

Análisis de Datos Categóricos

Ayudantía 8

Felipe Olivares

Contenido

1 Regresión multinomial

Regresión Multinomial

¿Qué es una regresión Multinomial?

Es una generalización de la regresión logística binomial, que busca hacerse cargo de la situación de una variable discreta que no necesariamente pueden estar ordenada y que tiene más de dos valores (cuestión que la regresión logística no puede cubrir). Un ejemplo clásico es cuándo se hace investigación sobre participación electoral. Tomemos el caso del voto por candidatos (cuándo hay más de dos candidatos, por ejemplo, en una primera vuelta electoral).

La regresión logística multinomial es una extensión simple de la regresión logística binaria que permite más de dos categorías de la variable dependiente o de resultado. Al igual que regresión logística, la regresión logística multinomial utiliza la estimación de máxima verosimilitud (MLE)

Regresión Multinomial

Para efecto de la regresión que vamos a utilizar, vamos a modelar la función de las formas de control de la protesta para distintas categorías de esta. Por lo tanto, es importante que definamos nuestra variable de referencia (o de contraste) respecto de las categorías de interés. En una regresión multinomial las estimaciones se obtienen respecto de una categoría

```
# agregamos la variable de participantes a otro subset de datos
df2 <- df1 %>% dplyr::select(apolicial,disruptiva,trabajadores, p11) %>%
  mutate( disruptiva = if_else(disruptiva=="Si",1,0),
           trabajadores = if_else(trabajadores=="Si",1,0),
           participantes = p11) %>%
  mutate( participantes = as.numeric(participantes),
           disruptiva = as.numeric(disruptiva),
           trabajadores= as.numeric(trabajadores)) %>%
  na.omit(df2)
```

Regresión Multinomial

Revisemos los datos nuevamente...

```
head(df2)
```

```
# # A tibble: 6 x 5
#   apolicial      disruptiva trabajadores  p11 participantes
#   <fct>          <dbl>          <dbl> <dbl>          <dbl>
# 1 Presencia policial      0              0    81             81
# 2 violencia policial      0              0    40             40
# 3 Enfrentamientos directos  1              0    20             20
# 4 Presencia policial      1              1  2000           2000
# 5 Enfrentamientos directos  0              0    60             60
# 6 Enfrentamientos directos  1              1   150            150
```

Regresión Multinomial

Formalmente:

$$\underbrace{\ln \frac{p_{ij}}{1 - p_{ij}}}_{\text{logit}(p_{ij})} = \beta_{0j} + \beta_{j1}x_{i1} + \dots + \beta_{jk}x_{ik}$$

Importante: Los coeficientes y sus transformaciones entregan información sobre las probabilidades relativas de los diferentes j 'ts. Esto siempre respecto de una variable de referencia anteriormente seleccionada (para este caso presencia policial)

Regresión Multinomial

```
mlogit1 <- multinom(apolicial ~ participantes + trabajadores, trace=F, data=df2)
summary(mlogit1)
```

```
# Call:
# multinom(formula = apolicial ~ participantes + trabajadores,
#   data = df2, trace = F)
#
# Coefficients:
#               (Intercept) participantes trabajadores
# Acciones preventivas    -1.76407512  1.503145e-05   -0.4194244
# Enfrentamientos directos 0.07487663  5.924185e-06   -0.5319464
# violencia policial      -0.51631844  1.717192e-05   -0.1605411
#
# Std. Errors:
#               (Intercept) participantes trabajadores
# Acciones preventivas    4.083983e-11  4.180938e-06  1.958634e-11
# Enfrentamientos directos 9.965112e-11  3.981581e-06  4.849998e-11
# violencia policial      8.978793e-11  3.558233e-06  5.598286e-11
#
# Residual Deviance: 6691.673
# AIC: 6709.673
```

Regresión Multinomial

Interpretación de los coeficientes de la variable trabajadores:

- 1) El logaritmo de un control con acciones preventivas versus un control de presencia policial aumenta en 0.0001 puntos de log odds por cada aumento que exista en el número de participantes de una protesta.
- 2) El logaritmo de un control con enfrentamientos directos versus un control de presencia policial disminuye en -0,53 puntos de log odds si existe presencia de trabajadores durante una manifestación.
- 3) Un aumento en 1 unidad de número de participantes se traduce en un aumento de 0.0001 unidades en el logit de un control violento de la protesta versus un control de presencia policial. El logaritmo de un control violento de la protesta versus un control de presencia policial aumenta en 0.0001 puntos de log odds por cada aumento que exista en el número de participantes.

Regresión Multinomial

Cálculemos efectos marginales de los logit de un control con enfrentamientos directos y de un control violento.

```
#Efecto 100 participantes
c(E= 0.07 + 0.001*100 -0.42*1,
  V= -0.52+ 0.001*100 -0.16*1)
```

```
#      E      V
# -0.25 -0.58
```

```
#Efecto 10000 participantes
c(E= 0.07 + 0.001*10000 -0.42*1,
  V= -0.52 + 0.001*10000 -0.16*1)
```

```
#      E      V
# 0.65 0.32
```

```
#Beta enfrentamientos directos
0.65 - 0.25
```

```
# [1] 0.4
```

```
#Beta Violencia policial
0.32 - 0.58
```

```
# [1] -0.26
```

Manteniendo todo constante, el aumento en 1000 participantes se traducen un aumento de 0.4 puntos de log odds para el control con enfrentamientos directos respecto de un control con presencia policial.

Manteniendo todo constante, el aumento en 1000 participantes se traduce en una disminución de 0.26 en el logaritmo de las odds de un control violento de la protesta versus un control de presencia policial.

Regresión Multinomial - Efectos multiplicativos

La transformación de probabilidades a odds es monotónica, si la probabilidad aumenta también lo hacen los odds, y viceversa. El rango de valores que pueden tomar los odds es de $(0, \infty+)$. Dado que el valor de una probabilidad está acotado entre $(0,1)$

donde:

- $e^{\beta_{jk}}$ está restringido al rango $[0, \infty+)$. Es una constante que “comprime” o amplifica el ratio entre las probabilidades de j respecto de J .
- Si $\beta_{jk} < 0 \rightarrow (0 < e^{\beta_{jk}} < 1)$. Es decir, un aumento en x_k está asociado con una reducción (multiplicativa) del ratio entre las probabilidades de j versus J .
- Si $\beta_{jk} = 0 \rightarrow (e^{\beta_{jk}} = 1)$. Es decir, un cambio en x_k está asociado a un cambio nulo en el ratio entre las probabilidades de j versus J .
- Si $\beta_{jk} > 0 \rightarrow (e^{\beta_{jk}} > 1)$. Es decir, un aumento en x_k está asociado a aumento (multiplicativo) del ratio entre las probabilidades de j versus J .

Regresión Multinomial - Efectos multiplicativos

```
#Log odds
```

```
summary(mlogit1)$coefficients
```

```
#               (Intercept) participantes trabajadores
# Acciones preventivas    -1.76407512  1.503145e-05  -0.4194244
# Enfrentamientos directos 0.07487663  5.924185e-06  -0.5319464
# violencia policial      -0.51631844  1.717192e-05  -0.1605411
```

```
#Odds ratio
```

```
exp(summary(mlogit1)$coefficients)
```

```
#               (Intercept) participantes trabajadores
# Acciones preventivas    0.1713452    1.000015    0.6574251
# Enfrentamientos directos 1.0777512    1.000006    0.5874604
# violencia policial      0.5967133    1.000017    0.8516829
```

Regresión Multinomial

Precauciones en el uso de modelos multinomiales:

- ① Efectos marginales son heterogéneos (multiplicidad de efectos)
- ② Efectos heterogéneos se vuelven más complejo con la inclusión de predictores
- ③ Los efectos muchas veces no son monotónicos (pueden variar de signo)

Regresión Multinomial

Cambiando la categoría de referencia

```
df2$apolicial2 <- relevel(df2$apolicial, ref = "Enfrentamientos directos") # 1° de cambiar valor de referencia
```

```
mlogit2 <- multinom(apolicial2 ~ participantes + trabajadores, trace=F, data=df2)
summary(mlogit2)
```

```
# Call:
# multinom(formula = apolicial2 ~ participantes + trabajadores,
#   data = df2, trace = F)
#
# Coefficients:
#               (Intercept) participantes trabajadores
# Presencia policial -0.07494316 -5.924776e-06  0.5314497
# Acciones preventivas -1.84033887  9.117559e-06  0.1127366
# violencia policial -0.59112192  1.125851e-05  0.3712487
#
# Std. Errors:
#               (Intercept) participantes trabajadores
# Presencia policial  9.494736e-11  3.982924e-06  4.923754e-11
# Acciones preventivas 3.224457e-11  3.773817e-06  1.350665e-11
# violencia policial  5.809218e-11  3.065192e-06  2.885183e-11
#
# Residual Deviance: 6691.673
# AIC: 6709.673
```