

# Metodología I

## Magíster en Ciencias Sociales

Pablo Pérez Ahumada  
Universidad de Chile  
Departamento de sociología

# **Módulo 3**

## **Regresión lineal y logística binaria**

# REGRESIÓN LOGÍSTICA BINARIA

# ¿Qué es la regresión logística binaria?

- Tipo de regresión utilizada cuando la **variable dependiente** es **dicotómica** (dummy)
  - ¿Ha participado en una marcha en los últimos 12 meses? 1 = sí, 0 = no
  - ¿Se ha contagiado de COVID alguna vez? 1 = sí, 0 = no
  - ¿Practica deportes al menos una vez por semana? 1 = sí, 0 = no
  - Etc.
- Cuando tenemos este tipo de variables, la estimación estadística es, esencialmente, una **estimación de probabilidades** (ej., probabilidad de contagiarse de COVID)
- Al igual que en una regresión lineal (OLS), en una regresión logística las variables independientes pueden ser nominales, ordinales, o de intervalo/razón

## Ejemplo: Encuesta Mundial de Valores (Chile, 2006 – 2018)

- ¿Existe una relación entre la afiliación a sindicatos y la participación en marchas?

## Ejemplo: Encuesta Mundial de Valores (Chile, 2006 – 2018)

- ¿Existe una relación entre la afiliación a sindicatos y la participación en marchas?

**Participación en marchas según afiliación sindical**

<i>demonstr_dummy</i>	<i>Unionized</i>		<b><i>Total</i></b>
	No	Sí	
No	941 80.6 %	214 74 %	1155 79.3 %
Sí	227 19.4 %	75 26 %	302 20.7 %
<b><i>Total</i></b>	1168 100 %	289 100 %	1457 100 %

$$\chi^2=5.598 \cdot df=1 \cdot \phi=0.064 \cdot p=0.018$$

# Modelando una relación con variable dependiente dicotómica

- ¿Es posible llegar a la misma conclusión, pero controlando por otros factores relevantes?
- ¿Es posible aprovechar las herramientas de un modelo de regresión convencional (ej., obtener un coeficiente  $\beta$  que muestre el “efecto” de una variable X sobre una variable Y)?
- En principio, eso se *podría* hacer con una regresión lineal (MCO)
  - Esto se conoce como [Modelo de Probabilidad Lineal](#)

# MODELO DE PROBABILIDAD LINEAL



# Modelo de Probabilidad Lineal

- Utilización de una recta de regresión de MCO para estimar la probabilidad de ocurrencia de un fenómeno (en este caso, participar en marchas)
- ¿Qué nos dice el coeficiente  $\beta$  para “Sindicalizado/a”?

```
=====
                        Modelo de Prob Lineal
-----
(Intercept)      0.167 ***
                  (0.022)
Unionized         0.072 **
                  (0.027)
WaveWave 6       0.079 **
                  (0.027)
WaveWave 7       -0.005
                  (0.027)
-----
R^2              0.013
Adj. R^2         0.011
Num. obs.        1457
=====
*** p < 0.001; ** p < 0.01; * p < 0.05; † p < 0.1
```

# Modelo de Probabilidad Lineal: problemas

1. Se pueden incumplir dos supuestos de la regresión MCO:
  1. Los residuos pueden presentar problemas de **heterocedasticidad** (su varianza no es constante). Con ello, el coeficiente es insesgado pero su error estándar será incorrecto
  2. Los **residuos** pueden no estar distribuidos normalmente (lo cual afecta a la prueba T y a la estimación de los intervalos de confianza)

# Modelo de Probabilidad Lineal: problemas

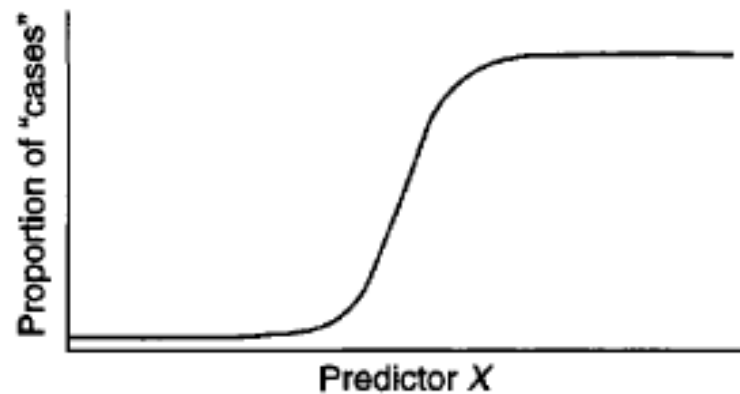
1. Se pueden incumplir dos supuestos de la regresión MCO:
  1. Los residuos pueden presentar problemas de **heterocedasticidad** (su varianza no es constante). Con ello, el coeficiente es insesgado pero su error estándar será incorrecto
  2. Los **residuos** pueden no estar distribuidos normalmente (lo cual afecta a la prueba T y a la estimación de los intervalos de confianza)
2. Además, **valores predichos** por la recta pueden caer **fuera del rango** de una probabilidad (ej., pueden ser mayores a 1).
  - Cuando esto pasa, el modelo producido no es un buen estimador de la probabilidad a nivel poblacional
  - ¿Por qué ocurre esto? Porque la distribución de probabilidades *no* es lineal

## Diferencia entre una predicción lineal y una de probabilidades

(A) For a continuous outcome variable  $Y$ , the numerical value of  $Y$  at each value of  $X$ .



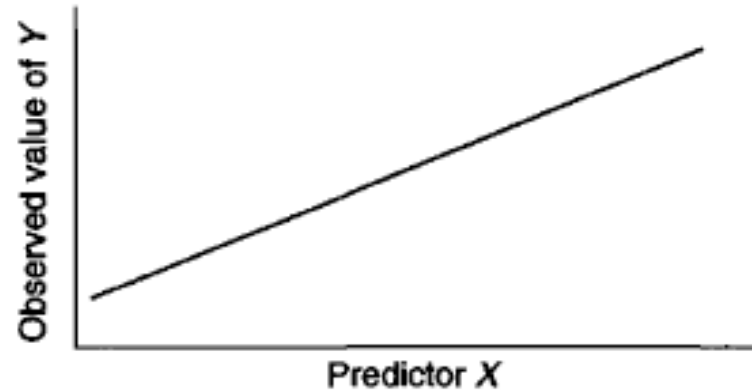
(B) For a binary outcome variable, the proportion of individuals who are "cases" (exhibit a particular outcome property) at each value of  $X$ .



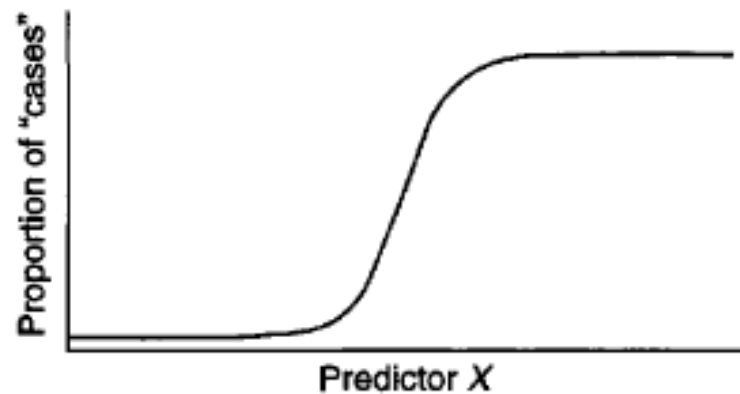
Jacob Cohen, et al. (2014). Applied multiple regression/correlation analysis for the behavioral sciences. Third edition. LEA Publishers, p. 482.

## Diferencia entre una predicción lineal y una de probabilidades

(A) For a continuous outcome variable  $Y$ , the numerical value of  $Y$  at each value of  $X$ .



(B) For a binary outcome variable, the proportion of individuals who are “cases” (exhibit a particular outcome property) at each value of  $X$ .



En este gráfico, la tasa del cambio de  $Y$  respecto a los cambios de  $X$  *no* es siempre la misma

# MODELO DE REGRESIÓN LOGÍSTICA

# Regresión logística

- La regresión logística resuelve estos problemas.
- Para ello, su **punto de partida** es transformación de los coeficientes  $\beta$  en **coeficientes *logit***

# Coeficientes *logit* y *odds* en la regresión logística

- Se conoce como “logit” a la transformación logarítmica de los *odds* (traducidos comúnmente como “chances”)
- ¿Qué son los *odds*? Una razón de probabilidades



# Coeficientes *logit* y *odds* en la regresión logística

- Por lo tanto, para estimar probabilidades a través de una regresión logística hay que seguir estos pasos
  1. Estimar los *odds* o razón de probabilidades
  2. Estimar *odds ratios* (razones entre odds)
  3. Aplicar una transformación logarítmica a esos odds ratios para obtener coeficientes *logit*
  4. Calcular las *probabilidades*

# 1) Odds

- Se define como la probabilidad de que ocurra un evento dividido por la probabilidad de que dicho evento no ocurra

$$Odds = \frac{p}{1 - p}$$

# Ejemplo participación en marchas

## Participación en marchas según afiliación sindical

<i>demonstr_dummy</i>	<i>Unionized</i>		<b><i>Total</i></b>
	No	Sí	
No	941 80.6 %	214 74 %	1155 79.3 %
Sí	227 19.4 %	75 26 %	302 20.7 %
<b><i>Total</i></b>	1168 100 %	289 100 %	1457 100 %

$$\chi^2=5.598 \cdot df=1 \cdot \phi=0.064 \cdot p=0.018$$

**Participación en marchas según afiliación sindical**

<i>demonstr_dummy</i>	<i>Unionized</i>		<b><i>Total</i></b>
	No	Sí	
No	941 80.6 %	214 74 %	1155 79.3 %
Sí	227 19.4 %	75 26 %	302 20.7 %
<b><i>Total</i></b>	1168 100 %	289 100 %	1457 100 %

$$\chi^2=5.598 \cdot df=1 \cdot \phi=0.064 \cdot p=0.018$$

$$Odds_{participar} = \frac{0,207}{0,793} = 0,26$$

Las chances de participar en una marcha son de 0,26, respecto a las chances de no participar

### Participación en marchas según afiliación sindical

<i>demonstr_dummy</i>	<i>Unionized</i>		<b><i>Total</i></b>
	No	Sí	
No	941 80.6 %	214 74 %	1155 79.3 %
Sí	227 19.4 %	75 26 %	302 20.7 %
<b><i>Total</i></b>	1168 100 %	289 100 %	1457 100 %

$$\chi^2 = 5.598 \cdot df = 1 \cdot \phi = 0.064 \cdot p = 0.018$$

$$Odds_{participar} = \frac{0,207}{0,793} = 0,26$$

Las chances de participar en una marcha son de 0,26, respecto a las chances de no participar

**En otras palabras:** por cada 1 persona, hay sólo 0,26 personas que participan en marchas.

O más **intuitivamente**, por cada 100 personas, hay sólo 26 personas que participan

**Participación en marchas según afiliación sindical**

<i>demonstr_dummy</i>	<i>Unionized</i>		<b><i>Total</i></b>
	No	Sí	
No	941 80.6 %	214 74 %	1155 79.3 %
Sí	227 19.4 %	75 26 %	302 20.7 %
<b><i>Total</i></b>	1168 100 %	289 100 %	1457 100 %

$$\chi^2=5.598 \cdot df=1 \cdot \phi=0.064 \cdot p=0.018$$

¿Cambian las chances de participar según se esté afiliado/a a un sindicato?

**Participación en marchas según afiliación sindical**

<i>demonstr_dummy</i>	<i>Unionized</i>		<b><i>Total</i></b>
	No	Sí	
No	941 80.6 %	214 74 %	1155 79.3 %
Sí	227 19.4 %	75 26 %	302 20.7 %
<b><i>Total</i></b>	1168 100 %	289 100 %	1457 100 %

$$\chi^2 = 5.598 \cdot df = 1 \cdot \phi = 0.064 \cdot p = 0.018$$

¿Cambian las chances de participar según se esté afiliado/a a un sindicato?

$$Odds_{sindical} = \frac{0,26}{0,74} = 0,35$$

$$Odds_{Nosindical} = \frac{0,194}{0,806} = 0,24$$

# Implicancias para la interpretación de los odds

- Valores bajo 1 indican que las chances de que ocurra un evento son negativas
- Valores iguales a 1 indican chances iguales
- Valores sobre 1 indican chances positivas



# Implicancias para la interpretación de los odds

- Valores bajo 1 indican que las chances de que ocurra un evento son negativas
- Valores iguales a 1 indican chances iguales
- Valores sobre 1 indican chances positivas
  - Ejemplo, imaginemos que el 65% de los/as afiliados/as a sindicatos ha participado en marchas, y sólo el 35% no lo hecho.

# Implicancias para la interpretación de los odds

- Valores bajo 1 indican que las chances de que ocurra un evento son negativas
- Valores iguales a 1 indican chances iguales
- Valores sobre 1 indican chances positivas
  - Ejemplo, imaginemos que el 65% de los/as afiliados/as a sindicatos ha participado en marchas, y sólo el 35% no lo hecho.

$$Odds = \frac{p}{1 - p}$$

$$Odds_{sindical} = \frac{0,65}{0,35} = 1,86$$

# Implicancias para la interpretación de los odds

- Valores bajo 1 indican que las chances de que ocurra un evento son negativas
- Valores iguales a 1 indican chances iguales
- Valores sobre 1 indican chances positivas
  - Ejemplo, imaginemos que el 65% de los/as afiliados/as a sindicatos ha participado en marchas, y sólo el 35% no lo hecho.

$$Odds = \frac{p}{1 - p}$$

$$Odds_{sindical} = \frac{0,65}{0,35} = 1,86$$

Por cada 100 personas sindicalizadas, hay 186 que participan

## 2) Odds ratios (razones de chances)

- Cálculo que permite reflejar **asociación** entre dos variables dicotómicas, a partir de una comparación entre chances
- Siguiendo con el ejemplo anterior, ¿tienen los/as sindicalizados más chances de participar en marchas *que quienes no están sindicalizados/as*?

## 2) Odds ratios

- Cálculo que permite reflejar asociación entre dos variables dicotómicas, a partir de una comparación entre chances
- Siguiendo con el ejemplo anterior, ¿tienen los/as sindicalizados más chances de participar en marchas *que quienes no están sindicalizados/as*?

$$OR = \frac{p_{sindical}/(1 - p_{sindical})}{p_{NOsindical}/(1 - p_{NOsindical})}$$

## 2) Odds ratios

- Cálculo que permite reflejar asociación entre dos variables dicotómicas, a partir de una comparación entre chances
- Siguiendo con el ejemplo anterior, ¿tienen los/as sindicalizados más chances de participar en marchas *que quienes no están sindicalizados/as*?

$$OR = \frac{p_{sindical}/(1 - p_{sindical})}{p_{NOsindical}/(1 - p_{NOsindical})}$$

$$OR = \frac{0,26/0,74}{0,194/0,806} = \frac{0,35}{0,24} = 1,46$$

Las chances de participar en marchas de los/as sindicalizados/as son 1,5 veces más que las de quienes no están sindicalizados/as

## 2) Odds ratios: implicancias

- El odds ratio o razones de chances es útil porque nos permite expresar en un número la relación entre dos variables categóricas
- En las regresiones logísticas, el odds ratio es la primera manera de aproximarnos a relación entre variables
- Sin embargo, falta un paso más necesario para construir modelos de regresión logística

### 3) Logit

- Es una unidad de medida de la relación entre dos variables (VD: dicotómica), que en regresión logística se calcula a partir del *logaritmo natural de los odds*



### 3) Logit

- Es una unidad de medida de la relación entre dos variables (VD: dicotómica), que en regresión logística se calcula a partir del *logaritmo natural de los odds*
- Esta transformación logarítmica es la base de la estimación de parámetros en la regresión logística:
  - Con ella se puede modelar la probabilidad de ocurrencia de un fenómeno como la función logística de una *combinación lineal de las variables independientes o predictores*

$$\ln\left(\frac{p}{1-p}\right) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n$$

### 3) Logit

- Es una unidad de medida de la relación entre dos variables (VD: dicotómica), que en regresión logística se calcula a partir del *logaritmo natural de los odds*
- Esta transformación logarítmica es la base de la estimación de parámetros en la regresión logística:
  - Con ella se puede modelar la probabilidad de ocurrencia de un fenómeno como la función logística de una *combinación lineal de las variables independientes o predictores*

$$\ln\left(\frac{p}{1-p}\right) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n$$

La *mejor combinación lineal* de predictores no se obtiene a través de MCO, sino a través del procedimiento de **máxima verosimilitud**

### 3) Logit

- La estimación de coeficientes *logit* es un gran avance, porque permite obtener coeficientes similares al  $\beta$  (beta) en la regresión MCO
- De hecho, la mayoría de los softwares estadísticos reportan resultados en coeficientes logit (también llamados *log odds*)
- A diferencia de los odds ratio, los coeficientes logit tienen valores que van de  $-\infty$  a  $+\infty$ 
  - Así, una relación negativa puede ser directamente interpretable (por el signo)

## REGRESIÓN LOGÍSTICA: ejemplo

# Ejemplo regresión logística (resultados en log odds)

- ¿Existe una relación estadísticamente significativa entre afiliación sindical y participación en marchas?
- Datos: Encuesta Mundial de Valores (Chile 2006, 2012 y 2018)
  - Variable dependiente: ¿Ha participado en marchas en los últimos 12 meses? 1 = sí; 0 = no
  - Variable independiente de interés: ¿Está afiliado/a a un sindicato?
  - Controles demográficos: género (female), edad en años (X003), educación (3 tramos), sector privado (private\_sector) y nivel de politización medido en escala de 0 a 6 (politicization)
  - Otro control: ola de aplicación de la encuesta. Wave: 2006, 2012, 2018

¿Existe una relación estadísticamente significativa entre afiliación sindical y participación en marchas?

	M1 (m prob lineal)	M1 (log odds)	M2 (log odds)	M3 (log odds)
(Intercept)	0.167 *** (0.022)	-1.601 *** (0.140)	-1.023 * (0.424)	-1.379 ** (0.437)
Unionized	0.072 ** (0.027)	0.418 ** (0.155)	0.390 * (0.157)	0.312 † (0.161)
WaveWave 6	0.079 ** (0.027)	0.470 ** (0.168)	0.498 ** (0.171)	0.496 ** (0.174)
WaveWave 7	-0.005 (0.027)	-0.031 (0.174)	-0.102 (0.181)	-0.121 (0.184)
Female			-0.020 (0.136)	0.100 (0.139)
X003			-0.007 (0.006)	-0.012 * (0.006)
Educ2			0.042 (0.268)	-0.053 (0.272)
Educ3			0.521 † (0.283)	0.259 (0.289)
private_sector			-0.516 ** (0.175)	-0.445 * (0.180)
politicization				0.282 *** (0.042)
R^2	0.013			
Adj. R^2	0.011			
Num. obs.	1457	1457	1457	1457
AIC		1475.887	1460.059	1415.850
BIC		1497.024	1507.616	1468.691
Log Likelihood		-733.944	-721.030	-697.925
Deviance		1467.887	1442.059	1395.850
*** p < 0.001; ** p < 0.01; * p < 0.05; † p < 0.1				

¿Existe una relación estadísticamente significativa entre afiliación sindical y participación en marchas?

Ej. Modelo 2:

	M1 (m prob lineal)	M1 (log odds)	M2 (log odds)	M3 (log odds)
(Intercept)	0.167 *** (0.022)	-1.601 *** (0.140)	-1.023 * (0.424)	-1.379 ** (0.437)
Unionized	0.072 ** (0.027)	0.418 ** (0.155)	0.390 * (0.157)	0.312 † (0.161)
WaveWave 6	0.079 ** (0.027)	0.470 ** (0.168)	0.498 ** (0.171)	0.496 ** (0.174)
WaveWave 7	-0.005 (0.027)	-0.031 (0.174)	-0.102 (0.181)	-0.121 (0.184)
Female			-0.020 (0.136)	0.100 (0.139)
X003			-0.007 (0.006)	-0.012 * (0.006)
Educ2			0.042 (0.268)	-0.053 (0.272)
Educ3			0.521 † (0.283)	0.259 (0.289)
private_sector			-0.516 ** (0.175)	-0.445 * (0.180)
politicization				0.282 *** (0.042)
R^2	0.013			
Adj. R^2	0.011			
Num. obs.	1457	1457	1457	1457
AIC		1475.887	1460.059	1415.850
BIC		1497.024	1507.616	1468.691
Log Likelihood		-733.944	-721.030	-697.925
Deviance		1467.887	1442.059	1395.850
*** p < 0.001; ** p < 0.01; * p < 0.05; † p < 0.1				

¿Existe una relación estadísticamente significativa entre afiliación sindical y participación en marchas?

### Ej. Modelo 2:

- En comparación a los no sindicalizados (categoría de referencia), el log-odds de participación en marchas para afiliados a sindicatos aumenta en 0,39 ( $p < 0,05$ )

	M1 (m prob lineal)	M1 (log odds)	M2 (log odds)	M3 (log odds)
(Intercept)	0.167 *** (0.022)	-1.601 *** (0.140)	-1.023 * (0.424)	-1.379 ** (0.437)
Unionized	0.072 ** (0.027)	0.418 ** (0.155)	0.390 * (0.157)	0.312 † (0.161)
WaveWave 6	0.079 ** (0.027)	0.470 ** (0.168)	0.498 ** (0.171)	0.496 ** (0.174)
WaveWave 7	-0.005 (0.027)	-0.031 (0.174)	-0.102 (0.181)	-0.121 (0.184)
Female			-0.020 (0.136)	0.100 (0.139)
X003			-0.007 (0.006)	-0.012 * (0.006)
Educ2			0.042 (0.268)	-0.053 (0.272)
Educ3			0.521 † (0.283)	0.259 (0.289)
private_sector			-0.516 ** (0.175)	-0.445 * (0.180)
politicization				0.282 *** (0.042)
R^2	0.013			
Adj. R^2	0.011			
Num. obs.	1457	1457	1457	1457
AIC		1475.887	1460.059	1415.850
BIC		1497.024	1507.616	1468.691
Log Likelihood		-733.944	-721.030	-697.925
Deviance		1467.887	1442.059	1395.850

\*\*\*  $p < 0.001$ ; \*\*  $p < 0.01$ ; \*  $p < 0.05$ ; †  $p < 0.1$



¿Existe una relación estadísticamente significativa entre afiliación sindical y participación en marchas?

Ej. Modelo 2:

- En comparación a los/as trabajadores/as del sector público (categoría de referencia), el log-odds de participación en marchas para los/as del sector privado *disminuye* en 0,52 (valor – p < 0,01)

	M1 (m prob lineal)	M1 (log odds)	M2 (log odds)	M3 (log odds)
(Intercept)	0.167 *** (0.022)	-1.601 *** (0.140)	-1.023 * (0.424)	-1.379 ** (0.437)
Unionized	0.072 ** (0.027)	0.418 ** (0.155)	0.390 * (0.157)	0.312 † (0.161)
WaveWave 6	0.079 ** (0.027)	0.470 ** (0.168)	0.498 ** (0.171)	0.496 ** (0.174)
WaveWave 7	-0.005 (0.027)	-0.031 (0.174)	-0.102 (0.181)	-0.121 (0.184)
Female			-0.020 (0.136)	0.100 (0.139)
X003			-0.007 (0.006)	-0.012 * (0.006)
Educ2			0.042 (0.268)	-0.053 (0.272)
Educ3			0.521 † (0.283)	0.259 (0.289)
private_sector			-0.516 ** (0.175)	-0.445 * (0.180)
politicization				0.282 *** (0.042)
R^2	0.013			
Adj. R^2	0.011			
Num. obs.	1457	1457	1457	1457
AIC		1475.887	1460.059	1415.850
BIC		1497.024	1507.616	1468.691
Log Likelihood		-733.944	-721.030	-697.925
Deviance		1467.887	1442.059	1395.850
*** p < 0.001; ** p < 0.01; * p < 0.05; † p < 0.1				

¿Existe una relación estadísticamente significativa entre afiliación sindical y participación en marchas?

Ej. Modelo 3

Politización: variable de intervalo  
¿Cómo interpretarla?

	M1 (m prob lineal)	M1 (log odds)	M2 (log odds)	M3 (log odds)
(Intercept)	0.167 *** (0.022)	-1.601 *** (0.140)	-1.023 * (0.424)	-1.379 ** (0.437)
Unionized	0.072 ** (0.027)	0.418 ** (0.155)	0.390 * (0.157)	0.312 † (0.161)
WaveWave 6	0.079 ** (0.027)	0.470 ** (0.168)	0.498 ** (0.171)	0.496 ** (0.174)
WaveWave 7	-0.005 (0.027)	-0.031 (0.174)	-0.102 (0.181)	-0.121 (0.184)
Female			-0.020 (0.136)	0.100 (0.139)
X003			-0.007 (0.006)	-0.012 * (0.006)
Educ2			0.042 (0.268)	-0.053 (0.272)
Educ3			0.521 † (0.283)	0.259 (0.289)
private_sector			-0.516 ** (0.175)	-0.445 * (0.180)
politicization				0.282 *** (0.042)
R^2	0.013			
Adj. R^2	0.011			
Num. obs.	1457	1457	1457	1457
AIC		1475.887	1460.059	1415.850
BIC		1497.024	1507.616	1468.691
Log Likelihood		-733.944	-721.030	-697.925
Deviance		1467.887	1442.059	1395.850

\*\*\* p < 0.001; \*\* p < 0.01; \* p < 0.05; † p < 0.1

¿Existe una relación estadísticamente significativa entre afiliación sindical y participación en marchas?

Ej. Modelo 3

Politización: variable de intervalo  
¿Cómo interpretarla?

“Por cada unidad de aumento en la escala de politización, el log-odds de participación en marchas aumenta en 0,28 (valor- $p < 0,001$ )”.

	M1 (m prob lineal)	M1 (log odds)	M2 (log odds)	M3 (log odds)
(Intercept)	0.167 *** (0.022)	-1.601 *** (0.140)	-1.023 * (0.424)	-1.379 ** (0.437)
Unionized	0.072 ** (0.027)	0.418 ** (0.155)	0.390 * (0.157)	0.312 † (0.161)
WaveWave 6	0.079 ** (0.027)	0.470 ** (0.168)	0.498 ** (0.171)	0.496 ** (0.174)
WaveWave 7	-0.005 (0.027)	-0.031 (0.174)	-0.102 (0.181)	-0.121 (0.184)
Female			-0.020 (0.136)	0.100 (0.139)
X003			-0.007 (0.006)	-0.012 * (0.006)
Educ2			0.042 (0.268)	-0.053 (0.272)
Educ3			0.521 † (0.283)	0.259 (0.289)
private_sector			-0.516 ** (0.175)	-0.445 * (0.180)
politicization				0.282 *** (0.042)
R^2	0.013			
Adj. R^2	0.011			
Num. obs.	1457	1457	1457	1457
AIC		1475.887	1460.059	1415.850
BIC		1497.024	1507.616	1468.691
Log Likelihood		-733.944	-721.030	-697.925
Deviance		1467.887	1442.059	1395.850
*** p < 0.001; ** p < 0.01; * p < 0.05; † p < 0.1				

Medidas de bondad de ajuste de los modelos  
(hablaremos sobre ellas más adelante)

	M1 (m prob lineal)	M1 (log odds)	M2 (log odds)	M3 (log odds)
(Intercept)	0.167 *** (0.022)	-1.601 *** (0.140)	-1.023 * (0.424)	-1.379 ** (0.437)
Unionized	0.072 ** (0.027)	0.418 ** (0.155)	0.390 * (0.157)	0.312 † (0.161)
WaveWave 6	0.079 ** (0.027)	0.470 ** (0.168)	0.498 ** (0.171)	0.496 ** (0.174)
WaveWave 7	-0.005 (0.027)	-0.031 (0.174)	-0.102 (0.181)	-0.121 (0.184)
Female			-0.020 (0.136)	0.100 (0.139)
X003			-0.007 (0.006)	-0.012 * (0.006)
Educ2			0.042 (0.268)	-0.053 (0.272)
Educ3			0.521 † (0.283)	0.259 (0.289)
private_sector			-0.516 ** (0.175)	-0.445 * (0.180)
politicization				0.282 *** (0.042)
R^2	0.013			
Adj. R^2	0.011			
Num. obs.	1457	1457	1457	1457
AIC		1475.887	1460.059	1415.850
BIC		1497.024	1507.616	1468.691
Log Likelihood		-733.944	-721.030	-697.925
Deviance		1467.887	1442.059	1395.850

\*\*\* p < 0.001; \*\* p < 0.01; \* p < 0.05; † p < 0.1

# Resultado M1 (log odds, comando *summary*)

```
> summary(m1log)

Call:
glm(formula = demonstr_dummy ~ Unionized + Wave, family = binomial(link = "logit"),
    data = WVS_2005_2022_Ch1_log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.8930  -0.7478  -0.6062  -0.5976   1.9030

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.60103    0.14045  -11.399  < 2e-16 ***
Unionized    0.41774    0.15458   2.702  0.00689 **
WaveWave 6   0.46979    0.16849   2.788  0.00530 **
WaveWave 7  -0.03108    0.17372  -0.179  0.85803
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1487.1  on 1456  degrees of freedom
Residual deviance: 1467.9  on 1453  degrees of freedom
AIC: 1475.9

Number of Fisher Scoring iterations: 4
```

# Log-odds – problemas de interpretación

- A pesar de sus ventajas, los coeficientes logit son difíciles de interpretar
  - Los coef. logit son el resultado de una transformación de la escala original
  - Ellos *no* muestran directamente probabilidades (éstas siempre tienen valores entre 0 y 1)

# Log-odds – problemas de interpretación

- ¿Qué hacer?
  - Volver a la escala original; *odds ratios*
- ¿Cómo?
  - Mediante la *exponenciación de los coeficientes* (la función exponencial es la inversa del logaritmo)

# Paso de log-odds a odds ratios

	M1 (m prob lineal)	M1 (log odds)	M2 (log odds)	M3 (log odds)
(Intercept)	0.167 *** (0.022)	-1.601 *** (0.140)	-1.023 * (0.424)	-1.379 ** (0.437)
Unionized	0.072 ** (0.027)	0.418 ** (0.155)	0.390 * (0.157)	0.312 † (0.161)
WaveWave 6	0.079 ** (0.027)	0.470 ** (0.168)	0.498 ** (0.171)	0.496 ** (0.174)
WaveWave 7	-0.005 (0.027)	-0.031 (0.174)	-0.102 (0.181)	-0.121 (0.184)
Female			-0.020 (0.136)	0.100 (0.139)
X003			-0.007 (0.006)	-0.012 * (0.006)
Educ2			0.042 (0.268)	-0.053 (0.272)
Educ3			0.521 † (0.283)	0.259 (0.289)
private_sector			-0.516 ** (0.175)	-0.445 * (0.180)
politicization				0.282 *** (0.042)
-----				
R^2	0.013			
Adj. R^2	0.011			
Num. obs.	1457	1457	1457	1457
AIC		1475.887	1460.059	1415.850
BIC		1497.024	1507.616	1468.691
Log Likelihood		-733.944	-721.030	-697.925
Deviance		1467.887	1442.059	1395.850
=====				
*** p < 0.001; ** p < 0.01; * p < 0.05; † p < 0.1				

$$\text{logit}_x = \log(\text{Odds}_x)$$

$$e^{\text{logit}} = \text{Odds}_x$$

$$e^{0.39} = \text{Odds}_x = 1.477$$



# Paso de log-odds a odds ratios

	M1 (m prob lineal)	M1 (log odds)	M2 (log odds)	M3 (log odds)
(Intercept)	0.167 *** (0.022)	-1.601 *** (0.140)	-1.023 * (0.424)	-1.379 ** (0.437)
Unionized	0.072 ** (0.027)	0.418 ** (0.155)	0.390 * (0.157)	0.312 † (0.161)
WaveWave 6	0.079 ** (0.027)	0.470 ** (0.168)	0.498 ** (0.171)	0.496 ** (0.174)
WaveWave 7	-0.005 (0.027)	-0.031 (0.174)	-0.102 (0.181)	-0.121 (0.184)
Female			-0.020 (0.136)	0.100 (0.139)
X003			-0.007 (0.006)	-0.012 * (0.006)
Educ2			0.042 (0.268)	-0.053 (0.272)
Educ3			0.521 † (0.283)	0.259 (0.289)
private_sector			-0.516 ** (0.175)	-0.445 * (0.180)
politicization				0.282 *** (0.042)
-----				
R^2	0.013			
Adj. R^2	0.011			
Num. obs.	1457	1457	1457	1457
AIC		1475.887	1460.059	1415.850
BIC		1497.024	1507.616	1468.691
Log Likelihood		-733.944	-721.030	-697.925
Deviance		1467.887	1442.059	1395.850
=====				
*** p < 0.001; ** p < 0.01; * p < 0.05; † p < 0.1				

$$\text{logit}_x = \log(\text{Odds})$$

$$e^{\text{logit}} = \text{Odds}_x$$

$$e^{0,39} = \text{Odds}_x = 1,477$$

Las chances (odds) de participar en marchas de los/as sindicalizados/as son 1,5 veces más que las de quienes no están sindicalizados/as, *controlando por las otras variables incluidas en el modelo*

# Comparación coeficientes logit v/s odds ratios

Comando básico de R `exp(coef())`

## Log odds

	m1 (log odds)	m2 (log odds)	m3 (log odds)
(Intercept)	-1.601 *** (0.140)	-1.023 * (0.424)	-1.379 ** (0.437)
Unionized	0.418 ** (0.155)	0.390 * (0.157)	0.312 † (0.161)
WaveWave 6	0.470 ** (0.168)	0.498 ** (0.171)	0.496 ** (0.174)
WaveWave 7	-0.031 (0.174)	-0.102 (0.181)	-0.121 (0.184)
Female		-0.020 (0.136)	0.100 (0.139)
X003		-0.007 (0.006)	-0.012 * (0.006)
Educ2		0.042 (0.268)	-0.053 (0.272)
Educ3		0.521 † (0.283)	0.259 (0.289)
private_sector		-0.516 ** (0.175)	-0.445 * (0.180)
politicization			0.282 *** (0.042)
AIC	1475.887	1460.059	1415.850
BIC	1497.024	1507.616	1468.691
Log Likelihood	-733.944	-721.030	-697.925
Deviance	1467.887	1442.059	1395.850
Num. obs.	1457	1457	1457

\*\*\* p < 0.001; \*\* p < 0.01; \* p < 0.05; † p < 0.1

## Odds ratio

	m1 (OR)	m2 (OR)	m3 (OR)
(Intercept)	0.202 ***	0.359 *	0.252 **
Unionized	1.519 **	1.477 *	1.366 †
WaveWave 6	1.600 **	1.646 **	1.642 **
WaveWave 7	0.969	0.903	0.886
Female		0.980	1.106
X003		0.993	0.988 *
Educ2		1.042	0.949
Educ3		1.684 †	1.296
private_sector		0.597 **	0.641 *
politicization			1.326 ***
AIC	1475.887	1460.059	1415.850
BIC	1497.024	1507.616	1468.691
Log Likelihood	-733.944	-721.030	-697.925
Deviance	1467.887	1442.059	1395.850
Num. obs.	1457	1457	1457

\*\*\* p < 0.001; \*\* p < 0.01; \* p < 0.05; † p < 0.1

¿Qué cambia y no cambia en las tablas?

## Comparación coeficientes logit v/s odds ratios

Log odds

	m1 (log odds)	m2 (log odds)	m3 (log odds)
(Intercept)	-1.601 *** (0.140)	-1.023 * (0.424)	-1.379 ** (0.437)
Unionized	0.418 ** (0.155)	0.390 * (0.157)	0.312 † (0.161)
WaveWave 6	0.470 ** (0.168)	0.498 ** (0.171)	0.496 ** (0.174)
WaveWave 7	-0.031 (0.174)	-0.102 (0.181)	-0.121 (0.184)
Female		-0.020 (0.136)	0.100 (0.139)
X003		-0.007 (0.006)	-0.012 * (0.006)
Educ2		0.042 (0.268)	-0.053 (0.272)
Educ3		0.521 † (0.283)	0.259 (0.289)
private_sector		-0.516 ** (0.175)	-0.445 * (0.180)
politicization			0.282 *** (0.042)
AIC	1475.887	1460.059	1415.850
BIC	1497.024	1507.616	1468.691
Log Likelihood	-733.944	-721.030	-697.925
Deviance	1467.887	1442.059	1395.850
Num. obs.	1457	1457	1457

\*\*\* p < 0.001; \*\* p < 0.01; \* p < 0.05; † p < 0.1

Odds ratio

	m1 (OR)	m2 (OR)	m3 (OR)
(Intercept)	0.202 ***	0.359 *	0.252 **
Unionized	1.519 **	1.477 *	1.366 †
WaveWave 6	1.600 **	1.646 **	1.642 **
WaveWave 7	0.969	0.903	0.886
Female		0.980	1.106
X003		0.993	0.988 *
Educ2		1.042	0.949
Educ3		1.684 †	1.296
private_sector		0.597 **	0.641 *
politicization			1.326 ***
AIC	1475.887	1460.059	1415.850
BIC	1497.024	1507.616	1468.691
Log Likelihood	-733.944	-721.030	-697.925
Deviance	1467.887	1442.059	1395.850
Num. obs.	1457	1457	1457

\*\*\* p < 0.001; \*\* p < 0.01; \* p < 0.05; † p < 0.1

¿Qué cambia y no cambia en las tablas?

## Comparación coeficientes logit v/s odds ratios

Log odds

	m1 (log odds)	m2 (log odds)	m3 (log odds)
(Intercept)	-1.601 *** (0.140)	-1.023 * (0.424)	-1.379 ** (0.437)
Unionized	0.418 ** (0.155)	0.390 * (0.157)	0.312 † (0.161)
WaveWave 6	0.470 ** (0.168)	0.498 ** (0.171)	0.496 ** (0.174)
WaveWave 7	-0.031 (0.174)	-0.102 (0.181)	-0.121 (0.184)
Female		-0.020 (0.136)	0.100 (0.139)
X003		-0.007 (0.006)	-0.012 * (0.006)
Educ2		0.042 (0.268)	-0.053 (0.272)
Educ3		0.521 † (0.283)	0.259 (0.289)
private_sector		-0.516 ** (0.175)	-0.445 * (0.180)
politicization			0.282 *** (0.042)
AIC	1475.887	1460.059	1415.850
BIC	1497.024	1507.616	1468.691
Log Likelihood	-733.944	-721.030	-697.925
Deviance	1467.887	1442.059	1395.850
Num. obs.	1457	1457	1457

\*\*\* p < 0.001; \*\* p < 0.01; \* p < 0.05; † p < 0.1

Odds ratio

	m1 (OR)	m2 (OR)	m3 (OR)
(Intercept)	0.202 ***	0.359 *	0.252 **
Unionized	1.519 **	1.477 *	1.366 †
WaveWave 6	1.600 **	1.646 **	1.642 **
WaveWave 7	0.969	0.903	0.886
Female		0.980	1.106
X003		0.993	0.988 *
Educ2		1.042	0.949
Educ3		1.684 †	1.296
private_sector		0.597 **	0.641 *
politicization			1.326 ***
AIC	1475.887	1460.059	1415.850
BIC	1497.024	1507.616	1468.691
Log Likelihood	-733.944	-721.030	-697.925
Deviance	1467.887	1442.059	1395.850
Num. obs.	1457	1457	1457

\*\*\* p < 0.001; \*\* p < 0.01; \* p < 0.05; † p < 0.1

¿Qué cambia y no cambia en las tablas?

A pesar de que tenemos los elementos suficientes (coeficientes) para construir tablas de reg. logística, siguen existiendo algunas limitaciones

A pesar de que tenemos los elementos suficientes (coeficientes) para construir tablas de reg. logística, siguen existiendo algunas limitaciones





# Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It

Carina Mood

Logistic regression estimates do not behave like linear regression estimates in one important respect: They are affected by omitted variables, even when these variables are unrelated to the independent variables in the model. This fact has important implications that have gone largely unnoticed by sociologists. Importantly, we cannot straightforwardly interpret log-odds ratios or odds ratios as effect measures, because they also reflect the degree of unobserved heterogeneity in the model. In addition, we cannot compare log-odds ratios or odds ratios for similar models across groups, samples, or time points, or across models with different independent variables in a sample. This article discusses these problems and possible ways of overcoming them.

---

# Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It

Carina Mood

Logistic regression estimates do not behave like linear regression estimates in one important respect: They are affected by omitted variables, even when these variables are unrelated to the independent variables in the model. This fact has important implications that have gone largely unnoticed by sociologists. Importantly, we cannot straightforwardly interpret log-odds ratios or odds ratios as effect measures, because they also reflect the degree of unobserved heterogeneity in the model. In addition, we cannot compare log-odds ratios or odds ratios for similar models across groups, samples, or time points, or across models with different independent variables in a sample. This article discusses these problems and possible ways of overcoming them.

---

**Problema central:** Los **coeficientes** de un modelo de reg. logística (log-odds u odds-ratios) **no son comparables** con los coeficientes de otro modelo



# Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It

Carina Mood

Logistic regression estimates do not behave like linear regression estimates in one important respect: They are affected by omitted variables, even when these variables are unrelated to the independent variables in the model. This fact has important implications that have gone largely unnoticed by sociologists. Importantly, we cannot straightforwardly interpret log-odds ratios or odds ratios as effect measures, because they also reflect the degree of unobserved heterogeneity in the model. In addition, we cannot compare log-odds ratios or odds ratios for similar models across groups, samples, or time points, or across models with different independent variables in a sample. This article discusses these problems and possible ways of overcoming them.

---

Pero hay una **solución**: calcular las **probabilidades predichas** (el equivalente a los puntajes predichos en la regresión MCO)

**Problema central:** Los **coeficientes** de un modelo de reg. logística (log-odds u odds-ratios) **no son comparables** con los coeficientes de otro modelo