# Análisis de Datos Categóricos

Ayudantía 5

Felipe Olivares

# Contenido

# LPM

El modelo lineal de probabilidad, se puede interpretar en términos probabilísticos, en el sentido de que un valor concreto de la recta de regresión mide la probabilidad de que ocurra el hecho objetivo de estudio. Es decir, nuestra variable dependiente se puede considerar como la estimación de la probabilidad de que ocurra el hecho objetivo de estudio $Yi = 1$ siguiendo el siguiente criterio: Valores próximos a cero se corresponde con una baja probabilidad de ocurrencia del hecho estudiado (menor cuanto más próximos a cero); mientras que a valores próximos a uno se les asigna una probabilidad elevada de ocurrencia (mayor cuanto más próximos a uno).

$$y_i = \beta_0 + \beta_1 x_{1i} + \ldots \beta_k x_{ki} + e_i$$

El modelo de regresión más simple aplicado a datos categóricos es el modelo lineal de probabilidad (LPM). Esto es básicamente lo mismo que un modelo de regresión lineal.

*Variable dependiente* $=$ *intercepto* $+$ *predictores* $+$ *error aleatorio*

# LPM

Revisemos los mismos datos de la ayudantía anterior para crear un modelo de regresión lineal en base a predictores que revisamos

```
head(df1)
```

```
# # A tibble: 6 x 17
#   ano   participa~1 region educa~2 indig~3 laboral salud pacif~4 disru~5 viole~6
#   <fct>       <dbl> <fct>  <fct>   <fct>   <fct>   <fct> <fct>   <fct>   <fct>
# 1 2009           NA Metro~ No      No      No      No    Sí      No      No
# 2 2009           NA Tarap~ Sí      No      No      No    Sí      No      No
# 3 2009            7 Tarap~ No      No      Sí      No    Sí      Sí      No
# 4 2009           10 O´Hig~ No      No      Sí      No    No      Sí      No
# 5 2009           81 Arauc~ No      Sí      No      No    Sí      No      No
# 6 2009           NA Arauc~ No      No      No      No    No      Sí      No
# # ... with 7 more variables: organizacion <fct>, nacional <fct>,
# #   macrozona <chr>, estudiantes <fct>, trabajadores <fct>, ppolicial <fct>,
# #   apolicial <fct>, and abbreviated variable names 1: participantes,
# #   2: educacion, 3: indigena, 4: pacifica, 5: disruptiva, 6: violenta
# # i Use `colnames()` to see all variable names
```

# LPM

La regresión lineal que vamos a construir la haremos sobre la base de la pregunta acerca del tipo de control policial que realizan las policías (Control negociado o control violento de la protesta) en el caso de las manifestaciones sobre educación en Chile para los años 2009-2019. Para realizar esto, utilizaremos como variable dependiente la acción policial y variables independientes: grupos sociales estudiantes, demandas sobre educación y tipo de protesta disruptiva.

```
str (df1) # siempre importante revisar la estructura de las codificaciones
```

```
# tibble [23,398 x 17] (S3: tbl_df/tbl/data.frame)
#  $ ano          : Factor w/ 11 levels "2009","2010",..: 1 1 1 1 1 1 1 1 1 1 ...
#   ..- attr(*, "label")= chr "Años"
#  $ participantes: num [1:23398] NA NA 7 10 81 NA NA NA NA NA ...
#   ..- attr(*, "label")= chr "N° de participantes"
#  $ region       : Factor w/ 16 levels "Tarapacá","Antofagasta",..: 13 1 1 6 9 9 9 13 13 13 ...
#   ..- attr(*, "label")= chr "Regiones"
#  $ educacion    : Factor w/ 2 levels "No","Sí": 1 2 1 1 1 1 1 1 1 1 ...
#   ..- attr(*, "label")= chr "Demanda - Educacional"
#  $ indigena     : Factor w/ 2 levels "No","Sí": 1 1 1 1 2 1 2 1 1 1 ...
#   ..- attr(*, "label")= chr "Demanda - Indígenas"
#  $ laboral      : Factor w/ 2 levels "No","Sí": 1 1 2 2 1 1 1 1 1 1 ...
#   ..- attr(*, "label")= chr "Demanda - Laborales"
#  $ salud        : Factor w/ 2 levels "No","Sí": 1 1 1 1 1 1 1 1 1 1 ...
#   ..- attr(*, "label")= chr "Demanda - Salud"
#  $ pacifica     : Factor w/ 2 levels "No","Sí": 2 2 2 1 2 1 2 2 2 2 ...
#   ..- attr(*, "label")= chr "Táctica - Pacífica"
#  $ disruptiva   : Factor w/ 2 levels "No","Sí": 1 1 2 2 1 2 2 1 1 1 ...
#   ..- attr(*, "label")= chr "Táctica - Disruptiva"
#  $ violenta     : Factor w/ 2 levels "No","Sí": 1 1 1 1 1 1 1 1 1 1 ...
#   ..- attr(*, "label")= chr "Táctica - Violenta"
#  $ organizacion : Factor w/ 3 levels "Sin organizaciones",..: 2 2 2 1 1 1 2 2 1 1 ...
#   ..- attr(*, "label")= chr "Organizaciones sociales presentes"
#  $ nacional     : Factor w/ 2 levels "No","Sí": 1 1 1 2 1 1 1 1 1 1 ...
#   ..- attr(*, "label")= chr "Protesta nacional"
#  $ macrozona    : chr [1:23398] "RM" "Norte" "Norte" "Centro" ...
#   ..- attr(*, "label")= chr "Macrozonas"
#  $ estudiantes  : Factor w/ 2 levels "No","Sí": 1 2 1 1 1 1 1 1 1 1 ...
```

# LPM

Creamos un subset de datos que contenga solo aquellos casos en que existe acción policial. Así mismo, y de acuerdo a lo que observamos previamente en la estructura de los datos, recodificamos las variables de interés para colocarlas en nuestra regresión.

Para efecto de la regresión que modelaremos la función de un control violento de la protesta respecto de un control negociado. Por lo tanto, es importante que el control negociado de la protesta=0 y control violento de la protesta=1 (consideren lo dicho previamente respecto del sentido del efecto que queremos estimar)

```
df2 <- df1 %>% select(apolicial,ppolicial,educacion,estudiantes,,disruptiva,trabajadores) %>%
  mutate(apolicial = if_else(apolicial=="Violencia Policial",1,0),
         educacion = if_else(educacion=="Sí",1,0),
                  estudiantes= if_else(estudiantes=="Sí",1,0),
         disruptiva = if_else(disruptiva=="Sí",1,0),
         trabajadores = if_else(trabajadores=="Sí",1,0))  %>%
  mutate(apolicial = as.numeric(apolicial),
         educacion = as.numeric(educacion),
                 estudiantes = as.numeric(estudiantes),
         disruptiva = as.numeric(disruptiva),
         trabajadores= as.numeric(trabajadores))  %>%
  na.omit(df2)
```

# LPM

```r
#Modelo de regresión propuesto
lm1 <-lm(apolicial ~ educacion + estudiantes + disruptiva, data=df2)
summary(lm1)
```

```
#
# Call:
# lm(formula = apolicial ~ educacion + estudiantes + disruptiva,
#     data = df2)
#
# Residuals:
#     Min      1Q  Median      3Q     Max
# -0.6878 -0.5488  0.3213  0.4512  0.5238
#
# Coefficients:
#              Estimate Std. Error t value Pr(>|t|)
# (Intercept)  0.485365   0.010487  46.282  < 2e-16 ***
# educacion   -0.009148   0.022350  -0.409    0.682
# estudiantes  0.139072   0.021381   6.504 8.42e-11 ***
# disruptiva   0.063414   0.012770   4.966 7.03e-07 ***
# ---
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 0.4926 on 6075 degrees of freedom
# Multiple R-squared:  0.01742, Adjusted R-squared:  0.01694
# F-statistic: 35.91 on 3 and 6075 DF,  p-value: < 2.2e-16
```

# LPM

**Interpretación**

recordemos..

$$y_i = \beta_0 + \beta_1 x_{1i} + \ldots \beta_k x_{ki} + e_i$$

Entonces decimos que el efecto de $x_k$ sobre $y$ es $\beta_k$. ¿Qué significa?

"Un cambio en $\Delta$ unidades de $x_k$ se traduce en un cambio en $\Delta\ \beta_k$ unidades en el valor esperado de $y_i$"

# LPM

**Lectura de coeficientes**

1- Efecto de educación sobre control violento de la protesta:

**R:** En promedio, la presencia de demandas por educación disminuye la probabilidad de un control violento de la protesta en 0.009, es decir, la presencia de demandas educativas durante la protesta disminuye un 9% controlando por el resto de la covariables del modelo. Sin embargo, este efecto no es significativo.

2- Efecto de la presencia de estudiantes sobre control violento de la protesta:

**R:**En promedio, la presencia de estudiantes durante una manifestación aumenta la probabilidad de un control violento de la protesta en 0.14, es decir, la presencia de estudiantes aumenta un 14% controlando por nuestras covariables del modelo. Este efecto es estadísticamente significativo a un 99,9% de confianza y un valor $p=$ 8.42e-11

3- Efecto de tácticas disruptivas sobre el control violento de la protesta:

**R:**En promedio, la presencia de tácticas disruptivas durante la protesta aumenta la probabilidad de un control violento de la protesta en 0.06, es decir, la presencia de tácticas disruptivas aumenta un 6%, controlando por el resto de la covariables del modelo. Este efecto es estadísticamente significativo a un 99,9% de confianza y un valor $p=$ -7.03e07

# LPM

*Limitaciones. . .*

**Distribución y rango**: Los modelos de regresión lineal asumen que la variable dependiente son manifestaciones de distribuciones normales, pero que ocurre cuándo nuestras observaciones no son normales (toma valores 0 y 1). Esto puede provocar que nuestras estimaciones escapen de los rangos 0 y 1 (que son el rango natural de una probabilidad).

**Sesgados e inconsistentes**: Coeficiente no da en el blanco y además consistentemente convergen en un valor erróneo.

**Varianza**: Supuesto de varianza constante en caso de las regresiones lineales no se cumple para variables categóricas $= p_i(1 - p_i)$ Sabemos que usar una variable Bernoulli o Binomial no tiene varianzas constantes dada la distribución que tienen los datos.

# Regresión Logística

Para ver regresión logística trabajaremos con los mismos datos que utilizamos para la regresión lineal que vimos previamente.

```
head(df2)
```

```
# # A tibble: 6 x 6
#   apolicial ppolicial educacion estudiantes disruptiva trabajadores
#       <dbl> <fct>         <dbl>       <dbl>      <dbl>        <dbl>
# 1         0 No                0           0          0            0
# 2         0 No                0           0          1            0
# 3         0 No                0           0          1            0
# 4         0 No                0           0          0            0
# 5         1 No                0           0          0            0
# 6         1 No                0           0          1            0
```

# Regresión Logística

Es importante recordar la recodificación y/o conversión de variables que se estimen necesarias. Las variables dicotómicas se leen mejor cuándo tienen una codificación 0 y 1 dependiendo de la categoría de interés. Este procedimiento ya lo realizamos para la regresión lineal.

**1- Ajuste un modelo de regresión logística que predice el control violento de la protesta como función de distinas covariables: educación, estudiantes y tácticas disruptivas.**

Formalmente:

$$\underbrace{\ln \frac{p_i}{1 - p_i}}_{\text{logit}(p_i)} = \beta_0 + \beta_1 \text{demandas por educación}_i + \beta_2 \text{presencia de estudiantes}_i + \beta_3 \text{presencia de tácticas disruptivas}_i + \beta_4 \text{ N° de manifestantes}_i$$

# Regresión Logística

Recordamos del modelo de regresión lineal que el valor de un coeficiente significaba el cambio en unidades de la variable dependiente por cada unidad de la variable independiente a que se refiere el coeficiente, manteniendo el resto de la covariables del modelo constantes. A nivel de coeficientes estimados exponencialmente la interpretación es muy similar y la diferencia estriba en que en este caso no se trata del cambio (incremento o disminución) de la probabilidad de la variable dependiente por cada unidad de cambio en las independientes, sino del incremento o disminución que se produce en el cociente entre $P(Y = 1)$ / $P(Y = 0)$

Utilizar este procedimiento no cambia en modo alguno la forma de interpretar el signo del coeficiente. Un coeficiente positivo aumenta la probabilidad, mientras que un valor negativo disminuye la probabilidad. Así pues si $\beta$ es positivo, su transformación (antilog) será mayor a 1, y el odds ratio aumentará. Este aumento se produce cuando la probabilidad prevista de ocurrencia de un suceso aumenta y la probabilidad prevista de su no ocurrencia disminuye. Por lo tanto, el modelo tiene una elevada probabilidad de ocurrencia. De la misma forma, si $\beta$ es negativo, el antilogaritmo es menor que 1 y el odds ratio disminuye. Un valor de cero equivale a un valor de 1, lo que no produce cambio en el odds.

# Regresión Logística

```
logit1 <- glm(apolicial ~ educacion + estudiantes + disruptiva + participantes, family = binomial(link=logit), data=df3)
summary(logit1)
```

```
#
# Call:
# glm(formula = apolicial ~ educacion + estudiantes + disruptiva +
#     participantes, family = binomial(link = logit), data = df3)
#
# Deviance Residuals:
#     Min       1Q   Median       3Q      Max
# -2.5818  -1.3051   0.7954   1.0546   1.2964
#
# Coefficients:
#                Estimate Std. Error z value Pr(>|z|)
# (Intercept)   -2.754e-01  6.664e-02  -4.133 3.58e-05 ***
# educacion      1.899e-01  1.332e-01   1.426   0.1539
# estudiantes    5.029e-01  1.289e-01   3.901 9.58e-05 ***
# disruptiva     5.707e-01  7.964e-02   7.166 7.70e-13 ***
# participantes  5.007e-06  2.465e-06   2.031   0.0422 *
# ---
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# (Dispersion parameter for binomial family taken to be 1)
#
#     Null deviance: 3736.6  on 2732  degrees of freedom
# Residual deviance: 3621.2  on 2728  degrees of freedom
# AIC: 3631.2
#
# Number of Fisher Scoring iterations: 4
```

# Regresión Logística

**2- Interprete el intercepto y los efectos del modelo en términos de log-odds.**

Formalmente:

$$\frac{\partial \text{logit}(p_i)}{\partial x_k} = \beta_k$$

```
#Estimación de GLM (Regresión logística)
summary(logit1)$coefficients[,1]
```

```
#   (Intercept)     educacion   estudiantes    disruptiva participantes
# -2.754453e-01  1.899092e-01  5.028775e-01  5.707099e-01  5.007220e-06
```

**Respuesta:** El coeficiente asociado a educación indica que las log-odd esperadas de un control violento de la protesta cuándo existen demandas por educación durante la protestas son 0.189 puntos mayores que cuándo no hay demandas de este tipo presentes. El coeficiente asociado a estudiantes indica que las log-odd esperadas de un control violento de la protesta cuándo existe presencia de estudiantes durante una manifestación son de 0.502 puntos mayores que cuándo no hay presencia de estudiantes. El coeficiente de táctica disruptiva nos indica que la presencia de este tipo de repertorio de la protesta aumenta las log-odd de un control violento de la protestsa en 0.57 puntos. Finalmente el número de participantes presenta un coeficiente log-odd muy pequeño 0.00005 por cada participante adicional a la protesta.

# Regresión Logística

**3- Considerando el modelo ajustado ¿Cuál sería la probabilidad esperada sobre una marcha de 500 personas respecto del control violento de la protesta**

Formalmente:

$$\mathbb{P}(\text{Control violento}|\ \text{Educación}=1,\ \text{Estudiantes}=1,\ \text{Disruptiva}=1,\ \text{N° participantes}=500)$$

$$\mathbb{P} = \frac{1}{1 + e^{-(\beta_0 + \beta_{Educación} + \beta_{Estudiantes} + \beta_{Disruptiva} + \beta_{Nparticipantes})}}$$

# Regresión Logística

```
summary(logit1)$coefficients[,1]
```

```
#   (Intercept)     educacion   estudiantes    disruptiva participantes
# -2.754453e-01  1.899092e-01  5.028775e-01  5.707099e-01  5.007220e-06
```

$$\mathbb{P} = \frac{1}{1 + e^{-(-0.275+0.189*1+0.502*1+0.570*1+0.00005*500)}}$$

```
# efecto multiplicativo - odd de una marcha de 500 personas
exp(-0.275  + 0.189*1 +  0.502*1 + 0.570*1 + 0.00005*500 )
```

```
# [1] 2.748348
```

# Regresión Logística

**4- Ajuste un modelo que evalue la violencia en el control de la protesta en función de otras variables. Esta vez incorporando a trabajadores en grupos sociales y generando un efecto multiplicativo (interacción) respecto del número de trabajadores y las tácticas disruptivas. Interprete los coeficientes en términos de odds-ratio**.

Formalmente:

$$\frac{p_i}{1-p_i} = e^{\beta_0} \cdot e^{\beta_1 x_{i1}} \dots e^{\beta_k x_{ik}}$$

La trasformación 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:

- $e^{\beta_k}$ está restringido al rango $[0, \infty+)$. Es una constante que "comprime" o amplifica las odds de éxito.
- Si $\beta_k < 0 \rightarrow (0 < e^{\beta_k} < 1)$. Es decir, un aumento en $x_k$ está asociado con una reducción (multiplicativa) de las odds de éxito.
- Si $\beta_k = 0 \rightarrow (e^{\beta_k} = 1)$. Es decir, un cambio en $x_k$ está asociado a un cambio nulo en las odds de éxito.
- Si $\beta_k > 0 \rightarrow (e^{\beta_k} > 1)$. Es decir, un aumento en $x_k$ está asociado a aumento (multiplicativo) en de las odds de éxito.

# Regresión Logística

Revisemos primero el modelo estimado anteriormente para entender algunas cosas..

```
summary(logit1)$coefficients[,1]
```

```
#   (Intercept)     educacion   estudiantes    disruptiva participantes
# -2.754453e-01  1.899092e-01  5.028775e-01  5.707099e-01  5.007220e-06
```

```
exp(summary(logit1)$coefficients[,1])
```

```
#   (Intercept)     educacion   estudiantes    disruptiva participantes
#      0.759234      1.209140      1.653472      1.769523      1.000005
```

```
#cambio entre una marcha de 500 personas y una de 600 personas
exp(-0.275  + 0.189*1 +  0.502*1 + 0.570*1 + 0.00005*500 )/exp(-0.275  + 0.189 *1 +  0.502*1 + 0.570*1 + 0.00005*600)
```

```
# [1] 0.9950125
```

# Regresión Logística

```
#Estimación de GLM (Regresión logística)
logit2 <- glm(apolicial ~ trabajadores*disruptiva + participantes + estudiantes, family = binomial(link=logit), data=df3)
exp(summary(logit2)$coefficients[,1])
```

```
#           (Intercept)           trabajadores              disruptiva
#             0.9023479              0.5897822               1.5741600
#         participantes            estudiantes trabajadores:disruptiva
#             1.0000059              1.9109697               1.4672480
```

# Regresión Logística

**Veamos el ejemplo anterior en términos de log-odds para entender el sentido de los efectos..**

Lo que controla la dirección del efecto es el coeficiente en la regresión logística. De ahí la importancia de mirar la regresión en términos de log-odds para ayudar a entender el efecto que estamos obteniendo de nuestros coeficientes para el modelo estimado.

```r
# Modelo en log-odds
summary(logit2)
```

```
#
# Call:
# glm(formula = apolicial ~ trabajadores * disruptiva + participantes +
#     estudiantes, family = binomial(link = logit), data = df3)
#
# Deviance Residuals:
#     Min       1Q   Median       3Q      Max
# -2.8073  -1.2662   0.7919   1.0324   1.4543
#
# Coefficients:
#                          Estimate Std. Error z value Pr(>|z|)
# (Intercept)            -1.028e-01  7.593e-02  -1.353   0.1760
# trabajadores           -5.280e-01  1.268e-01  -4.163 3.14e-05 ***
# disruptiva              4.537e-01  9.759e-02   4.649 3.33e-06 ***
# participantes           5.871e-06  2.643e-06   2.221   0.0263 *
# estudiantes             6.476e-01  8.696e-02   7.448 9.50e-14 ***
# trabajadores:disruptiva  3.834e-01  1.701e-01   2.253   0.0242 *
# ---
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# (Dispersion parameter for binomial family taken to be 1)
#
#     Null deviance: 3736.6  on 2732  degrees of freedom
# Residual deviance: 3604.0  on 2727  degrees of freedom
# AIC: 3616
#
# Number of Fisher Scoring iterations: 4
```

# Regresión Logística

Veamos la intepretación:

```
exp(summary(logit2)$coefficients[,1])
```

```
#          (Intercept)          trabajadores            disruptiva
#            0.9023479             0.5897822             1.5741600
#         participantes           estudiantes trabajadores:disruptiva
#            1.0000059             1.9109697             1.4672480
```

El coeficiente asociado al grupo social de los trabajadores indica que las odds de un control violento de la protesta en presencia de este grupo social disminuye en un 0.59 las odds respecto de que no haya trabajadores presentes, con un efecto significativo al 99,9% de confianza, controlando por todas las covariables. El coeficiente asociado a las tácticas disruptivas de un control violento de la protesta aumenta en un 1.57 odds respecto de la no presencia de este tipo de tácticas, controlando por todas las covariables. En tercer lugar, el coeficiente asociado a al número de participantes nos indica que por cada participante adicional durante una manifestación las odds de que un control violento de la protesta aumenta en 1.00 puntos, con un efecto significativo de un 95 % de confianza, controlando por el resto de las covariables del modelo.

El coeficiente asociado a estudiantes indica que las odds de presencia de estudiante durante un control violento de la protesta aumenta un 1.91 las odds respecto de la no presencia de este grupo social en una manifestación. Esto tiene un efecto significativo al 99% de confianza, controlando por todas las covariables. Finalmente, el coeficiente de la interacción entre trabajdores y tácticas disruptivas nos indica que las odds sobre la presencia de trabajadores y el uso de tácticas disruptivas durante un control violento de la protesta aumenta en 1.47, estadísticamente significativo a un 95% de confianza, controlando por el resto de nuestras variables independientes del modelo.

# Regresión Logística - AME, MEM, MER

El efecto marginal en la media (MEM) se calcula ajustando los valores de todas las covariables a sus medias dentro de la muestra. Es decir, el MEM es el efecto parcial de la variable dependiente ($Y$) condicionado a un regresor ($X$) después de establecer todas las demás covariables en sus medias. En otras palabras, el MEM es la diferencia en el efecto de $X$ sobre $Y$ cuando todas las demás covariables están en su media.

```
# margins package
library(margins) # AME, EM y MER
library(marginaleffects) #M AME, EM y MER

AME<-summary(margins(logit2))

AME
```

```
#        factor     AME     SE       z      p   lower   upper
#    disruptiva  0.1363 0.0179  7.6015 0.0000  0.1012  0.1714
#   estudiantes  0.1513 0.0195  7.7490 0.0000  0.1130  0.1896
# participantes  0.0000 0.0000  2.2282 0.0259  0.0000  0.0000
#   trabajadores -0.0748 0.0194 -3.8527 0.0001 -0.1129 -0.0368
```

Lo que obtenemos es el efecto promedio de las variables del modelo.

# Regresión Logística - AME, MEM, MER

El MEM (efecto marginal a la media) es muy similar al AME, excepto que en lugar de mantenerse en sus valores observados, las covariables se mantienen en sus valores medios.

```
df4 <- df3 %>% mutate(disruptiva = 1*(disruptiva==1)) # esto ya se encuentra así, pero es importante hacerlo en caso de tener o
logit_disruptiva <- glm(apolicial ~ disruptiva + trabajadores, family=binomial(link="logit"),
                             data=df4)
mean_disruptiva = mean(df3$disruptiva);  mean_trabajadores = mean(df3$trabajadores)
c("Promedio trab"=mean_trabajadores, "Promedio dis"=mean_disruptiva)
```

```
# Promedio trab  Promedio dis
#      0.334431      0.553970
```

```
mem<-summary(margins(logit_disruptiva, at=list(disruptiva=mean_disruptiva,trabajadores=mean_trabajadores)))

MEM<-marginaleffects(logit_disruptiva, newdata =  typical(trabajadores=mean_trabajadores, disruptiva=mean_disruptiva))

mem
```

```
#        factor disruptiva trabajadores     AME     SE       z      p   lower
#    disruptiva     0.5540       0.3344  0.1303 0.0192  6.7838 0.0000  0.0927
#  trabajadores     0.5540       0.3344 -0.0799 0.0202 -3.9535 0.0001 -0.1195
#    upper
#   0.1680
#  -0.0403
```

# Regresión Logística - AME, MEM, MER

Así mismo, podemos obtener los efectos para casos específicos que estamos buscando investigar. Tratemos de observar las diferencias por quartiles de acuerdo a la variable participantes.

```
#Modelo
logit_participantes <- glm(apolicial ~ disruptiva + participantes, family=binomial(link="logit"), data=df3)

mer <-margins(logit_participantes,at=list(disruptiva=c(1,0),participantes=quantile(df3$participantes,p=c(0.25,0.5,0.75))))

mer
```

```
# Average marginal effects at specified values


# glm(formula = apolicial ~ disruptiva + participantes, family = binomial(link = "logit"),     data = df3)


#  at(disruptiva) at(participantes) disruptiva participantes
#              0                40     0.1338     1.892e-06
#              1                40     0.1261     1.783e-06
#              0               150     0.1338     1.892e-06
#              1               150     0.1261     1.782e-06
#              0              1000     0.1338     1.892e-06
#              1              1000     0.1259     1.780e-06
```

Lo que podemos observar es la diferencia respecto de 6 casos específicos según criterior que estamos dando para investigar. Para este caso en específico, observamos las diferencias en distintos quartiles para los casos de presencia de táctica disruptiva o no.