

Proyecto Final

Rotación de asesores de venta

Paola Mejía

Karla Alfaro

Rodrigo Suárez

Ana Bertha Coronel

Diciembre 2019

1 Introducción

1.1 Contexto del problema

La rotación de empleados . . .

La empresa tiene presencia a nivel nacional a través de 174 sucursales.

Toda vez que un nuevo colaborador entra a la empresa, recibe un curso de inducción para familiarizarse con la misión y visión de la compañía y, dependiendo del área, se les da también una capacitación. En el caso del departamento de ventas, la capacitación tiene una duración de 6 meses y los colaboradores tienen interacción con personal de otras sucursales.

En algunas ocasiones, una vez terminado el periodo de capacitación, el colaborador decide dar por terminada la relación laboral y esto conlleva a un nuevo proceso de reclutamiento e inversión en capacitación.

La base de datos con la que se trabajará está conformada por `nrow(data)` asesores de ventas de la empresa y las variables se describen a continuación:

- fecha de alta (F_ALTA): fecha de ingreso a la empresa
- sexo: sexo del asesor
- edad: edad del asesor en años
- sucursal: sucursal u oficina en la que está asignado el asesor
- div: división o región a la que pertenece la sucursal. Las divisiones son: Metropolitana, Noreste, Norte, Norte/Sureste y Sur.
- número de ventas: total de ventas acumuladas en tres meses
- sueldo promedio: pago por las comisiones generadas según el número de ventas
- antigüedad: antigüedad laboral. La variable se calcula a partir de la fecha de ingreso al Seguro Social
- trabajo previos: cantidad de trabajos previos
- edo_civil: estado civil del asesor
- experiencia: experiencia en ventas (sin experiencia, 1 año, 2 años, más de 2 años)
- escolaridad: nivel de escolaridad del asesor (secundaria, bachillerato, licenciatura, maestría)
- salario diario anterior: salario anterior del asesor. Para fines de los análisis, la variable se multiplica por un factor=30.4 y se obtiene el salario mensual anterior.
- rotación o target: el asesor renunció después de terminar la capacitación o no

1.2 Objetivos

Bajo el contexto antes mencionado, el objetivo del modelo es calcular la probabilidad de que un nuevo asesor de ventas renuncie una vez terminada la capacitación.

De manera particular, se desarrollarán dos modelos:

- modelo lineal generalizado
- modelo jerárquico

El trabajo se divide en XX secciones: (1) análisis exploratorio de datos en donde se da una breve explicación de las variables así como la representación gráfica de algunas de las variables y relaciones bivariadas; en la sección *análisis inferencial* se detalla la especificación de los modelos así como los resultados de cada uno de ellos y las variables que resultaron significativas. Finalmente, se presentan las conclusiones del proyecto.

2 Análisis Exploratorio de Datos

3 Análisis inferencial

3.1 Definición de los modelos

Como se menciona en la sección del análisis exploratorio de los datos, resulta de interés conocer las variables que tienen un efecto en la baja voluntaria de los asesores de ventas.

Si consideramos la baja voluntaria del asesor como $\{1=\text{renuncia}, 0=\text{permanece}\}$, podemos decir que la variable de interés sigue una distribución Bernoulli. Si quisiéramos considerar a la variable como la tasa de rotación, podríamos pensar en una distribución beta para la variable de interés. En este último caso, necesitaríamos calcular el número de bajas y el total de asesores expuestos bajo algún criterio.

Para considerar los datos de manera desagregada, definimos a la variable $Y_i|\pi_i \sim Be(\pi_i)$; es decir,

$$Y_i|\pi_i = \begin{cases} 1 & \text{si el asesor renuncia después de la capacitación} \\ 0 & \text{si el asesor permanece después de la capacitación} \end{cases} \quad (1)$$

Consideraremos las siguientes ligas:

- liga logística:

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

- liga normal:

$$\pi_i = \phi(X_i' \beta)$$

- liga log log complementaria:

$$\pi_i = 1 - e^{-e^{X_i' \beta}}$$

En cualquiera de los casos anteriores, definimos $\eta_i = X_i' \beta$

Usando la información anterior, se realizarán 3 modelos: 1. un modelo simple con controles, 2. un modelo simple con controles e interacciones y 3. un modelo jerárquico.

3.1.1 1.Modelo simple con controles

$$\eta_i = \alpha + \beta_1 * \text{sueldo.prom}_i \text{edad}_i + \beta_2 * \text{sueldo.men.ant}_i + \beta_3 * \text{incremento}_i + \beta_4 * \text{antiguedad}_i + \beta_5 * \text{tprevios} + \beta_6 * \text{nventas}_i + \beta_7 * \text{edad}_i$$

en donde

- z_{1i} división: $z_{1i} \in \{1, 2\}$
- z_{2i} mes: $z_{2i} \in \{1, \dots, 6\}$
- z_{3i} sexo del asesor: $z_{3i} \in \{1, 2\}$
- z_{4i} estado civil: $z_{4i} \in \{1, 2\}$
- z_{5i} dependientes: $z_{5i} \in \{1, 2, 3\}$
- z_{6i} escolaridad: $z_{6i} \in \{1, 2, 3\}$
- z_{7i} experiencia: $z_{7i} \in \{1, 2, 3\}$

3.1.2 2.Modelo simple con controles e interacciones

$$\eta_i = \alpha + \beta_1 * \text{sueldo.prom}_i \text{edad}_i + \beta_2 * \text{sueldo.men.ant}_i + \beta_3 * \text{incremento}_i + \beta_4 * \text{antiguedad}_i + \beta_5 * \text{tprevios} + \beta_6 * \text{nventas}_i + \beta_7 * \text{edad}_i$$

en donde

- z_{8i} edad_cat: $z_{8i} \in \{1, 2, 3\}$
- z_{9i} sueldo_promedio_cat: $z_{9i} \in \{1, 2, 3\}$

Al tener variables explicativas categóricas, impondremos restricciones de estimabilidad sobre algunos parámetros:

$$\sum_{i=1}^2 \delta 1 = \sum_{i=1}^6 \delta 2 = \sum_{i=1}^2 \delta 3 = \sum_{i=1}^2 \delta 4 = \sum_{i=1}^3 \delta 5 = \sum_{i=1}^3 \delta 6 = \sum_{i=1}^3 \delta 7 = \sum_{i=1} \sum_{j=1} \nu 1_{ij} = 0$$

3.2 Estimación de los modelos

Para estimar los modelos, definimos el número de iteraciones y cadenas a correr, el periodo de calentamiento y el adelgazamiento de las cadenas. Además, para que el intercepto tuviera una interpretación correcta, escalamos las variables numéricas. No es necesario centrar las variables categóricas pues se centran al imponer las restricciones de estimabilidad definidas anteriormente.

-Iteraciones: Se corrieron 50,000 iteraciones para asegurar que las estimaciones de los coeficientes variaran lo menos posible si se calculaba nuevamente el modelo. Con menos iteraciones, había variación en las estimaciones del modelo.

-Cadenas: Se corrieron dos cadenas para asegurar que la convergencia fuera correcta y que no se haya atorado alguna cadena en alguna región inusual del espacio de parámetros.

-Periodo de calentamiento: Se tiraron las primeras 5,000 iteraciones (10% de las iteraciones en total), pues nos interesa el valor de convergencia y así disminuimos la probabilidad de que las iteraciones se encuentren influenciadas por el valor inicial especificado.

-Adelgazamiento: Se utilizó un adelgazamiento de 1 para reducir la correlación entre iteraciones consecutivas.

4 Resultados

Los valores del criterio de información de devianza (DIC) para los 9 modelos: 3 simples, 3 con interacciones y 3 jerárquicos son los siguientes:

Liga	DIC simple	DIC interacciones	DIC jerárquico
logit	611.2663	629.725	
probit	620.1496	636.8094	
clog log	627.9551	619.7625	

De los 3 modelos simples estimados, el que tuvo mejor ajuste de acuerdo con el DIC fue el modelo con la liga logística. Para los modelos con interacciones, el mejor modelo fue el de la liga cloglog. Y, para los modelos jerárquicos, es el modelo. . .

4.0.1 1.Modelo simple con controles

El resultado del modelo simple con controles con liga logística se muestra a continuación:

Variables	Media	Desviación estándar	Límite inferior	Límite superior	Significativa
Constante	-1.4732	5.7006	-12.6980	9.7265	
Devianza	587.2223	6.9345	575.8497	602.5498	
Información laboral					
Sueldo promedio	-0.4163	0.1681	-0.7493	-0.0916	*
Sueldo mensual anterior	0.1015	0.1127	-0.1188	0.3237	
Incremento en sueldo	-0.1141	0.1287	-0.4163	0.0884	
Antigüedad	0.3068	0.2911	-0.2635	0.8761	
Número de trabajos previos	0.3365	0.1765	-0.0064	0.6851	
Número de ventas	-3.0174	0.2299	-3.4809	-2.5793	*
<i>División territorial</i>					
Norte	-0.1518	0.1470	-0.4391	0.1368	
Sur	0.1518	0.1470	-0.1368	0.4391	
<i>Mes de alta</i>					
Julio	0.3794	0.5195	-0.6242	1.4117	
Agosto	0.2290	0.2170	-0.1959	0.6568	
Septiembre	0.6671	0.2908	0.0994	1.2388	*
Octubre	-0.4513	0.2716	-0.9855	0.0790	
Noviembre	-0.4439	0.2567	-0.9518	0.0575	
Diciembre	-0.3802	0.2837	-0.9361	0.1760	
Información personal					
Edad	-0.7205	0.2536	-1.2192	-0.2269	*
<i>Sexo</i>					
Masculino	0.2997	0.1153	0.0761	0.5283	*
Femenino	-0.2997	0.1153	-0.5283	-0.0761	*
<i>Estado civil</i>					
Casado	0.0980	0.1597	-0.2143	0.4109	
Soltero	-0.0980	0.1597	-0.4109	0.2143	
<i>Dependientes económicos</i>					

Variables	Media	Desviación estándar	Límite inferior	Límite superior	Significativa
Con hijos	0.1239	0.2190	-0.3032	0.5540	
Sin hijos	-0.2707	0.1909	-0.6435	0.1059	
Otros	0.1468	0.2316	-0.3071	0.6013	
<i>Escolaridad</i>					
Bachillerato/carrera técnica	0.2266	0.4369	-0.6426	1.0710	
Licenciatura	0.4081	0.4278	-0.4428	1.2339	
Maestría	-0.6347	0.8329	-2.2360	1.0245	
<i>Experiencia</i>					
Sin experiencia	0.0289	0.3128	-0.5778	0.6500	
1 año	0.1257	0.2126	-0.2941	0.5401	
2 años o más	-0.1547	0.2379	-0.6212	0.3097	

Las variables significativas del modelo son:

- El sueldo promedio: Caeteris paribus, ante un aumento de una unidad en el sueldo promedio, la razón de momios disminuye en 34%.
- El número de ventas: Caeteris paribus, al aumentar en una unidad el número de ventas, la razón de momios disminuyen en 95%.
- Darse de alta en septiembre: Caeteris paribus, darse de alta en septiembre aumenta la razón de momios en 95%.
- La edad: Caeteris paribus, aumentar la edad en una unidad disminuye en 51% la razón de momios.
- El sexo: Caeteris paribus, ser hombre aumenta en 35% la razón de momios mientras que ser mujer la disminuye en 26%.

4.0.2 2.Modelo simple con controles e interacciones

El resultado del modelo con interacciones y liga cloglog se muestra a continuación:

Variables	Media	Desviación estándar	Límite inferior	Límite superior	Significativa
Constante	-4.4648	4.1003	-11.5675	4.2955	
Devianza	590.7545	7.4514	578.1950	607.2148	
Información laboral					
<i>Sueldo promedio</i>					
<2000	1.8014	1.6716	-0.3501	5.8659	
200010000	-1.2750	0.8809	-3.3629	-0.0197	*
Sueldo mensual anterior	0.0257	0.0671	-0.1079	0.1557	
Incremento en sueldo	-0.1138	0.1091	-0.3725	0.0469	
Antigüedad	-0.0430	0.1324	-0.3048	0.2144	
Número de trabajos previos	0.0905	0.1007	-0.1074	0.2873	
Número de ventas	-2.0743	0.1506	-2.3799	-1.7883	*
<i>División territorial</i>					
Norte	-0.2004	0.0862	-0.3699	-0.0328	*
Sur	0.2004	0.0862	0.0328	0.3699	*

Variables	Media	Desviación estándar	Límite inferior	Límite superior	Significativa
<i>Mes de alta</i>					
Julio	0.1681	0.2757	-0.3634	0.7165	
Agosto	0.0817	0.1255	-0.1675	0.3257	
Septiembre	0.2629	0.1730	-0.0794	0.5992	
Octubre	-0.2366	0.1460	-0.5252	0.0495	
Noviembre	-0.1582	0.1448	-0.4430	0.1250	
Diciembre	-0.1178	0.1683	-0.4502	0.2092	
Información personal					
<i>Edad</i>					
<30	-0.6163	0.8627	-2.6676	0.5904	
3050	1.6044	1.7025	-0.6922	5.6903	
<i>Sexo</i>					
Masculino	0.0627	0.0657	-0.0672	0.1908	
Femenino	-0.0627	0.0657	-0.1908	0.0672	
<i>Estado civil</i> [^]					
Casado	0.1211	0.0943	-0.0622	0.3067	
Soltero	-0.1211	0.0943	-0.3067	0.0622	
<i>Dependientes económicos</i>					
Con hijos	0.0787	0.1266	-0.1696	0.3253	
Sin hijos	-0.2435	0.1107	-0.4622	-0.0283	*
Otros	0.1648	0.1319	-0.0909	0.4252	
<i>Escolaridad</i>					
Bachillerato/carrera técnica	0.2583	0.2625	-0.2336	0.7905	
Licenciatura	0.1709	0.2602	-0.3203	0.6969	
Maestría	-0.4292	0.5055	-1.4576	0.5168	
<i>Experiencia</i>					
Sin experiencia	0.1288	0.1831	-0.2271	0.4880	
1 año	-0.1543	0.1249	-0.3989	0.0911	
2 años o más	0.0256	0.1378	-0.2451	0.2941	
Interacciones					
Edad <30 y sueldo <2000	-2.0149	1.6705	-6.0653	0.1251	
3050 y sueldo <2000	4.3841	3.3321	0.1593	12.4773	*
Edad <30 y 200050 y 200010000	1.1960	0.8827	-0.0708	3.2867	
3010000	1.4183	0.8839	0.1564	3.5113	*
Edad >50 y sueldo >10000	-2.6143	1.7452	-6.7602	-0.1761	*

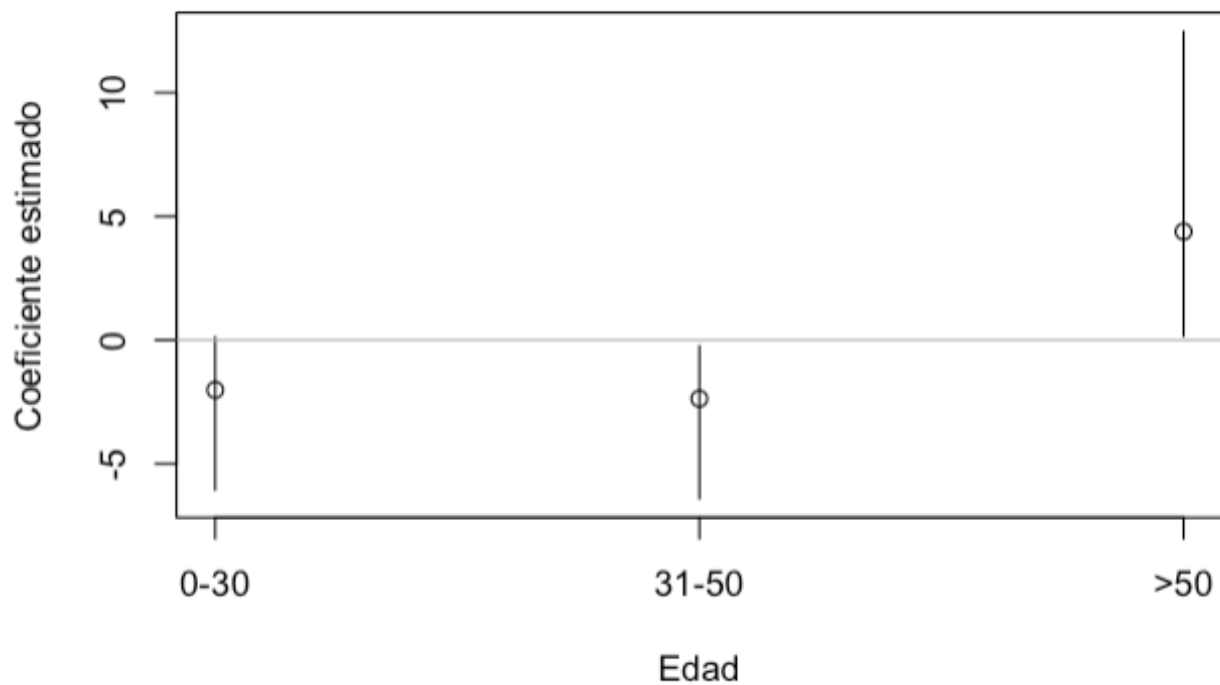
Las variables significativas del modelo son:

- Tener un sueldo >1000: Caeteris paribus, tener un sueldo promedio mayor a \$10000 disminuye el momio en 0.2794.
- Número de ventas: Caeteris paribus, al aumentar en una unidad el número de ventas, el momio disminuye en 0.1256.
- División territorial: Trabajar en el norte disminuye el momio en 0.8184 mientras que, trabajar en el sur lo aumenta en 1.2219.
- No tener hijos: No tener hijos disminuye el momio en 0.7839.

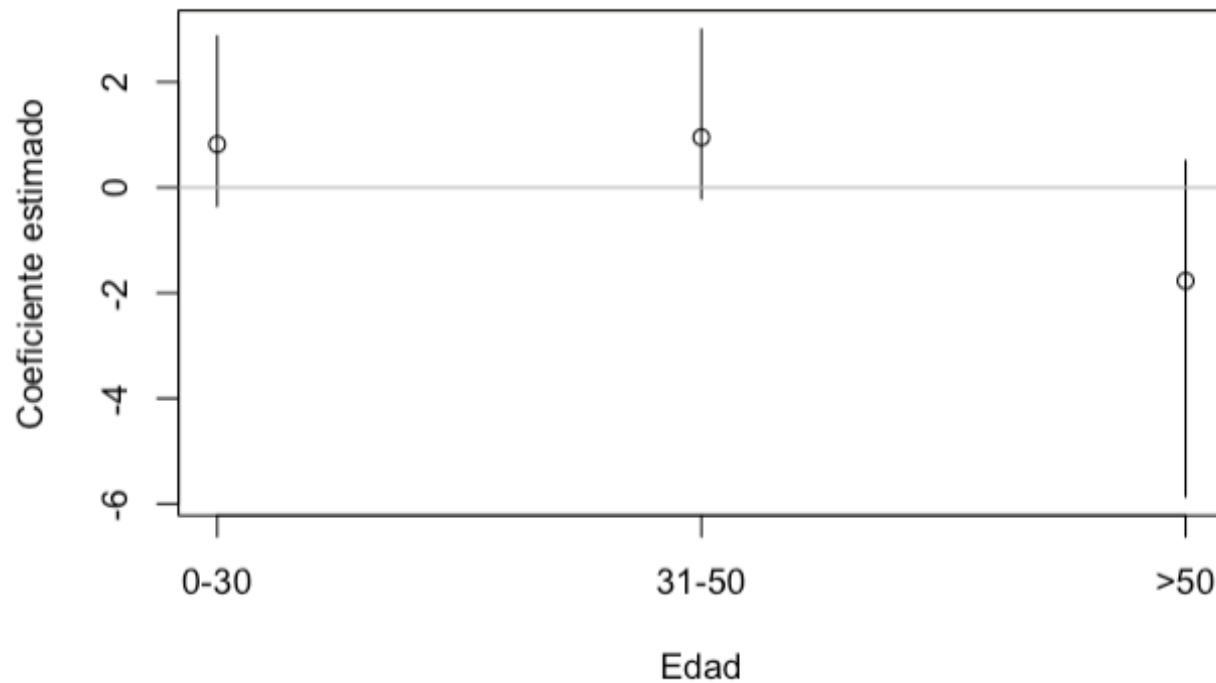
- Tener entre 30 y 50 años y un sueldo menor a 2000: Tener entre 30 y 50 años y un sueldo menor a \$2000 disminuye el momio en 0.0936.
- Tener más de 50 años y un sueldo menor a 2000: Tener más 50 años y un sueldo menor a \$2000 aumenta el momio en 80.1665.
- Tener entre 30 y 50 años y un sueldo mayor a 10000: Tener entre 30 y 50 años y un sueldo mayor a \$10000 aumenta el momio en 4.1300.
- Tener más de 50 años y un sueldo mayor a 10000: Tener más de 50 años y un sueldo mayor a \$10000 disminuye el momio en 0.0732.

Los efectos de las interacciones pueden observarse en las siguientes gráficas:

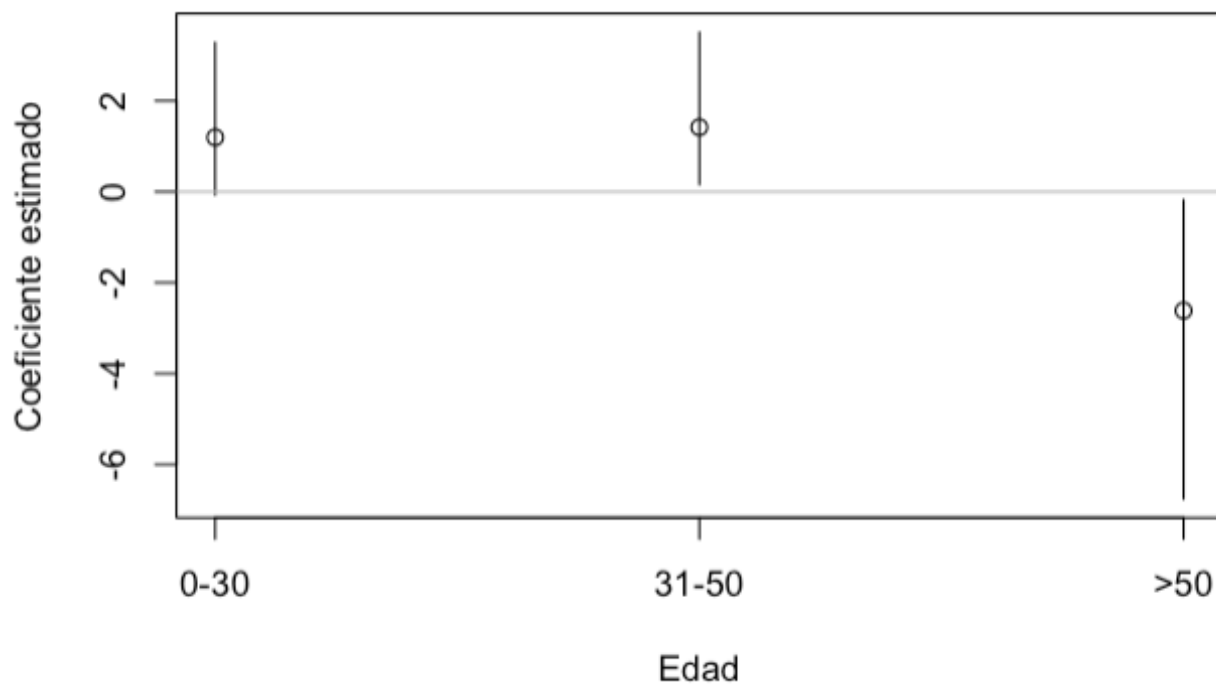
Interacción del efecto de tener un sueldo bajo y la edad



Interacción del efecto de tener un sueldo medio y la edad



Interacción del efecto de tener un sueldo alto y la edad



4.0.3 3.Modelo jerárquico

El resultado del modelo con interacciones y liga cloglog se muestra a continuación:

Las variables significativas del modelo son:

5 Conclusiones

6 Bibliografía

7 Anexo

En la siguiente sección se muestra el código utilizado tanto para realizar el análisis exploratorio de los datos como para el desarrollo de los modelos.