

Trabajo Práctico N° 3: **Modelos para Variables Categóricas No Ordenadas.**

Ejercicio 1: Alternativas de Pesca.

La variable dependiente *y* toma el valor 1, 2, 3 o 4, dependiendo de cuál de los cuatro modos alternativos de pesca, respectivamente, playa, muelle, barco privado y barco chárter, se elija. En la base de datos, estos son *beach*, *pier*, *private* o *charter*. Los datos provienen de Herriges, J. A. y Kling, C. L. (1999): "Nonlinear Income Effects in Random Utility Models", *Review of Economics and Statistics*, 81, 62-72.

(a) Abrir la base y describir las categorías.

Fishing mode	N(income)	mean(income)	sd(income)
beach	134	4.051617	2.50542
pier	178	3.387172	2.340324
private	418	4.654107	2.777898
charter	452	3.880899	2.050029

Fishing mode	mean(pbeach)	mean(ppier)	mean(pprivate)	mean(pcharter)
beach	35.69949	35.69949	97.80914	125.0032
pier	30.57133	30.57133	82.42908	109.7634
private	137.5271	137.5271	41.60681	70.58408
charter	120.6483	120.6483	44.56376	75.09694

Fishing mode	mean(qbeach)	mean(qpier)	mean(qprivate)	mean(qcharter)
beach	.2791948	.2190015	.1593985	.5176089
pier	.2614444	.2025348	.1501489	.4980798
private	.2082868	.1297646	.1775412	.6539167
charter	.2519077	.1595341	.1771628	.6914998

(b) Estimar un modelo logit multinomial.

Multinomial logistic regression	Number of obs = 1,182
	LR chi2(3) = 41.14
	Prob > chi2 = 0.0000
Log likelihood = -1477.1506	Pseudo R2 = 0.0137

Logit multinomial (relative-risk ratios):

Multinomial logistic regression	Number of obs = 1,182
	LR chi2(3) = 41.14
	Prob > chi2 = 0.0000
Log likelihood = -1477.1506	Pseudo R2 = 0.0137

Note: cons estimates baseline relative risk for each outcome.

(c) *Estimar un modelo logit condicional.*

Alternative-specific conditional logit	Number of obs	=	4,728
Case ID variable: id	Number of cases	=	1182
Alternatives variable: fishmode	Alts per case: min	=	4
	avg	=	4.0
	max	=	4
	Wald chi2(5)	=	252.98
Log likelihood = -1215.1376	Prob > chi2	=	0.0000

	d	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
fishmode							
	p	-.0251166	.0017317	-14.50	0.000	-.0285106	-.0217225
	q	.357782	.1097733	3.26	0.001	.1426302	.5729337
beach							
		(base alternative)					
charter							
	income	-.0332917	.0503409	-0.66	0.508	-.131958	.0653745
	_cons	1.694366	.2240506	7.56	0.000	1.255235	2.133497
pier							
	income	-.1275771	.0506395	-2.52	0.012	-.2268288	-.0283255
	_cons	.7779593	.2204939	3.53	0.000	.3457992	1.210119
private							
	income	.0894398	.0500671	1.79	0.074	-.0086898	.1875694
	_cons	.5272788	.2227927	2.37	0.018	.0906132	.9639444

Ejercicio 2: Predicción de Calificaciones de Clientes.

Net Promoter Score®, o *NPS®*, mide la experiencia del cliente y predice el crecimiento del negocio. Es utilizada por empresas que brindan servicios al consumidor final (bancos, telefónicas, etc). EL NPS se calcula usando la respuesta a una pregunta usando una escala de 0 a 10: ¿Qué tan probable es que recomiende a un amigo o colega? Los encuestados se agrupan de la siguiente manera:

- Los promotores (puntuación 9-10) son entusiastas leales que seguirán comprando y recomendarán a otros, lo que impulsará el crecimiento.
- Los neutrales (puntuación 7-8) son clientes satisfechos pero poco entusiastas que son vulnerables a las ofertas de la competencia.
- Los detractores (puntuación 1-6) son clientes insatisfechos que pueden dañar su marca e impedir el crecimiento a través del boca a boca negativo.

Al restar el porcentaje de detractores del porcentaje de promotores, se obtiene el puntaje neto del promotor, que puede oscilar entre un mínimo de -100 (si todos los clientes son detractores) y un máximo de 100 (si todos los clientes son promotores). Estas encuestas se utilizan para generar estrategias de originación (nuevos clientes) y de reducción de churn (fuga de clientes). La base con la que se va a hacer la primera parte de la práctica consiste en la encuesta de NPS que se le hace a los clientes de un Banco luego de efectuar una transacción en caja. En base a esto, utilizando la base “NPS.dta”, responder las siguientes preguntas.

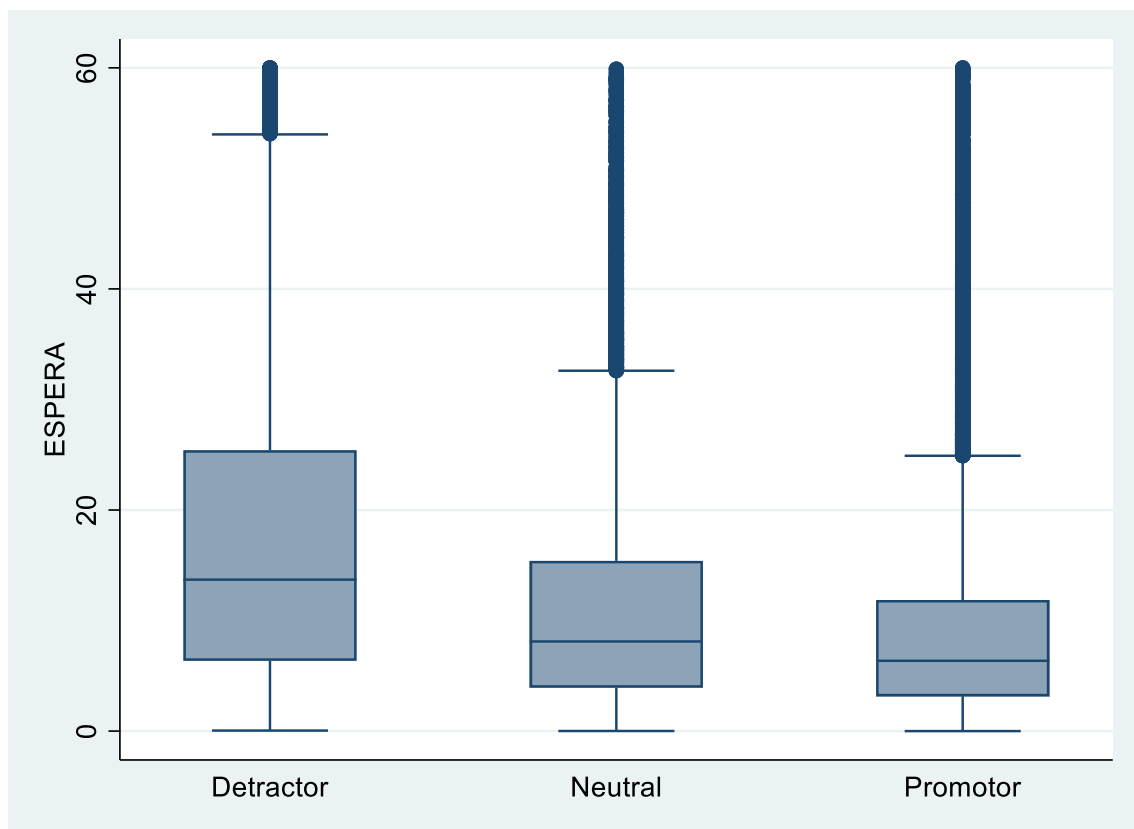
(a) Abrir y describir la base.

Variable	Obs	Mean	Std. dev.	Min	Max
nps	42,019	8.369975	2.263878	1	10
marital_status	0				
gender_code	0				
edad	42,020	52.16497	12.56996	19	101
branch_desc	0				
segmento	0				
operaciones	42,020	1.728439	1.476585	1	31
mes	42,020	6.736292	3.241668	1	12
nps_anterior	0				
hora	42,020	11.7812	1.743031	7	18
dia	42,020	14.91792	8.634796	1	31
dia_semana	0				
espera	42,020	10.89938	10.70589	0	60
cliente	42,020	21372.36	12335.51	1	42760

(b) Generar una variable que clasifique a los clientes en función de si son promotores, detractores o neutrales.

clasificaci on	Freq.	Percent	Cum.
Detractor	6,265	14.91	14.91
Neutral	9,579	22.80	37.71
Promotor	26,175	62.29	100.00
Total	42,019	100.00	

(c) Analizar cómo cambia la variable de espera en función de la clasificación de los clientes.



(d) Tomar una muestra del 10% de los datos. Estimar un logit multinomial para predecir cómo cambian las clasificaciones en función de la espera, condicionando en explicativas que se considere relevantes.

Logit (betas):

Multinomial logistic regression

Number of obs = 4,202

LR chi2(14) = 418.26

Prob > chi2 = 0.0000

Pseudo R2 = 0.0542

Log likelihood = -3647.5859

clasificacion	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
Detractor	(base outcome)					
Neutral						
_Igender_co_2	-.0022797	.1117659	-0.02	0.984	-.2213368	.2167774
_edad	.0106823	.0042832	2.49	0.013	.0022873	.0190772
_Isegmento_2	12.80348	730.9035	0.02	0.986	-1419.741	1445.348
_Isegmento_3	.0192837	.1868698	0.10	0.918	-.3469745	.3855418
_Isegmento_4	-.7049277	.1983862	-3.55	0.000	-1.093758	-.3160979
_Isegmento_5	-.5423917	.2023154	-2.68	0.007	-.9389226	-.1458607
_espera	-.0234117	.0044156	-5.30	0.000	-.032066	-.0147573
_cons	.4567806	.2819115	1.62	0.105	-.0957557	1.009317
Promotor						
_Igender_co_2	-.0740685	.0991182	-0.75	0.455	-.2683366	.1201995
_edad	.0222569	.0038062	5.85	0.000	.0147969	.0297169
_Isegmento_2	13.38895	730.903	0.02	0.985	-1419.155	1445.933
_Isegmento_3	.254493	.1689136	1.51	0.132	-.0765715	.5855575
_Isegmento_4	-.6899248	.1774649	-3.89	0.000	-1.03775	-.3421
_Isegmento_5	-.7035198	.1827513	-3.85	0.000	-1.061706	-.3453338
_espera	-.0479308	.0040826	-11.74	0.000	-.0559326	-.039929
_cons	1.070479	.2520943	4.25	0.000	.5763835	1.564575

Logit multinomial (relative-risk ratios):

Multinomial logistic regression

Number of obs = 4,202

LR chi2(14) = 418.26

Prob > chi2 = 0.0000

Pseudo R2 = 0.0542

Log likelihood = -3647.5859

clasificacion	RRR	Std. err.	z	P> z	[95% conf. interval]	
Detractor	(base outcome)					
Neutral						
_Igender_co_2	.9977229	.1115114	-0.02	0.984	.8014467	1.242068
_edad	1.01074	.0043292	2.49	0.013	1.00229	1.01926
_Isegmento_2	363481.5	2.66e+08	0.02	0.986	0	.
_Isegmento_3	1.019471	.1905084	0.10	0.918	.7068233	1.470411
_Isegmento_4	.4941443	.0980314	-3.55	0.000	.3349555	.7289881
_Isegmento_5	.5813562	.1176173	-2.68	0.007	.3910489	.8642781
_espera	.9768603	.0043134	-5.30	0.000	.9684427	.985351
_cons	1.578982	.4451333	1.62	0.105	.9086859	2.743726
Promotor						
_Igender_co_2	.9286081	.0920419	-0.75	0.455	.7646504	1.127722
_edad	1.022506	.0038919	5.85	0.000	1.014907	1.030163
_Isegmento_2	652751.9	4.77e+08	0.02	0.985	0	.
_Isegmento_3	1.289808	.217866	1.51	0.132	.9262867	1.795992
_Isegmento_4	.5016138	.0890188	-3.89	0.000	.354251	.7102772
_Isegmento_5	.4948405	.0904327	-3.85	0.000	.3458654	.707984
_espera	.9531997	.0038915	-11.74	0.000	.9456029	.9608576
_cons	2.916777	.7353029	4.25	0.000	1.779591	4.780643

Note: _cons estimates baseline relative risk for each outcome.

(e) Calcular los efectos marginales.

Efectos marginales en Logit multinomial (detractor):

Marginal effects after mlogit

```
y = Pr(clasificacion==Detractor) (predict, pr outcome(1))
= .13172136
```

variable	dy/dx	Std. err.	z	P> z	[95% C.I.]	X
-----+-----						
_Igend~2*	.0062526	.01127	0.56	0.579	-.015827 .028332	.678486
edad	-.0021919	.0012	-1.83	0.067	-.004541 .000157	52.2109
_Isegm~2*	-.1331684	.00569	-23.41	0.000	-.144317 -.12202	.000952
_Isegm~3*	-.0220274	.02219	-0.99	0.321	-.065524 .021469	.567587
_Isegm~4*	.0931274	.05089	1.83	0.067	-.006608 .192863	.183246
_Isegm~5*	.0890482	.04974	1.79	0.073	-.008432 .186529	.148263
espera	.0047328	.00246	1.92	0.055	-.000097 .009562	11.1349
-----+-----						

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Efectos marginales en Logit multinomial (neutral):

Marginal effects after mlogit

```
y = Pr(clasificacion==Neutral) (predict, pr outcome(2))
= .23194672
```

variable	dy/dx	Std. err.	z	P> z	[95% C.I.]	X
-----+-----						
_Igend~2*	.01049	.01435	0.73	0.465	-.017644 .038624	.678486
edad	-.001382	.00072	-1.91	0.056	-.002801 .000037	52.2109
_Isegm~2*	-.0628502	.1635	-0.38	0.701	-.383304 .257604	.000952
_Isegm~3*	-.034214	.02341	-1.46	0.144	-.08009 .011662	.567587
_Isegm~4*	-.02724	.02669	-1.02	0.307	-.079548 .025068	.183246
_Isegm~5*	.0021924	.02992	0.07	0.942	-.056453 .060838	.148263
espera	.0029036	.00123	2.37	0.018	.000501 .005306	11.1349
-----+-----						

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Efectos marginales en Logit multinomial (promotor):

Marginal effects after mlogit

```
y = Pr(clasificacion==Promotor) (predict, pr outcome(3))
= .63633192
```

variable	dy/dx	Std. err.	z	P> z	[95% C.I.]	X
-----+-----						
_Igend~2*	-.0167426	.01648	-1.02	0.310	-.049037 .015551	.678486
edad	.0035739	.0009	3.99	0.000	.001817 .005331	52.2109
_Isegm~2*	.1960187	.16356	1.20	0.231	-.124551 .516589	.000952
_Isegm~3*	.0562415	.02657	2.12	0.034	.004172 .108311	.567587
_Isegm~4*	-.0658873	.04487	-1.47	0.142	-.153828 .022054	.183246
_Isegm~5*	-.0912406	.04211	-2.17	0.030	-.173769 -.008712	.148263
espera	-.0076364	.0016	-4.77	0.000	-.010776 -.004497	11.1349
-----+-----						

(*) dy/dx is for discrete change of dummy variable from 0 to 1

(f) Repetir el análisis con un Probit multinomial y comparar.

Probit multinomial:

Multinomial probit regression

Number of obs = 4,202

Wald chi2(14) = 416.83

Log likelihood = -3635.6144

Prob > chi2 = 0.0000

clasificacion	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
Detractor	(base outcome)					
Neutral						
_Igender_co_2	-.0536078	.0798336	-0.67	0.502	-.2100787	.1028631
_edad	.0092218	.003062	3.01	0.003	.0032205	.0152231
_Isegmento_2	-.352632	.5757431	-0.61	0.540	-1.481068	.7758037
_Isegmento_3	.0867023	.1308623	0.66	0.508	-.1697831	.3431876
_Isegmento_4	-.7015738	.1429056	-4.91	0.000	-.9816635	-.421484
_Isegmento_5	-.3109711	.1472973	-2.11	0.035	-.5996685	-.0222737
_espera	-.0138713	.0033331	-4.16	0.000	-.020404	-.0073386
_cons	.2119086	.2034099	1.04	0.298	-.1867675	.6105848
Promotor						
_Igender_co_2	-.097611	.0738029	-1.32	0.186	-.242262	.0470399
_edad	.012833	.002822	4.55	0.000	.007302	.018364
_Isegmento_2	-1.411541	.6475008	-2.18	0.029	-2.680619	-.1424626
_Isegmento_3	.2629534	.1220016	2.16	0.031	.0238348	.5020721
_Isegmento_4	-.6144694	.1313595	-4.68	0.000	-.8719294	-.3570095
_Isegmento_5	-.4984651	.1378136	-3.62	0.000	-.7685749	-.2283554
_espera	-.0350071	.0031476	-11.12	0.000	-.0411763	-.0288379
_cons	1.035228	.1878494	5.51	0.000	.6670502	1.403406

Efectos marginales en Probit multinomial (detractor):

Marginal effects after mprobit

y = Pr(clasificacion==Detractor) (predict, pr outcome(1))
= .13404784

variable	dy/dx	Std. err.	z	P> z	[95% C.I.]		X
_Igend~2*	.0136968	.01125	1.22	0.223	-.008346	.03574	.677297
_edad	-.0019418	.00044	-4.41	0.000	-.002806	-.001078	52.1844
_Isegm~2*	.2216672	.14863	1.49	0.136	-.06965	.512984	.002618
_Isegm~3*	-.0345801	.01966	-1.76	0.079	-.07312	.00396	.578058
_Isegm~4*	.1251906	.02726	4.59	0.000	.071753	.178628	.183484
_Isegm~5*	.0823211	.02707	3.04	0.002	.02926	.135382	.140171
_espera	.0046788	.00048	9.74	0.000	.003737	.005621	11.1349

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Efectos marginales en Probit multinomial (neutral):

Marginal effects after mprobit

y = Pr(clasificacion==Neutral) (predict, pr outcome(2))
= .23112599

variable	dy/dx	Std. err.	z	P> z	[95% C.I.]		X
_Igend~2*	.0048601	.01424	0.34	0.733	-.023054	.032774	.677297
_edad	-.00013	.00055	-0.24	0.812	-.0012	.00094	52.1844
_Isegm~2*	.1369606	.15124	0.91	0.365	-.159457	.433378	.002618
_Isegm~3*	-.0261368	.0229	-1.14	0.254	-.071021	.018747	.578058
_Isegm~4*	-.0571084	.02392	-2.39	0.017	-.103993	-.010224	.183484
_Isegm~5*	.0120419	.02754	0.44	0.662	-.041943	.066027	.140171
_espera	.0029577	.00066	4.52	0.000	.001674	.004242	11.1349

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Efectos marginales en Probit multinomial (promotor):

Marginal effects after mprobit

```
y = Pr(clasificacion==Promotor) (predict, pr outcome(3))
= .63482617
```

variable	dy/dx	Std. err.	z	P> z	[95% C.I.]	X
-----+-----								
_Igend~2*	-.0185569	.01634	-1.14	0.256	-.05058	.013466		.677297
edad	.0020718	.00063	3.31	0.001	.000845	.003298		52.1844
_Isegm~2*	-.3586278	.15141	-2.37	0.018	-.655388	-.061868		.002618
_Isegm~3*	.0607169	.0265	2.29	0.022	.008771	.112663		.578058
_Isegm~4*	-.0680822	.0313	-2.18	0.030	-.12943	-.006734		.183484
_Isegm~5*	-.0943629	.03271	-2.88	0.004	-.158476	-.03025		.140171
espera	-.0076366	.00076	-10.01	0.000	-.009132	-.006141		11.1349
-----+-----								

(*) dy/dx is for discrete change of dummy variable from 0 to 1

(g) Realizar un test de la significatividad de las variables.

Stata.

Ejercicio 3.

Utilizando la EPH del cuarto trimestre de 2016, estimar un modelo multinomial que permita predecir la condición de actividad de una persona, entre inactivo, ocupado o desocupado.

Stata.