

SEMINARIO DE DOCTORADO

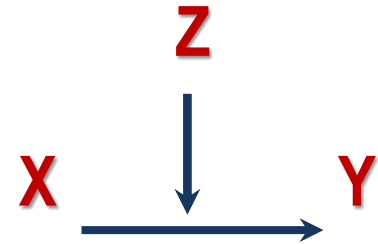
TÉCNICAS AVANZADAS DE INVESTIGACIÓN SOCIAL

Módulo 3 D MODELOS DE REGRESIÓN EFECTOS DE INTERACCIÓN

EFECTO DE INTERACCIÓN

Si ocurriera que el efecto de la educación sobre el ingreso horario fuera diferente según el sexo estaríamos hablando de un efecto de interacción entre educación y sexo de tal manera que la relación entre educación y el ingreso estaría modulado

(interferido/condicionado) por la variable sexo. Por ejemplo, en el caso que ocurriera que para valores altos/bajos de ingresos el sexo se mostrara muy relevante pero no para valores altos/bajos de la misma



X = Años de escolaridad

Y = Ingreso laboral horario

Z = Sexo

XZ= Interacción

Ecuación de Regresión

$$Y = b_0 + b_1X + b_2Z + b_3XZ$$

REGRESIONES CON EFECTOS DE INTERACCIÓN

SI ASUMIMOS A TRAVÉS DE LOS TEST DE TOLERANCIA Y ANÁLISIS DE CORRELACIÓN PARCIAL QUE LA RELACIÓN XY ESTÁ MODULADA/CONDICIONADA POR Z, EL EFECTO DE X SOBRE Y NO ES EL MISMO DEPENDIENDO DE LOS VALORES DE Z. EN LOS CASOS EN QUE EL EFECTO DE X-Y DEPENDE DEL VALOR O NIVEL DE Z, DEBE CONTROLARSE EL EFECTO INTERACCIÓN XZ

- 1) CREAR LA VARIABLE INTERACCIÓN COMO PRODUCTO XZ (ASEGURARSE DE QUE LA VARIABLE ES MÉTRICA O DUMMY).
- 2) EVALUACIÓN EL PESO/SIGNIFICANCIA DE LOS CAMBIOS EN LOS COEFICIENTES PRINCIPALES Y EL COEFICIENTE INTERACCIÓN.
- 3) LECTURA / INTERPRETACIÓN DE LOS EFECTOS DEPENDIENDO DEL DISEÑO.

En caso de COLINEALIDAD, debe evaluarse el efecto diferencial que la variable Z ejerce en la relación de X con Y. A este respecto, podemos calcular el efecto de X sobre Y para los distintos valores de Z. Para ello, reestructuramos la ecuación de regresión de la siguiente forma:

$$Y = (b_0) + (b_1x) + (b_2z) + (b_3zx)$$

$$Y = (b_0) + (b_1 * X) + (b_2 * Z) + (b_1 * X) (b_2 * Z)$$

DIFERENTES MODELOS:

X Y Z AMBAS SON DUMMY

X ES METRICA Y Z ES DUMMY

X Y Z AMBAS SON MÉTRICAS

¿Cómo interpretar los distintos coeficientes de regresión de la ecuación cuando se introduce XZ siendo ambas VARIABLES DUMMY?

- El COEF. b_0 es el efecto en Y cuando $X=0$ y $Z=0$ (Constante).
- El COEF. b_1X es el efecto en Y dado b_0 cuando $X=1$ y $Z=0$ (en ausencia de Z).
- EL COEF. b_2Z es el efecto en Y dado b_0 cuando $X=0$ y $Z=1$ (en ausencia de X).
- El COEF. b_3XZ es el efecto NETO sobre Y que agrega XY, dado $b_0+b_1+b_2$ cuando $X=1$ y $Z=1$

MODELO DONDE X Y Z SON DUMMY

Si las variables X y Z son cualitativas con dos categorías (0 y 1), la ecuación de regresión para determinar los valores de Y (INGRESOS) con un efecto interacción para X=1 o X=0 (EDUCACIÓN ALTA / BAJA) y Z=1 o Z=0 (SEXO VARON / MUJER) será:

$$Y = b_0 + b_1x + b_2z + b_3xz$$

| | | X EDUCACIÓN (b1) | |
|----------------|---------|------------------|----------|
| Z SEXO (b2) | | BAJO=0 | ALTO=1 |
| | MUJER=0 | 0, 0 (b0) | 0,1 (b1) |
| | VARÓN=1 | 1,0 (b2) | 1,1 (b3) |

$$Y = b_0 (MB) + b_1x (MA) + b_2z (VB) + INTb_3xz (VA)$$

$$Y \text{ VARON ALTA} = (b_0 + b_1*1 + b_2*1 + b_3*1*1)$$

$$Y \text{ VARON BAJA} = (b_0 + b_1*0 + b_2*1 + b_3*0*1)$$

$$Y \text{ MUJER ALTA} = (b_0 + b_1*1 + b_2*0 + b_3*1*0)$$

$$Y \text{ MUJER BAJA} = (b_0 + b_1*0 + b_2*0 + b_3*0*0)$$

Correlación XY, ZY y XZ

Correlaciones

| | | LNINGRESO H | D_UnivCom | D_Varon |
|------------|------------------------|----------------|-----------|---------|
| LNINGRESOH | Correlación de Pearson | 1 | ,313** | ,015 |
| | Sig. (bilateral) | | ,000 | ,092 |
| | N | 13155 | 13155 | 13155 |
| D_UnivCom | Correlación de Pearson | ,313** | 1 | -,155** |
| | Sig. (bilateral) | ,000 | | ,000 |
| | N | 13155 | 13155 | 13155 |
| D_Varon | Correlación de Pearson | ,015 | -,155** | 1 |
| | Sig. (bilateral) | ,092 | ,000 | |
| | N | 13155 | 13155 | 13155 |

** . La correlación es significativa al nivel 0,01 (bilateral).

CORRELACIÓN PARCIAL

Correlación XY controlada por la variable Z

Correlaciones

| Variables de control | | | LNINGRESO H | D_UnivCom |
|----------------------|------------|---------------------------|----------------|-----------|
| D_Varon | LNINGRESOH | Correlación | 1,000 | ,320 |
| | | Significación (bilateral) | . | ,000 |
| | | gl | 0 | 13152 |
| D_UnivCom | D_UnivCom | Correlación | ,320 | 1,000 |
| | | Significación (bilateral) | ,000 | |
| | | gl | 13152 | 0 |

Resumen del modelo

| Modelo | R | R cuadrado | R cuadrado corregida | Error típ. de la estimación | Estadísticos de cambio | | | | |
|--------|-------------------|------------|----------------------|-----------------------------|------------------------|-------------|-----|-------|------------------|
| | | | | | Cambio en R cuadrado | Cambio en F | gl1 | gl2 | Sig. Cambio en F |
| 1 | ,313 ^a | ,098 | ,098 | ,76281 | ,098 | 1432,428 | 1 | 13153 | ,000 |
| 2 | ,320 ^b | ,102 | ,102 | ,76110 | ,004 | 60,252 | 1 | 13152 | ,000 |
| 3 | ,323 ^c | ,104 | ,104 | ,76040 | ,002 | 25,427 | 1 | 13151 | ,000 |

a. Variables predictoras: (Constante), D_UnivCom

b. Variables predictoras: (Constante), D_UnivCom, D_Varon

c. Variables predictoras: (Constante), D_UnivCom, D_Varon, INTERACCION VARONES C/UNIVERSITARIO

Coefficientes^a

| Modelo | | Coefficients no estandarizados | | Coefficients tipificados | t | Sig. |
|--------|-------------------------------------|--------------------------------|------------|--------------------------|---------|------|
| | | B | Error típ. | Beta | | |
| 1 | (Constante) | 3,457 | ,008 | | 460,502 | ,000 |
| | D_UnivCom | ,613 | ,016 | ,313 | 37,847 | ,000 |
| 2 | (Constante) | 3,392 | ,011 | | 301,566 | ,000 |
| | D_UnivCom | ,632 | ,016 | ,323 | 38,678 | ,000 |
| | D_Varon | ,106 | ,014 | ,065 | 7,762 | ,000 |
| 3 | (Constante) | 3,369 | ,012 | | 278,617 | ,000 |
| | D_UnivCom | ,710 | ,022 | ,363 | 31,601 | ,000 |
| | D_Varon | ,142 | ,015 | ,087 | 9,228 | ,000 |
| | INTERACCION VARONES C/UNIVERSITARIO | -,165 | ,033 | -,060 | -5,043 | ,000 |

a. Variable dependiente: LNINGRESOH

Variables introducidas/eliminadas^b

| Modelo | Variables introducidas | Variables eliminadas | Método |
|--------|------------------------|----------------------|------------|
| 1 | MUJER_UNI ^a | . | Introducir |
| 2 | VARON_NOUNI | . | Introducir |
| 3 | VARON_UNI ^a | . | Introducir |

a. Todas las variables solicitadas introducidas.

b. Variable dependiente: LNINGRESOH

Resumen del modelo

| Modelo | R | R cuadrado | R cuadrado corregida | Error típ. de la estimación | Estadísticos de cambio | | | | |
|--------|-------------------|------------|----------------------|-----------------------------|------------------------|-------------|-----|-------|------------------|
| | | | | | Cambio en R cuadrado | Cambio en F | gl1 | gl2 | Sig. Cambio en F |
| 1 | ,228 ^a | ,052 | ,052 | ,78206 | ,052 | 723,175 | 1 | 13153 | ,000 |
| 2 | ,229 ^b | ,052 | ,052 | ,78204 | ,000 | 1,818 | 1 | 13152 | ,178 |
| 3 | ,323 ^c | ,104 | ,104 | ,76040 | ,052 | 760,440 | 1 | 13151 | ,000 |

a. Variables predictoras: (Constante), MUJER_UNI

b. Variables predictoras: (Constante), MUJER_UNI, VARON_NOUNI

c. Variables predictoras: (Constante), MUJER_UNI, VARON_NOUNI, VARON_UNI

Coefficientes^a

| Modelo | | Coefficients no estandarizados | | Coefficients tipificados | t | Sig. |
|--------|-------------|--------------------------------|------------|--------------------------|---------|------|
| | | B | Error típ. | Beta | | |
| 1 | (Constante) | 3,520 | ,007 | | 483,593 | ,000 |
| | MUJER_UNI | ,559 | ,021 | ,228 | 26,892 | ,000 |
| 2 | (Constante) | 3,531 | ,011 | | 324,692 | ,000 |
| | MUJER_UNI | ,548 | ,022 | ,224 | 24,580 | ,000 |
| | VARON_NOUNI | -,020 | ,015 | -,012 | -1,348 | ,178 |
| 3 | (Constante) | 3,359 | ,012 | | 278,617 | ,000 |
| | MUJER_UNI | ,710 | ,022 | ,290 | 31,601 | ,000 |
| | VARON_NOUNI | ,142 | ,015 | ,088 | 9,228 | ,000 |
| | VARON_UNI | ,687 | ,025 | ,248 | 27,576 | ,000 |

a. Variable dependiente: LNINGRESOH

MODELO DONDE X ES MÉTRICA Y Z ES DUMMY

Si la variable X es métrica con N valores y Z es cualitativa con dos categorías (0 y 1), la ecuación de regresión para determinar los valores de Y (INGRESOS) con un efecto de interacción entre X (AÑOS DE ESCOLARIDAD) y Z (VARON / MUJER) será:

$$Y = b_0 + b_1x + b_2z + b_3xz$$

$$Y = b_0 (M/AE=0) + b_1x (M*AE) + b_2z (V/AE=0) + b_3xz (V*AE)$$

- El COEF. b_0 es el efecto en Y cuando $X=0$ y $Z=0$ (Constante).
- El COEF. b_1X es el efecto en Y por cada unidad de X, dado b_0 cuando $Z=0$ y $X \neq 0$.
- EL COEF. b_2Z es el efecto sobre Y que agrega Z, dado b_0 cuando $X=0$ y $Z \neq 0$.
- El COEF. b_3XZ es el efecto NETO en Y que agrega Z por cada unidad de X, dado $b_0+b_1+b_2$ cuando $X \neq 0$ y $Z=1$

Correlaciones

Correlación XY, ZY y XZ

| | | LNINGRESO H | Educac_años | D_Varon |
|-------------|------------------------|----------------|-------------|---------|
| LNINGRESOH | Correlación de Pearson | 1 | ,362** | ,015 |
| | Sig. (bilateral) | | ,000 | ,092 |
| | N | 13155 | 13155 | 13155 |
| Educac_años | Correlación de Pearson | ,362** | 1 | -,176** |
| | Sig. (bilateral) | ,000 | | ,000 |
| | N | 13155 | 13155 | 13155 |
| D_Varon | Correlación de Pearson | ,015 | -,176** | 1 |
| | Sig. (bilateral) | ,092 | ,000 | |
| | N | 13155 | 13155 | 13155 |

** . La correlación es significativa al nivel 0,01 (bilateral).

Correlación XY controlada por la variable Z

Correlaciones

| Variables de control | | | LNINGRESO H | Educac_años |
|----------------------|------------|---------------------------|----------------|-------------|
| D_Varon | LNINGRESOH | Correlación | 1,000 | ,370 |
| | | Significación (bilateral) | . | ,000 |
| | | gl | 0 | 13152 |
| Educac_años | LNINGRESOH | Correlación | ,370 | 1,000 |
| | | Significación (bilateral) | ,000 | |
| | | gl | 13152 | 0 |

Resumen del modelo

| Modelo | R | R cuadrado | R cuadrado corregida | Error típ. de la estimación | Estadísticos de cambio | | | | |
|--------|-------------------|------------|----------------------|-----------------------------|------------------------|-------------|-----|-------|------------------|
| | | | | | Cambio en R cuadrado | Cambio en F | gl1 | gl2 | Sig. Cambio en F |
| 1 | ,362 ^a | ,131 | ,131 | ,74889 | ,131 | 1980,017 | 1 | 13153 | ,000 |
| 2 | ,370 ^b | ,137 | ,137 | ,74617 | ,006 | 96,885 | 1 | 13152 | ,000 |
| 3 | ,372 ^c | ,139 | ,139 | ,74555 | ,002 | 23,017 | 1 | 13151 | ,000 |

a. Variables predictoras: (Constante), Educac_años

b. Variables predictoras: (Constante), Educac_años, D_Varon

c. Variables predictoras: (Constante), Educac_años, D_Varon, VARON_EDANOS

Coefficientes^a

| Modelo | | Coefficients no estandarizados | | Coefficients tipificados | t | Sig. |
|--------|--------------|--------------------------------|------------|--------------------------|---------|------|
| | | B | Error típ. | Beta | | |
| 1 | (Constante) | 2,709 | ,021 | | 130,201 | ,000 |
| | Educac_años | ,075 | ,002 | ,362 | 44,497 | ,000 |
| 2 | (Constante) | 2,599 | ,024 | | 110,177 | ,000 |
| | Educac_años | ,078 | ,002 | ,376 | 45,695 | ,000 |
| | D_Varon | ,132 | ,013 | ,081 | 9,843 | ,000 |
| 3 | (Constante) | 2,482 | ,034 | | 73,131 | ,000 |
| | Educac_años | ,087 | ,003 | ,421 | 33,712 | ,000 |
| | D_Varon | ,328 | ,043 | ,202 | 7,614 | ,000 |
| | VARON_EDANOS | -,017 | ,003 | -,128 | -4,798 | ,000 |

a. Variable dependiente: LNINGRESOH

MODELO DONDE X ES MÉTRICA Y Z ES METRICA

Si la variable X y Z son variables métricas con N valores, la ecuación de regresión para determinar los valores de Y (INGRESOS) con un efecto de interacción entre X (AÑOS DE ESCOLARIDAD) y Z (GRUPOS DE EDAD) será:

$$Y = b_0 + b_1x + b_2z + b_3xz$$

$$Y = b_0(GE=0/AE=0) + b_1x(AE/GE=0) + b_2z(GE/AE=0) + b_3xz(GE*AE)$$

- El COEF. b_0 es el efecto en Y cuando $X=0$ y $Z=0$ (Constante).
- El COEF. b_1X es el efecto en Y por cada unidad de X, dado b_0 cuando $Z=0$ y $X \neq 0$.
- EL COEF. b_2Z es el efecto sobre Y que agrega cada cambio en la unidad de Z, dado b_0 cuando $X=0$ y $Z \neq 0$.
- El COEF. b_3XZ es el efecto NETO en Y que agrega cada unidad de Z por cada unidad de X, dado $b_0+b_1+b_2$ cuando $X \neq 0$ y $Z \neq 0$

Correlación XY, ZY y XZ

Correlaciones

| | | LNINGRESO H | Educac_años | Grupos de Edad_en 3 categorías |
|-----------------------------------|------------------------|----------------|-------------|--------------------------------------|
| LNINGRESOH | Correlación de Pearson | 1 | ,362** | ,058** |
| | Sig. (bilateral) | | ,000 | ,000 |
| | N | 13155 | 13155 | 13084 |
| Educac_años | Correlación de Pearson | ,362** | 1 | -,094** |
| | Sig. (bilateral) | ,000 | | ,000 |
| | N | 13155 | 13155 | 13084 |
| Grupos de Edad_en 3 categorías | Correlación de Pearson | ,058** | -,094** | 1 |
| | Sig. (bilateral) | ,000 | ,000 | |
| | N | 13084 | 13084 | 13084 |

** . La correlación es significativa al nivel 0,01 (bilateral).

Correlación XY controlada por la variable Z

Correlaciones

| Variables de control | | | LNINGRESO H | Educac_años |
|-----------------------------------|-------------|---------------------------|----------------|-------------|
| Grupos de Edad_en 3 categorías | LNINGRESOH | Correlación | 1,000 | ,369 |
| | | Significación (bilateral) | . | ,000 |
| | | gl | 0 | 13081 |
| | Educac_años | Correlación | ,369 | 1,000 |
| | | Significación (bilateral) | ,000 | . |
| | | gl | 13081 | 0 |

Regresión de X, Z y XZ sobre el Log de los Ingresos

Resumen del modelo

| Modelo | R | R cuadrado | R cuadrado corregida | Error típ. de la estimación | Estadísticos de cambio | | | | |
|--------|-------------------|------------|----------------------|-----------------------------|------------------------|-------------|-----|-------|------------------|
| | | | | | Cambio en R cuadrado | Cambio en F | gl1 | gl2 | Sig. Cambio en F |
| 1 | ,362 ^a | ,131 | ,131 | ,74772 | ,131 | 1969,282 | 1 | 13081 | ,000 |
| 2 | ,373 ^b | ,139 | ,139 | ,74404 | ,009 | 130,491 | 1 | 13080 | ,000 |
| 3 | ,374 ^c | ,140 | ,139 | ,74401 | ,000 | 2,138 | 1 | 13079 | ,144 |

a. Variables predictoras: (Constante), Educac_años

b. Variables predictoras: (Constante), Educac_años, Grupos de Edad_en 3 categorías

c. Variables predictoras: (Constante), Educac_años, Grupos de Edad_en 3 categorías, EDUCA_GEDAD3

Coefficientes^a

| Modelo | | Coefficients no estandarizados | | Coefficients tipificados | t | Sig. |
|--------|--------------------------------|--------------------------------|------------|--------------------------|---------|------|
| | | B | Error típ. | Beta | | |
| 1 | (Constante) | 2,713 | ,021 | | 130,253 | ,000 |
| | Educac_años | ,075 | ,002 | ,362 | 44,377 | ,000 |
| 2 | (Constante) | 2,493 | ,028 | | 88,107 | ,000 |
| | Educac_años | ,077 | ,002 | ,370 | 45,474 | ,000 |
| | Grupos de Edad_en 3 categorías | ,117 | ,010 | ,093 | 11,423 | ,000 |
| 3 | (Constante) | 2,409 | ,064 | | 37,451 | ,000 |
| | Educac_años | ,084 | ,005 | ,405 | 16,221 | ,000 |
| | Grupos de Edad_en 3 categorías | ,162 | ,033 | ,129 | 4,963 | ,000 |
| | EDUCA_GEDAD3 | -,004 | ,003 | -,049 | -1,462 | ,144 |

a. Variable dependiente: LNINGRESOH

Regresión de X, Z y XZ sobre los Ingresos

Resumen del modelo^d

| Modelo | R | R cuadrado | R cuadrado corregida | Error típ. de la estimación | Estadísticos de cambio | | | | |
|--------|-------------------|------------|----------------------|-----------------------------|------------------------|-------------|-----|-------|------------------|
| | | | | | Cambio en R cuadrado | Cambio en F | gl1 | gl2 | Sig. Cambio en F |
| 1 | ,312 ^a | ,097 | ,097 | 41,43012 | ,097 | 1406,783 | 1 | 13081 | ,000 |
| 2 | ,341 ^b | ,117 | ,116 | 40,98261 | ,019 | 288,255 | 1 | 13080 | ,000 |
| 3 | ,343 ^c | ,118 | ,117 | 40,96227 | ,001 | 13,994 | 1 | 13079 | ,000 |

a. Variables predictoras: (Constante), Educac_años

b. Variables predictoras: (Constante), Educac_años, Grupos de Edad_en 3 categorías

c. Variables predictoras: (Constante), Educac_años, Grupos de Edad_en 3 categorías, EDUCA_GEDAD3

d. Variable dependiente: Ingreso Horario de la ocupacion principal

Coefficientes^a

| Modelo | | Coefficients no estandarizados | | Coefficients tipificados | t | Sig. |
|--------|--------------------------------|--------------------------------|------------|--------------------------|--------|------|
| | | B | Error típ. | Beta | | |
| 1 | (Constante) | 6,866 | 1,154 | | 5,949 | ,000 |
| | Educac_años | 3,509 | ,094 | ,312 | 37,507 | ,000 |
| 2 | (Constante) | -11,149 | 1,559 | | -7,153 | ,000 |
| | Educac_años | 3,657 | ,093 | ,325 | 39,348 | ,000 |
| | Grupos de Edad_en 3 categorías | 9,559 | ,563 | ,140 | 16,978 | ,000 |
| 3 | (Constante) | ,747 | 3,541 | | ,211 | ,833 |
| | Educac_años | 2,651 | ,285 | ,235 | 9,310 | ,000 |
| | Grupos de Edad_en 3 categorías | 3,168 | 1,799 | ,046 | 1,761 | ,078 |
| | EDUCA_GEDAD3 | ,547 | ,146 | ,127 | 3,741 | ,000 |

a. Variable dependiente: Ingreso Horario de la ocupacion principal