

Talleres de Análisis Político I

Sesión 6

18/12/2023

Pau Vall-Prat

pau.vall@uc3m.es

La semana pasada

Regresión multivariante

- Los modelos de regresión multivariante modelizan una relación lineal entre
 - Una variable dependiente
 - Dos o más variables independientes

$$Y = \alpha + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \dots + \epsilon$$

- Recordad, en la práctica

$$\hat{Y} = \hat{\alpha} + \hat{\beta}_1 \cdot x_1 + \hat{\beta}_2 \cdot x_2$$

Interpretación

- Al añadir más elementos en los modelos de regresión la interpretación de los coeficientes cambia ligeramente
 - Constante: Valor de y cuando x_1 y x_2 son iguales a 0
 - β_1 : efecto de x_1 sobre y siempre que x_2 se mantenga constante
 - β_2 : efecto de x_2 sobre y siempre que x_1 se mantenga constante
- * se mantenga constante = cláusula *ceteris paribus*
- Hay que cuidar el lenguaje y distinguir la interpretación de coeficientes en función de regresiones bivariadas o multivariantes

Ejemplo

Queremos entender por qué hay
variación en el número de
consejerías de las CCAA

H1: A mayor número de partidos,
más consejerías
*Para acomodar cargos para todos
los partidos*

H2: Los gobiernos de izquierdas
tendrán más consejerías
Tienden a gastar más

```
Call:
lm(formula = cabinet_size ~ num_parties + left_cabinet, data = rcs)
```

Residuals:

Min	1Q	Median	3Q	Max
-4.6269	-1.6269	-0.2862	1.1358	5.3731

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	10.2050	0.3061	33.343	<2e-16 ***
num_parties	0.4220	0.1872	2.254	0.0249 *
left_cabinet	0.2373	0.2504	0.948	0.3441

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.08 on 278 degrees of freedom

Multiple R-squared: 0.02088, Adjusted R-squared: 0.01384

F-statistic: 2.964 on 2 and 278 DF, p-value: 0.05324

Fuente de los datos: Vall-Prat & Rodon (2017)

Ejercicios

- Haced regresiones con la base de datos de resultados electorales de 2019 combinada con datos sociodemográficos
 1. Una regresión simple
 2. Una regresión múltiple

Efectos heterogéneos

Estimar el impacto de una variable

- Hasta ahora interpretábamos siempre los coeficientes β como la pendiente en la relación entre x e y .
- Las variables de control, modifican el pendiente de una variable
 - Ajustan por la variación entre x e y que no se debe a la influencia de x
 - El impacto de una variable siempre es constante sobre y
- Hay veces que nos puede interesar distinguir un efecto diferencial de x_1 sobre y entre diferentes grupos de x_2

Flexibilizar la asociación entre variables

- Quedaría definido de este modo

$$Y = \alpha + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_1 \times x_2 + \epsilon$$

- Recordad, en la práctica

$$\hat{Y} = \hat{\alpha} + \hat{\beta}_1 \cdot x_1 + \hat{\beta}_2 \cdot x_2 + \hat{\beta}_3 \cdot x_1 \times x_2$$

- En este escenario

- β_1 : Pendiente de x_1 cuando

$$x_2 = 0$$

- β_2 : Diferencia en el valor de \hat{y} por cada punto de aumento de x_2 cuando

$$x_1 = 0$$

- β_3 : Variación en el pendiente de x_1 respecto a β_1 para cada punto que aumenta x_2

Efectos heterogéneos

- Es el concepto para referirse a un efecto diferente para una misma variable para distintos valores de una tercera variable
- A veces es más informativo que, simplemente, controlar por una tercera variable
- La influencia de esta tercera variable debe estar justificada teóricamente
- Las variables no necesariamente deben ser dicotómicas...

Por ejemplo

- VD: N.º de consejerías en un gobierno
- Tenemos dos variables independientes binarias
 - x_1 : tener mayoría parlamentaria (1) o no, ergo, minoría (0)
 - x_2 : gobierno de izquierdas (1) de derechas (0)

$$\hat{Y} = \alpha + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_1 \cdot x_2$$

	Derechas	Izquierdas
Minoría	6	8
Mayoría	7	12

Ejemplo (1)

Call:

```
lm(formula = cabinet_size ~ majority * left_cabinet, data = rcs)
```

Residuals:

Min	1Q	Median	3Q	Max
-5.0492	-1.5758	-0.1798	1.3776	5.8202

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	11.0492	0.2675	41.306	<2e-16 ***
majority	-0.4267	0.3407	-1.252	0.2115
left_cabinet	-0.4734	0.4515	-1.049	0.2953
majority:left_cabinet	1.0307	0.5453	1.890	0.0598 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.089 on 277 degrees of freedom

Multiple R-squared: 0.0157, Adjusted R-squared: 0.005043

F-statistic: 1.473 on 3 and 277 DF, p-value: 0.2221

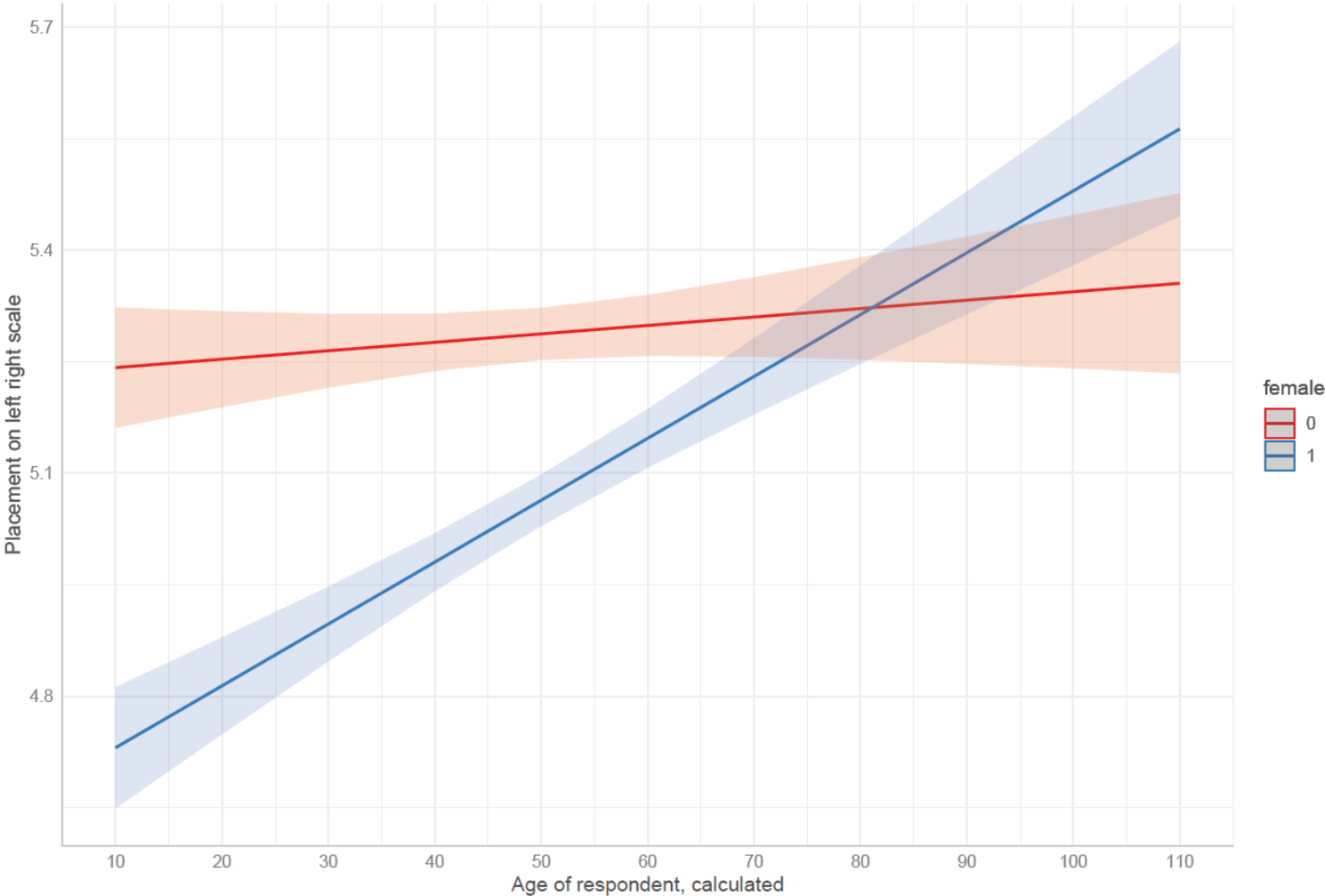
Ejemplo (2)

- Calcula la posición ideológica estimada

Sexo	Edad	Ideología
Mujer	20	
Hombre	20	
Mujer	45	
Hombre	45	
Mujer	80	
Hombre	80	

<i>Dependent variable:</i>	
	lrscale
agea	0.001 (0.001)
female1	−0.583*** (0.071)
agea:female1	0.007*** (0.001)
Constant	5.231*** (0.050)
Observations	31,064
R ²	0.005
Adjusted R ²	0.005
Residual Std. Error	2.197 (df = 31060)
F Statistic	53.277*** (df = 3; 31060)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

Predicted values of Placement on left right scale



Ejercicio

Con los datos municipales de 2019 (elecciones y sociodemográficos)

- Crea variables de participación, paro y viviendas turísticas
 - Variables continuas (%) y dicotómicas
- Investiga el impacto del paro y de las viviendas turísticas en la participación
- ¿Estos factores se moderan entre si?