Variables Categoricas e Interacciones

Modelando Heterogeneidad en Retornos Salariales

EC3003B - Economía Laboral Aplicada

Tecnológico de Monterrey, Campus Puebla

Martes 10 de febrero, 2025 | 3-5pm

Contenido de la Sesión

Conexión con la Sesión Anterior

En M01 aprendimos:

- La ecuación de Mincer: $\ln(w) = \beta_0 + \beta_1 S + \beta_2 X + \beta_3 X^2$
- Retorno a educación $\approx 9\,\%$ por año
- Retornos decrecientes a experiencia

Conexión con la Sesión Anterior

En M01 aprendimos:

- La ecuación de Mincer: $\ln(w) = \beta_0 + \beta_1 S + \beta_2 X + \beta_3 X^2$
- Retorno a educación $\approx 9\,\%$ por año
- Retornos decrecientes a experiencia

Pero... £es el mismo retorno para todos?

- £Ganan lo mismo hombres y mujeres con igual educación?
- £El retorno es igual en sector formal e informal?
- £Preparatoria y licenciatura tienen el mismo "valor por año"?

Conexión con la Sesión Anterior

En M01 aprendimos:

- La ecuación de Mincer: $\ln(w) = \beta_0 + \beta_1 S + \beta_2 X + \beta_3 X^2$
- Retorno a educación $\approx 9\,\%$ por año
- Retornos decrecientes a experiencia

Pero... £es el mismo retorno para todos?

- £Ganan lo mismo hombres y mujeres con igual educación?
- £El retorno es igual en sector formal e informal?
- £Preparatoria y licenciatura tienen el mismo "valor por año"?

Hoy aprenderemos

A modelar **heterogeneidad** usando variables categoricas e interacciones.

Objetivos de Aprendizaje

Al finalizar está sesión, podrás:

1. Crear e interpretar variables dummy
2. Elegir correctamente la categoría base
3. Modelar interacciones entre variables
4. Interpretar coeficientes de interaccion
5. Realizar pruebas de diferencias entre grupos

£Que es una Variable Dummy?

Definicion

Una **variable dummy** (indicadora o binaria) toma solo dos valores:

- 1 si la condicion se cumple
- 0 si no se cumple

£Que es una Variable Dummy?

Definicion

Una **variable dummy** (indicadora o binaria) toma solo dos valores:

- 1 si la condicion se cumple
- 0 si no se cumple

Ejemplos en datos salariales:

- `mujer` $= 1$ si es mujer, 0 si es hombre
- `formal` $= 1$ si empleo formal, 0 si informal
- `sindicalizado` $= 1$ si tiene sindicato
- `cdmx` $= 1$ si trabaja en Ciudad de México

£Que es una Variable Dummy?

Definicion

Una **variable dummy** (indicadora o binaria) toma solo dos valores:

- 1 si la condicion se cumple
- 0 si no se cumple

Ejemplos en datos salariales:

- `mujer` $= 1$ si es mujer, 0 si es hombre
- `formal` $= 1$ si empleo formal, 0 si informal
- `sindicalizado` $= 1$ si tiene sindicato
- `cdmx` $= 1$ si trabaja en Ciudad de México

En Stata

```
gen mujer = (sexo == 2)
gen formal = (tipo_empleo == 1)
```

Interpretación de Coeficientes Dummy

Modelo:

$$\ln(w) = \beta_0 + \beta_1 \cdot \text{mujer} + \beta_2 S + \beta_3 X + \varepsilon$$

Interpretación de Coeficientes Dummy

Modelo:

$$\ln(w) = \beta_0 + \beta_1 \cdot \text{mujer} + \beta_2 S + \beta_3 X + \varepsilon$$

Para hombres (mujer = 0):

$$E[\ln(w)|\text{hombre}] = \beta_0 + \beta_2 S + \beta_3 X$$

Para mujeres (mujer = 1):

$$E[\ln(w)|\text{mujer}] = (\beta_0 + \beta_1) + \beta_2 S + \beta_3 X$$

Interpretación de Coeficientes Dummy

Modelo:

$$\ln(w) = \beta_0 + \beta_1 \cdot \text{mujer} + \beta_2 S + \beta_3 X + \varepsilon$$

Para hombres (mujer = 0):

$$E[\ln(w)|\text{hombre}] = \beta_0 + \beta_2 S + \beta_3 X$$

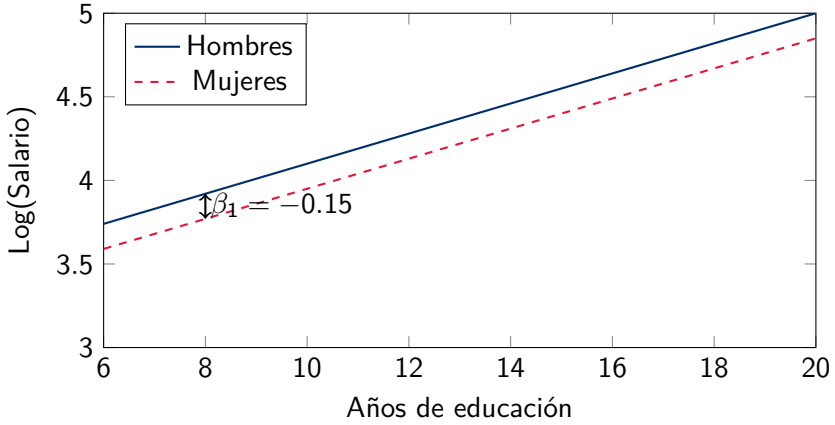
Para mujeres (mujer = 1):

$$E[\ln(w)|\text{mujer}] = (\beta_0 + \beta_1) + \beta_2 S + \beta_3 X$$

Interpretación de β_1

β_1 es la **diferencia** en log-salario entre mujeres y hombres, *controlando por* educación y experiencia.
Si $\beta_1 = -0.15$: mujeres ganan $\approx 15\,\%$ menos que hombres comparables.

Visualización: Diferencia de Interceptos



Nota: La dummy desplaza el intercepto, pero las pendientes son iguales.

De Binario a Categorico

Problema: £Cómo modelar variables con más de 2 categorías?

Ejemplos:

- Nivel educativo: secundaria, preparatoria, licenciatura, posgrado
- Region: Norte, Centro, Sur, CDMX
- Sector: Manufactura, Servicios, Construccion, Comercio

De Binario a Categorico

Problema: £Cómo modelar variables con más de 2 categorías?

Ejemplos:

- Nivel educativo: secundaria, preparatoria, licenciatura, posgrado
- Region: Norte, Centro, Sur, CDMX
- Sector: Manufactura, Servicios, Construccion, Comercio

Regla de oro

Para una variable con K categorías, creamos $K - 1$ dummies.
La categoría omitida es la **categoría base** o de referencia.

Ejemplo: Nivel Educativo

4 categorías $\rightarrow$ 3 dummies

Nivel	prepa	licenciatura	posgrado	(base)
Secundaria o menos	0	0	0	✓
Preparatoria	1	0	0	
Licenciatura	0	1	0	
Posgrado	0	0	1	

Ejemplo: Nivel Educativo

4 categorías → 3 dummies

Nivel	prepa	licenciatura	posgrado	(base)
Secundaria o menos	0	0	0	✓
Preparatoria	1	0	0	
Licenciatura	0	1	0	
Posgrado	0	0	1	

Modelo:

$$\ln(w) = \beta_0 + \beta_1 \cdot \text{prepa} + \beta_2 \cdot \text{lic} + \beta_3 \cdot \text{pos} + \gamma X + \varepsilon$$

Interpretación:

- β_1: Premio de prepa vs secundaria
- β_2: Premio de licenciatura vs secundaria
- β_3: Premio de posgrado vs secundaria

Dummies en Stata: Factor Variables

Manera antigua (manual):

```
gen prepa = (nivel_educ == 2)
gen lic = (nivel_educ == 3)
gen pos = (nivel_educ == 4)
reg ln_salario prepa lic pos experiencia
```

Dummies en Stata: Factor Variables

Manera antigua (manual):

```
gen prepa = (nivel_educ == 2)
gen lic = (nivel_educ == 3)
gen pos = (nivel_educ == 4)
reg ln_salario prepa lic pos experiencia
```

Manera moderna (automática):

```
* i. crea dummies automaticamente
reg ln_salario i.nivel_educ experiencia

* Cambiar categoría base
reg ln_salario ib3.nivel_educ experiencia  // base = licenciatura
```

Ventajas de i.

- Automático, menos errores
- Fácil cambiar categoría base
- Funciona con interacciones y predicciones

Eleccion de la Categoría Base

£Cuál categoría omitir?

- **Por defecto:** La primera categoría (menor valor numerico)
- **Recomendado:** La categoría más relevante para comparación

Eleccion de la Categoría Base

£Cuál categoría omitir?

- **Por defecto:** La primera categoría (menor valor numerico)
- **Recomendado:** La categoría más relevante para comparación

Ejemplos de eleccion estrategica:

Variable	Base recomendada	Razon
Nivel educativo	Secundaria	Nivel minimo legal
Género	Hombres	Referencia tradicional
Sector	Manufactura	Sector más grande
Region	CDMX	Mercado laboral central

Importante

La eleccion de base NO cambia el ajuste del modelo (R^2), solo la interpretación de coeficientes.

Motivación: Retornos Heterogeneos

Hasta ahora asumimos:

- Mismo retorno a educación para hombres y mujeres
- Mismo retorno en sector formal e informal
- Misma brecha de género en todos los niveles educativos

Motivación: Retornos Heterogeneos

Hasta ahora asumimos:

- Mismo retorno a educación para hombres y mujeres
- Mismo retorno en sector formal e informal
- Misma brecha de género en todos los niveles educativos

Pero la realidad es más compleja:

- El retorno a licenciatura puede ser mayor para hombres
- La brecha de género puede crecer con la educación
- El sector formal puede premiar más la experiencia

Motivación: Retornos Heterogeneos

Hasta ahora asumimos:

- Mismo retorno a educación para hombres y mujeres
- Mismo retorno en sector formal e informal
- Misma brecha de género en todos los niveles educativos

Pero la realidad es más compleja:

- El retorno a licenciatura puede ser mayor para hombres
- La brecha de género puede crecer con la educación
- El sector formal puede premiar más la experiencia

Solución: Terminos de Interaccion

Permiten que el efecto de una variable **dependa** de otra variable.

Interaccion Dummy $\times$ Continua

Modelo con interaccion:

$$\ln(w) = \beta_0 + \beta_1 \cdot \text{mujer} + \beta_2 \cdot S + \beta_3 \cdot (\text{mujer} \times S) + \varepsilon$$

Interaccion Dummy $\times$ Continua

Modelo con interaccion:

$$\ln(w) = \beta_0 + \beta_1 \cdot \text{mujer} + \beta_2 \cdot S + \beta_3 \cdot (\text{mujer} \times S) + \varepsilon$$

Para hombres:

$$E[\ln(w)] = \beta_0 + \beta_2 \cdot S$$

Para mujeres:

$$E[\ln(w)] = (\beta_0 + \beta_1) + (\beta_2 + \beta_3) \cdot S$$

Interaccion Dummy $\times$ Continua

Modelo con interaccion:

$$\ln(w) = \beta_0 + \beta_1 \cdot \text{mujer} + \beta_2 \cdot S + \beta_3 \cdot (\text{mujer} \times S) + \varepsilon$$

Para hombres:

$$E[\ln(w)] = \beta_0 + \beta_2 \cdot S$$

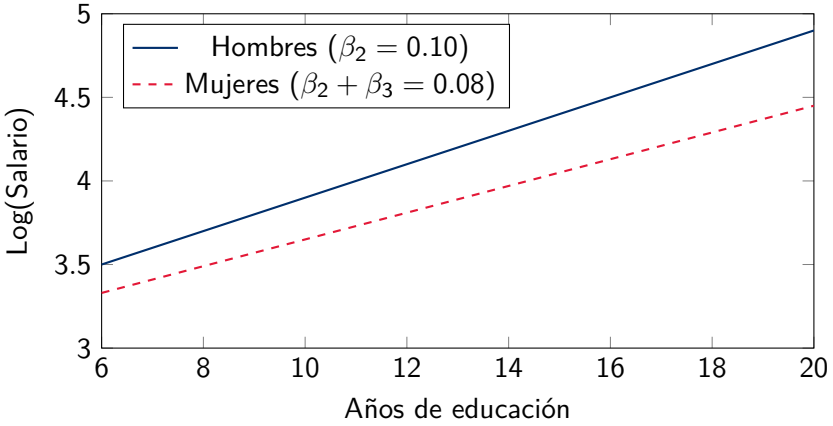
Para mujeres:

$$E[\ln(w)] = (\beta_0 + \beta_1) + (\beta_2 + \beta_3) \cdot S$$

Interpretación de β_3

Diferencia en el retorno a educación entre mujeres y hombres.
Si $\beta_3 = -0.02$: cada año de educación rinde 2 puntos porcentuales **menos** para mujeres.

Visualización: Diferentes Pendientes



Nota: Ahora tanto intercepto como pendiente difieren.
La brecha de género **crece** con la educación.

Interacciones en Stata

```
* Interaccion manual
gen mujer_educ = mujer * escolaridad
reg ln_salario mujer escolaridad mujer_educ

* Interaccion con operador # (recomendado)
reg ln_salario i.mujer##c.escolaridad

* Descomponer:
* i.mujer          = efecto principal de mujer
* c.escolaridad    = efecto principal de educación
* i.mujer#c.escolaridad = interaccion
```

Notacion Stata

- # = solo interaccion
- ## = efectos principales + interaccion
- c. = variable continua
- i. = variable categorica

Interaccion Dummy $\times$ Dummy

Modelo:

$$\ln(w) = \beta_0 + \beta_1 \cdot \text{mujer} + \beta_2 \cdot \text{formal} + \beta_3 \cdot (\text{mujer} \times \text{formal}) + \varepsilon$$

Interaccion Dummy $\times$ Dummy

Modelo:

$$\ln(w) = \beta_0 + \beta_1 \cdot \text{mujer} + \beta_2 \cdot \text{formal} + \beta_3 \cdot (\text{mujer} \times \text{formal}) + \varepsilon$$

	Informal	Formal
Hombre	β_0	$\beta_0 + \beta_2$
Mujer	$\beta_0 + \beta_1$	$\beta_0 + \beta_1 + \beta_2 + \beta_3$

Interaccion Dummy $\times$ Dummy

Modelo:

$$\ln(w) = \beta_0 + \beta_1 \cdot \text{mujer} + \beta_2 \cdot \text{formal} + \beta_3 \cdot (\text{mujer} \times \text{formal}) + \varepsilon$$

	Informal	Formal
Hombre	β_0	$\beta_0 + \beta_2$
Mujer	$\beta_0 + \beta_1$	$\beta_0 + \beta_1 + \beta_2 + \beta_3$

Interpretación de β_3:

- Efecto **adicional** de ser mujer en sector formal
- Si $\beta_3 > 0$: la brecha de género es **menor** en sector formal
- Si $\beta_3 < 0$: la brecha de género es **mayor** en sector formal

Pruebas con Variables Categoricas

Preguntas tipicas:

1. £Hay diferencia significativa entre grupos?
2. £Son iguales todos los coeficientes de las dummies?
3. £La interaccion es significativa?

Pruebas con Variables Categoricas

Preguntas tipicas:

1. £Hay diferencia significativa entre grupos?
2. £Son iguales todos los coeficientes de las dummies?
3. £La interaccion es significativa?

Herramienta: Test F conjunto

H_0: Todos los coeficientes de un grupo $= 0$

```
* Test de significancia conjunta de dummies educativas
reg ln_salario i.nivel_educ experiencia
testparm i.nivel_educ

* Test de que licenciatura = posgrado
test 3.nivel_educ = 4.nivel_educ
```

Margins: Predicciones por Grupo

```
* Modelo con interaccion
reg ln_salario i.mujer##c.escolaridad experiencia, robust

* Predicción promedio por género
margins mujer

* Retorno a educación por género
margins, dydx(escolaridad) at(mujer=(0 1))

* Efecto marginal de ser mujer en diferentes niveles de educación
margins, dydx(mujer) at(escolaridad=(9 12 16 18))
```

£Por que usar margins?

- Predicciones en escala original (no log)
- Errores estándar correctos
- Fácil de graficar con `marginsplot`

Conexión con el Proyecto de Compensaciones

Variables categoricas relevantes para la empresa cliente:

- **Área funcional:** Operativo, Técnico, Coordinacion, Direccion, Admin
- **Nivel jerárquico:** 1-5
- **Tipo de puesto:** Campo vs Oficina
- **Turno:** Diurno, Nocturno, Mixto

Conexión con el Proyecto de Compensaciones

Variables categoricas relevantes para la empresa cliente:

- **Área funcional:** Operativo, Técnico, Coordinacion, Direccion, Admin
- **Nivel jerárquico:** 1-5
- **Tipo de puesto:** Campo vs Oficina
- **Turno:** Diurno, Nocturno, Mixto

Preguntas a responder con dummies e interacciones:

1. £Cuánto más gana un Coordinador vs Técnico (controlando por educación)?
2. £El retorno a experiencia es igual en campo y oficina?
3. £La brecha entre áreas crece con la antiguedad?

Ejemplo: Estructura de la Empresa

Nivel	Dummy	Salario esperado	Premio vs N1
1 - Operativo	(base)	β_0	–
2 - Técnico	d_2	$\beta_0 + \beta_2$	β_2
3 - Supervision	d_3	$\beta_0 + \beta_3$	β_3
4 - Coordinacion	d_4	$\beta_0 + \beta_4$	β_4
5 - Direccion	d_5	$\beta_0 + \beta_5$	β_5

Ejemplo: Estructura de la Empresa

Nivel	Dummy	Salario esperado	Premio vs N1
1 - Operativo	(base)	β_0	–
2 - Técnico	d_2	$\beta_0 + \beta_2$	β_2
3 - Supervision	d_3	$\beta_0 + \beta_3$	β_3
4 - Coordinacion	d_4	$\beta_0 + \beta_4$	β_4
5 - Direccion	d_5	$\beta_0 + \beta_5$	β_5

Para el tabulador

Los coeficientes $\beta_2, \beta_3, \beta_4, \beta_5$ informan los **diferenciales** entre niveles que debe reflejar la estructura salarial.

Resumen de la Sesión

Variables Dummy:

- Codifican categorías
- K categorías $\rightarrow$ $K - 1$ dummies
- Coeficiente = diferencia vs base

Interacciones:

- Permiten efectos heterogeneos
- Dummy $\times$ Continua: pendientes diferentes
- Dummy $\times$ Dummy: efectos condicionales

Comandos Stata clave:

- `i.var` para dummies
- `ib#.var` cambiar base
- `##` para interacciones
- `testparm` pruebas F
- `margins` predicciones

£Preguntas?

Próxima Sesión:

M03: Diagnósticos OLS y Errores Robustos

Jueves 12 de febrero, 3-5pm

Entrega E1: Jueves 12 feb, 11:59pm