Profesor: Graciela Martínez Sánchez, graciela.mtz@ciencias.unam.mx
Ayudante: José de Jesús Ojeda Gónzalez, jesusojeda@ciencias.unam.mx
Alumno: Cristobal Bautista Hernández.
Fecha de entrega: jueves 30 de abril de 2020.

Instrucciones: La tarea se entrega de manera individual, ordenada y bien escrita. La solución de la tarea debe enviarse en un archivo `.tex`

1. Considere que se ajustó un modelo multinomial logístico para predecir las preferencias para presidente (demócrata, repúblicanos e independientes), en función del ingreso anual x (en dólares):

$$\log\left(\frac{\hat{\pi}_D}{\hat{\pi}_I}\right) = 3.3 - 0.2x, \qquad \log\left(\frac{\hat{\pi}_R}{\hat{\pi}_I}\right) = 1.0 + 0.3x.$$

(i) Encuentre una expresión para $\log\left(\frac{\hat{\pi}_D}{\hat{\pi}_R}\right)$. Interprete la pendiente en la ecuación.
(ii) Encuentre el rango de x para el cual $\hat{\pi}_D > \hat{\pi}_R$.
(iii) Obtenga una expresión para la probabilidad $\hat{\pi}_I$.

(i) *Solución:* Observe que $\log\left(\frac{\pi_i}{\pi_j}\right) = \log(\left(\frac{\pi_i}{\pi_k}\right)\left(\frac{\pi_k}{\pi_j}\right)) = log\left(\frac{\pi_i}{\pi_k}\right) + log\left(\frac{\pi_k}{\pi_j}\right) = log\left(\frac{\pi_i}{\pi_k}\right) - log\left(\frac{\pi_j}{\pi_k}\right)$ para algún $k = 1, 2, \ldots, J$
De esta manera, aplicando a nuestro problema:

$$\log\left(\frac{\widehat{\pi}_D}{\widehat{\pi}_R}\right) = \log\left(\frac{\widehat{\pi}_D}{\widehat{\pi}_I}\right) - \log\left(\frac{\widehat{\pi}_R}{\widehat{\pi}_I}\right) = 3.3 - 0.2x - (1.0 + 0.3x) = 2.3 - 0.5x \Rightarrow \log\left(\frac{\widehat{\pi}_D}{\widehat{\pi}_R}\right) = 2.3 - 0.5x$$

De está ecuación se puede decir que entre mayor sea el ingreso anual dolares para una persona, está tendrá con mayor probabilidad a preferir a un presidente republicano que a un demócrata. □

(ii) *Solución:* Utilizando la hipótesis y el hecho de que la función *log* es monótona creciente.

$$\widehat{\pi}_D > \widehat{\pi}_R \Leftrightarrow \widehat{\pi}_D > \widehat{\pi}_R\left(\frac{\widehat{\pi}_I}{\widehat{\pi}_I}\right) \Leftrightarrow \log\left(\widehat{\pi}_D\right) > \log\left(\widehat{\pi}_I\left(\frac{\widehat{\pi}_R}{\widehat{\pi}_I}\right)\right) = log\left(\widehat{\pi}_I\right) + log\left(\frac{\widehat{\pi}_R}{\widehat{\pi}_I}\right) \Leftrightarrow \log\left(\widehat{\pi}_D\right) - log\left(\widehat{\pi}_I\right) > log\left(\frac{\widehat{\pi}_R}{\widehat{\pi}_I}\right)$$

$$\Leftrightarrow log\left(\frac{\widehat{\pi}_D}{\widehat{\pi}_I}\right) > log\left(\frac{\widehat{\pi}_R}{\widehat{\pi}_I}\right) \Leftrightarrow 3.3 - 0.2x > 1.0 + 0.3x \Leftrightarrow 2.3 > 0.5x \Leftrightarrow 4.6 > x$$

Por lo tanto, el intervalo para x cuando $\widehat{\pi}_D > \widehat{\pi}_R$ es $x \in (-\infty, 4.6)$
De otra manera, este intervalo significa el ingreso anual en dolares de una persona con una mayor probabilidad de preferencia por un presidente demócrata que por un republicano □

(iii) *Solución:* Si se toma $\log\left(\frac{\widehat{\pi}_D}{\widehat{\pi}_I}\right) = 3.3 - 0.2x$ de la hipótesis. Esto lleva a que:

$$\frac{\widehat{\pi}_D}{\widehat{\pi}_I} = exp(3.3 - 0.2x) \Rightarrow \widehat{\pi}_D = \widehat{\pi}_I exp(3.3 - 0.2x) \Rightarrow \widehat{\pi}_I = \widehat{\pi}_D exp(0.2x - 3.3)$$

Por lo tanto, una expresión para $\widehat{\pi}_I$ es $\widehat{\pi}_I = \widehat{\pi}_D exp(0.2x - 3.3)$ □

2. Encuentre la devianza para el modelo logístico multinomial.

Solución: Observe que $f(y_i,\pi_i) = P[Y_{i1} = y_{i1}, \ldots, Y_{iJ} = y_{iJ}] = \pi_{i1}^{y_{i1}} \cdots \pi_{iJ}^{iJ}$. Entonces:

$$f(\overline{y},\overline{\pi}) = \prod_{i=1}^{n} f(y_i,\pi_i) = \prod_{i=1}^{n}\prod_{k=1}^{J}\pi_{ik}^{y_{ik}} \Rightarrow l(y,\pi) = log\left(\prod_{i=1}^{n}\prod_{k=1}^{J}\pi_{ik}^{y_{ik}}\right) = \sum_{i=1}^{n} log\left(\prod_{k=1}^{J}\pi_{ik}^{y_{ik}}\right) = \sum_{i=1}^{n}\sum_{k=1}^{J} log\left(\pi_{ik}^{y_{ik}}\right)$$

$$\Rightarrow l(y,\pi) = \sum_{i=1}^{n}\sum_{k=1}^{J} y_{ik} log\,(\pi_{ik})$$

$$\Rightarrow D = 2[l(y,\pi) - l(y,y)] = 2[\sum_{i=1}^{n}\sum_{k=1}^{J} y_{ik} log\,(\pi_{ik}) - \sum_{i=1}^{n}\sum_{k=1}^{J} y_{ik} log\,(y_{ik})] = \sum_{i=1}^{n}\sum_{k=1}^{J} y_{ik} log\left(\frac{\pi_{ik}}{y_{ik}}\right)$$

$\therefore D = \sum_{i=1}^{n}\sum_{k=1}^{J} y_{ik} log\left(\frac{\pi_{ik}}{y_{ik}}\right)$

□

3. Un modelo logístico acumulativo se ajustó a los datos de una encuesta social, donde se considera la ideología política (poco liberal a muy liberal) como respuesta, y variable explicativa la religión profesada (protestante, católicos, judíos y otros). Considerando variables dummys para las primeras tres religiones, el ajuste del modelo fue: $\hat{\alpha}_1 = -1.03$, $\hat{\alpha}_2 = -0.13$, $\hat{\alpha}_3 = 1.57$, $\hat{\alpha}_4 = 2.41$, $\hat{\beta}_1 = -1.27$, $\hat{\beta}_2 = -1.22$ y $\hat{\beta}_3 = -0.44$. Conteste lo siguiente

(i) ¿Cuántas categorías tiene la respuesta Y?. ¿Cuál grupo religioso se estima si se es a) el más liberal y b) el más conservador?.

(ii) Utilizando cociente de momios compare la ideología política para los grupos religiosos protestantes y católicos. Interprete.

(i) *Solución:* Y resulta tener 4 categorías derivadas de la región que profesan. De esta manera, $P[Y_i \leq j] = e^{\alpha_j + x_1\beta_1 + x_2\beta_2 + x_3\beta_3}$
Por lo cual, si se quiere saber el grupo religioso el cuál es más liberal, y el más conservador. Entonces, evaluando en cada categoría:

$$P[Y_i = 4] = P[Y_i \leq 4] - P[Y_i \leq 3] = e^{\alpha_4} - e^{\alpha 3 - \beta_3} = e^{2.41} - e^{1.57+(-0.44)} = e^{2.41} - e^{1.13} \approx 8.0383$$

$$P[Y_i = 3] = P[Y_i \leq 3] - P[Y_i \leq 2] = e^{\alpha_3+\beta_3} - e^{\alpha_2+\beta_2} = e^{1.57+(-0.44)} - e^{-0.13+(-1.22)} = e^{1.13} - e^{-1.35} \approx 2.8364$$

$$P[Y_i = 2] = P[Y_i \leq 2] - P[Y_i \leq 1] = e^{\alpha_2+\beta_2} - e^{\alpha_1+\beta_1} = e^{-0.13+(-1.22)} - e^{-1.03+(-1.27)} = e^{-1.35} - e^{-2.30} \approx 0.1589$$

$$P[Y_i = 1] = P[Y_i \leq 1] = e^{-2.30} \approx 0.1002$$

De lo cual, se puede notar que una persona más liberal se estima que profesa otra religión no señalada, mientras que los menos liberales serían las personas que profesan el Protestantismo.

□

(ii) *Solución:* Para encontrar el cociente de momios se necesita tener π_C y π_P, de lo cual se obtiene los siguiente:

$$\pi_C = P[Y_i = 2] = P[Y_i \leq 2] - P[Y_i \leq 1] = e^{\alpha_2+\beta_2} - e^{\alpha_1+\beta_1} = e^{-0.13-1.22} - e^{-1.03-1.27} = e^{-1.35} - e^{-2.30} \approx 0.1589$$

De igual manera se obtiene que:

$$\pi_P = P[Y_i = 1] = P[Y_i \leq 1] - P[Y_i \leq 0] = e^{-2.30} = 0.1002 \Rightarrow \pi_P \approx 0.1002$$

Entonces, el cociente de momios entre protestantes y católicos es:

$$\frac{\pi_P/(1-\pi_P)}{\pi_C/(1-\pi_C)} = \frac{0.1002/(1-0.1002)}{0.1590/(1-0.1590)} = \frac{0.1002/(0.8998)}{0.1590/(0.8410)} \approx \frac{0.1113}{0.1890} \approx 0.5888$$

Esto nos quiere decir que si se escoge una persona que es poco liberal hay hasta un 58% más de probabilidad de que sea protestante a que sea católico.

□

4. (R) La tabla presenta los resultados que tuvieron pasajeros en vehículos accidentados, de acuerdo con el género, el lugar del accidente y el uso de cinturón de seguridad. Las categorías de respuesta son: (1) no se lastimó, (2) lastimado, pero sin necesidad de ser transportado en ambulancia, (3) lastimado, transportado por ambulancia pero no hospitalizado, (4) lastimado, hospitalizado, pero sobrevivió y (5) muerto. Ajusta un modelo considerando la variable respuesta no ordenada y un modelo considerándola ordenada. Comenta tus resultados.

			respuesta				
género	localización	cinturón	1	2	3	4	5
femenino	urbana	no	7287	175	720	91	10
		sí	11587	126	577	48	8
	rural	no	3246	73	710	159	31
		sí	6134	94	564	82	17
masculino	urbana	no	10381	136	566	96	14
		sí	10969	83	259	37	1
	rural	no	6123	141	710	188	45
		sí	6693	74	353	74	12