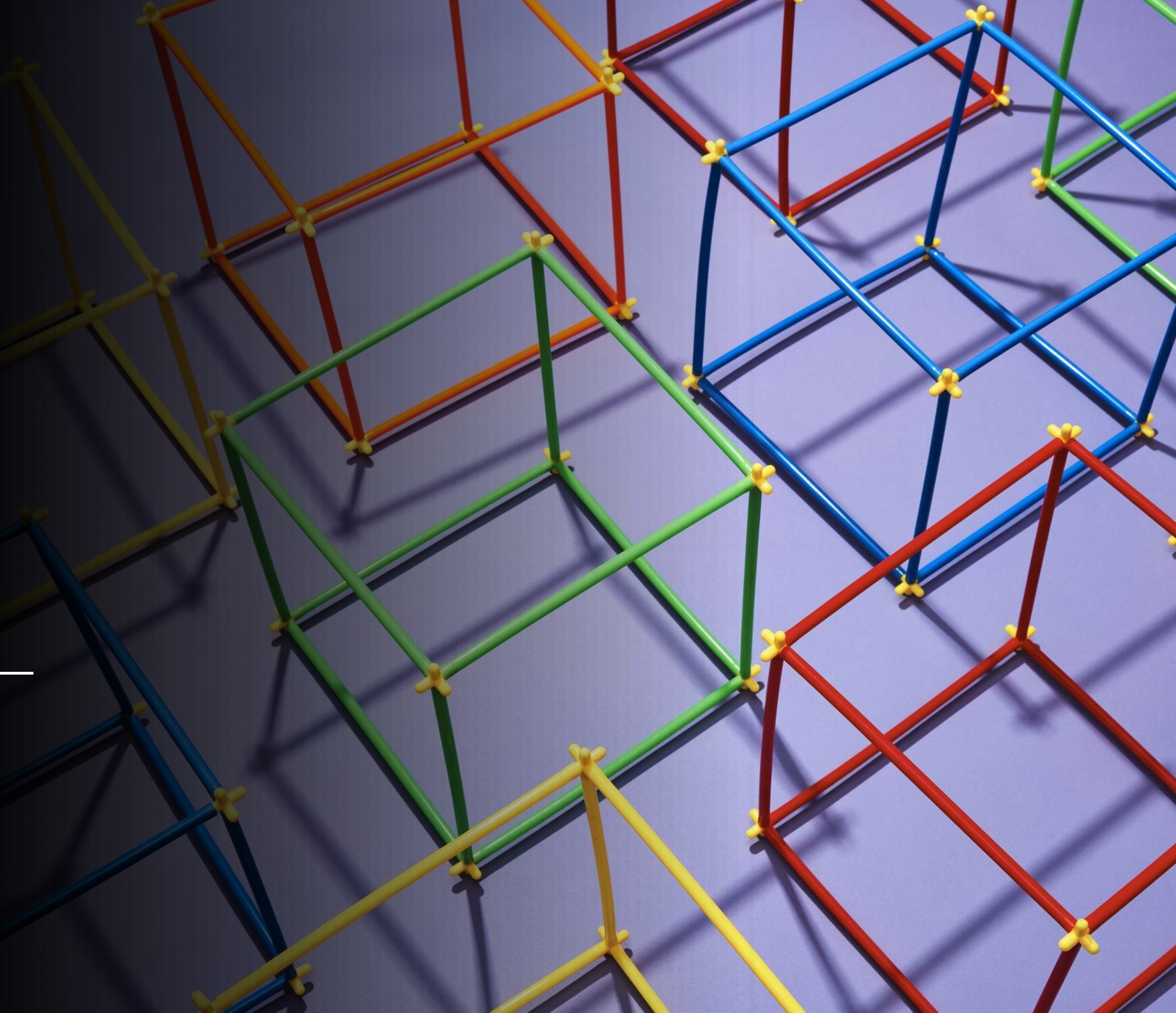
Evaluación de impacto: Tema 2

Francesco Bogliacino

OLS y Mean Independence



Primero una terminología

- Y: LHS, variable dependiente, outcome, variable explicada, variable predecida
- X: RHS, variable independiente, explicativa, control, predictor, regresor

1. ¿Cuál es la relación funcional entre x & y?
2. ¿Cómo permitimos que otras variables aparte de x afecten y?
3. ¿Cómo hacemos para identificar una relación ceteris paribus?

OLS

- La regresión lineal es un modelo sobre la **población;**

$$\boldsymbol{Y = \beta_0 + \beta_1 x + u}$$

- para poder estimar $\boldsymbol{\beta_0}$ y $\boldsymbol{\beta_1}$necesitamos caracterizar las relación entre x y u **en la población**

Supuestos

- Hacemos primero el supuesto que el valor esperado de u, que escribimos E[u]=0
- ¿Qué es el valor esperado? sería la media calculada sobre la población entera que es $\int u(i)p(u(i))di$ o $\sum u_i p(u_i)$
- Es un problema ese supuesto? NO, asuma que no sea cero sino α_0

Puedo redefinir $\boldsymbol{Y} = (\boldsymbol{\beta_0} + \alpha_0) + \boldsymbol{\beta_1 x} + (\boldsymbol{u} - \alpha_0)$

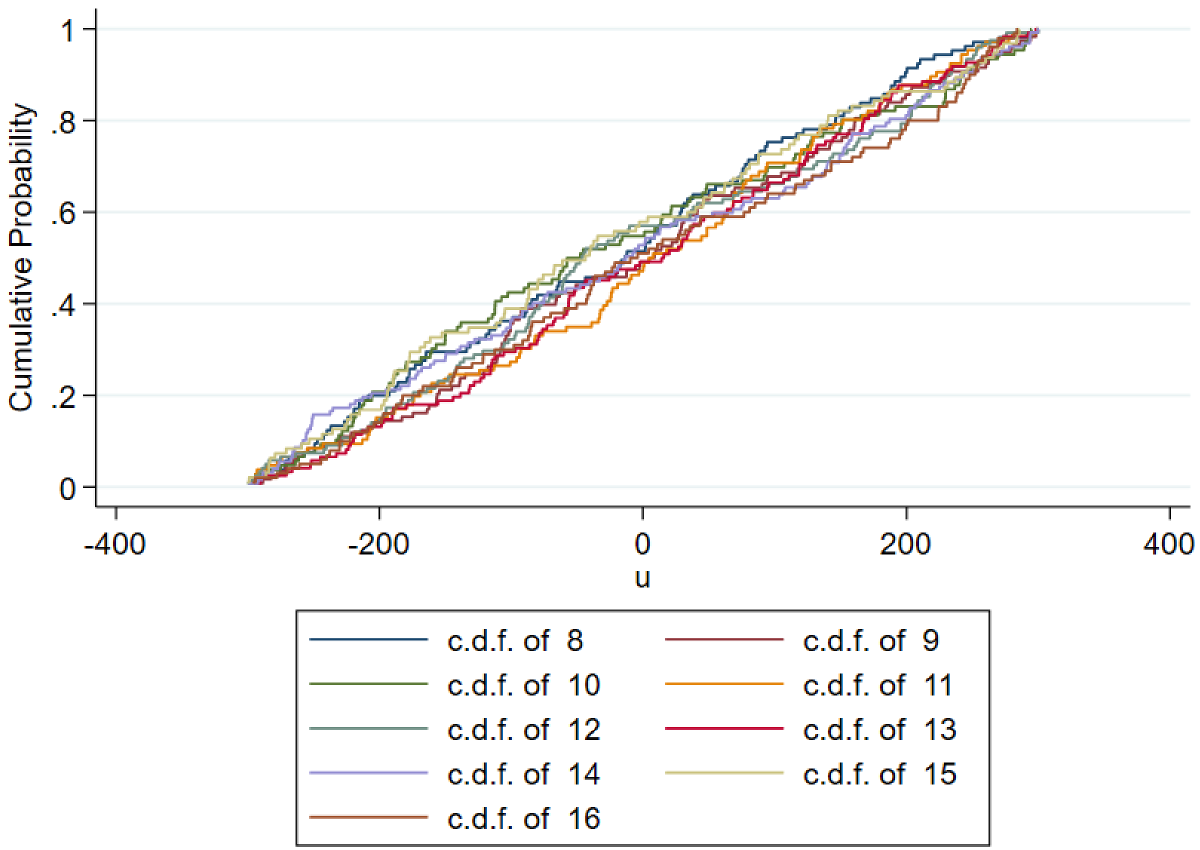
Supuestos

- Un supuesto clave (con el cual lidiaremos todo el curso) es este:
- $E[u|x] = E[u]$

	Altas habilidades blandas	**Bajas habilidades blandas**	
Graduado	0.16	0.14	0.30
No graduado	0.37	0.33	0.70
	0.53	0.47	

- $E[u] = 2 * 0.53 + 1 * 0.47 = 1.53$
- $E[u|G] = 2 * 0.53 + 1 * 0.47 = 1.53 = E[u|NG]$

	Altas habilidades blandas=2	**Bajas habilidades blandas=1**	
Graduado	P(u=2|G)=0.16/0.30=0.53	P(u=1|G)=0.14/0.30=0.47	0.30
No graduado	0.37	0.33	0.70
	0.53	0.47	



Cumulative Probability
1
.8
.6
.4
.2
0
-400
-200
0
200
400
u
c.d.f. of 8
c.d.f. of 9
c.d.f. of 10
c.d.f. of 11
c.d.f. of 12
c.d.f. of 13
c.d.f. of 14
c.d.f. of 15
c.d.f. of 16

Mean independence

- Obviamente en muchos casos (ej educación) este es un mal supuesto;
- Este segundo supuesto es lo que llamamos independencia en promedio (mean Independence en inglés)

OLS

- La regresión lineal es un modelo sobre la **población;**

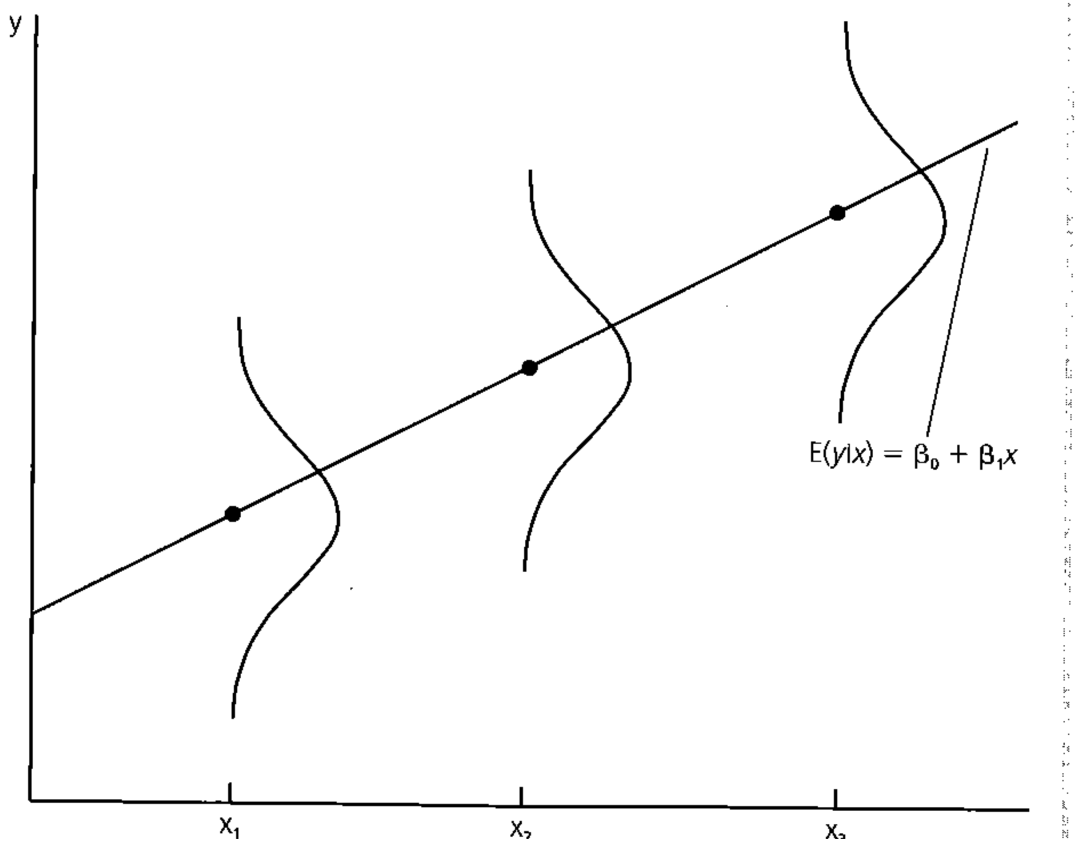
$$\boldsymbol{Y = \beta_0 + \beta_1 x + u}$$

- Qué pasa con el supuesto de independencia en promedio?

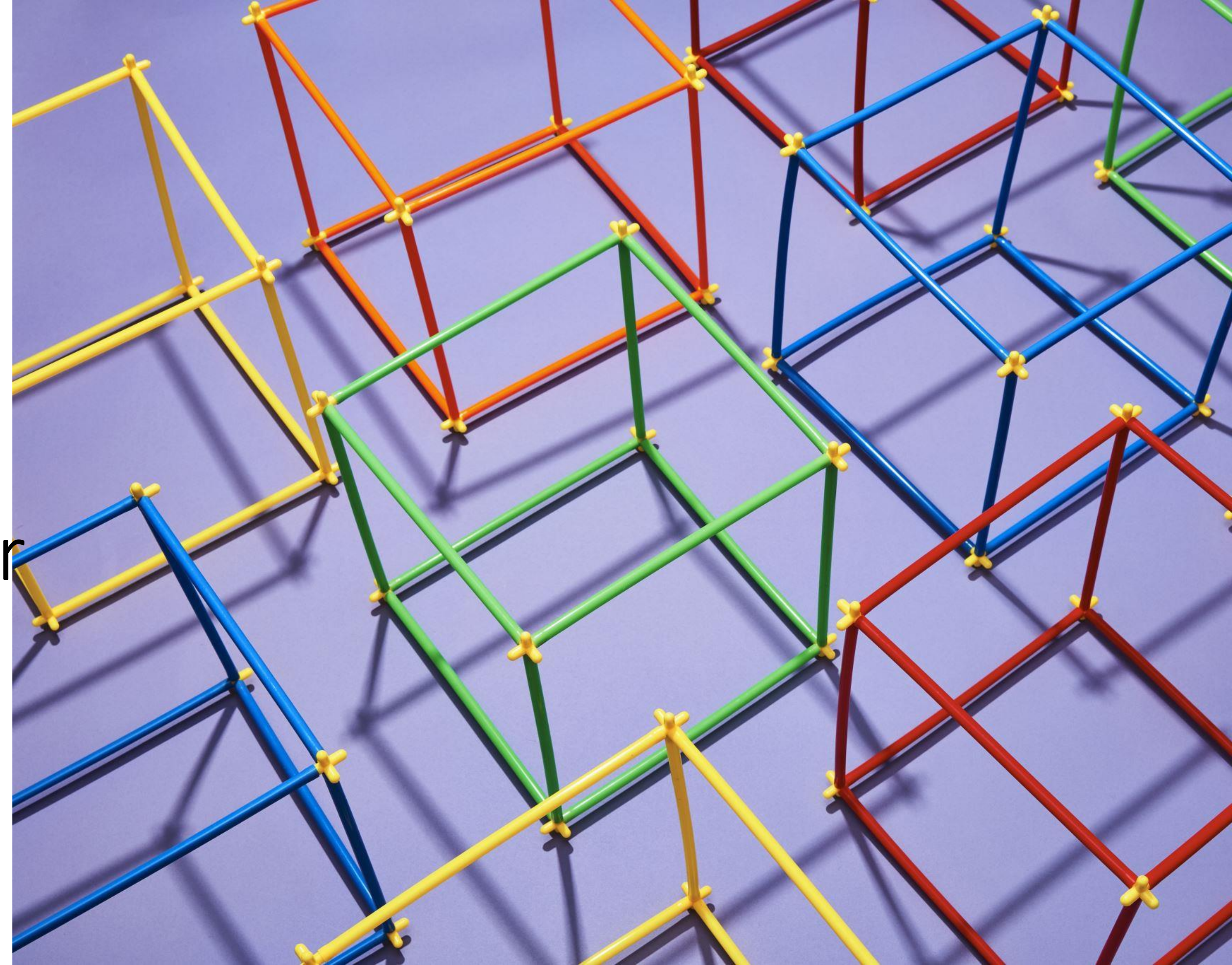
$$\boldsymbol{E[Y|x] = \beta_0 + \beta_1 x + E[u|x] = \beta_0 + \beta_1 x}$$

La CEF o PRF en práctica

$E(y|x)$ as a linear function of x.



El estimador



¿Cómo estimamos?

- Primero nosotros queremos "identificar" $\boldsymbol{\beta_0}, \boldsymbol{\beta_1}$ pero los "estimamos" porque no trabajamos con la población sino con una muestra $\{\boldsymbol{y_i}, \boldsymbol{x_i}\}$ y no observamos $\boldsymbol{u_i}$;
- El estimador nos dará un valor $\widehat{\boldsymbol{\beta}}$ que depende de la muestra, es decir a muestras diferentes sacaremos valores diferentes;
- Esta es la "variabilidad" de la muestra:
  - Piensen en Lalonde, el valor estaba alrededor de los 800 USD, pero muestras diferentes me dan valores diferentes
  - Quisiéramos pegarle en "promedio"

Algunas implicaciones de mean independence

- Si $\boldsymbol{E}[\boldsymbol{u}|\boldsymbol{x}] = \boldsymbol{0}$ es obvio que $\boldsymbol{E}[\boldsymbol{u}] = \boldsymbol{0}$
- Si $\boldsymbol{E}[\boldsymbol{u}|\boldsymbol{x}] = \boldsymbol{0}$ es obvio que $\boldsymbol{Cov}(\boldsymbol{x}, \boldsymbol{u}) = \boldsymbol{E}[\boldsymbol{xu}] = \boldsymbol{0}$

Llamamos $x_i', x_i'', \ldots, x_i^n$ los valores que asume la variable x

- $\frac{1}{N}\sum_{i=1}^{N} x_i\, u_i = \frac{1}{N}[N' x_i' \left(\frac{1}{N'}\sum_{i=1}^{N'} u_i\right) + \cdots + Nn\; x_i^N(\frac{1}{Nn}\sum_{i=1}^{N\prime} u_i)]$
- Pero sabemos que todos esos promedios son ceros

- Noten que acabamos de hacer $\boldsymbol{E}[\boldsymbol{xu}] = \boldsymbol{E}[\boldsymbol{xE}[\boldsymbol{u}|\boldsymbol{x}]]$ que tiene el nombre de Ley de Expectativas Iteradas

Algunas implicaciones de mean independence

- Si $\boldsymbol{E[u|x] = 0}$ es obvio que $\boldsymbol{E[u] = E[y - \beta_0 + \beta_1 x] = 0}$
- Si $\boldsymbol{E[u|x] = 0}$ es obvio que $\boldsymbol{Cov(x, u) = E[x(y - \beta_0 - \beta_1 x)] = 0}$

- En la muestra:
- $\frac{1}{N}\sum_{i=1}^{N}(y_i - \widehat{\beta_0} - \widehat{\beta_1} x_i) = 0$
- $\frac{1}{N}\sum_{i=1}^{N} x_i(y_i - \widehat{\beta_0} - \widehat{\beta_1} x_i) = 0$

Ahora lo que hay que hacer es resolver esas dos ecuaciones, en dos incognitas

Algunas implicaciones de mean independence

- $\frac{1}{N}\sum_{i=1}^{N} y_i - \widehat{\beta_0} - \widehat{\beta_1}\frac{1}{N}\sum_{i=1}^{N} x_i = 0 \rightarrow \widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\bar{x}$
- $\frac{1}{N}\sum_{i=1}^{N} x_i(y_i - \widehat{\beta_0} - \widehat{\beta_1}x_i) = 0 \rightarrow$

$$\frac{1}{N}\sum_{i=1}^{N} x_i(y_i - [\bar{y} - \widehat{\beta_1}\bar{x}] - \widehat{\beta_1}x_i) = 0$$

$$\frac{1}{N}\sum_{i=1}^{N} x_i(y_i - \bar{y}) = \widehat{\beta_1}\frac{1}{N}\sum_{i=1}^{N} x_i(x_i - \bar{x})$$

$$\frac{1}{N}\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y}) = \widehat{\beta_1}\frac{1}{N}\sum_{i=1}^{N} (x_i - \bar{x})^2$$

Algunas implicaciones de mean independence

- Si $\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 > 0$ entonces

$$\widehat{\beta_1} = \frac{covarianza.m(x,y)}{varianza.m(x)}$$

OLS

- OLS es lo que acabamos de ver
- La lógica es que usamos la variabilidad en la muestra de x para poder identificar el impacto sobre y
- Pero ojo que:
  - Estamos USANDO un SUPUESTO sobre la población
  - Siempre podemos **calcular** OLS (a menos que X no varíe) pero no necesariamente **identificamos** el beta de la población
- Para calcular OLS en esta manera usamos condiciones sobre los MOMENTOS
- Una alternativa es la siguiente

OLS

- Define $\widehat{y_i}$ como la predicción. Puedo querer predecir el sueldo de la unidad i sabiendo su educación x_i
- $\widehat{y_i} = \widehat{\beta_0} - \widehat{\beta_1} x_i$
- El error de predicción es $\widehat{u_i} = y_i - \widehat{\beta_0} - \widehat{\beta_1} x_i$
- Definamos un criterio a minimizar,

$$\frac{1}{N} \sum_{1}^{N} \widehat{u_i}^{2}$$

- Minimizando eso logramos exactamente lo mismo

OLS

- OJO! Los errores de predicción o residuales NO son los términos u_i que nunca observamos;
- Los residuales por construcción suman a cero

$$\frac{1}{N}\sum_{i=1}^{N}(y_i - \widehat{\beta_0} - \widehat{\beta_1}x_i) = 0$$

- Y por construcción $\frac{1}{N}\sum_{i=1}^{N} x_i(y_i - \widehat{\beta_0} - \widehat{\beta_1}x_i) = 0$ son ortogonales a x_i
- No usen la ortogonalidad de los residuales para inferir independencia de los u_i de x_i (si lo hacen me va a caer una lagrimita)

OLS

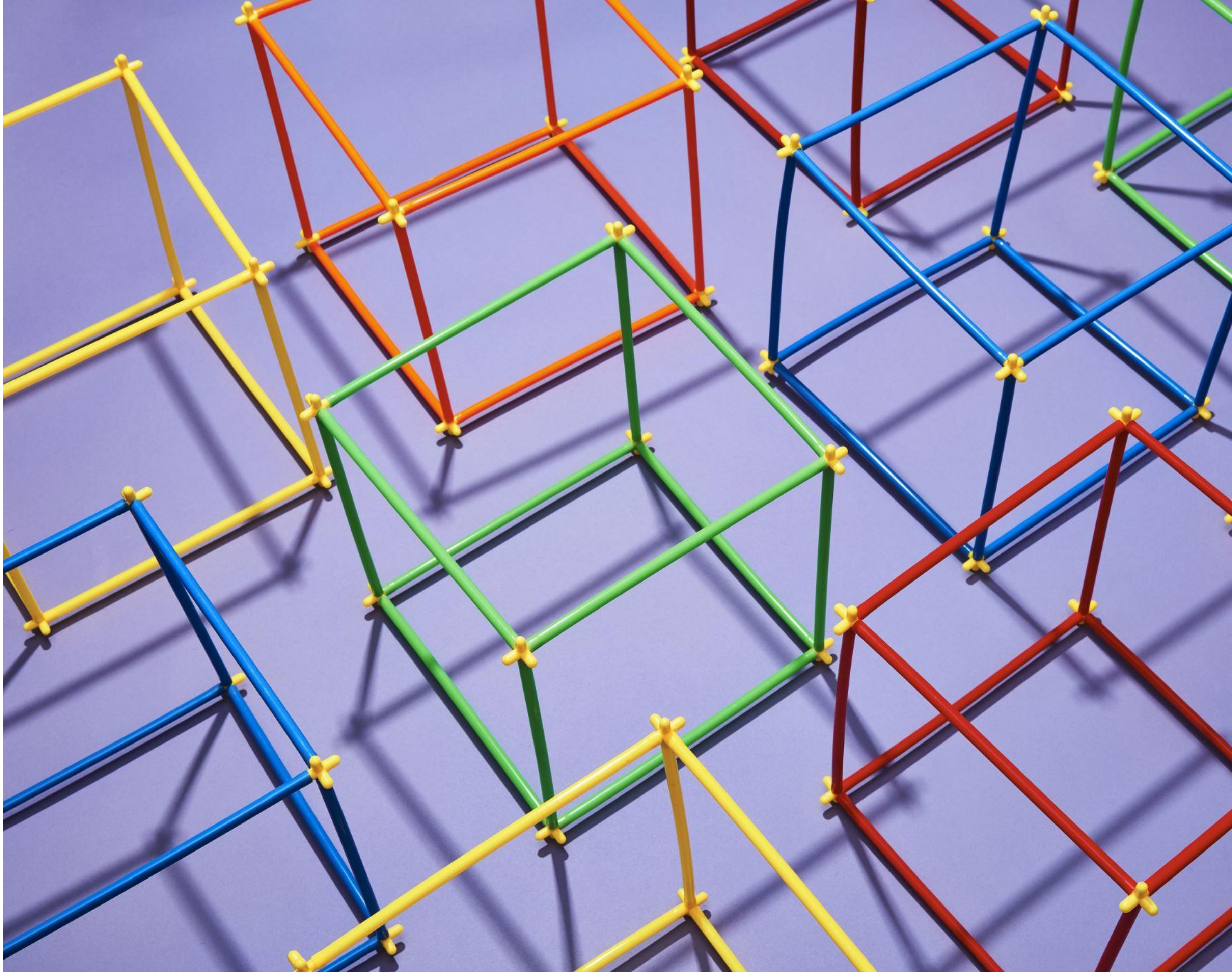
- Recuerden también que $\widehat{y_i} = \widehat{\beta_0} - \widehat{\beta_1} x_i$
- $\frac{1}{N}\sum_{i=1}^{N} \widehat{y_i}\, \widehat{u_i} = \widehat{\beta_0}\frac{1}{N}\sum_{i=1}^{N} \widehat{u_i} - \widehat{\beta_1}\frac{1}{N}\sum_{i=1}^{N} x_i \widehat{u_i} = 0$
- Es decir las predicciones son ortogonales a los errores, y así cualquier función lineal

- Finalmente recuerden que $\widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\bar{x}$ o sea que si hago la predicción para el promedio saco el promedio. Esto es porque decidimos que así fuera

Resumen: propriedades algébricas de OLS

- $(\bar{x}, \bar{y})$ está en la recta de regresión
- Las predicciones y los residuales no están correlacionados
- Los residuales y el predictor no están correlacionados
- Los residuales suman a cero

OLS
y su valor esperado



Tengamos presente el objetivo

- Nosotros observamos una muestra
- Quisiéramos decir algo sobre la población

- Vamos a ver como nos va con OLS en Stata

OLS

- Como ocurrió con el ejemplo simulado, OLS tiene variabilidad, cada muestra nos restituye un valor diferente que depende de la características de los datos
- En el ejemplo simulado, en promedio acertábamos, pero se ve una campana alrededor del valor verdadero;
- Vimos también que dada la varianza de los datos, la cantidad de observaciones y el efecto que queremos identificar, hay un cierto número de veces que nos equivocamos en una dirección o en la otra

Supuestos (1)

- El modelo para la población se puede escribir $\boldsymbol{Y = \beta_0 + \beta_1 x + u}$

- Implicaciones: nuestros parámetros de interés son $\boldsymbol{\beta_0, \beta_1}$
- Recuerden donde queremos ir: nosotros vamos a tener un modelo así por definición de Outcome Alternativos Potenciales
- **X** y **u** son realizaciones de variables aleatorias por lo tanto lo es **y**
- Este supuesto es el de ***linealidad en los parámetros***

Supuestos (2)

- El muestreo es aleatorio e independiente
- Para cada i podemos escribir $\boldsymbol{y_i = \beta_0 + \beta_1 x_i + u_i}$

- Ojo que $\boldsymbol{u_i}$ es la componente no observable. No es el error de predicción

Supuestos (3): obvio

- Tenemos variabilidad en la $\boldsymbol{x_i}$
- Esto en práctica es obvio, sin variabilidad en la x podemos cerrar todo e irnos

Supuestos (4)

- Este es el supuesto clave: la componente no observable tiene media condicional cero

$$E[u|x] = 0$$

Recuerden: yo (o Stata, o R) podemos calcular OLS que eso se cumpla o menos y de hecho independientemente hasta del supuesto 1

OLS es no sesgado

- Queremos mostrar que $E\left[\widehat{\beta_1}\right] = \beta_1$

$$\widehat{\beta_1} = \frac{\sum_i (y_i - \bar{y})(x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

Llamemos $\sum_i (x_i - \bar{x})^2 = TOT_x$

Sabemos que $TOT_x > 0$ por el supuesto (3)

OLS es no sesgado

- Remplacemos (por supuesto 1 y 2) $\boldsymbol{y_i = \beta_0 + \beta_1 x_i + u_i}$

$$\sum_i (y_i - \bar{y})(x_i - \bar{x}) = \sum_i (x_i - \bar{x})(\beta_0 + \beta_1 x_i + u_i) =$$

$$\beta_0 \sum_i (x_i - \bar{x}) + \beta_1 \sum_i (x_i - \bar{x})\, x_i + \sum_i (x_i - \bar{x})\, u_i =$$

$$\beta_1 TOT_x + \sum_i (x_i - \bar{x})\, u_i$$

$$\widehat{\beta_1} = \beta_1 + \frac{\sum_i (x_i - \bar{x})\, u_i}{\sum_i (x_i - \bar{x})^2}$$

¿Qué es ese último término?

OLS es no sesgado

- Definan $w_i = \frac{x_i - \bar{x}}{\sum_i (x_i - \bar{x})^2}$
- Podemos escribir el último término como $\sum_i w_i u_i$
- recuerden que:
  - cada w_i es función de los x_i
  - El estimador es entonces función lineal de los términos no observables

OLS es no sesgado

- Vamos a usar el mismo trick que hemos aprendido

$$E\left[\widehat{\beta_1}\right] = \beta_1 + E\left[\sum_i w_i u_i\right] =$$

$$\beta_1 + \sum_i E[w_i u_i] = \beta_1 + \sum_i E[w_i E[u_i|x_i]$$

- Luego usamos el supuesto (4) y $E\left[\widehat{\beta_1}\right] = \beta_1$
- Es clave que en la población x y u no sean correlacionados o esto no ocurriría

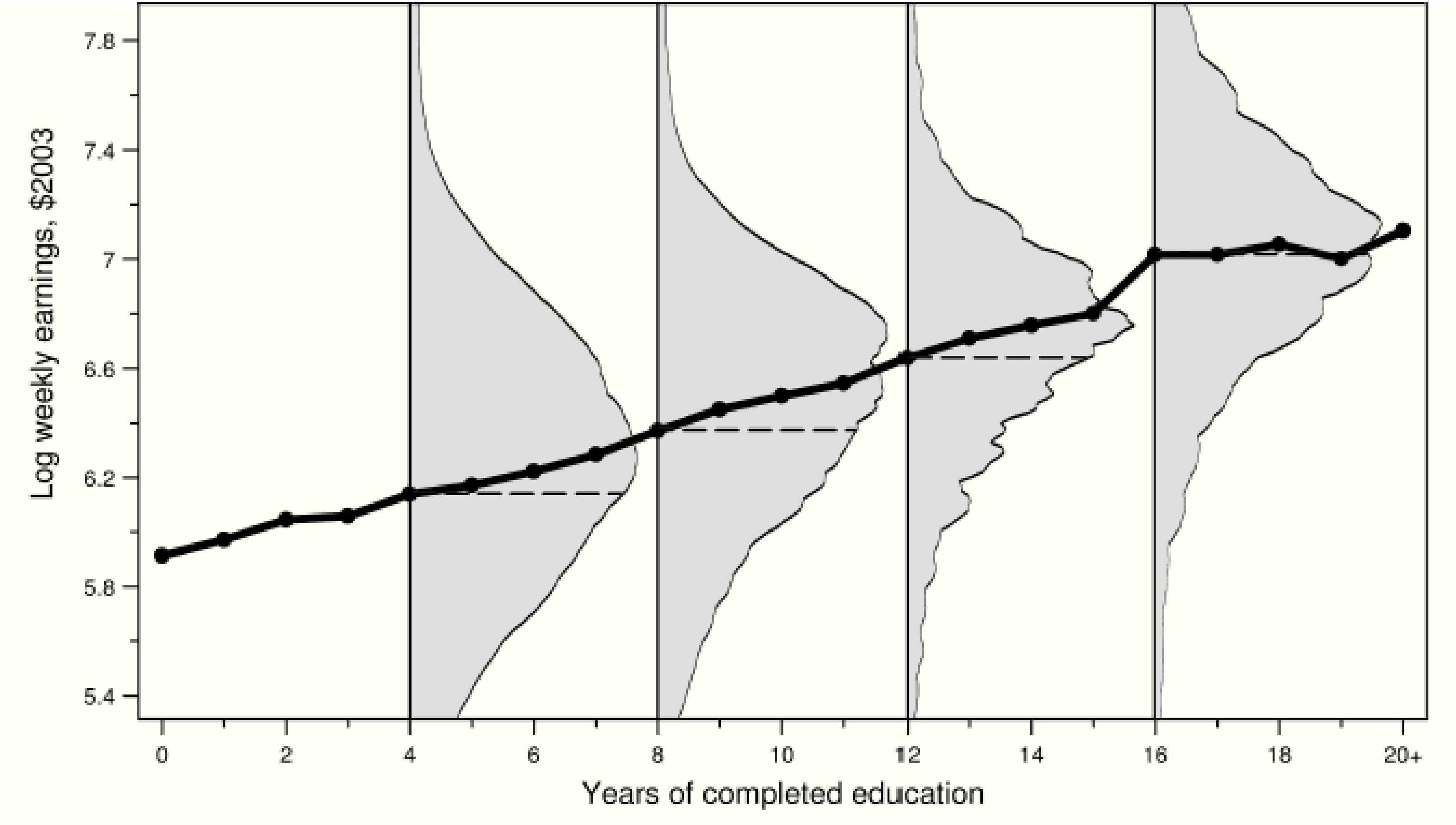
Resumen

- Con los supuestos (1)-(4) OLS no es sesgado
- Problema: cada muestra nos dará un valor diferentes. En general nos concentramos cerca del valor del parámetro
- Nosotros podemos esperar que nuestra muestra sea “típica” pero no sabemos porque
- Recuerden también:
  - Los u_i no son observables, son las distancias de las observaciones respeto a la CEF o RPF
  - Los residuales son las distancias verticales entre las observaciones y la relación estimada. Eso si se observan (más precisamente se calculan)

W LINEAL

Back to CEF

- Trabajemos con una población finita, sueldo y educación, hay un número de estados del mundo $x_i = \{1, \ldots, 16\}$ que son variables aleatorias
- $y_i = I(x_i = 1)n^{-1} \sum_i y_i^{cuando\ tienen\ 1}$+…+ $I(x_i = 16)n^{-1} \sum_i y_i^{cuando\ tienen\ 16} + algo$
- En otras palabras puedo decir que una variable es la suma de dos componentes: lo que puedo predecir dado que está en el estado x_i más un algo que ya no depende [=es independiente] de x_i



Source: MHE, p. 24

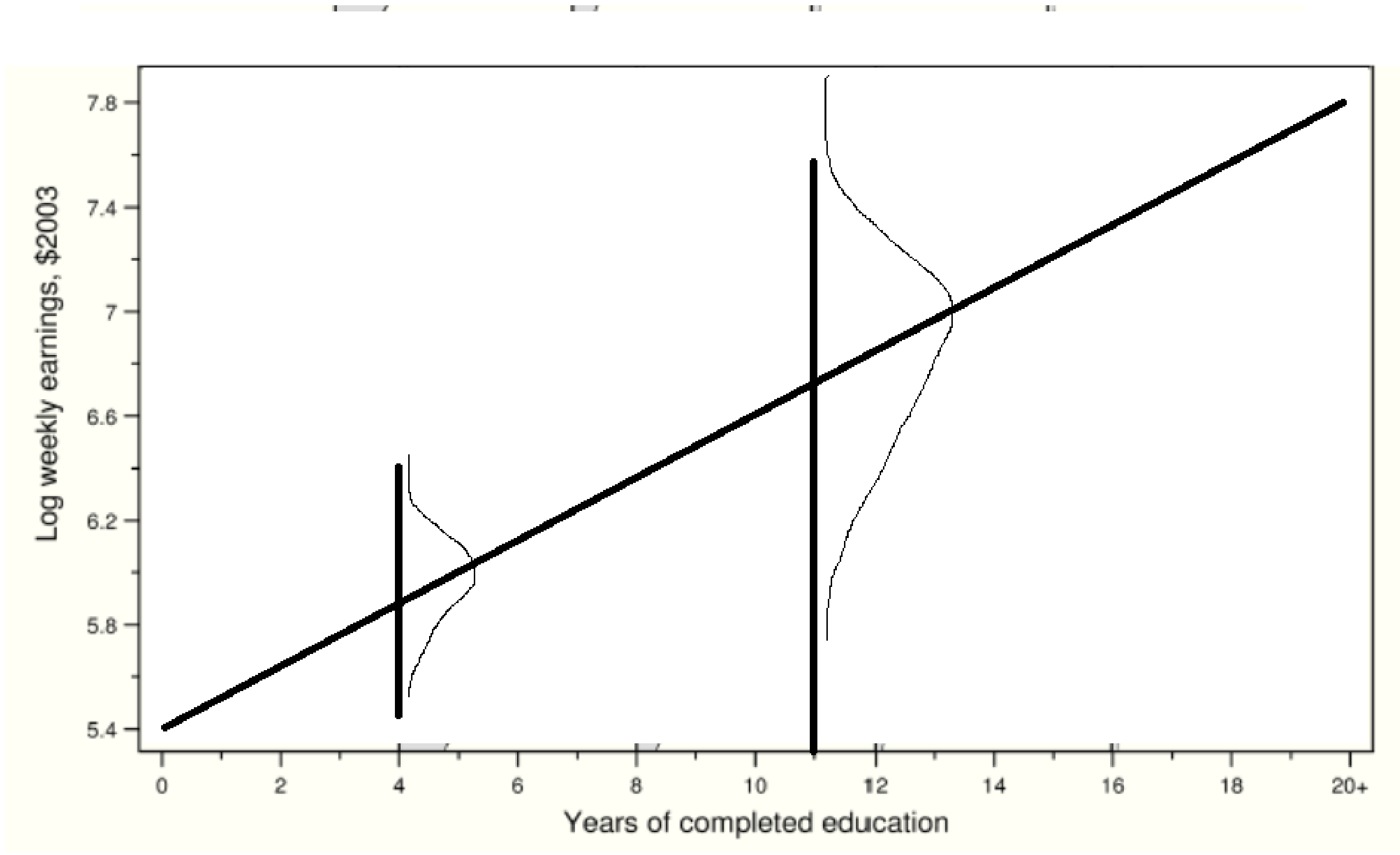
Dos interpretaciones

- Una variable y_i la puedo “romper” en dos componentes:
  - Lo que me explica la x_i y lo que no me explica la x_i
- Una variable la puedo predecir de la siguiente manera:
  - Lo que me ayude a predecir x_i y lo que no me ayuda a predecir la x_i

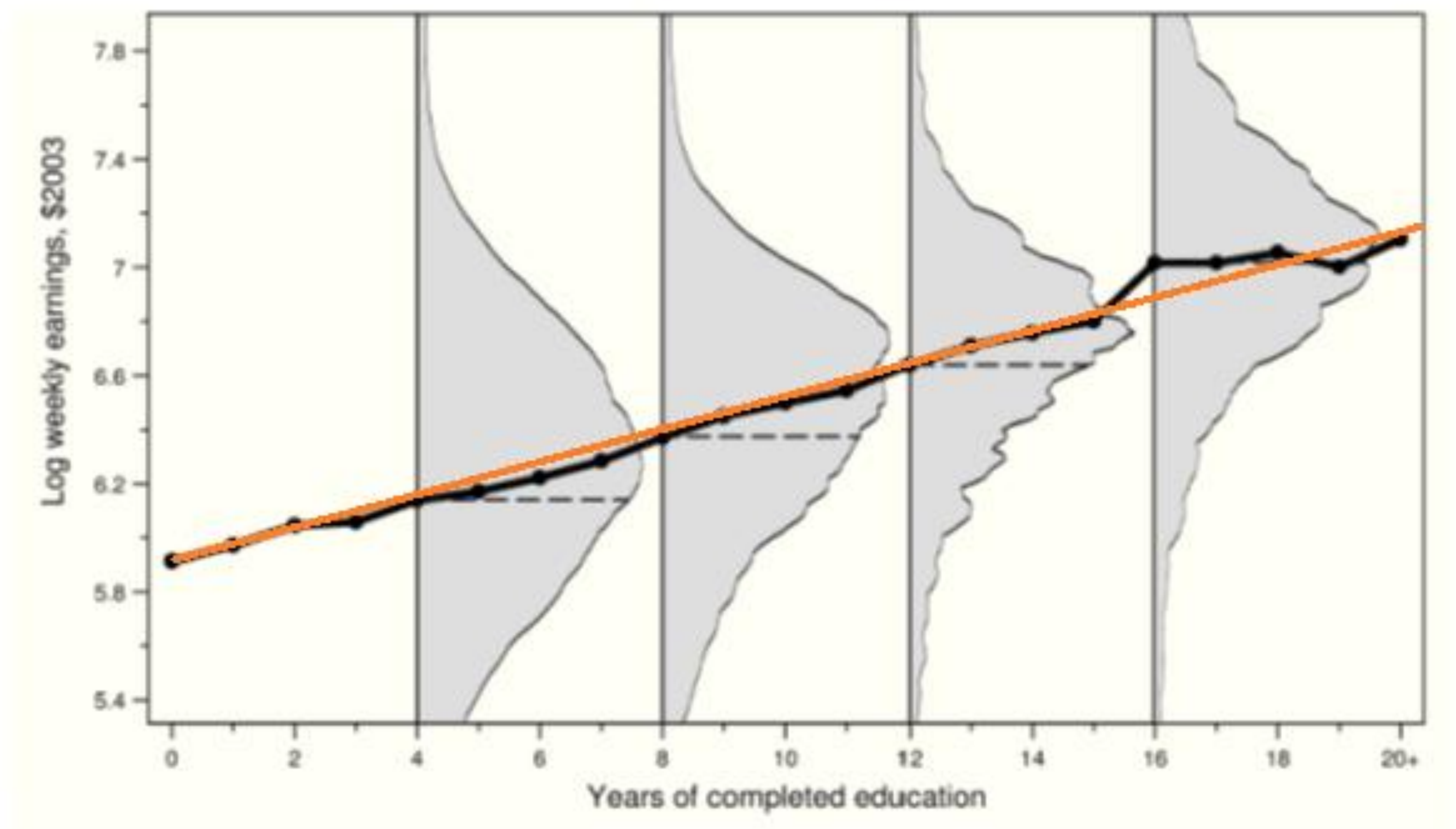
- Esa primera componente es la CEF, la $E[y_i|x_i]$, por lo cual la relación en la población siempre es

$$y_i = E[y_i|x_i] + \varepsilon_i = CEF + algo\ ortogonal$$

Por qué Lineal es vida ☺



Por qué Lineal es vida ☺



Recuerden hacia siempre donde vamos

- Vamos a usar la Regresión Lineal en la población para escribir la CEF

$$\boldsymbol{E[Y|x] = \beta_0 + \beta_1 x}$$

- Porque al final lo que nos interesa es la CEF (y como esa se conecta a nuestro caso especifico donde la x_i toma los dos valores 1 y 0), no las características especificas del soporte de una variable
- Eso nos permite una cuantificación "sencilla" y "transparente" de los coeficientes, que no *depende* de otras variables

Dos formulas importantes

- La anatomía de la regresión
- La variable omitida

La anatomía de la regresión

- Supongamos que la educación sea tan buena como aleatoriamente asignada una vez tengamos en cuenta la discriminación por género y el estrato socio-económico

- Miremoslo en una base de datos (-> Stata)

¿Qué quiere decir?

- Cuando introducimos variables de control además de nuestra x_i estamos usando la variabilidad *que queda* después de "igualar" las personas en término de estas variables de control
- Por esto es importante:
  - No controlar por cualquier cosa, sino por variables que me ayuden en eliminar la selección

La variable omitida

- No controlar por una variable que explica la selección impide recuperar el valor del parámetro, en otras palabras no nos permite estimar sin sesgo -> hay algo más, aparte de la variabilidad de la muestra, que nos afecta
- Pensemos en la educación, y el hecho que la habilidad también cuente en la educación y en el mercado del trabajo
- Back to the data....

La variable omitida

- Se lo voy a derivar para un caso más sencillo

$$y_i = \beta_0 + \beta_1 S_i + \beta_2 A_i + u_i$$

¿Qué pasa si no observo A?

$$\widehat{\beta_1} = \frac{Cov(S_i, y_i)}{Var(S_i)} = \frac{E[S_i, (\beta_0 + \beta_1 S_i + \beta_2 A_i + u_i)] - E[S_i]E[\beta_0 + \beta_1 S_i + \beta_2 A_i + u_i]}{Var(S_i)} =$$

$$= \frac{\beta_0 E[S_i] + \beta_1 E\left[S_i^2\right] + \beta_2 E[S_i A_i] + E[S_i u_i] - E[S_i]\beta_0 - \beta_1 E[S_i]^2 - \beta_2 E[S_i]E[A_i] + E[S_i]E[u_i]}{Var(S_i)}$$

$$= \frac{\beta_1 Var\left[S_i^2\right] + \beta_2 Cov[S_i A_i] + Cov[S_i u_i]}{Var(S_i)} = \beta_1 + \frac{\beta_2 Cov[S_i A_i]}{Var(S_i)}$$

La variable omitida

- Si $Cov[S_i A_i] > 0$ y A_i no la observo, yo puedo tener muchos datos, muy limpios y hacer un montón de regresiones pero nunca identificaré el efecto de la educación (a lo mejor se en qué dirección estoy sobre estimando)
- Esta va a ser la norma cuando trabajemos con datos observables
- Para poder identificar β_1 necesito una estrategia de identificación

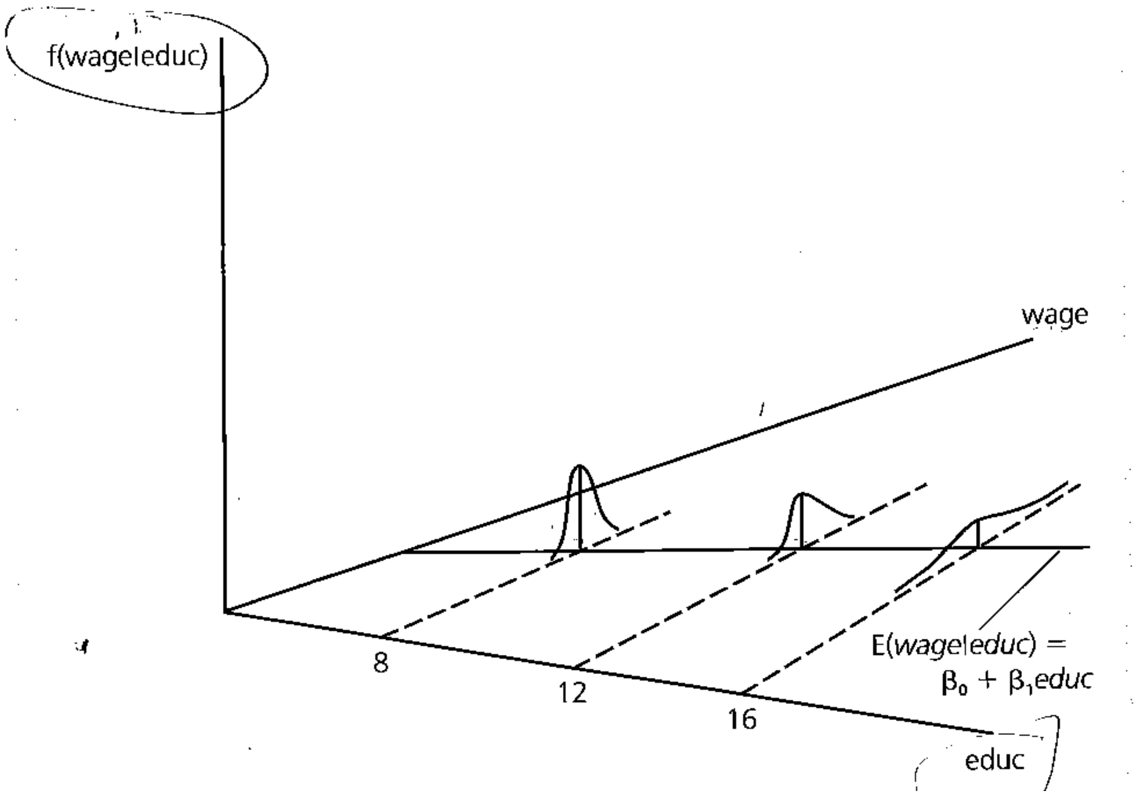
¿Y la varianza?

Un último esfuerzo

heteroschedasticity

Figure 2.9

Var (*wage*|*educ*) increasing with *educ*.



Dependencia en los datos

- A veces hay dependencia en los datos, cuando las observaciones de un grupo (cluster) se “contaminan”

- Como nos va con ,rob -> back to Stata