



Guía de ejercicios 4

1. ¿Cuál es la teoría detrás de los modelos de elección discreta MNL y HL? ¿En qué se diferencian ambos modelos? ¿Cuándo debe preferirse un modelo HL frente a un modelo MNL?
2. ¿Cuáles condiciones debe cumplir un parámetro estructural \emptyset en un modelo jerárquico para que sea válido?
3. ¿Qué relación debe cumplirse entre dos parámetros de una función de utilidad en un modelo de partición modal, en el que uno de ellos acompaña al tiempo de espera y el segundo acompaña al tiempo de espera considerando variación sistemática de gustos con respecto al sexo? Considere que la variable dummy asociada al sexo es 1 cuando es mujer y cero en otro caso. ¿Cuáles serían los signos de ambos parámetros?
4. Considere un viaje que puede realizarse en los modos auto, bus, metro y bicicleta. Proponga dos estructuras jerárquicas para modelar los viajes, y por qué pueden resultar adecuadas.
5. En un modelo agregado de Distribución – Partición modal, ¿qué relación debe existir entre los parámetros β y λ ?
6. Si tiene un modelo de partición modal con tres alternativas: auto, bus y tren. Escriba la matriz de covarianzas para un MNL, un HL con un nido de transporte público y un probit.
7. Considere un modelo HL con tres alternativas modales A, B y C, donde B y C están correlacionadas (existe un nido N). Las funciones de utilidad son:

$$\begin{array}{ll}
 U_A = V_A + \varepsilon_A & \varepsilon_A, \varepsilon_N + \varepsilon_B, \varepsilon_N + \varepsilon_C \quad \text{Gumbel}(\beta) \\
 U_B = V_B + \varepsilon_N + \varepsilon_B & \varepsilon_B, \varepsilon_C \quad \text{Gumbel}(\lambda) \\
 U_C = V_C + \varepsilon_N + \varepsilon_C & \varepsilon_N \quad \text{Logística}
 \end{array}$$

Sean ρ_{BC} la correlación entre las utilidades de B y C; y $\phi_N = \beta/\lambda$ la razón entre los parámetros de escala asociada al nido N. Demuestre que $\rho_{BC} + \phi_N^2 = 1$.

8. Usted estima un Logit Jerárquico y obtiene un parámetro de escala del nido de 0,8 con un test $t = 4$. ¿Qué puede concluir sobre los nidos de este modelo si el valor del test-t es al 95%?

9. Si tuviera que elegir entre dos modelos de partición modal, un MNL y un HL, de los que obtuvo sus valores FPR; ER y E(CR), señale qué modelo escogería si:

- a) Para el modelo MNL, FPR y ER no son similares, y para el modelo HL, FPR, ER y E(CR) son similares.
- b) Para ambos modelos, los valores FPR y ER son similares.

Si faltara información, explique que otros factores consideraría para escoger entre estos modelos.

10. Si tuviera que evaluar la localización de una nueva línea de Metro en una zona antes inexistente, ¿qué niveles de servicio evaluaría en esa zona? ¿Por qué? A nivel de partición modal de la ciudad, ¿Cómo evaluaría el efecto de la nueva infraestructura para ver si es adecuada o no?

11. Recuerde que la varianza de una variable aleatoria Gumbel está dada por la expresión $\sigma^2 = \pi^2/6\beta^2$. Suponga un modelo Logit jerárquico de elección conjunta Destino-Modo. ¿Cuáles son los límites que puede adoptar la razón entre el parámetro de escala del nido superior y el parámetro de escala del nido inferior (elección de modo)? Haga los supuestos que considere necesarios.

12. ¿Por qué en el contexto de modelos de elección de ruta, a diferencia de otros modelos de elección discreta como elección de modo u horario de viaje, es necesario utilizar algoritmos para generar un conjunto de alternativas disponibles? Considere en particular el Método de Eliminación visto en clases, ¿qué problema posee utilizar únicamente este algoritmo para generar el conjunto de alternativas disponibles?

13. En el contexto de modelos de localización residencial, explique conceptual los enfoques de elección (*choice*) y de remate (*bid*) identificando claramente: el tomador de decisiones, las variables relevantes al momento de decidir y el proceso de decisión.

14. ¿Cómo caracterizaría la correlación entre alternativas de elección de ruta? ¿Por qué en general no puede usar un modelo Logit Jerárquico para modelar dicha correlación?

15. ¿Qué supuesto simplificador realiza el modelo *stop & go* de frecuencia de viaje al plantear su estructura jerárquica de probabilidades? ¿Le parece razonable dicho supuesto?

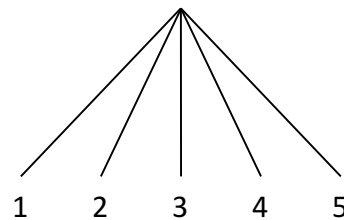
16. ¿Por qué el modelo *stop – go* para modelar frecuencia de viajes es mejor que el ACM?

17. ¿Cómo se integran los modelos de localización residencial y los modelos de transporte, en términos de datos de entrada y salida?

18. Usted obtiene un modelo de partición modal que tiene un coeficiente para el tiempo de viaje distinto si el modo es de transporte público o privado. ¿Cómo tomaría la decisión de aceptar o rechazar el modelo, dada la siguiente información? Determine cuál sería su decisión. Los datos son:

	Tv_TPúblico	Tv_TPrivado
Parámetro	-0,0303	-0,09986
Desviación estándar	0,0137	0,0287

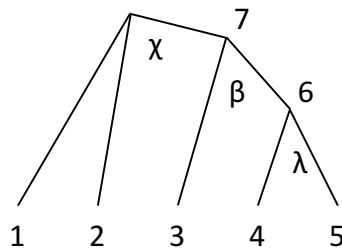
19. Considere el siguiente MNL:



$$U_i = V_i + \varepsilon_i$$

con: $V_i = (\theta_t t_i + \theta_c C_i + \theta_i) \dots$ y considere que $\theta_1 = 0$
y $\varepsilon_i \sim \text{IID Gumbel}(0, \pi^2/6\alpha^2)$

Considere ahora el siguiente HL:



Escriba las ecuaciones equivalentes para U_i y V_i ($i = 1, \dots, 7$) en este caso. Ud. debiera saber que la probabilidad de escoger la opción 4 en este modelo está dada por:

$$P_4 = \frac{\exp(\chi V_7)}{\exp(\chi V_1) + \exp(\chi V_2) + \exp(\chi V_7)} \cdot \frac{\exp(\beta V_6)}{\exp(\beta V_3) + \exp(\beta V_6)} \cdot \frac{\exp(\lambda V_4)}{\exp(\lambda V_4) + \exp(\lambda V_5)}$$

Señale qué relación debe existir entre χ , β y λ para que este HL sea consistente. ¿Puede decir algo interesante respecto a la relación de los parámetros anteriores con α ?

20. Usted ha estimado un modelo de elección de horario de viaje con cuatro alternativas, utilizando una muestra aleatoria de 500 individuos con las siguientes características:

- Todos los individuos tienen disponibles las cuatro alternativas de horario.
 - 60% de la muestra son trabajadores, 30% son estudiantes y 10% son desempleados.
 - El número de individuos que elige cada alternativa es 150, 200, 40 y 110 respectivamente.
- a) Ha estimado un modelo Logit Multinomial (MNL) con seis parámetros: tres atributos (tiempo, costo y ocupación) y tres constantes específicas. $\rho^2 = 0,224$. Además estimó un modelo Logit Jerárquico (HL) donde genera dos nidos: uno que agrupa los dos horarios punta y otro que agrupa los dos horarios fuera de punta. La log-verosimilitud del modelo es $l(\theta_{HL}) = -535,7$. ¿Cuál modelo prefiere, entre el MNL y el HL?

Respuesta: Al comparar con el valor de tabla, no es posible rechazar la hipótesis nula, se elige el modelo MNL.

- b) Determine el valor del índice $\bar{\rho}^2$ del modelo MNL y del modelo HL.

Respuesta: MNL: 0,148; HL: 0,151.

En caso de necesitar valores Chi-Cuadrado:	$\chi^2_{1,95\%} = 3,84$	$\chi^2_{2,95\%} = 5,99$
	$\chi^2_{3,95\%} = 7,81$	
	$\chi^2_{4,95\%} = 9,49$	
	$\chi^2_{5,95\%} = 11,07$	$\chi^2_{6,95\%} = 12,59$
	$\chi^2_{7,95\%} = 14,07$	$\chi^2_{8,95\%} = 15,51$

21. Considere la siguiente estructura de datos que corresponde a un modelo logit jerárquico:

$$\begin{aligned}
U_{11} &= V_{11} + e_1 + e_{11} \\
U_{12} &= V_{12} + e_1 + e_{12} \\
U_{21} &= V_{21} + e_2 + e_{21} \\
U_{22} &= V_{22} + e_2 + e_{22}
\end{aligned}$$

La siguiente expresión entrega la relación entre el parámetro de escala y la varianza de una distribución Gumbel: $\beta^2 = \pi^2/6\sigma^2$

- ¿Qué representan V_{ij} , e_i , e_{ij} ? En relación a los errores e , describa que distribución adopta cada uno de ellos y la suma de ambos.
- Obtenga la covarianza entre U_{ij} y U_{ij}
- Obtenga la correlación entre U_{ij} y U_{ij} . Determine los valores límites de la correlación.
- Deduzca los límites que puede adoptar el cociente $\phi = \beta/\lambda$

22. Considere la siguiente estructura para un HL:

$$\begin{aligned}
U_{111} &= V_{111} + e_1 + e_{11} + e_{111} \\
U_{112} &= V_{112} + e_1 + e_{11} + e_{112} \\
U_{121} &= V_{121} + e_1 + e_{12} + e_{121} \\
U_{222} &= V_{222} + e_2 + e_{22} + e_{222}
\end{aligned}$$

Suponga que los errores e_{ijm} son Gumbel iid con parámetro λ , los errores $e_{ij} + e_{ijm}$ son errores Gumbel iid con parámetro β y los errores $e_i + e_{ij} + e_{ijm}$ son Gumbel iid con parámetro γ . La relación entre el parámetro de escala y la varianza de una distribución Gumbel es $\sigma^2 = \pi^2/6\beta^2$.

- Dadas las estructuras de los errores, dibuje el árbol jerárquico.
- ¿Cuál es la covarianza entre U_{111} y U_{112} y entre U_{111} y U_{121} ?
- ¿Cuál es la matriz de varianzas y covarianzas de este modelo?
- Obtenga la correlación entre U_{111} y U_{112} y entre U_{111} y U_{121}

23. Usted ha estimado cuatro modelos de elección de modo de transporte, y desea escoger el mejor. Los cuatro modelos son tanto del tipo MNL como HL y contienen cuatro modos: Bicicleta Pública (BP), Bus (B), Metro (M) y Caminata (C). En la Tabla 1 se presentan los cuatro modelos (M1, M2, M3 y M4), con todos sus parámetros, test- t (con respecto a cero), e indicadores de ajuste global.

Tabla 1: Modelos Estimados

Atributo	Modo	M1 (MNL)		M2 (HL)		M3 (HL)		M4 (HL)	
		Valor	Test-t	Valor	Test-t	Valor	Test-t	Valor	Test-t
Costo	BP	-0,005	-3,80	-	-	-0,005	-3,80	-0,005	-3,80
	B	-	-	-	-	-	-	-0,006	-3,90
	M	-	-	-	-	-	-	-0,001	-1,96
	B, M	-0,002	-2,71	-	-	-0,002	-2,71	-	-
	BP, B, M	-	-	-0,003	-1,96	-	-	-	-
Tiempo de Viaje	BP	-0,112	-3,73	-	-	-0,095	-2,59	-0,095	-2,59
	B	-	-	-	-	-	-	-0,135	-2,54
	M	-	-	-	-	-	-	-0,142	-2,52
	B, M	-0,125	-4,25	-	-	-0,137	-3,48	-	-
	BP, B, M	-	-	-0,118	-3,95	-	-	-	-
Tiempo de Caminata	C	-0,489	-2,10	-0,125	-1,35	-0,489	-2,10	-0,489	-2,10
Tiempo de Acceso	BP	-	-	-	-	-0,221	-2,18	-0,221	-2,18
	B	-	-	-	-	-	-	-0,304	-3,54
	M	-	-	-	-	-	-	-0,291	-1,99
	B, M	-	-	-	-	-0,287	-2,18	-	-
	BP, B, M	-0,256	-1,99	0,015	2,15	-	-	-	-
Sexo Mujer	B, M	-	-	-	-	-0,150	-2,40	0,211	2,00
ϕ		-	-	0,32	1,97	0,62	3,65	0,76	8,16
Modos en el Nido		-		BP, B, M		B, M		B, M	
Log-verosimilitud		-1.664		-1.639		-1.660		-1.657	

Elija el mejor modelo y justifique.

R: Se prefiere el Modelo 3

24. Entre dos localidades en que solo compiten Auto y Bus, se calibró un modelo MNL Logit con las siguientes funciones de utilidad representativas para los individuos de tipo q:

$$V_{Aq} = -0.03 * C_{Aq} - 0.06 * t_{Aq} + 2.05$$

$$V_{Bq} = -0.03 * C_{Bq} - 0.06 * t_{Bq} + 0.05 * E_q + 0.25 * S_q$$

Donde C_{mq} y t_{mq} denotan el costo y tiempo de viaje en el modo m para el individuo q y E_q y S_q corresponden a la edad y sexo del individuo (1 indica mujer). Considerando los valores para cada observación junto con sus elecciones (presentados en la tabla 2).

Tabla 2 – Atributos por individuo

Persona	C _{Auto}	C _{Bus}	t _{Auto}	t _{Bus}	Edad	Sexo	Elección
1	250	200	26	30	30	1	Bus
2	180	200	12	24	33	1	Auto
3	170	150	11	23	40	0	Auto
4	120	150	22	30	46	0	Bus
5	125	125	11	15	52	1	Auto
6	200	150	20	20	62	1	Bus
7	240	150	15	30	47	0	Bus
8	150	125	19	21	37	0	Auto

- a) Calcule la probabilidad de elección de ambas alternativas para todos los individuos.

Persona	VA	VB	PA	PB
1	-7.01	-6.05	28%	72%
2	-4.07	-5.54	81%	19%
3	-3.71	-3.88	54%	46%
4	-2.87	-4	76%	24%
5	-2.36	-1.8	36%	64%
6	-5.15	-2.35	6%	94%
7	-6.05	-3.95	11%	89%
8	-3.59	-3.16	39%	61%

Respuesta:

- b) Calcule los indicadores de “*First Preference Recovery*” (FPR), “*Chance Recovery*” (CR) y “*Expected Recovery*” (E(FPR)=ER) y concluya sobre los modelos.

Respuesta: FPR = 5 (1,4); ER = 5,91; CR = 4 (2)

25. En el Taller 3 usted estimó cuatro modelos, todos ellos con signos correctos y buenos test-*t*. Usted desea comparar la capacidad predictiva de los modelos usando la base de validación. Para esto aplica los tests vistos en clase, obteniendo los siguientes resultados que se indican en la tabla (en el caso de FPR y CR se indica se intervalo de confianza al 95%). El valor crítico de tabla para el Test de Bandas es $\chi^2 = 23,7$. ¿Qué modelo elegiría usted?

	FPR	ER	CR	Test de Bandas	Verosimilitud Directa
Modelo 1	75 \pm 10	80	50 \pm 10	27,9	Tercero
Modelo 2	90 \pm 7	75	50 \pm 10	17,4	Mejor
Modelo 3	65 \pm 9	70	50 \pm 10	32,1	Peor
Modelo 4	80 \pm 10	85	50 \pm 10	20,1	Segundo

R: Partimos comparando FPR, ER y CR, descartamos los modelos 2 y 3. Aplicamos el test de bandas de probabilidad y elegimos el modelo 4

26. Considere un modelo combinado de elección de horario y modo (donde los individuos eligen primero a qué hora viajar y luego en qué modo viajar) con dos alternativas horarias (punta y fuera de punta) y dos alternativas modales (auto y metro). Ambas decisiones son modeladas mediante modelos Logit. El parámetro de escala para el modelo de elección de horario de viaje es 0,05 y el parámetro de escala para el modelo de elección de modo de viaje es 0,1. Los costos de viaje son:

	Punta	Fuera de Punta
Auto	40	20
Metro	30	30

De acuerdo al modelo, ¿qué proporción de los viajes se realizarán en período punta y en período fuera de punta? ¿Qué proporción se realizará en auto y en metro?

Respuesta:

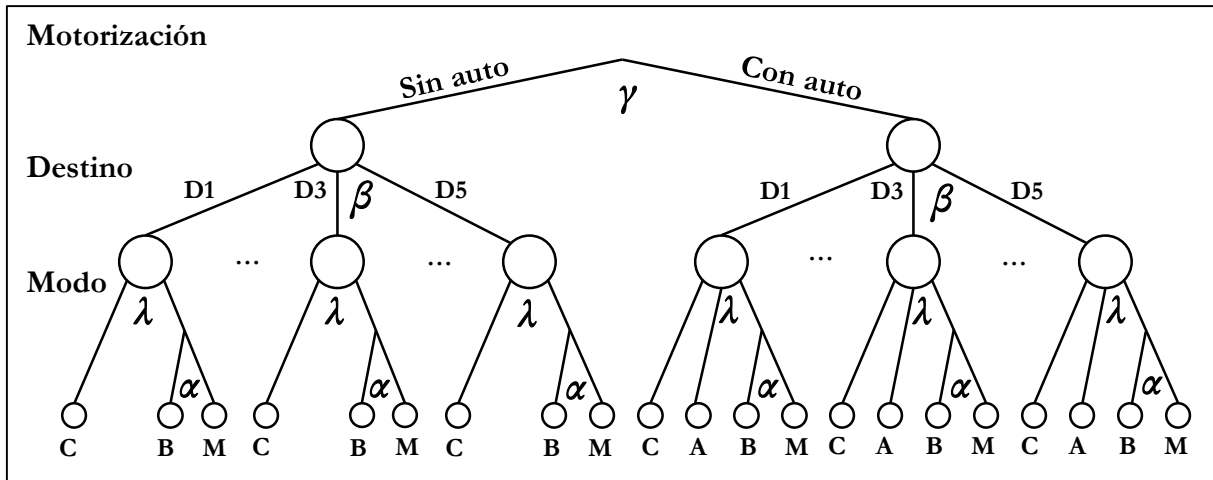
$$P_{metro/Punta} = 0,731 \quad P_{metro/FdP} = 0,269$$

$$P_{FdP} = 0,622 \quad P_{metro} = 0,443$$

26. Considere un modelo combinado de elección de tasa de motorización (p), destino (d) y modo (m), como el del Taller 3, con las siguientes características:

- La elección de tasa de motorización corresponde a un modelo Logit Binomial con parámetro de escala γ , con dos opciones: tener o no tener auto.
- La elección de destino corresponde a un modelo Logit Multinomial con parámetro de escala β , con cinco opciones macrozonales: centro, oriente, poniente, norte y sur.
- La elección de modo corresponde a un modelo Logit Jerárquico, donde el bus y metro están correlacionados, pero el auto y la caminata son independientes. El parámetro de escala λ define la elección Multinomial entre auto, caminata y transporte público; mientras que el parámetro de escala α define la elección Binomial entre bus y metro.

Cada alternativa está caracterizada por una utilidad V_{pdm} igual para todos los individuos.



Escriba las funciones de utilidad de los cuatro modos para individuo que posee auto y que viaja al destino 4. Esto es: $U_{1,4,C}$, $U_{1,4,A}$, $U_{1,4,B}$ y $U_{1,4,M}$. Identifique claramente las componentes de error asociadas a cada función de utilidad, señalando cómo distribuyen (en función de los parámetros de escala del modelo).

27. Se proponen tres modelos combinados jerárquicos de elección de modo y destino para la ciudad de Santiago. El tipo de usuario representa estudiantes. En la elección de modo, se consideran tres: auto (A), bus (B) y metro (M), mientras que en la elección de destino hay cinco opciones disponibles: norte (N), oriente (O), sur (S), poniente (P) y centro (C). Las funciones de utilidad para las 15 alternativas son las mismas en los tres modelos:

$$U^{md} = \theta_{cte}^m + \theta_{costo}^m * Costo + \theta_{viaje}^m * TViaje + \theta_{esp}^m * TEsp + \theta_{caminata}^m * TCam + \theta_{com}^d * DensCom + \theta_{ofi}^d * DensOfi + \theta_{ed}^d * DensEd + \theta_{salud}^d * DensSalud + \theta_{costoIngAlto}^m * Costo * ZIngAlto + \theta_{costoIngAlto}^m * Costo * ZIngMedio$$

Los modelos combinados propuestos son: (a) un modelo en que primero se elige el destino y luego el modo; (b) un modelo en que primero se elige el modo y luego el destino; y (c) un modelo en que primero se elige el destino y luego el modo, y la elección de metro y bus está correlacionada. Los parámetros son los siguientes para cada modelo:

	Modelo A		Modelo B		Modelo C	
	Valor	Test t	Valor	Test t	Valor	Test t
θ_{cte}	-4,85	-3,15	-3,5	-4,10	-4,41	-3,17
θ_{costo}	-6,15	-5,22	-5,87	-2,15	-6,22	-5,41
θ_{viaje}	-1,21	-2,04	-1,88	-2,88	-1,47	-2,25
θ_{esp}	-1,81	-1,85	-2,91	-2,91	-2,19	-2,01

θ_{cam}	-2,25	-1,99	-3,55	-1,92	-2,76	-2,02
θ_{Com}	3,14	2,30	2,55	5,44	3,44	2,45
θ_{Ofi}	-0,92	-2,99	-1,01	-2,45	0,51	2,01
θ_{Ed}	8,21	5,41	4,21	4,19	7,88	5,28
θ_{salud}	-0,32	-2,55	-0,55	-2,07	-1,01	-2,44
$\theta_{costoIngAlto}$	2,33	1,81	1,52	2,41	2,22	1,99
$\theta_{costoIngMedio}$	1,15	2,12	0,31	2,02	1,00	1,95
$\alpha_{Destino}$	0,42	1,99	-	-	0,50	2,02
$\beta_{Modo-Norte}$	0,49	2,05	-	-	0,44	2,26
$\beta_{Modo-Oriente}$	0,51	2,22	-	-	0,47	2,47
$\beta_{Modo-Poniente}$	0,53	3,48	-	-	0,46	3,31
$\beta_{Modo-Sur}$	0,55	2,51	-	-	0,47	2,70
$\beta_{Modo-Centro}$	0,49	2,33	-	-	0,44	2,28
γ_{Modo}	-	-	0,44	3,55	-	-
$\delta_{Destino-Auto}$	-	-	0,45	2,21	-	-
$\delta_{Destino-Bus}$	-	-	0,46	2,25	-	-
$\delta_{Destino-Metro}$	-	-	0,55	3,44	-	-
$\varepsilon_{Modo-Norte-TP}$	-	-	-	-	0,65	2,22
$\varepsilon_{Modo-Oriente-TP}$	-	-	-	-	0,76	2,44
$\varepsilon_{Modo-Poniente-TP}$	-	-	-	-	0,73	2,56
$\varepsilon_{Modo-Sur-TP}$	-	-	-	-	0,68	3,01
$\varepsilon_{Modo-Centro-TP}$	-	-	-	-	0,72	2,55

- Dibuje los 3 árboles de decisión de los modelos, y plantee una expresión para el cálculo de las utilidades de cada una de las opciones.
- Para ellos, plantee el cálculo de la probabilidad de elegir viajar al norte en bus y de elegir viajar al sur en metro.
- Dados los parámetros de escala, discuta cuál de los 3 es el mejor modelo.

28. Considere el enfoque *Stop & Go* para estimar frecuencia de viajes mediante un modelo de elección discreta jerárquico. Demuestre que si la probabilidad de no viajar es $(1 - p)$ en la primera etapa del modelo, y la probabilidad de escoger la opción de seguir viajando en cada etapa subsiguiente es q , entonces el número esperado de viajes estaría dado simplemente por:

$$\text{Viajes} = p/(1 - q)$$