3. REGRESIÓN LOGÍSTICA MULTINOMIAL

REGRESION LOGISTICA MULTINOMIAL

La regresión logística Multinomial es un modelo de regresión para variables dependientes o de respuesta con tres o más categorías con escala de medición Nominal.

Si tenemos una variable respuesta Y categórica con j categorías, la distribución de probabilidad de las categorías se puede representar a través de una Distribución Multinomial.

REGRESION LOGISTICA MULTINOMIAL

El Modelo logístico Multinomial (llamado también Multicategórico o *Politómico)* para variables nominales utiliza la transformación logit, describiendo de manera simultanea el logaritmo del momio para todas las *j* categorías.

Se seleccionan de manera conveniente *j*-1 de las categorías, la que sobra es redundante (es el complemento del conjunto).

Categoría Base:

Estos modelos crean pares de cada una de las j-1 categorías con una categoría de base seleccionada, comúnmente la última o la más frecuente.

Como ejemplo, si tomamos la última categoría (J) como la base, entonces la transformación logit es:

$$log\left(\frac{\pi_j}{\pi_J}\right), \qquad j = 1,\ldots,J\text{-}1$$

Que es el logaritmo del momio de que la categoría correspondiente sea j.

Para el caso de J = 3, el modelo utiliza:

$$log\left(\frac{\pi_1}{\pi_3}\right) \text{y} \qquad log\left(\frac{\pi_2}{\pi_3}\right)$$

De modo que el modelo general con **Categoría Base** es:

$$log\left(\frac{\pi_j}{\pi_J}\right) = \alpha_j + \beta_j x, \quad j = 1,\ldots,J\text{-}1 \tag{1}$$

De modo que se obtienen J-1 ecuaciones con parámetros independientes.

Los efectos varían según la categoría que se contrasta con la Categoría Base.

Las ecuaciones para los pares de categorías en (1), determinan ecuaciones para todos los pares de categorías posibles. Por ejemplo, para un para arbitrario de categorías a y b, tenemos que:

$$log\left(\frac{\pi_a}{\pi_b}\right)=log\left(\frac{\pi_a/\pi_J}{\pi_b/\pi_J}\right)=log\left(\frac{\pi_a}{\pi_J}\right)-log\left(\frac{\pi_b}{\pi_J}\right)$$

$$=(\alpha_a+\beta_a x)-(\alpha_b+\beta_b x)$$

$$=(\alpha_a-\alpha_b)+(\beta_a-\beta_b)x$$

De modo que la ecuación para a y b es de forma $\alpha+\beta x$ donde la ordena al origen es $\alpha=(\alpha_a+\alpha_b)$ y la pendiente es $\beta=(\beta_a+\beta_b)$

ESTIMACION DE PARAMETROS

Dado que el procedimiento de estimación no tiene una forma analítica cerrada, los paquetes estadísticos utilizan la Función de Verosimilitud y ajustan los parámetros con un procedimiento iterativo (de aproximaciones sucesivas) conocido como: Algoritmo Newton–Raphson

En la mayoría de los casos el procedimiento converge rápidamente a valores de parámetros que Maximizan la Función de Verosimilitud

ESTIMACION DE PARAMETROS

Las estimaciones de parámetros en modelos logísticos multinomiales tienen errores de estimación mas pequeños que cuando se ajustan modelos logísticos binarios para cada componente de la ecuación (1) por separado.

En estimaciones simultaneas, se obtienen los mismos parámetros para cada par de categorías sin importar que categoría se utilizó como base.

La elección de la categoría base es arbitraria, por lo que es recomendable escogerla según la interpretación que se busca.

ESTIMACION DE LAS PROBABILIDADES DE RESPUESTA

El modelo logístico multinomial se puede representar alternativamente en términos de la probabilidad de respuesta con la siguiente expresión:

$$\pi_j = \frac{e^{\alpha_j + \beta_j x}}{1 + \sum_{j=1}^{J} e^{\alpha_j + \beta_j x}}, \quad j = 1,\ldots,J\text{-}1 \qquad (2)$$

El denominador es el mismo para cada probabilidad y el numerador son los sumandos del denominador por lo que:

$$\sum_{j=1}^{J} \pi_j = 1$$

Ejemplo:

ESTUDIO SOBRE PREFERENCIAS ALIMENTICIAS EN LAGARTOS

Se tomó una muestra de 59 lagartos del lago George en Florida.

Se analizó el contenido estomacal de los lagartos y se clasificó la preferencia alimenticia en tres grupos de especies.

- En éste estudio se pretende conocer si la talla de los lagartos (longitud total en metros) determina la preferencia alimenticia de estos depredadores en su estado natural.

Los grandes grupos alimenticios son:

- Peces
- Invertebrados
- Otros (incluye aves y reptiles)

```
. mlogit  alim longitud, baseout(3)

Iteration 0:   log likelihood = -57.570928
Iteration 1:   log likelihood =  -49.97414
Iteration 2:   log likelihood = -49.186349
Iteration 3:   log likelihood = -49.170647
Iteration 4:   log likelihood = -49.170622
Iteration 5:   log likelihood = -49.170622

Multinomial logistic regression                 Number of obs     =         59
                                                LR chi2(2)        =      16.80
                                                Prob > chi2       =     0.0002
Log likelihood = -49.170622                     Pseudo R2         =     0.1459
```

alim	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Peces						
longitud	-.110109	.517082	-0.21	0.831	-1.123571	.9033531
_cons	1.617731	1.307275	1.24	0.216	-.9444801	4.179943
Invertebrados						
longitud	-2.465446	.8996503	-2.74	0.006	-4.228728	-.702164
_cons	5.697444	1.793808	3.18	0.001	2.181644	9.213244
Otros	(base outcome)					

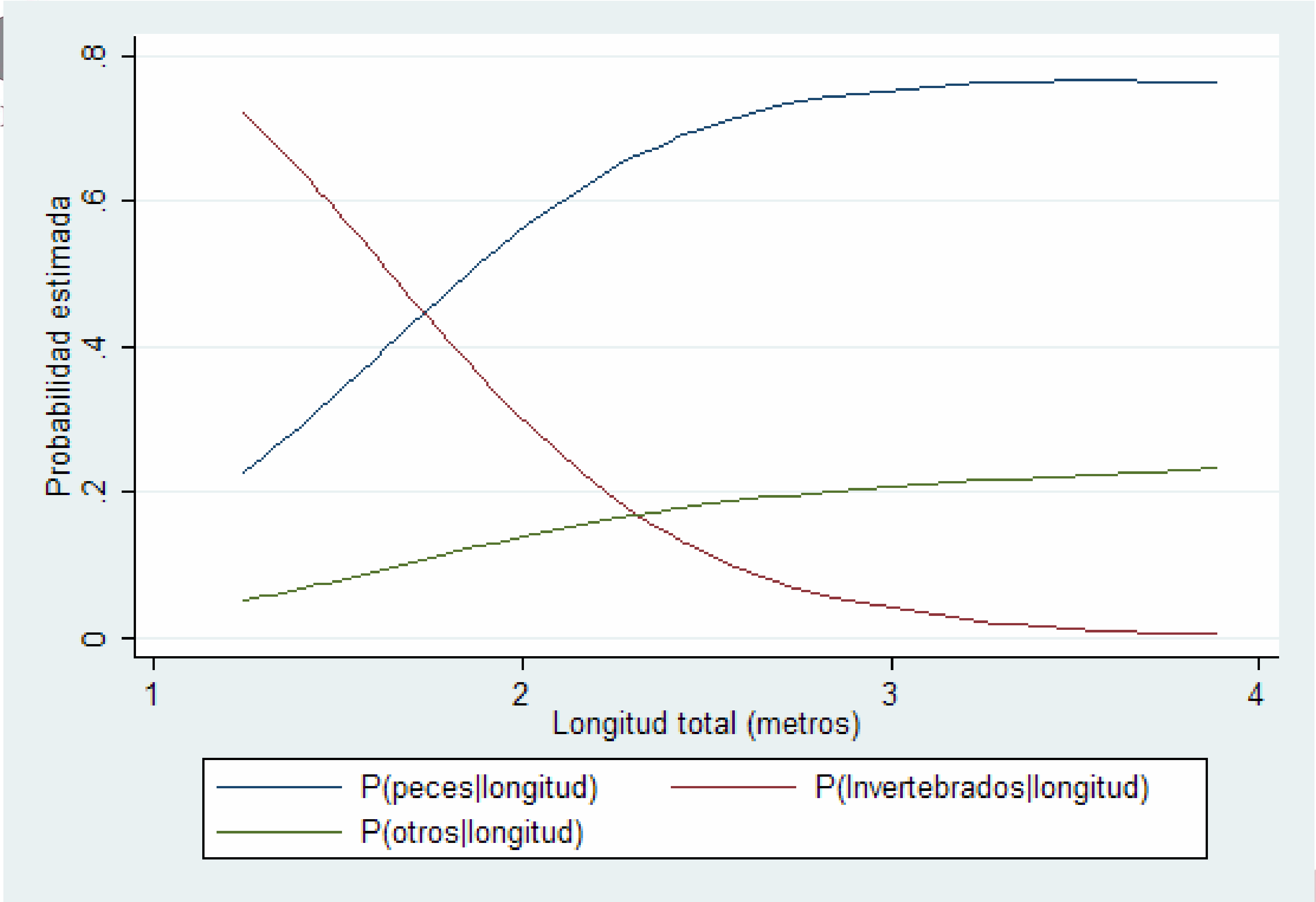
ESTIMACION DE PROBABILIDADES

A partir de la estimación de parámetros del modelo, se estiman las probabilidades de cada grupo alimenticio como función de la longitud del lagarto, esto es:

$$\hat{\pi}_P = \frac{e^{1.62+-.11 \cdot x}}{1+e^{1.62+-.11 \cdot x}+e^{5.69+-2.46 \cdot x}}$$

$$\hat{\pi}_I = \frac{e^{5.69+-2.46 \cdot x}}{1+e^{1.62+-.11 \cdot x}+e^{5.69+-2.46 \cdot x}}$$

$$\hat{\pi}_0 = \frac{1}{1+e^{1.62+-.11 \cdot x}+e^{5.69+-2.46 \cdot x}}$$



Probabilidad estimada
0
.2
.4
.6
.8
1
2
3
4
Longitud total (metros)
P(peces|longitud)
P(Invertebrados|longitud)
P(otros|longitud)