

# Soporte examen

Modelos de elección discreta y modelos probabilísticos

**Profesor: Pedro Araneda**

Auxiliares: R. Cerda, G. Mora, A. Muñoz y R. Tiara

## Pregunta 1 - Modelos de elección discreta

Un grupo de ex-alumn-s del IN5602 que hoy se han dedicado a emprender están considerando crear un servicio de turismo basado en la pesca recreativa. Para entender el comportamiento de l-s clientes y así poder planificar las estregias de venta, una integrante de este grupo de emprendedores implementó una encuesta para entender las preferencias de l-s clientes. La encuesta fue contestada por una muestra de 1182 personas quienes eligieron su modalidad de pesca preferido entre 4 modos distintos. La encuesta incluye la opción escogida (beach, pier, boat, charter) el precio por alternativa de modo de pesca (price), la tasa de pesca por alternativa (catch) y el ingreso de cada individuo (income). El modelo de utilidad observable de los individuos puede ser descrita de la siguiente manera:

$$V_{i,j} = \beta_1 * Precio_{i,j} + \beta_2 * Income_i + \beta_3 * Catch_{i,j}$$

1. Escriba la Log-verosimilitud de un modelo logit.
2. Según los siguientes coeficientes y valores por alternativa, calcule la probabilidad de que una persona que gana 1.000.000 elija la opción pier.

price.beach	price.pier	price.boat	price.charter	catch.beach	catch.pier	catch.boat	catch.charter	income
157	157	157	182	0.06	0.05	0.2	0.5	7000

```
Call:
mlogit(formula = mode ~ price | income | catch, data = Fish,
        method = "nr")

Frequencies of alternatives:choice
  beach  boat charter  pier
0.11337 0.35364 0.38240 0.15059

nr method
7 iterations, 0h:0m:0s
g'(-H)^-1g = 2.54E-05
successive function values within tolerance limits

Coefficients :
              Estimate Std. Error z-value Pr(>|z|)
(Intercept):boat  8.4184e-01  2.9996e-01  2.8065 0.0050080 **
(Intercept):charter 2.1549e+00  2.9746e-01  7.2443 4.348e-13 ***
(Intercept):pier  1.0430e+00  2.9535e-01  3.5315 0.0004132 ***
price            -2.5281e-02  1.7551e-03 -14.4046 < 2.2e-16 ***
income:boat      5.5428e-05  5.2130e-05  1.0633 0.2876612
income:charter  -7.2337e-05  5.2557e-05 -1.3764 0.1687088
income:pier     -1.3550e-04  5.1172e-05 -2.6480 0.0080977 **
catch:beach     3.1177e+00  7.1305e-01  4.3724 1.229e-05 ***
catch:boat      2.5425e+00  5.2274e-01  4.8638 1.152e-06 ***
catch:charter   7.5949e-01  1.5420e-01  4.9254 8.417e-07 ***
catch:pier      2.8512e+00  7.7464e-01  3.6807 0.0002326 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood: -1199.1
McFadden R^2: 0.19936
Likelihood ratio test : chisq = 597.16 (p.value = < 2.22e-16)
```

- Interprete los coeficientes del modelo
- No satisfechos con los resultados del modelo anterior, el grupo decide hacer un modelo mixed-logit. Comente los resultados y compárelos.

```
Call:
mlogit(formula = mode ~ price | income | catch, data = Fish,
        rpar = c(`(Intercept):boat` = "n", `(Intercept):charter` = "n",
                `(Intercept):pier` = "n", price = "n",
                `income:boat` = "n", `income:charter` = "n",
                `income:pier` = "n", `catch:boat` = "n",
                `catch:charter` = "n", `catch:pier` = "n",
                `catch:beach` = "n"), correlation = FALSE)

Frequencies of alternatives:choice
  beach boat charter pier
0.11337 0.35364 0.38240 0.15059

bfgs method
4 iterations, 0h:0m:22s
g'(-H)^-1g = -3.8E+05
last step couldn't find higher value

Coefficients :
                Estimate Std. Error z-value Pr(>|z|)
(Intercept):boat  -1.5089e+02  2.8971e-01 -520.8380 < 2.2e-16 ***
(Intercept):charter -1.5441e+02  3.5344e-01 -436.8713 < 2.2e-16 ***
(Intercept):pier   -1.6625e+02  6.1991e-01 -268.1768 < 2.2e-16 ***
price              -1.4201e-02  7.6941e-04 -18.4576 < 2.2e-16 ***
income:boat        1.7013e-05  2.7108e-05  0.6276  0.53027
income:charter     -2.7678e-04  4.4036e-05 -6.2852  3.274e-10 ***
income:pier        -2.4199e-03  7.3552e-05 -32.8999 < 2.2e-16 ***
catch:beach        -6.1368e+02  9.5911e+00 -63.9845 < 2.2e-16 ***
catch:boat         3.5255e+00  3.0017e-01  11.7453 < 2.2e-16 ***
catch:charter      6.1963e-01  1.0667e-01  5.8088  6.291e-09 ***
catch:pier         -6.1459e+00  1.1445e+00 -5.3700  7.872e-08 ***
sd.(Intercept):boat -1.6447e-01  1.7513e-01 -0.9391  0.34766
sd.(Intercept):charter -3.5573e-01  1.8687e-01 -1.9036  0.05696 .
sd.(Intercept):pier  9.1900e-01  4.1182e-01  2.2315  0.02565 *
sd.price           9.7907e-02  1.3117e-03  74.6424 < 2.2e-16 ***
sd.income:boat     1.0126e-01  6.4443e-05 1571.3162 < 2.2e-16 ***
sd.income:charter  1.0274e-01  7.8987e-05 1300.7819 < 2.2e-16 ***
sd.income:pier     1.0987e-01  1.7808e-04 616.9420 < 2.2e-16 ***
sd.catch:beach    -4.8204e+02  6.1681e+00 -78.1510 < 2.2e-16 ***
sd.catch:boat     -1.0889e+00  5.1103e-01 -2.1308  0.03310 *
sd.catch:charter  -9.7577e-01  2.0395e-01 -4.7845  1.714e-06 ***
sd.catch:pier     9.5216e+00  1.9370e+00  4.9155  8.854e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood: -1919.3
McFadden R^2: -0.2815
Likelihood ratio test : chisq = -843.22 (p.value = 1)
```

## Pregunta 2 - Modelos probabilísticos

*TikTok* fue la aplicación más descargada durante el primer trimestre del 2020 a nivel mundial. El Home de su App consiste en un único módulo que muestra sin interrupción, y uno tras otro, videos cortos publicados por otros usuarios. En cualquier momento el usuario puede deslizar el video fuera de la pantalla para pasar al siguiente video. El modelo de negocios de *TikTok* se beneficia mientras los usuarios pasen más tiempo usando la App. *TikTok* realizó un proyecto experimental para que los usuarios pasaran más tiempo viendo videos en una sesión, pero sospecha que hubo un efecto secundario al implementar el proyecto y no está seguro de cómo afecta. Si bien el proyecto podría hacer que los usuarios pasen más tiempo por sesión, *TikTok* no está seguro si es que esto afectaría la **cantidad de veces** que un usuario ingresa a la App. Por ejemplo, el ingreso de un usuario

podría aumentar si los videos mostrados son de mayor interés que antes o disminuir si los usuarios sienten que la aplicación les consume más tiempo que antes. Gracias a tu experiencia con modelos probabilísticos, *TikTok* te ha llamado para evaluar dicho fenómeno. Para esto le ha proporcionado la siguiente información:

Tabla 1: Descripción de variables

Id	Identificador del usuario
Mujer	1 si usuari@ es mujer, 0 si es hombre
Heavy	1 si el usuario es frecuente en la aplicación, 0 si es casual
Proyecto	1 si el usuario fue parte del proyecto de TikTok implementado, 0 si no fue parte
Heavy* Proyecto	Variable de interacción entre Heavy y Proyecto
Covariables	Conjunto (vector) de variables de control de características sociodemográficas del usuario

- Utilizando la información de la Tabla 1. Escriba la log-verosimilitud de cuatro modelos probabilísticos que modelen el fenómeno de interés. Un modelo debe ser homogéneo, otro heterogéneo continuo, otro heterogéneo observable y, el último, un modelo híbrido.

Observe las siguientes estimaciones de distintos modelos en la siguiente tabla:

	Modelo Poisson (1)	Modelo Poisson-Gamma (2)	Regresión Poisson (3)	Modelo Binomial Negativo (4)
$\lambda$	0.008 (0.001)	-	0.0439 (0.01)	-
$\alpha$	-	16.4 (3.5)	-	8.1979 (2.7)
$r$	-	0.10 (0.009)	-	0.1388 (0.3)
$\beta_{Mujer}$	-	-	0.0043 (0.006)	0.0093 (0.0065)
$\beta_{Heavy}$	-	-	0.243 (0.09)	0.222 (0.01)
$\beta_{Proyecto}$	-	-	0.15 (0.02)	0.13 (0.02)
$\beta_{Heavy*Proyecto}$	-	-	0.05 (0.08)	0.10 (0.03)
$\vec{\delta}_{covariables}$	-	-	✓ <sup>1</sup>	✓
<i>AIC</i>	3760	3659	3658	3541
<i>BIC</i>	3850	3700	3699	3680
<i>RMSE</i>	4.5	4.2	4.2	2.45

- Para explicar el fenómeno de interés ¿Qué desventajas presenta el modelo (1) respecto a los restantes? Indique el supuesto sobre los usuarios que haría que este modelo fuera mejor por sobre otros
- Según el modelo (3) ¿Cuál es el efecto del proyecto sobre un usuario Heavy que fue parte del proyecto? ¿y sobre un usuario casual con proyecto? ¿Hubiera esperado que estos fueran así? Recuerde interpretar según magnitud y significancia.
- ¿Qué modelo usaría para predecir la frecuencia de los usuarios de la plataforma?

5. Concluya sobre los resultados del proyecto implementado por TikTok en esta pregunta.