

# Variables Categoricas e Interacciones

## Modelando Heterogeneidad en Retornos Salariales

EC3003B - Economia Laboral Aplicada

Tecnologico de Monterrey, Campus Puebla

Martes 10 de febrero, 2025 | 3-5pm

# Contenido de la Sesión

- 1 Introducción
- 2 Variables Dummy
- 3 Variables Categoricas con Multiples Niveles
- 4 Interacciones
- 5 Pruebas de Hipótesis
- 6 Aplicación al Proyecto
- 7 Resumen

# Conexion con la Sesion Anterior

## En M01 aprendimos:

- La ecuacion de Mincer:  $\ln(w) = \beta_0 + \beta_1S + \beta_2X + \beta_3X^2$
- Retorno a educacion  $\approx 9\%$  por año
- Retornos decrecientes a experiencia

## En M01 aprendimos:

- La ecuacion de Mincer:  $\ln(w) = \beta_0 + \beta_1S + \beta_2X + \beta_3X^2$
- Retorno a educacion  $\approx 9\%$  por año
- Retornos decrecientes a experiencia

## Pero... ¿es el mismo retorno para todos?

- ¿Ganan lo mismo hombres y mujeres con igual educacion?
- ¿El retorno es igual en sector formal e informal?
- ¿Preparatoria y licenciatura tienen el mismo “valor por año”?

# Conexion con la Sesion Anterior

En M01 aprendimos:

- La ecuacion de Mincer:  $\ln(w) = \beta_0 + \beta_1S + \beta_2X + \beta_3X^2$
- Retorno a educacion  $\approx 9\%$  por año
- Retornos decrecientes a experiencia

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

- ¿Ganan lo mismo hombres y mujeres con igual educacion?
- ¿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 esta sesion, podras:

- ① Crear e interpretar variables dummy
- ② Elegir correctamente la categoria 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 Mexico

# ¿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 Mexico

## En Stata

```
gen mujer = (sexo == 2)
```

```
gen formal = (tipo_empleo == 1)
```

# Interpretacion de Coeficientes Dummy

Modelo:

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

# Interpretacion de Coeficientes Dummy

Modelo:

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

Para hombres ( $\text{mujer} = 0$ ):

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

Para mujeres ( $\text{mujer} = 1$ ):

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

# Interpretacion de Coeficientes Dummy

Modelo:

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

Para hombres ( $\text{mujer} = 0$ ):

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

Para mujeres ( $\text{mujer} = 1$ ):

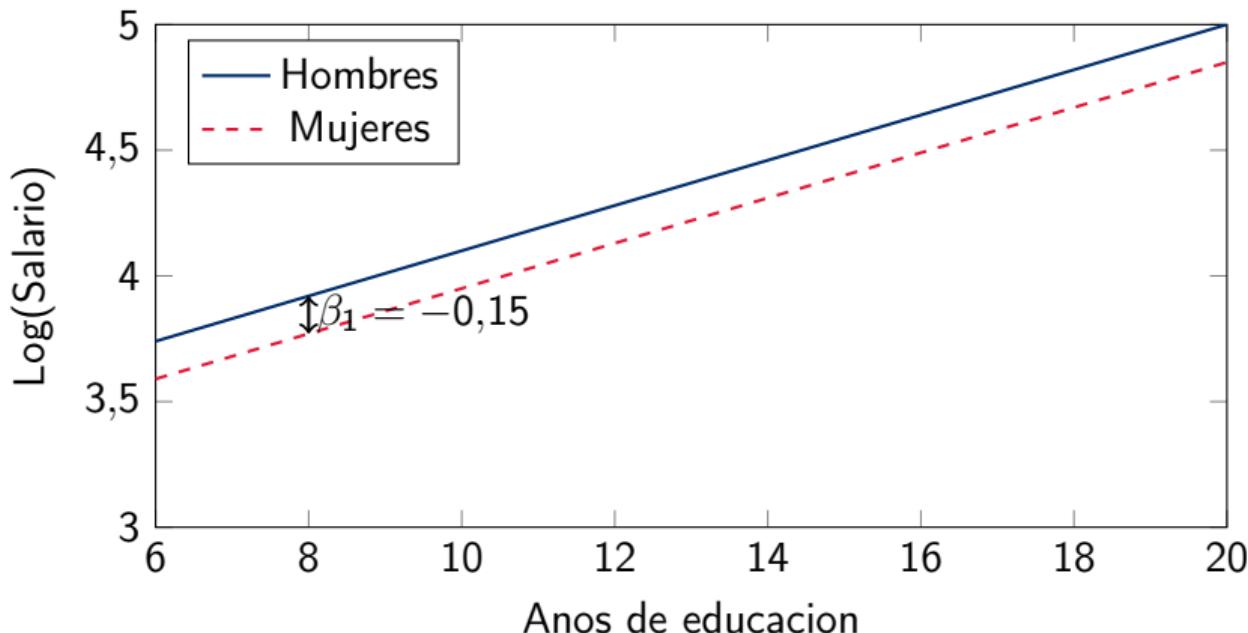
$$E[\ln(w)|\text{mujer}] = (\beta_0 + \beta_1) + \beta_2 S + \beta_3 X$$

## Interpretacion de $\beta_1$

$\beta_1$  es la **diferencia** en log-salario entre mujeres y hombres, *controlando por* educacion y experiencia.

Si  $\beta_1 = -0,15$ : mujeres ganan  $\approx 15\%$  menos que hombres comparables.

# Visualizacion: Diferencia de Interceptos



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

# De Binario a Categorico

**Problema:** ¿Como modelar variables con mas de 2 categorias?

**Ejemplos:**

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

# De Binario a Categorico

**Problema:** ¿Como modelar variables con mas de 2 categorias?

**Ejemplos:**

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

## Regla de oro

Para una variable con  $K$  categorias, creamos  $K - 1$  dummies.

La categoria omitida es la **categoria base** o de referencia.

# 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        |        |

## 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:

- $\beta_1$ : Premio de prepa vs secundaria
- $\beta_2$ : Premio de licenciatura vs secundaria
- $\beta_3$ : 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 (automatica):

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

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

### Ventajas de i.

- Automatico, menos errores
- Facil cambiar categoria base

# Eleccion de la Categoria Base

## ¿Cual categoria omitir?

- **Por defecto:** La primera categoria (menor valor numerico)
- **Recomendado:** La categoria mas relevante para comparacion

# Eleccion de la Categoria Base

## ¿Cual categoria omitir?

- **Por defecto:** La primera categoria (menor valor numerico)
- **Recomendado:** La categoria mas relevante para comparacion

## Ejemplos de eleccion estrategica:

| Variable        | Base recomendada | Razon                   |
|-----------------|------------------|-------------------------|
| Nivel educativo | Secundaria       | Nivel minimo legal      |
| Genero          | Hombres          | Referencia tradicional  |
| Sector          | Manufactura      | Sector mas grande       |
| Region          | CDMX             | Mercado laboral central |

## Importante

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

## Hasta ahora asumimos:

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

**Hasta ahora asumimos:**

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

**Pero la realidad es mas compleja:**

- El retorno a licenciatura puede ser mayor para hombres
- La brecha de genero puede crecer con la educacion
- El sector formal puede premiar mas la experiencia

# Motivacion: Retornos Heterogeneos

Hasta ahora asumimos:

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

Pero la realidad es mas compleja:

- El retorno a licenciatura puede ser mayor para hombres
- La brecha de genero puede crecer con la educacion
- El sector formal puede premiar mas la experiencia

Solucion: 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:

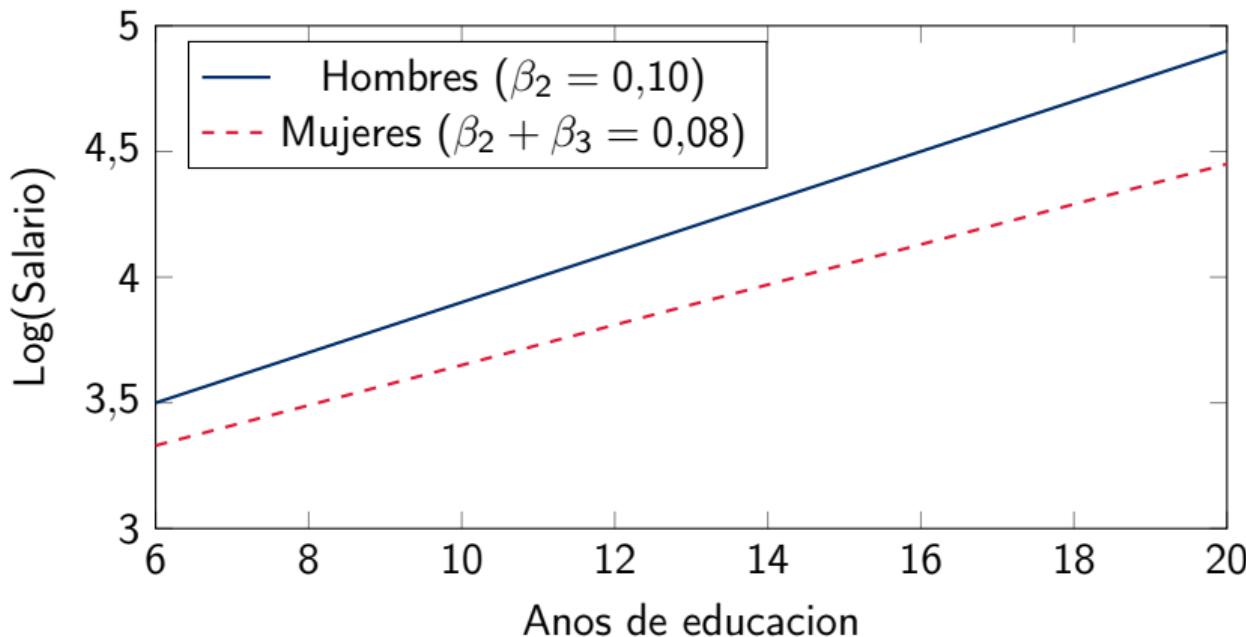
$$E[\ln(w)] = (\beta_0 + \beta_1) + (\beta_2 + \beta_3) \cdot S$$

Interpretacion de  $\beta_3$

Diferencia en el retorno a educacion entre mujeres y hombres.

Si  $\beta_3 = -0,02$ : cada año de educacion rinde 2 puntos porcentuales **menos** para mujeres.

# Visualizacion: Diferentes Pendientes



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

# 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 educacion  
* i.mujer#c.escolaridad = interaccion
```

## Notacion Stata

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

# Interaccion Dummy × 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 | $\beta_0$           | $\beta_0 + \beta_2$                     |
| 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 | $\beta_0$           | $\beta_0 + \beta_2$                     |
| Mujer  | $\beta_0 + \beta_1$ | $\beta_0 + \beta_1 + \beta_2 + \beta_3$ |

Interpretacion de  $\beta_3$ :

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

## 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  
  
* Prediccion promedio por genero  
margins mujer  
  
* Retorno a educacion por genero  
margins , dydx(escolaridad) at(mujer=(0 1))  
  
* Efecto marginal de ser mujer en diferentes niveles de educacion  
margins , dydx(mujer) at(escolaridad=(9 12 16 18))
```

## ¿Por que usar margins?

- Predicciones en escala original (no log)
- Errores estandar correctos
- Facil de graficar con marginsplot

## Variables categoricas relevantes para la empresa cliente:

- **Area funcional:** Operativo, Tecnico, Coordinacion, Direccion, Admin
- **Nivel jerarquico:** 1-5
- **Tipo de puesto:** Campo vs Oficina
- **Turno:** Diurno, Nocturno, Mixto

## Variables categoricas relevantes para la empresa cliente:

- **Area funcional:** Operativo, Tecnico, Coordinacion, Direccion, Admin
- **Nivel jerarquico:** 1-5
- **Tipo de puesto:** Campo vs Oficina
- **Turno:** Diurno, Nocturno, Mixto

## Preguntas a responder con dummies e interacciones:

- ① ¿Cuanto mas gana un Coordinador vs Tecnico (controlando por educacion)?
- ② ¿El retorno a experiencia es igual en campo y oficina?
- ③ ¿La brecha entre areas crece con la antiguedad?

## Ejemplo: Estructura de la Empresa

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

## Ejemplo: Estructura de la Empresa

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

Para el tabulador

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

## Variables Dummy:

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

## Interacciones:

- Permiten efectos heterogéneos
- Dummy × Continua: pendientes diferentes
- Dummy × Dummy: efectos condicionales

## Comandos Stata clave:

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

# ¿Preguntas?

Proxima Sesion:

**M03: Diagnosticos OLS y Errores Robustos**

Jueves 12 de febrero, 3-5pm

**Entrega E1: Jueves 12 feb, 11:59pm**