

Untitled

Ana B. Coronel

15 de diciembre de 2019

Modelo jerárquico

Para los modelos jerárquicos, se considerarán dos casos, y como se menciona en la sección del Análisis Exploratorio de Datos, resulta de interés considerar al género como la variable de interés de agrupamiento por las diferencias observadas.

Para la construcción de los modelos jerárquicos, sea:

y_i la observación del individuo i donde $i \in \{1, \dots, 1242\}$. Representaremos el género del individuo i mediante el subíndice $j \in \{1, 2\}$ y X_i representa la i -ésima covariable.

Modelo de intercepto variable sin covariables:

El modelo sin covariables se utilizó para comparar las diferentes ligas y seleccionar a la de menor DIC para la construcción del modelo con de intercepto variable con pendientes fijas. Las ligas consideradas son: logística, probit, lognormal y cloglog.

La especificación del modelo es la siguiente:

$$Y_i \sim Be(\pi_i), i \in \{1, \dots, 1242\}$$

$$\pi_i = f(\theta_j), j \in \{1, 2\} \text{ y } f(\theta_j) \text{ se define como en la sección de mlg}$$

$$\theta_j \sim N(\phi, \lambda) \text{ con } \lambda \text{ conocida}$$

$$\phi \sim N(\mu_0, \lambda_0) \text{ ambos parámetros conocidos}$$

La distribución que se consideró para el hiperparámetro fue una distribución no informativa.

Para la construcción del modelo, consideraremos las cuatro ligas y se tomará aquella con el menor DIC.

liga	DIC
logística	1484
probit	1485
lognormal	1484
cloglog	1485

A partir de la tabla anterior, se podrían considerar dos ligas como las mejores -logística y lognormal- por ser las que menor DIC tienen. En este caso, por facilidad en la interpretación de los coeficientes, se escogerá la liga logística.

Modelo con intercepto variante y pendiente fija :

Con el objetivo de mejorar el modelo antes mencionado, se agregaron las siguientes variables explicativas:

- edad
- número de ventas

- antigüedad
- número de trabajos previos
- incremento

Y la especificación del modelo es la siguiente:

$$Y_i \sim Be(\pi_i), i \in \{1, \dots, 1242\}$$

$$\pi_i = \frac{e^{\alpha_j + X_i' \beta}}{1 + e^{\alpha_j + X_i' \beta}}$$

$$\alpha_j \sim N(\phi, \lambda) \text{ con } \lambda \text{ conocida}$$

$$\phi \sim N(\mu_0, \lambda_0) \text{ ambos parámetros conocidos}$$

El DIC que se obtiene con este modelo es 603.8. En comparación con el modelo sin covariables, mejoró bastante.

A continuación se muestran los intervalos de confianza para cada una de las variables antes mencionadas:

Variables	media	Desviación estándar	Límite inferior	Límite superior
número de ventas	-0.4269	0.02467	-0.47740	-0.37990
edad	-0.07566	0.03081	-0.13550	-0.015280
cant. trabajos previos	0.08900	0.06231	-0.03093	0.213400
antigüedad	0.0380	0.03331	-0.02823	0.10480
incremento	-0.00893	0.00705	-0.02518	0.002121

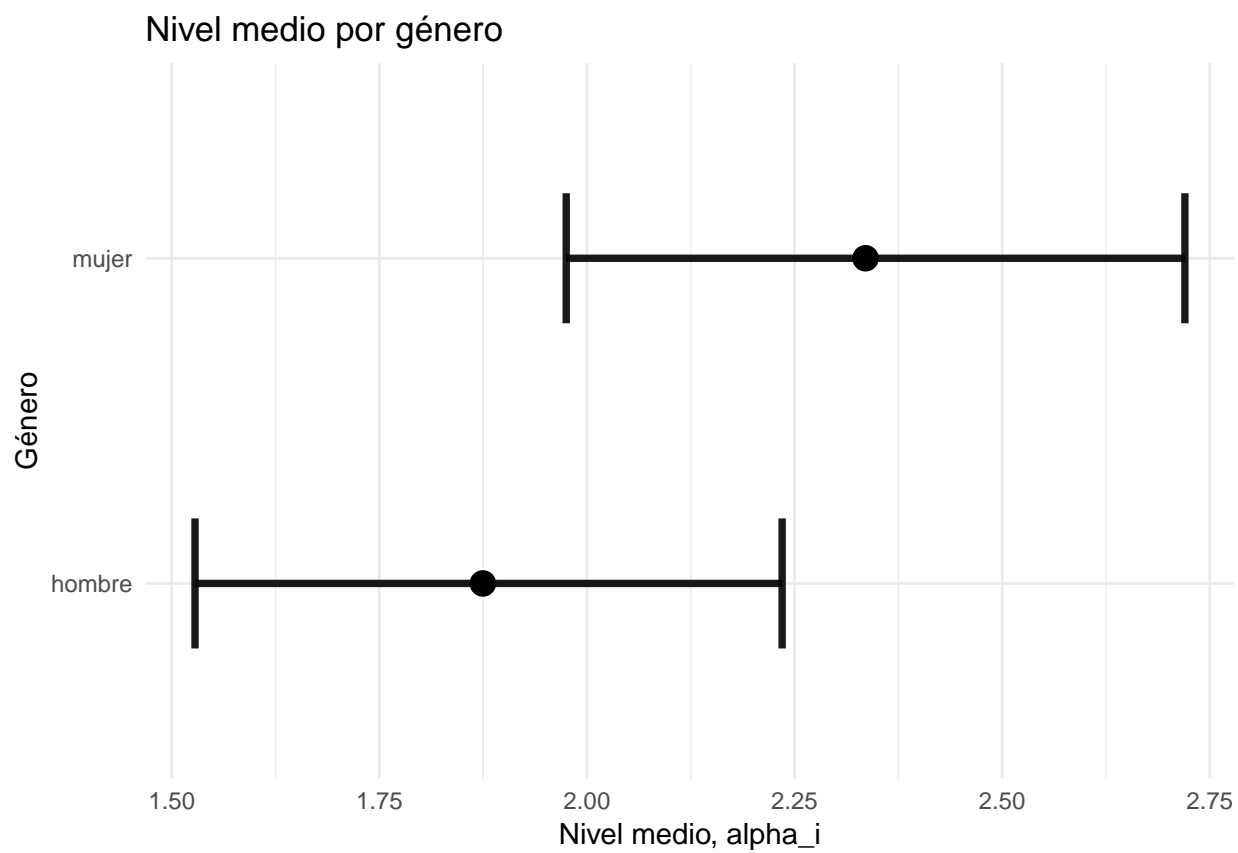
Se observa que en el caso de los coeficientes $\delta 1$ y $\delta 2$ que corresponden a las variables *antigüedad* e *incremento*, el intervalo contiene al cero, por lo que se concluye que las variables no son significativas.

En la siguiente tabla, se muestran los coeficientes del intercepto, α_j por género así como su intervalo de confianza.

Variables	media	Desviación estándar	Límite inferior	Límite superior
α_1	2.3354	0.1903	1.9750	2.7200
α_2	1.8746	0.1808	1.5280	2.2350

A partir de los niveles medios antes mencionados, se obtuvo la probabilidad de que el *individuo_i* con género *j* renuncie de manera voluntaria así como los intervalos de confianza; es decir:

$$\pi_{ij} = \frac{e^{\alpha_j}}{1 + e^{\alpha_j}}$$



Y se observan las probabilidades (θ_i) para los 50 primeros colaboradores con sus intervalos de confianza

Probabilidad de renuncia de los asesores de venta

