4. REGRESIÓN LOGÍSTICA ORDINAL

Regresión Logística Ordinal

- La **regresión logística Ordinal** es un modelo de regresión para variables dependientes o de respuesta con tres o más categorías con un orden específico, de modo que las categorías implican cambios graduales en los niveles de la variable en cuestión.

Modelos Logit para Respuestas Ordinales

- La variable respuesta es ordinal (categórica con orden natural)
- La variable(s) predictivas pueden ser numéricas o categóricas (variables dummy)
- Etiquetando las categorías ordinales de 1 (nivel inferior) a *c* (superior), podremos obtener las **probabilidades acumuladas**:

$$P(Y \leq j) = P(Y = 1) + \cdots + P(Y = j) \quad j = 1, \ldots, c$$

Regresión Logística Ordinal

El Momio (odd) para categoría *j* o inferior es:

$$\frac{P(Y \le j)}{P(Y > j)} \quad j = 1, \ldots, c-1 \qquad P(Y \le c) = 1$$

El Logit (logaritmo del momio) de las probabilidades acumuladas se modelan como funciones lineales de las variable(s) predictivas:

$$\operatorname{logit}\left[P(Y \le j)\right] = \log\left[\frac{P(Y \le j)}{P(Y > j)}\right] = \alpha_j - \beta X \quad j = 1, \ldots, c-1$$

Esto se llama **Modelo de Momios Proporcionales**, el cual supone que el efecto de las *X´s* es el mismo para cada probabilidad acumulada (supuesto de líneas paralelas).

Ejemplo conceptual

Si se mide la satisfacción de los electores sobre el sistema de justicia penal, usando las opciones de respuesta: mala(1), regular(2), buena(3) y excelente(4).

Para el ejemplo las probabilidades son:

$$P(Y \leq 1) \big/ P(Y > 1) = \text{ Probabilidad de mala}$$

$$P(Y \leq 2) \big/ P(Y > 2) = \text{ Probabilidad de mala o regular}$$

$$P(Y \leq 3) \big/ P(Y > 3) = \text{ Probabilidad de mala o regular o buena}$$

Se modela la categoría inferior respecto a las superiores, lo que explica el signo negativo en el predictor lineal, de modo que un coeficiente positivo implica menor probabilidad de la categoría inferior y mayor probabilidad de la o las superiores.

Regresión Logística Ordinal Generalizada

El supuesto de líneas paralelas no se cumple en un gran número de caos, por lo que se utilizan modelos de Regresión Logística Ordinal Generalizada:

$$\operatorname{logit}\left[P(Y \leq j)\right] = \log\left[\frac{P(Y \leq j)}{P(Y > j)}\right] = \alpha_j - \beta_j X \qquad j = 1, \ldots, c-1$$

Las funciones lineales de las variables predictivas no requieren cumplir el supuesto de paralelismo, por lo que se estiman funciones lineales de las variables predictivas para cada probabilidad acumulada ordinal.

Otra alternativa es usar **Modelos de Momios Proporcionales Parciales**, el cual supone que el efecto de algunas de las *X´s* es el mismo para cada probabilidad acumulada y para otras no.

Ejemplo - Actitudes Respecto a la Renovación Urbana

- Respuesta: Actitud respecto a un proyecto de renovación urbana (Negativa (Y=1), Moderada (Y=2), Positiva (Y=3))
- Variable predictiva: Raza del encuestado (Blanco, No Blanco)

Tabla de Contingencia:

Actitud\Raza	Blanco	No Blanco
Negativa (Y=1)	101	106
Moderada (Y=2)	91	127
Positiva (Y=3)	170	190

Ajuste del modelo con el paquete Rcmdr:

$$\operatorname{logit}[P(Y \le j)] = \log\left[\frac{P(Y \le j)}{P(Y > j)}\right] = \alpha_j - \beta X \qquad j = 1, \ldots, c-1$$

```
> summary(OrdRegModel.3)
Call:
polr(formula = Actitud ~ Raza, data = Actitudes, Hess = TRUE, method = "logistic")

Coefficients:
                   Value Std. Error   t value
Raza[T.Blanco] -0.001159     0.1337 -0.008673

Intercepts:
                  Value    Std. Error t value
Negativa|Moderada  -1.0274   0.1010   -10.1694
Moderada|Positiva   0.1655   0.0937     1.7664

Residual Deviance: 1671.743
AIC: 1677.743
```

Nótese que el efecto de Raza no es significativo.

La ecuación ajustada para cada grupo/categoría es:

$$\text{Negativa/Blanco: logit}\left[\frac{P(Y\leq 1|\text{Blanco})}{P(Y>1|\text{Blanco})}\right]=-1.027-(-0.001)=-1.026$$

$$\text{Negativa/No Blanco: logit}\left[\frac{P(Y\leq 1|\text{No Blanco})}{P(Y>1|\text{No Blanco})}\right]=-1.027$$

$$\text{Neg o Mod/Blanco: logit}\left[\frac{P(Y\leq 2|\text{Blanco})}{P(Y>2|\text{Blanco})}\right]=0.165-(-0.001)=0.166$$

$$\text{Neg o Mod/No Blanco: logit}\left[\frac{P(Y\leq 2|\text{No Blanco})}{P(Y>2|\text{No Blanco})}\right]=0.165$$

Para cada grupo, la probabilidad ajustada de estar en ese conjunto de categorías es $e^L/(1+e^L)$ donde L ese el valor logit estimado:

$P(Y\leq 1 \mid \text{Blanco}) = 0.264$, $P(Y\leq 1 \mid \text{No Blanco}) = 0.264$,

$P(Y\leq 2 \mid \text{Blanco}) = 0.541$, $P(Y\leq 2 \mid \text{No Blanco}) = 0.541$

Inferencia para los Coeficientes de Regresión

- Si $b = 0$, la respuesta (Y) es independiente de X
- Se pueden utilizar pruebas de Z (el estimador dividido entre su error estándar)
- La mayoría de los programas calculan la prueba de Wald, con el estadístico como Z-cuadrada, la cual tiene distribución Ji-cuadrada con 1 grado de libertad bajo la hipótesis de nulidad de efectos.
- Razones de momios para el incremento de X en 1 y su intervalo de confianza se obtienen usando el coeficiente de regresión como exponente con base e *(función inversa del logaritmo natural),* así como los límites inferiores y superiores de los intervalos.

Variables Predictivas Ordinales

- Crear variables dummy para categorías ordinales les confiere tratamiento como nominales
- Para crear una variable ordinal, se requiere una variable X numérica que represente los niveles de la variable ordinal.
- Cuando se tienen 6 o mas categorías ordinales, funciona bien usar la numeración de las categorías como variable numérica, pero 5 categorías o menos es cuestionable.

- La elección de los niveles depende de la definición de las categorías ordinales. La forma simple es $X=1, 2,...,c$ para las categorías ordenadas, lo que implica que los niveles de las categorías son equi-espaciados.

- Cuando los niveles de las categorías no son equi-espaciados, se puede utilizar números que representen de manera aproximada la dimensión de las distancias entre las categorías ordenadas.