Simulación con capacidad de ordenamiento de 4 atributos (q=4)

Media poblacional \bar{v}_i	Estimación \widehat{v}_i en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
7,85	7,05 *	6,43 *	7,52 *	7,84 *
6,16	5,68 *	5,03 *	6,15 *	6,45 *
4,08	3,05 *	2,65 *	3,59 *	4,00 *
2,20	-3,23	-3,71	-1,55 *	-2,38
Media poblacional $\bar{\alpha}_{i1}$	Estimación $\widehat{\alpha}_{i1}$ en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
3,46	3,49 *	3,49 *	3,49 *	3,49 *

Tabla A.7. Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 1 con simulación homogénea (20 preguntas, 125 individuos) con q=4 y estimación homogénea

Medida de desempeño	Resultados obtenidos en estimación con capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
MLL	-422,64	-423,84	-421,98	-422,40
Log-verosimilitud de orden	-464,95	-339,61	-239,46	-176,09
Tasa de éxito (%)	95,95	95,95	95,95	95,95
RMSE	0,16	0,16	0,15	0,16

Tabla A.8. Medidas de desempeño en diseño 1 con simulación homogénea (20 preguntas, 125 individuos) con q=4 y estimación homogénea

2. Simulación heterogénea (20 preguntas, 125 individuos) y enfoque de estimación heterogéneo

Media \bar{v}_i subgrupo con $q=1$	Estim. \hat{v}_i subgrupo con $\hat{q}=1$	Media \bar{v}_i subgrupo con $q=2$	Estim. \hat{v}_i subgrupo con $\hat{q}=2$	Media \bar{v}_i subgrupo con $q=3$	Estim. \hat{v}_i subgrupo con $\hat{q}=3$	Media \bar{v}_i subgrupo con $q=4$	Estim. \hat{v}_i subgrupo con $\hat{q}=4$
7,70	6,78 *	7,80	6,77 *	8,00	6,76 *	7,86	6,81 *
0,14	4,02 *	6,35	3,99 *	6,15	4,05 *	6,00	3,90 *
0,10	1,06 *	-0,05	1,10 *	4,29	1,06 *	3,98	1,11 *
0,52	-5,58	-0,18	-5,58	0,25	-5,58	2,20	-5,57
Media $\bar{\alpha}_{i1}$ subgrupo con $q=1$	Estim. $\hat{\alpha}_{i1}$ subgrupo con $\hat{q}=1$	Media $\bar{\alpha}_{i1}$ subgrupo con $q=2$	Estim. $\hat{\alpha}_{i1}$ subgrupo con $\hat{q}=2$	Media $\bar{\alpha}_{i1}$ subgrupo con $q=3$	Estim. $\hat{\alpha}_{i1}$ subgrupo con $\hat{q}=3$	Media $\bar{\alpha}_{i1}$ subgrupo con $q=4$	Estim. $\hat{\alpha}_{i1}$ subgrupo con $\hat{q}=4$
3,61	3,33 *	3,53	3,52 *	3,35	3,41 *	3,37	3,48 *

Tabla A.9. Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 1 con simulación heterogénea (20 preguntas, 125 individuos) y estimación heterogénea

Medida de desempeño	Resultados obtenidos en estimación con capacidad heterogénea de ordenamiento
MLL	-452,36
Log-verosimilitud de orden	-280,56
Tasa de éxito (%)	95,55
RMSE	0,20

Tabla A.10. Medidas de desempeño en diseño 1 con simulación heterogénea (20 preguntas, 125 individuos) y estimación heterogénea

ANEXO B: Resultados simulación con Diseño 2

1. Simulación homogénea (20 preguntas, 125 individuos) y enfoque de estimación homogéneo

Simulación con capacidad de ordenamiento de 1 atributo (q=1)

Media poblacional \bar{v}_i	Estimación \hat{v}_i en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
7,85	6,24 *	7,79 *	7,42 *	7,33 *
0,16	-0,95 *	1,67	-0,10 *	-0,39 *
0,08	-0,38 *	2,38	1,80	1,58
0,20	-4,58	0,17 *	-1,00	-0,64
Media poblacional $\bar{\alpha}_{i1}$	Estimación $\hat{\alpha}_{i1}$ en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
3,46	2,77	2,26	2,02	1,93

Tabla B.1. Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación homogénea (20 preguntas, 125 individuos) con q=1 y estimación homogénea

Medida de desempeño	Resultados obtenidos en estimación con capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
MLL	-478,94	-951,45	-1107,90	-1143,60
Log-verosimilitud de orden	-411,98	-347,26	-296,64	-278,86
Tasa de éxito (%)	76,84	81,83	85,13	86,10
RMSE	0,16	0,13	0,14	0,13

Tabla B.2. Medidas de desempeño en diseño 2 con simulación homogénea (20 preguntas, 125 individuos) con q=1 y estimación homogénea

Simulación con capacidad de ordenamiento de 2 atributos (q=2)

Media poblacional \bar{v}_i	Estimación \widehat{v}_i en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
7,85	10,60 *	9,87 *	8,85 *	8,88 *
6,16	9,11 *	8,04 *	7,21 *	7,26 *
0,08	-7,20	3,12	2,14	1,51
0,20	0,35 *	0,69 *	-1,20 *	0,45 *
Media poblacional $\bar{\alpha}_{i1}$	Estimación $\widehat{\alpha}_{i1}$ en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
3,46	2,59	3,44 *	2,95	2,51

Tabla B.3. Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación homogénea (20 preguntas, 125 individuos) con q=2 y estimación homogénea

Medida de desempeño	Resultados obtenidos en estimación con capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
MLL	-569,66	-471,63	-671,13	-799,60
Log-verosimilitud de orden	-449,15	-295,16	-223,69	-228,50
Tasa de éxito (%)	78,03	86,35	89,20	91,18
RMSE	0,13	0,10	0,11	0,08

Tabla B.4. Medidas de desempeño en diseño 2 con simulación homogénea (20 preguntas, 125 individuos) con q=2 y estimación homogénea

Simulación con capacidad de ordenamiento de 3 atributos (q=3)

Media poblacional \bar{v}_i	Estimación \widehat{v}_i en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
7,85	10,60 *	12,35 *	10,58 *	9,02 *
6,16	8,76 *	10,71 *	8,79 *	7,15 *
4,08	6,91 *	8,64	7,32 *	5,83
0,20	4,37	5,75	4,15 *	0,68 *
Media poblacional $\bar{\alpha}_{i1}$	Estimación $\widehat{\alpha}_{i1}$ en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
3,46	2,95	3,68	3,98	3,07

Tabla B.5. Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación homogénea (20 preguntas, 125 individuos) con q=3 y estimación homogénea

Medida de desempeño	Resultados obtenidos en estimación con capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
MLL	-468,59	-402,07	-388,25	-569,95
Log-verosimilitud de orden	-467,23	-347,21	-249,64	-207,03
Tasa de éxito (%)	79,30	87,98	91,70	94,28
RMSE	0,10	0,10	0,10	0,11

Tabla B.6. Medidas de desempeño en diseño 2 con simulación homogénea (20 preguntas, 125 individuos) con q=3 y estimación homogénea

Simulación con capacidad de ordenamiento de 4 atributos (q=4)

Media poblacional \bar{v}_i	Estimación \widehat{v}_i en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
7,85	7,87 *	9,23 *	9,15 *	7,62 *
6,16	6,53 *	7,84 *	7,76	6,28 *
4,08	3,88 *	5,57 *	1,62	3,94 *
2,20	-10,36	2,42 *	3,46	1,84 *
Media poblacional $\bar{\alpha}_{i1}$	Estimación $\widehat{\alpha}_{i1}$ en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
3,46	2,87	3,43 *	3,33 *	3,45 *

Tabla B.7. Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación homogénea (20 preguntas, 125 individuos) con q=4 y estimación homogénea

Medida de desempeño	Resultados obtenidos en estimación con capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
MLL	-501,25	-450,20	-455,68	-448,27
Log-verosimilitud de orden	-464,88	-344,39	-227,86	-229,02
Tasa de éxito (%)	79,47	87,76	94,42	95,81
RMSE	0,18	0,13	0,18	0,12

Tabla B.8. Medidas de desempeño en diseño 2 con simulación homogénea (20 preguntas, 125 individuos) con q=4 y estimación homogénea

2. Simulación heterogénea (20 preguntas, 125 individuos) y enfoque de estimación heterogéneo

Media \bar{v}_i subgrupo con $q=1$	Estim. \hat{v}_i subgrupo con $\hat{q}=1$	Media \bar{v}_i subgrupo con $q=2$	Estim. \hat{v}_i subgrupo con $\hat{q}=2$	Media \bar{v}_i subgrupo con $q=3$	Estim. \hat{v}_i subgrupo con $\hat{q}=3$	Media \bar{v}_i subgrupo con $q=4$	Estim. \hat{v}_i subgrupo con $\hat{q}=4$
7,70	9,54 *	7,80	9,42 *	8,00	9,48 *	7,86	9,15 *
0,14	6,70	6,35	6,90 *	6,15	6,89 *	6,00	7,22 *
0,10	4,25	-0,05	4,18	4,29	4,35 *	3,98	4,22 *
0,52	2,23	-0,18	2,14	0,25	1,92	2,20	2,31 *
Media $\bar{\alpha}_{i1}$ subgrupo con $q=1$	Estim. $\hat{\alpha}_{i1}$ subgrupo con $\hat{q}=1$	Media $\bar{\alpha}_{i1}$ subgrupo con $q=2$	Estim. $\hat{\alpha}_{i1}$ subgrupo con $\hat{q}=2$	Media $\bar{\alpha}_{i1}$ subgrupo con $q=3$	Estim. $\hat{\alpha}_{i1}$ subgrupo con $\hat{q}=3$	Media $\bar{\alpha}_{i1}$ subgrupo con $q=4$	Estim. $\hat{\alpha}_{i1}$ subgrupo con $\hat{q}=4$
3,61	3,38	3,53	3,42 *	3,35	3,65	3,37	3,51 *

Tabla B.9. Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación heterogénea (20 preguntas, 125 individuos) y estimación heterogénea

Medida de desempeño	Resultados obtenidos en estimación con capacidad heterogénea de ordenamiento
MLL	-475,96
Log-verosimilitud de orden	-296,94
Tasa de éxito (%)	87,09
RMSE	0,14

Tabla B.10. Medidas de desempeño en diseño 2 con simulación heterogénea (20 preguntas, 125 individuos) y estimación heterogénea

3. Simulación homogénea de (40 preguntas, 125 individuos) y enfoque de estimación homogéneo

Simulación con capacidad de ordenamiento de 1 atributo (q=1)

Media poblacional \bar{v}_i	Estimación \hat{v}_i en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
7,85	7,78 *	8,00 *	7,88 *	7,79 *
0,16	1,31 *	1,53	0,26 *	-0,35 *
0,08	0,58 *	1,38	0,82	0,30 *
0,20	-1,72 *	-0,04 *	-0,12 *	0,05 *
Media poblacional $\bar{\alpha}_{i1}$	Estimación $\hat{\alpha}_{i1}$ en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
3,46	2,41	2,19	1,96	1,87

Tabla B.11. Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación homogénea (40 preguntas, 125 individuos) con q=1 y estimación homogénea

Medida de desempeño	Resultados obtenidos en estimación con capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
MLL	-1022,60	-1864,00	-2235,50	-2278,40
Log-verosimilitud de orden	-407,22	-351,10	-329,95	-284,95
Tasa de éxito (%)	78,29	82,71	85,27	85,93
RMSE	0,13	0,12	0,11	0,11

Tabla B.12. Medidas de desempeño en diseño 2 con simulación homogénea (40 preguntas, 125 individuos) con q=1 y estimación homogénea

Simulación con capacidad de ordenamiento de 2 atributos (q=2)

Media poblacional \bar{v}_i	Estimación \widehat{v}_i en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
7,85	8,75 *	7,56 *	7,76 *	7,35 *
6,16	7,50 *	6,18 *	6,38 *	6,00 *
0,08	-10,94	0,72 *	1,19	0,57 *
0,20	-0,26 *	-0,12 *	0,17 *	0,34 *
Media poblacional $\bar{\alpha}_{i1}$	Estimación $\widehat{\alpha}_{i1}$ en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
3,46	2,46	3,05	2,64	2,47

Tabla B.13. Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación homogénea (40 preguntas, 125 individuos) con q=2 y estimación homogénea

Medida de desempeño	Resultados obtenidos en estimación con capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
MLL	-1051,00	-888,09	-1447,40	-1578,90
Log-verosimilitud de orden	-452,03	-297,75	-255,95	-251,49
Tasa de éxito (%)	79,71	86,91	89,98	91,32
RMSE	0,16	0,08	0,08	0,07

Tabla B.14. Medidas de desempeño en diseño 2 con simulación homogénea (40 preguntas, 125 individuos) con q=2 y estimación homogénea

Simulación con capacidad de ordenamiento de 3 atributos (q=3)

Media poblacional \bar{v}_i	Estimación \widehat{v}_i en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
7,85	12,59	7,42 *	7,76 *	7,45 *
6,16	10,84	5,76 *	6,03 *	5,73 *
4,08	8,97	4,11 *	4,26 *	3,97 *
0,20	3,92	-15,46	0,29 *	0,07 *
Media poblacional $\bar{\alpha}_{i1}$	Estimación $\widehat{\alpha}_{i1}$ en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
3,46	2,36	3,05	3,70	3,02

Tabla B.15. Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación homogénea (40 preguntas, 125 individuos) con q=3 y estimación homogénea

Medida de desempeño	Resultados obtenidos en estimación con capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
MLL	-1072,30	-847,97	-784,75	-1099,90
Log-verosimilitud de orden	-464,35	-347,72	-243,73	-226,45
Tasa de éxito (%)	79,40	87,50	92,68	94,53
RMSE	0,10	0,14	0,09	0,09

Tabla B.16. Medidas de desempeño en diseño 2 con simulación homogénea (40 preguntas, 125 individuos) con q=3 y estimación homogénea

Simulación con capacidad de ordenamiento de 4 atributos (q=4)

Media poblacional \bar{v}_i	Estimación \widehat{v}_i en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
7,85	6,61 *	8,41 *	12,31	8,27 *
6,16	4,77 *	6,76 *	10,67	6,60 *
4,08	2,99 *	5,07 *	8,85	4,78 *
2,20	1,75 *	2,86 *	7,13	2,64 *
Media poblacional $\bar{\alpha}_{i1}$	Estimación $\widehat{\alpha}_{i1}$ en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
3,46	2,27	3,24	3,52 *	3,43 *

Tabla B.17. Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación homogénea (40 preguntas, 125 individuos) con q=4 y estimación homogénea

Medida de desempeño	Resultados obtenidos en estimación con capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
MLL	-1168,50	-893,63	-863,44	-855,13
Log-verosimilitud de orden	-470,41	-364,19	-275,72	-222,50
Tasa de éxito (%)	79,39	87,71	92,99	95,94
RMSE	0,13	0,12	0,11	0,12

Tabla B.18. Medidas de desempeño en diseño 2 con simulación homogénea (40 preguntas, 125 individuos) con q=4 y estimación homogénea

4. Simulación heterogénea (40 preguntas, 125 individuos) y enfoque de estimación heterogéneo

Media \bar{v}_i subgrupo con $q=1$	Estim. \hat{v}_i subgrupo con $\hat{q}=1$	Media \bar{v}_i subgrupo con $q=2$	Estim. \hat{v}_i subgrupo con $\hat{q}=2$	Media \bar{v}_i subgrupo con $q=3$	Estim. \hat{v}_i subgrupo con $\hat{q}=3$	Media \bar{v}_i subgrupo con $q=4$	Estim. \hat{v}_i subgrupo con $\hat{q}=4$
7,70	9,03 *	7,80	8,84 *	8,00	8,68 *	7,86	8,82 *
0,14	6,16	6,35	6,41 *	6,15	6,57 *	6,00	6,36 *
0,10	3,93	-0,05	3,72	4,29	3,81 *	3,98	3,92 *
0,52	1,38 *	-0,18	1,46	0,25	1,47	2,20	1,43 *
Media $\bar{\alpha}_{i1}$ subgrupo con $q=1$	Estim. $\hat{\alpha}_{i1}$ subgrupo con $\hat{q}=1$	Media $\bar{\alpha}_{i1}$ subgrupo con $q=2$	Estim. $\hat{\alpha}_{i1}$ subgrupo con $\hat{q}=2$	Media $\bar{\alpha}_{i1}$ subgrupo con $q=3$	Estim. $\hat{\alpha}_{i1}$ subgrupo con $\hat{q}=3$	Media $\bar{\alpha}_{i1}$ subgrupo con $q=4$	Estim. $\hat{\alpha}_{i1}$ subgrupo con $\hat{q}=4$
3,61	2,59	3,53	2,70	3,35	3,02	3,37	2,97

Tabla B.19. Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación heterogénea (40 preguntas, 125 individuos) y estimación heterogénea

Medida de desempeño	Resultados obtenidos en estimación con capacidad heterogénea de ordenamiento
MLL	-926,44
Log-verosimilitud de orden	-298,11
Tasa de éxito (%)	88,79
RMSE	0,14

Tabla B.20. Medidas de desempeño en diseño 2 con simulación heterogénea (40 preguntas, 125 individuos) y estimación heterogénea

5. Simulación homogénea (20 preguntas, 250 individuos) y enfoque de estimación homogéneo

Simulación con capacidad de ordenamiento de 1 atributo (q=1)

Media poblacional \bar{v}_i	Estimación \hat{v}_i en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
7,97	6,17 *	8,45 *	7,80 *	7,62 *
-0,04	-1,49 *	0,68 *	-0,48 *	-0,54 *
-0,02	-0,69 *	1,46	1,15	0,87
0,01	-3,38	0,36 *	-0,22 *	0,07 *
Media poblacional $\bar{\alpha}_{i1}$	Estimación $\hat{\alpha}_{i1}$ en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
3,46	2,71	2,37	2,00	1,91

Tabla B.21. Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación homogénea (20 preguntas, 250 individuos) con q=1 y estimación homogénea

Medida de desempeño	Resultados obtenidos en estimación con capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
MLL	-947,53	-1902,90	-2178,10	-2247,00
Log-verosimilitud de orden	-811,63	-706,43	-622,05	-594,26
Tasa de éxito (%)	76,78	82,83	85,60	86,24
RMSE	0,12	0,10	0,12	0,11

Tabla B.22. Medidas de desempeño en diseño 2 con simulación homogénea (20 preguntas, 250 individuos) con q=1 y estimación homogénea

Simulación con capacidad de ordenamiento de 2 atributos (q=2)

Media poblacional \bar{v}_i	Estimación \widehat{v}_i en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
7,97	6,98	7,47 *	8,96 *	8,26 *
5,96	5,22 *	5,64 *	6,76 *	6,37 *
-0,02	-4,03	-1,30 *	2,06	1,56
0,01	-4,25	-0,24 *	-0,91 *	0,29 *
Media poblacional $\bar{\alpha}_{i1}$	Estimación $\widehat{\alpha}_{i1}$ en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
3,46	2,79	3,31	2,93	2,59

Tabla B.23. Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación homogénea (20 preguntas, 250 individuos) con q=2 y estimación homogénea

Medida de desempeño	Resultados obtenidos en estimación con capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
MLL	-961,88	-806,54	-1214,20	-1490,40
Log-verosimilitud de orden	-893,08	-580,63	-445,52	-458,88
Tasa de éxito (%)	78,20	86,81	89,85	91,90
RMSE	0,08	0,08	0,10	0,08

Tabla B.24. Medidas de desempeño en diseño 2 con simulación homogénea (20 preguntas, 250 individuos) con q=2 y estimación homogénea

Simulación con capacidad de ordenamiento de 3 atributos (q=3)

Media poblacional \bar{v}_i	Estimación \widehat{v}_i en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
7,97	8,07 *	10,17 *	8,25 *	8,25 *
5,96	6,10 *	8,44	6,34 *	6,38 *
3,98	4,77 *	5,86 *	5,02 *	4,97
0,01	1,85 *	3,02	0,91 *	0,31 *
Media poblacional $\bar{\alpha}_{i1}$	Estimación $\widehat{\alpha}_{i1}$ en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
3,46	3,07	3,71	3,63	3,00

Tabla B.25. Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación homogénea (20 preguntas, 250 individuos) con q=3 y estimación homogénea

Medida de desempeño	Resultados obtenidos en estimación con capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
MLL	-927,36	-809,58	-824,66	-1159,40
Log-verosimilitud de orden	-932,80	-666,76	-484,74	-430,54
Tasa de éxito (%)	78,65	87,36	91,10	94,01
RMSE	0,10	0,10	0,10	0,10

Tabla B.26. Medidas de desempeño en diseño 2 con simulación homogénea (20 preguntas, 250 individuos) con q=3 y estimación homogénea

Simulación con capacidad de ordenamiento de 4 atributos (q=4)

Media poblacional \bar{v}_i	Estimación \widehat{v}_i en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
7,97	7,29 *	7,92 *	8,55 *	7,74 *
5,96	5,51 *	6,03 *	6,56 *	5,74 *
3,98	2,86 *	3,45 *	1,59	3,62 *
2,01	-1,83 *	2,19 *	2,98	1,69 *
Media poblacional $\bar{\alpha}_{i1}$	Estimación $\widehat{\alpha}_{i1}$ en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
3,46	2,75	3,30	3,42 *	3,37 *

Tabla B.27. Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación homogénea (20 preguntas, 250 individuos) con q=4 y estimación homogénea

Medida de desempeño	Resultados obtenidos en estimación con capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
MLL	-1068,10	-922,01	-921,06	-923,20
Log-verosimilitud de orden	-914,09	-681,04	-480,65	-461,00
Tasa de éxito (%)	78,47	87,45	94,09	95,55
RMSE	0,14	0,12	0,17	0,12

Tabla B.28. Medidas de desempeño en diseño 2 con simulación homogénea (20 preguntas, 250 individuos) con q=4 y estimación homogénea

6. Simulación heterogénea de (20 preguntas, 250 individuos) y enfoque de estimación heterogéneo

Media \bar{v}_i subgrupo con $q=1$	Estim. \hat{v}_i subgrupo con $\hat{q}=1$	Media \bar{v}_i subgrupo con $q=2$	Estim. \hat{v}_i subgrupo con $\hat{q}=2$	Media \bar{v}_i subgrupo con $q=3$	Estim. \hat{v}_i subgrupo con $\hat{q}=3$	Media \bar{v}_i subgrupo con $q=4$	Estim. \hat{v}_i subgrupo con $\hat{q}=4$
7,79	8,68 *	7,84	8,66 *	8,20	8,65 *	8,03	8,56 *
0,05	6,23	6,01	6,35 *	5,92	6,34 *	5,86	6,40 *
-0,14	4,27	0,00	4,23	4,21	4,38 *	3,85	4,32 *
0,29	0,88 *	-0,09	0,78 *	-0,18	0,67 *	2,05	0,84
Media $\bar{\alpha}_{i1}$ subgrupo con $q=1$	Estim. $\hat{\alpha}_{i1}$ subgrupo con $\hat{q}=1$	Media $\bar{\alpha}_{i1}$ subgrupo con $q=2$	Estim. $\hat{\alpha}_{i1}$ subgrupo con $\hat{q}=2$	Media $\bar{\alpha}_{i1}$ subgrupo con $q=3$	Estim. $\hat{\alpha}_{i1}$ subgrupo con $\hat{q}=3$	Media $\bar{\alpha}_{i1}$ subgrupo con $q=4$	Estim. $\hat{\alpha}_{i1}$ subgrupo con $\hat{q}=4$
3,49	2,98	3,40	3,38	3,52	3,36	3,43	3,25

Tabla B.29. Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación heterogénea (20 preguntas, 250 individuos) y estimación heterogénea

Medida de desempeño	Resultados obtenidos en estimación con capacidad heterogénea de ordenamiento
MLL	-961,54
Log-verosimilitud de orden	-605,41
Tasa de éxito (%)	87,04
RMSE	0,16

Tabla B.30. Medidas de desempeño en diseño 2 con simulación heterogénea (20 preguntas, 250 individuos) y estimación heterogénea

7. Simulación homogénea de (40 preguntas, 250 individuos) y enfoque de estimación homogéneo

Simulación con capacidad de ordenamiento de 1 atributo (q=1)

Media poblacional \bar{v}_i	Estimación \hat{v}_i en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
7,97	8,53 *	8,48 *	7,39 *	7,72 *
-0,04	0,44 *	1,53	0,18 *	-0,49 *
-0,02	0,83 *	1,19	0,72	0,36 *
0,01	0,09 *	0,17 *	-0,07 *	-0,04 *
Media poblacional $\bar{\alpha}_{i1}$	Estimación $\hat{\alpha}_{i1}$ en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
3,46	2,38	2,10	1,93	1,83

Tabla B.31. Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación homogénea (40 preguntas, 250 individuos) con q=1 y estimación homogénea

Medida de desempeño	Resultados obtenidos en estimación con capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
MLL	-2096,10	-3818,30	-4516,20	-4629,30
Log-verosimilitud de orden	-811,19	-703,65	-663,54	-555,73
Tasa de éxito (%)	78,00	82,22	84,82	85,49
RMSE	0,09	0,11	0,11	0,12

Tabla B.32. Medidas de desempeño en diseño 2 con simulación homogénea (40 preguntas, 250 individuos) con q=1 y estimación homogénea

Simulación con capacidad de ordenamiento de 2 atributos (q=2)

Media poblacional \bar{v}_i	Estimación \widehat{v}_i en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
7,97	5,33	6,52 *	7,47 *	7,29 *
5,96	3,81	5,10 *	6,10 *	5,92 *
-0,02	-6,23	-0,92 *	1,14	0,56
0,01	-3,83	-0,52 *	0,34 *	0,24 *
Media poblacional $\bar{\alpha}_{i1}$	Estimación $\widehat{\alpha}_{i1}$ en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
3,46	2,42	3,07	2,63	2,50

Tabla B.33. Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación homogénea (40 preguntas, 250 individuos) con q=2 y estimación homogénea

Medida de desempeño	Resultados obtenidos en estimación con capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
MLL	-2051,20	-1690,00	-2789,70	-3038,30
Log-verosimilitud de orden	-908,96	-608,93	-539,39	-508,02
Tasa de éxito (%)	79,53	87,02	90,26	91,64
RMSE	0,09	0,08	0,08	0,08

Tabla B.34. Medidas de desempeño en diseño 2 con simulación homogénea (40 preguntas, 250 individuos) con q=2 y estimación homogénea

Simulación con capacidad de ordenamiento de 3 atributos (q=3)

Media poblacional \bar{v}_i	Estimación \widehat{v}_i en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
7,97	8,05 *	7,52 *	7,53 *	7,78 *
5,96	6,23 *	5,54 *	5,55 *	5,73 *
3,98	4,27 *	4,10 *	4,02 *	4,12 *
0,01	-5,17	-0,60 *	0,45 *	0,18 *
Media poblacional $\bar{\alpha}_{i1}$	Estimación $\widehat{\alpha}_{i1}$ en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
3,46	2,55	3,20	3,43 *	2,98

Tabla B.35. Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación homogénea (40 preguntas, 250 individuos) con q=3 y estimación homogénea

Medida de desempeño	Resultados obtenidos en estimación con capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
MLL	-2038,00	-1677,10	-1597,30	-2209,50
Log-verosimilitud de orden	-920,89	-694,32	-492,80	-446,26
Tasa de éxito	79,30	87,59	92,61	94,45
RMSE	0,12	0,10	0,09	0,09

Tabla B.36. Medidas de desempeño en diseño 2 con simulación homogénea (40 preguntas, 250 individuos) con q=3 y estimación homogénea

Simulación con capacidad de ordenamiento de 4 atributos (q=4)

Media poblacional \bar{v}_i	Estimación \widehat{v}_i en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
7,97	6,66	8,21 *	8,09 *	7,36 *
5,96	4,67	6,42 *	6,27 *	5,64 *
3,98	3,33 *	4,83 *	4,62 *	4,03 *
2,01	1,01 *	2,77 *	2,93 *	2,06 *
Media poblacional $\bar{\alpha}_{i1}$	Estimación $\widehat{\alpha}_{i1}$ en base a capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
3,46	2,41	3,12	3,66	3,55 *

Tabla B.37. Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación homogénea (40 preguntas, 250 individuos) con q=4 y estimación homogénea

Medida de desempeño	Resultados obtenidos en estimación con capacidad de ordenamiento "q"			
	q=1	q=2	q=3	q=4
MLL	-2062,70	-1639,20	-1520,00	-1535,40
Log-verosimilitud de orden	-938,50	-730,61	-568,63	-483,21
Tasa de éxito (%)	79,50	88,02	93,47	96,38
RMSE	0,13	0,12	0,11	0,11

Tabla B.38. Medidas de desempeño en diseño 2 con simulación homogénea (40 preguntas, 250 individuos) con q=4 y estimación homogénea

8. Simulación heterogénea (40 preguntas, 250 individuos) y enfoque de estimación heterogéneo

Media \bar{v}_i subgrupo con $q=1$	Estim. \hat{v}_i subgrupo con $\hat{q}=1$	Media \bar{v}_i subgrupo con $q=2$	Estim. \hat{v}_i subgrupo con $\hat{q}=2$	Media \bar{v}_i subgrupo con $q=3$	Estim. \hat{v}_i subgrupo con $\hat{q}=3$	Media \bar{v}_i subgrupo con $q=4$	Estim. \hat{v}_i subgrupo con $\hat{q}=4$
7,79	8,62 *	7,84	8,49 *	8,20	8,40 *	8,03	8,42 *
0,05	6,16	6,01	6,32 *	5,92	6,35 *	5,86	6,36 *
-0,14	3,97	0,00	3,93	4,21	3,09 *	3,85	4,00 *
0,29	1,50	-0,09	1,55	-0,18	1,47	2,05	1,56 *
Media $\bar{\alpha}_{i1}$ subgrupo con $q=1$	Estim. $\hat{\alpha}_{i1}$ subgrupo con $\hat{q}=1$	Media $\bar{\alpha}_{i1}$ subgrupo con $q=2$	Estim. $\hat{\alpha}_{i1}$ subgrupo con $\hat{q}=2$	Media $\bar{\alpha}_{i1}$ subgrupo con $q=3$	Estim. $\hat{\alpha}_{i1}$ subgrupo con $\hat{q}=3$	Media $\bar{\alpha}_{i1}$ subgrupo con $q=4$	Estim. $\hat{\alpha}_{i1}$ subgrupo con $\hat{q}=4$
3,49	2,56	3,40	2,84	3,52	3,03	3,43	3,15

Tabla B.39. Recuperación de parámetros \bar{v}_i y $\bar{\alpha}_{i1}$ en diseño 2 con simulación heterogénea (40 preguntas, 250 individuos) y estimación heterogénea

Medida de desempeño	Resultados obtenidos en estimación con capacidad heterogénea de ordenamiento
MLL	-1826,70
Log-verosimilitud de orden	-604,42
Tasa de éxito (%)	88,62
RMSE	0,14

Tabla B.40. Medidas de desempeño en diseño 2 con simulación heterogénea (40 preguntas, 250 individuos) y estimación heterogénea

ANEXO C: Tasa de éxito post estimación *out of sample*

Cantidad de preguntas de la encuesta	Capacidad de ordenamiento de la muestra simulada	Capacidad de ordenamiento en estimación	Tamaño de la muestra simulada	
			125	250
20	q=1	q=1 q=4	80,20% 79,40%	75,90% 75,50%
	q=2	q=2 q=4	71,80% 71,80%	74,10% 74,80%
	q=3	q=3 q=4	77,60% 75,40%	77,00% 77,90%
	q=4	q=4	77,00%	78,10%
40	q=1	q=1 q=4	80,70% 79,90%	79,80% 79,15%
	q=2	q=2 q=4	78,40% 78,00%	76,60% 75,95%
	q=3	q=3 q=4	78,70% 78,60%	78,95% 78,95%
	q=4	q=4	82,00%	79,45%

Tabla C.1. Tasa de éxito *out of sample* post estimación bajo el enfoque de estimación homogéneo

Cantidad de preguntas de la encuesta	Capacidad de ordenamiento de la muestra simulada	Capacidad de ordenamiento en estimación	Tamaño de la muestra simulada	
			125	250
20	Heterogénea	Heterogénea q=4	74,20% 74,00%	77,40% 75,70%
40	Heterogénea	Heterogénea q=4	79,40% 78,50%	79,10% 77,50%

Tabla C.2. Tasa de éxito *out of sample* post estimación bajo el enfoque de estimación heterogéneo