



Modelos de regresión: Introducción

Christoper A. Alarcon-Ruiz MD MSc(c)

(Parte de la presentación en coordinación con Álvaro Taype Rondan)



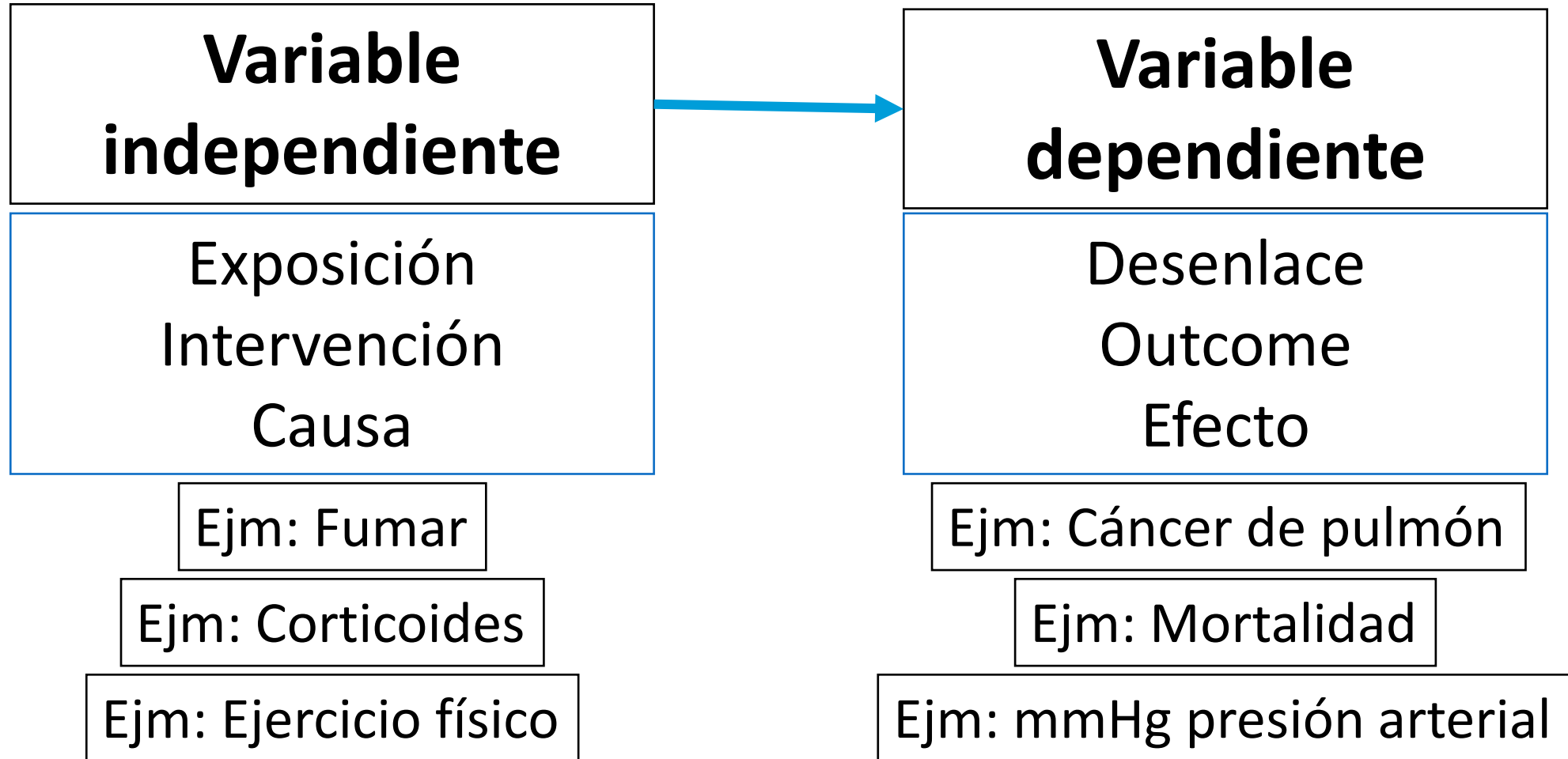
Christopher A. Alarcon Ruiz
MD, MSc(c)

christoper.alarconr20@gmail.com



- **Médico epidemiólogo**
- **Afiliaciones actuales:**
 - Consultor en Dirección de Investigación IETSI, EsSalud, Lima, Perú
 - Asistente de investigación en Universidad San Ignacio de Loyola, Lima, Perú
 - Docente en Universidad Científica del Sur, Lima, Perú
- **Potenciales conflictos de interés:**
 - Cursó estudios en: URP, UPCH, UChile
 - Ha recibido pagos/financiamientos de: USIL, UCSUR, IETSI-EsSalud, CMP, MINSA, EviSalud
 - **Financiamiento de la industria farmacéutica: ninguno**

Variable independiente y dependiente



- En estudios biomédicos, generalmente vamos a evaluar una relación de **causa-efecto**.
- Es decir, si la **VI** tiene un impacto en la **VD** (la aumenta o la disminuye)



Variable independiente y dependiente

En pacientes con diabetes, ¿la actividad física (minutos por día) está asociada a la glucosa en ayunas (mg/dL)?



Población		Pacientes con diabetes
Variable independiente	Nombre	Actividad física
	Tipo	Numérica
Variable dependiente	Nombre	Glucosa en ayunas
	Tipo	Numérica

Variable independiente y dependiente

En pacientes con forunculosis recurrente, ¿dar antibióticos profilácticos disminuye el número de recurrencias a los seis meses?



Población

Pacientes con forunculosis recurrente

Variable
independiente

Nombre

Recibir antibióticos profilácticos

Tipo

Dicotómica

Variable
dependiente

Nombre

Número de recurrencias a 6m

Tipo

Numérica

Variable independiente y dependiente

En pacientes con tuberculosis, ¿cuáles son los factores asociados a presentar síntomas depresivos?



Población

Pacientes con tuberculosis

Variable
independiente

Nombre

Varias

Tipo

-

Variable
dependiente

Nombre

Presencia de síntomas depresivos

Tipo

Dicotómica

Modelos de regresión

Determinan el impacto de una o más **variables independientes** en la **variable dependiente**

Tienen dos posibles objetivos:

Objetivo 1: Evaluar la asociación entre dos variables

- Asociación simple: ¿Fumar está asociado a cáncer de pulmón?
- Asociación causal: ¿Fumar causa cáncer de pulmón?

"A" causa "B"



Objetivo 2: Predecir el valor de la variable dependiente, para cierto valor de la variable independiente.

- Para una gestante de **70 Kg**, ¿cual es en promedio el **peso esperado de su bebé a término**?



Modelos de regresión

- Existen diversos modelos de regresión.
- Elegir un modelo dependerá de:
 - De qué tipo es la variable dependiente (numérica, dicotómica, tiempo a evento)
 - El cumplimiento de supuestos



Variable dependiente	Regresión	Calcularemos
Numérica	Lineal	Coeficiente (β)
Dicotómica	Logística Poisson	Odds Ratio Riesgo Relativo
Tiempo a evento	Cox	Hazard Ratio

Muchas gracias (:



MEJORES
DECISIONES
EN SALUD



📞 950 876 703
🌐 www.evisalud.net
📘 facebook.com/evisalud/
📷 @evisaludperu

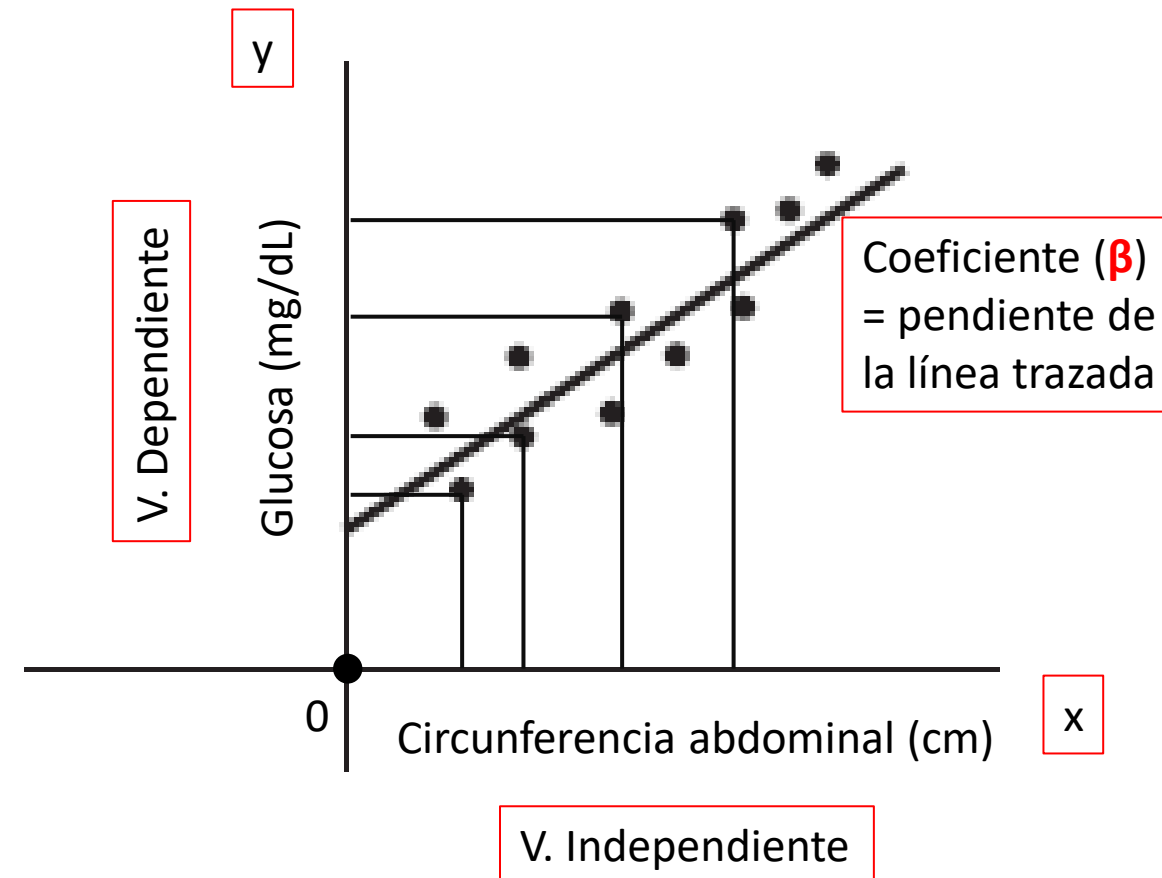


Introducción a la regresión lineal

Christopher A. Alarcon-Ruiz MD MSc(c)

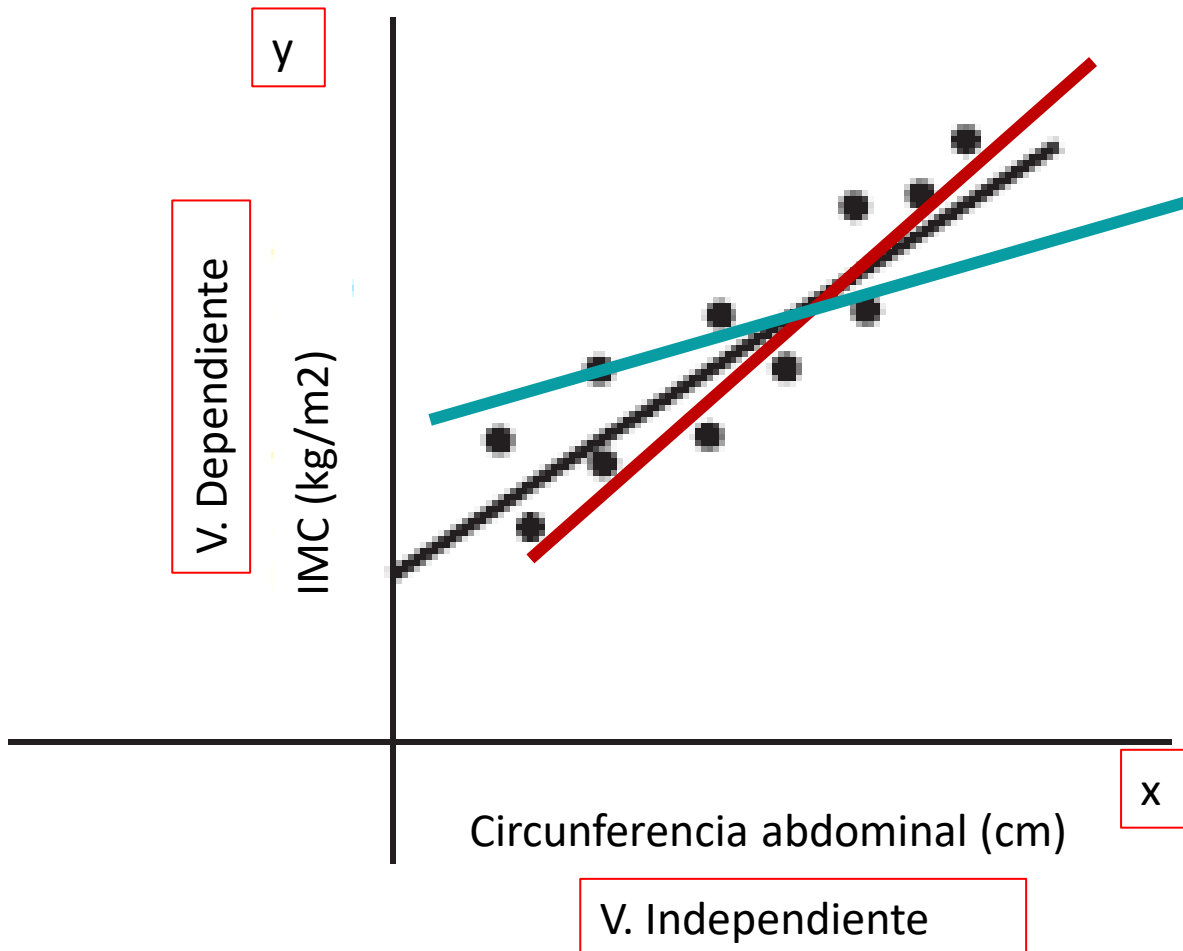
Asociación entre dos variables numéricas

Subject	X	Y
1	74.75	25.72
2	72.60	25.89
3	81.80	42.60
4	83.95	42.80
5	74.65	29.84
6	71.85	21.68
7	80.90	29.08
8	83.40	32.98
9	63.50	11.44
10	73.20	32.22
11	71.90	28.32
12	75.00	43.86
13	73.10	38.21
14	79.00	42.48
15	77.00	30.96
16	68.85	55.78
17	75.95	43.78
18	74.15	33.41



Para cada asociación pueden haber muchas pendientes

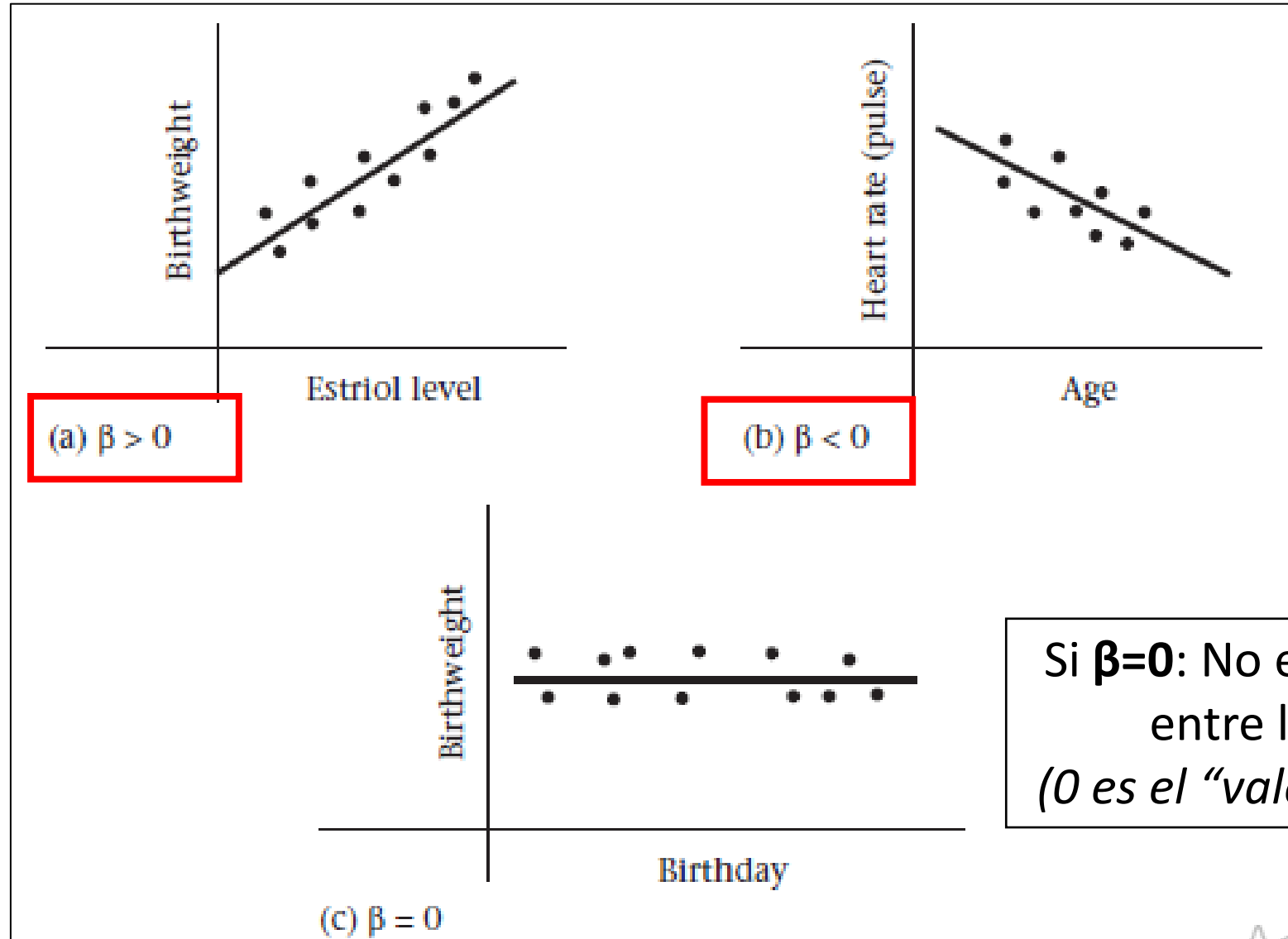
β = Coeficiente = pendiente



¿Cómo saber qué β explica mejor la asociación?

- Stata calculará un β usando la metodología de “mínimos cuadrados”
- Dicho β será la mejor pendiente (la que mejor explica la relación entre X e Y)

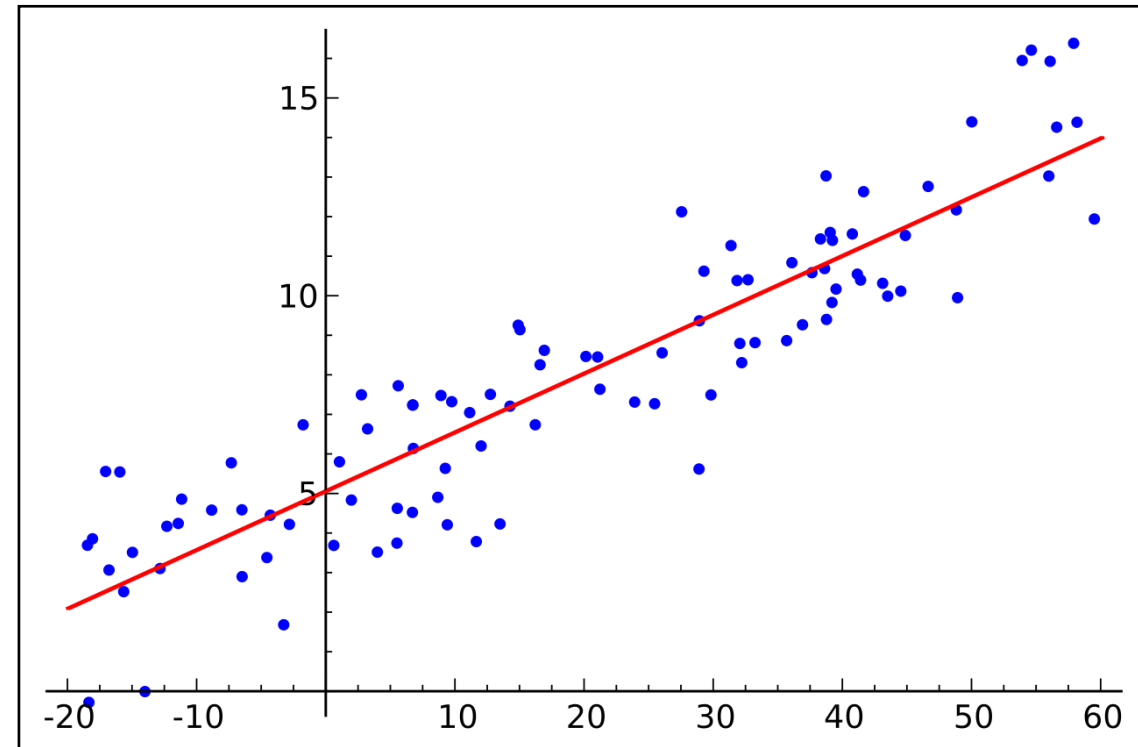
β puede ser positivo o negativo



Si $\beta=0$: No existe asociación entre la VI y la VD
(0 es el "valor de no efecto")

Regresión lineal

- Modelo matemático usado para evaluar la **asociación** entre una **variable dependiente** y una o más **variables independientes**
- La **variable dependiente** debe ser **numérica**
- Las **variables independientes** pueden ser **numéricas o categóricas**





- Para poder realizar una **regresión lineal**, la **VD debe ser numérica**.
- La VI puede ser numérica o categórica.
- Primero veremos qué sucede cuando la VI es **numérica**. Luego, veremos qué sucede cuando es **categórica**.

Regresión lineal cuando la variable independiente es numérica

$$y = \text{Cons} + \beta_1 * x$$

Variable
dependiente
(IMC)

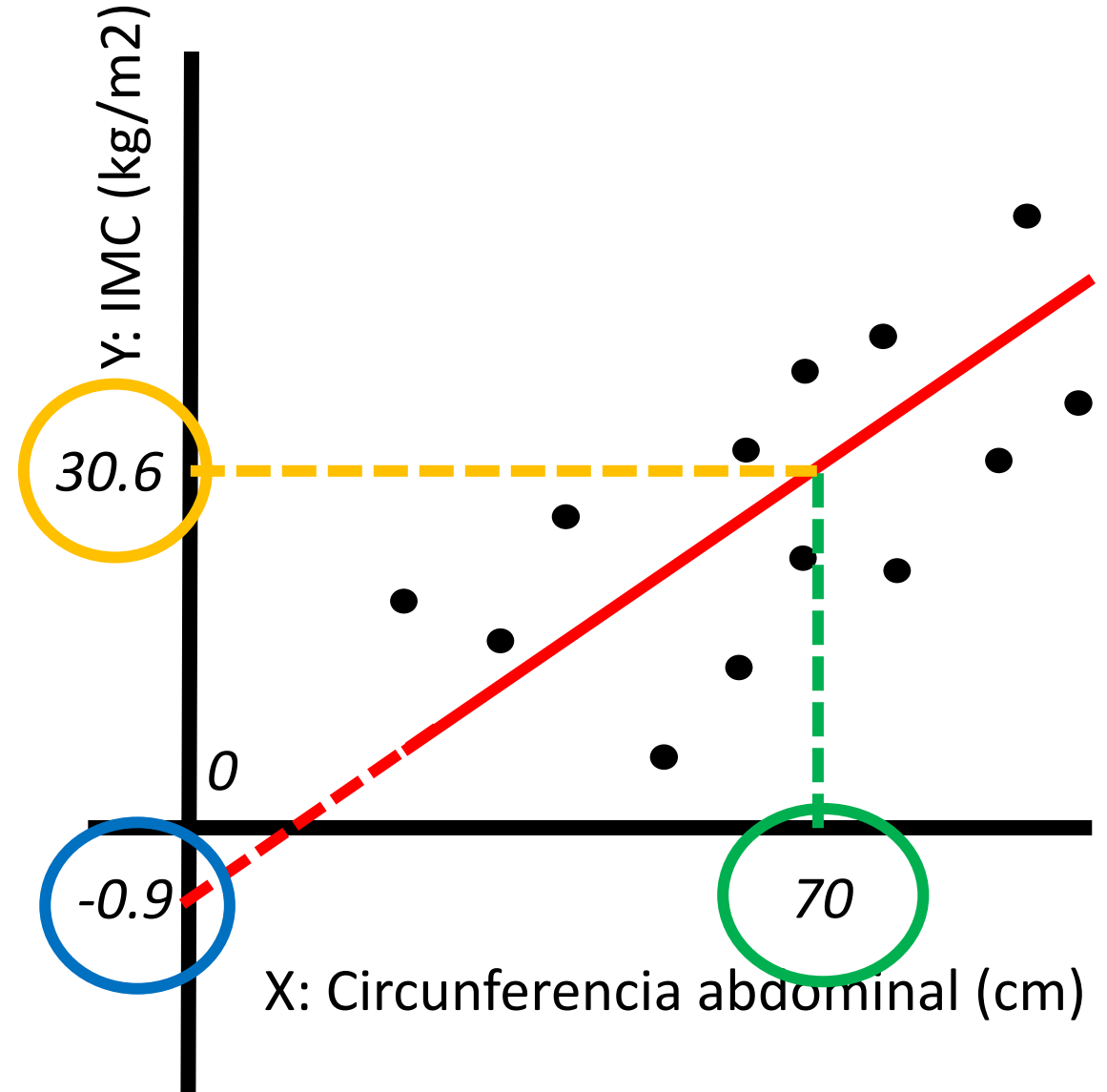
Intercepto (valor
de y, cuando x es
cero)

Coefficiente
o pendiente
(β)

Variable
independiente
(Circ. Abdomi.)

$$y = -0.9 + 0.45 * 70$$

$$30.6 = -0.9 + 0.45 * 70$$



Regresión lineal cuando la variable independiente es numérica

$$y = \text{Cons} + \beta_1 * x$$

Variable
dependiente
(FC)

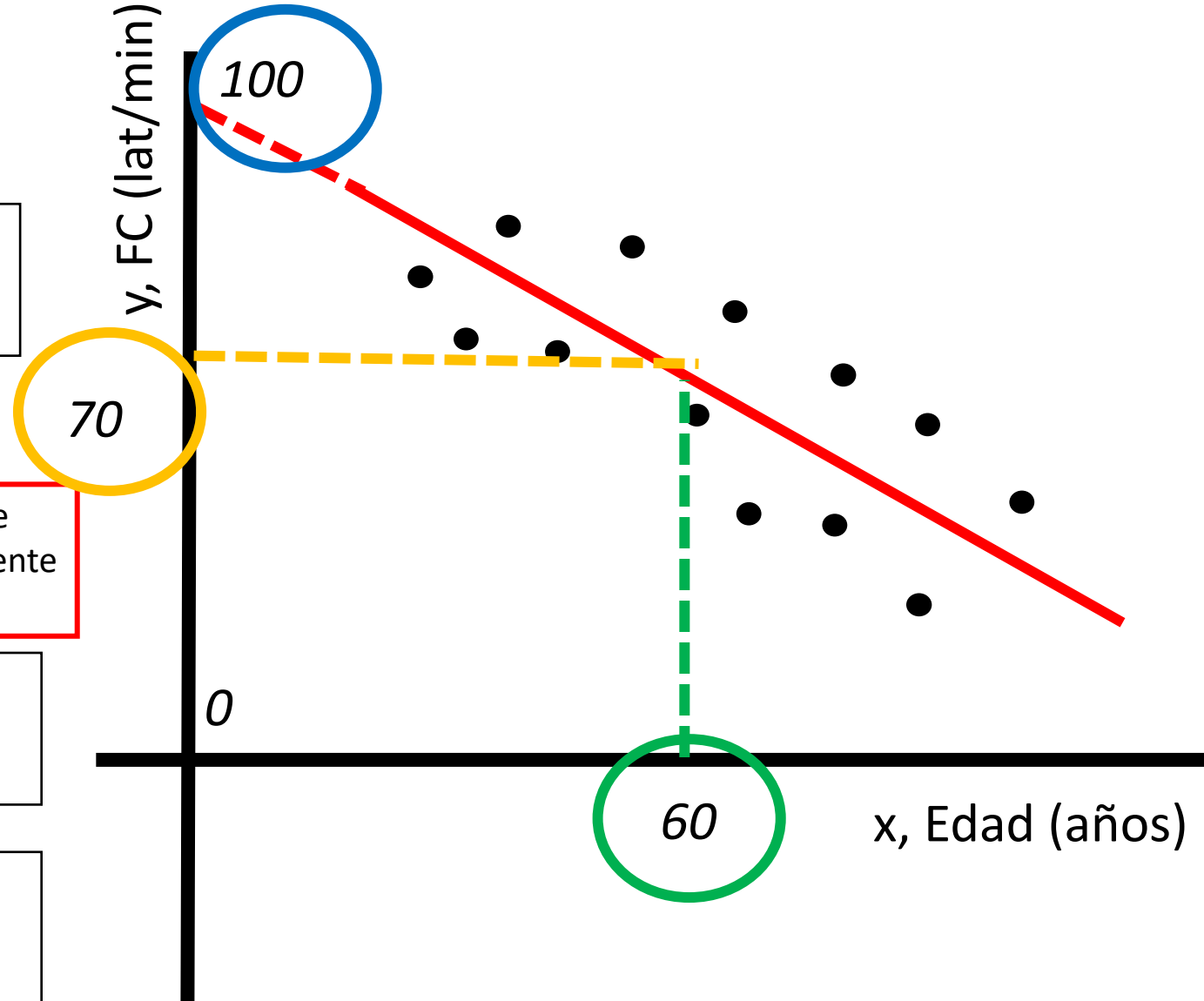
Intercepto (valor
de y, cuando x es
cero)

Coefficiente
o pendiente
(β)

Variable
independiente
(Edad)

$$y = 100 + -0.5 * 60$$

$$70 = 100 + -0.5 * 60$$



Regresión lineal cuando la variable independiente es numérica

Fórmula:

$$y = \text{Cons} + \beta_1 * x$$

- y: Valor numérico de v. dependiente
- x: Valor numérico de v. independiente
- β_1 : Pendiente o coeficiente
- Cons: β_0 o intercepto: Valor de “y” cuando “x” es cero

Regresión lineal cuando la variable independiente es numérica

Interpretación:

En

las gestantes
evaluadas

Población

, por cada

punto adicional en la
nota de secundaria

Unidad adicional de la variable independiente

,



la

nota obtenida en la
universidad

Desenlace

fue en promedio

1,41 puntos

Coefficiente de regresión lineal, y
unidad del desenlace

mayor

Mayor o Menor

. Este resultado

fue estadísticamente significativo

Fue estadísticamente significativo o
No fue estadísticamente significativo

Regresión lineal cuando la variable independiente es numérica

Comandos y resultados en STATA



Usaremos el siguiente ejemplo:

- **Población:** Universitarios
- **Variable independiente:** Nota promedio de la secundaria
- **Variable dependiente:** Nota en primer año de la universidad

Regresión lineal cuando la variable independiente es numérica

Comando en STATA

```
regress nota promsecu
```

Comando de regresión
lineal

Variable dependiente
(desenlace)

Variable independiente

$$\text{Nota} = \text{Cons} + \beta 1 * \text{Promsecu}$$

Regresión lineal cuando la variable independiente es numérica

Comandos y resultados en STATA

$$\text{Nota} = \text{Cons} + \beta_1 * \text{Promsecu}$$

`regress nota promsecu`

	nota	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
β_1	promsecu	1.414903	.0332339	42.57	0.000	1.349531	1.480275
	_cons	-6.511647	.4737255	-13.75	0.000	-7.443478	-5.579815

En los estudiantes evaluados, por cada punto adicional en el promedio de secundaria, la nota en el 1er año de la universidad fue en promedio 1.41 puntos mayor

Si $\beta_1=0$: No existe asociación

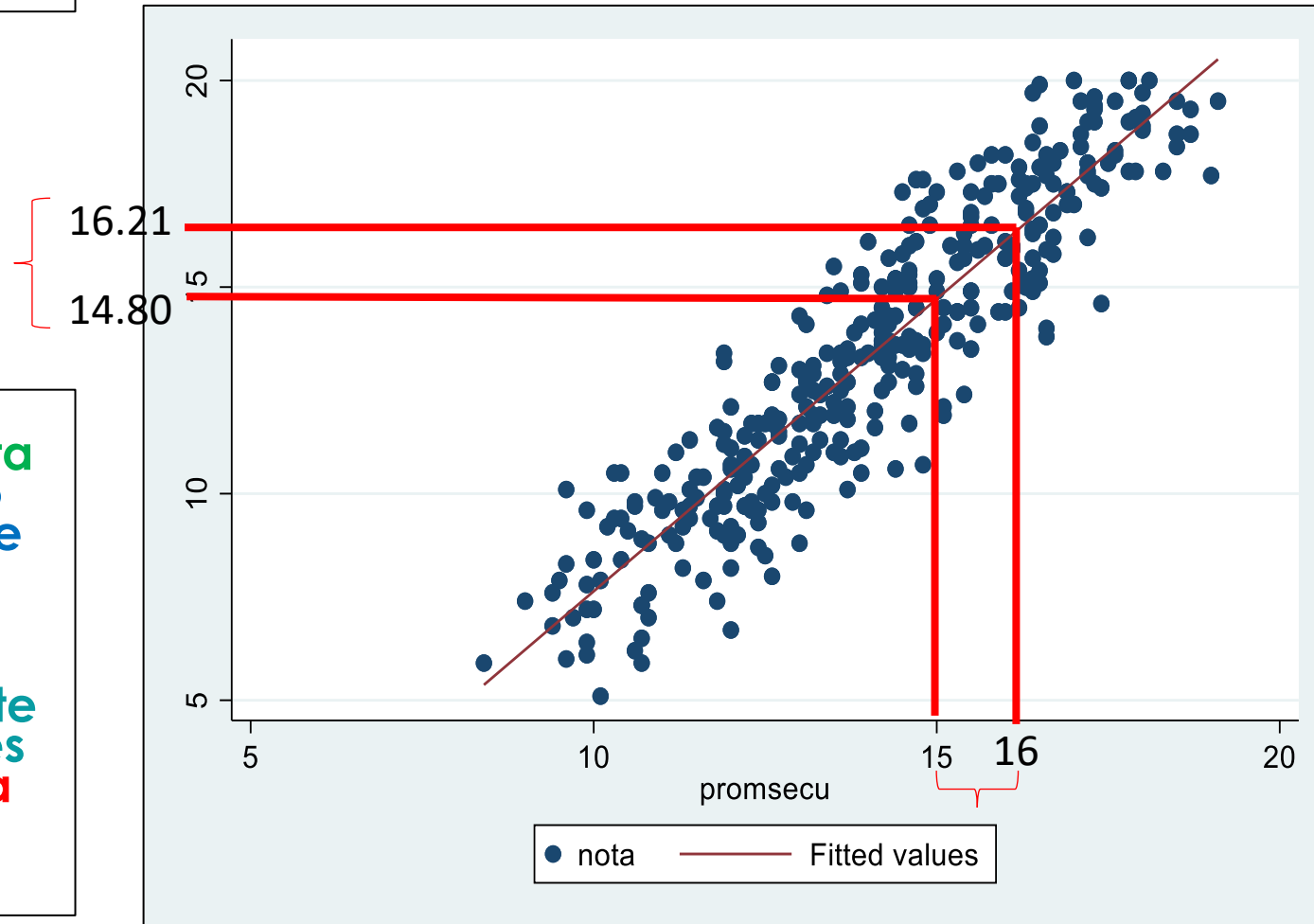
Regresión lineal cuando la variable independiente es numérica

lowess nota promsecu

Cuando la variable independiente aumenta una unidad (un punto más en el promedio de secundaria)



La variable dependiente aumenta 1.41 unidades (1.41 puntos más en la nota de 1er año de universidad)



Regresión lineal cuando la variable independiente es numérica

$$\text{Nota} = \text{Cons} + \beta_1 * \text{Promsecu}$$

`regress nota promsecu`

	nota	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
	promsecu	1.414903	.0332339	42.57	0.000	1.349531	1.480275
β_0	_cons	-6.511647	.4737255	-13.75	0.000	-7.443478	-5.579815

- **B0 = Coeficiente**
- **Interpretación:** “Cuando el promedio de secundaria es cero, la nota universitaria promedio es -6.51 puntos”
- No se suele mostrar ni interpretar en las tablas finales.

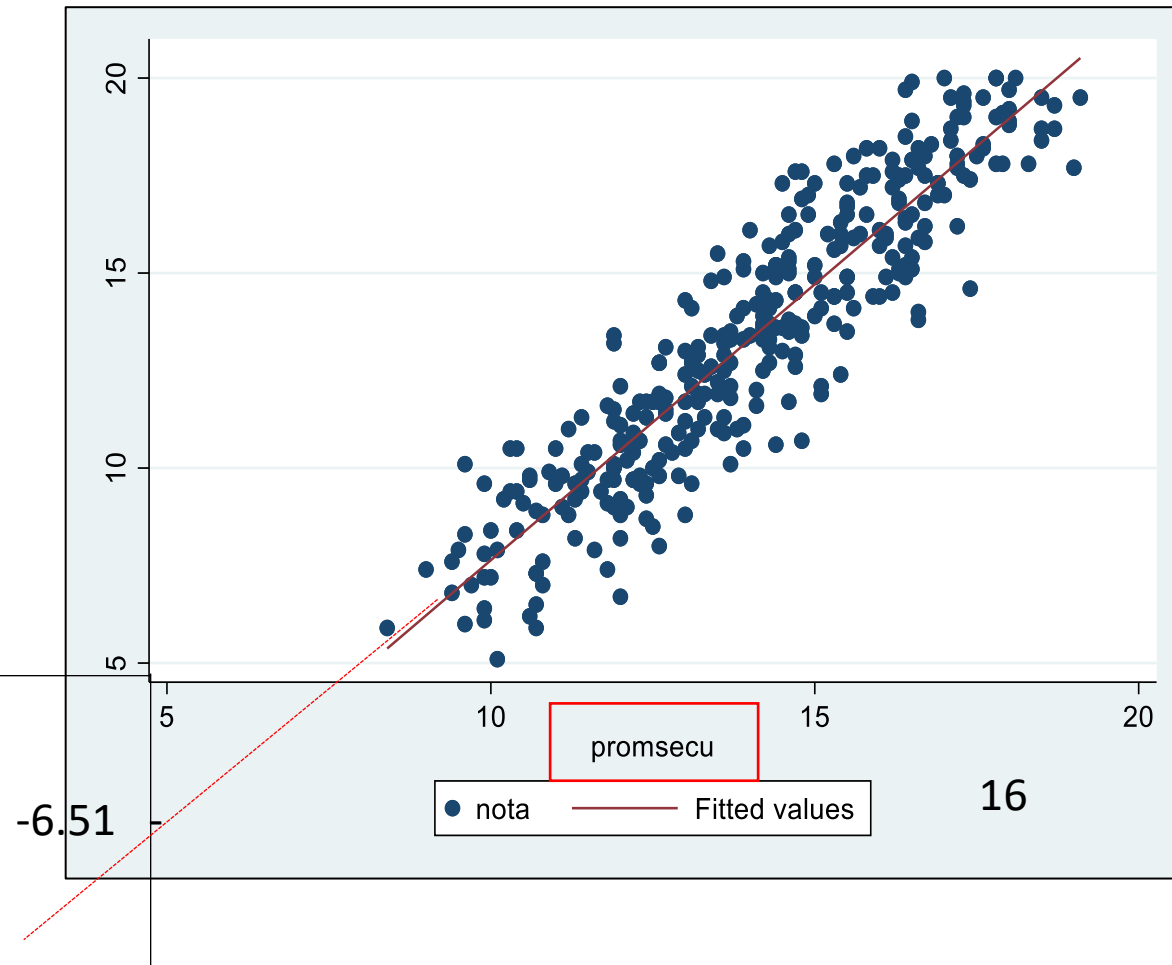
Regresión lineal cuando la variable independiente es numérica

lowess nota promsecu

Quando la variable independiente es cero (Cero en el promedio de secundaria)



La variable dependiente será -6.51 puntos (-6.51 en nota de primer año de universidad)



Regresión lineal cuando la variable independiente es numérica

regress nota promsecu

Error
estándar
de los
Coef.

Valor t:
Coef/SE

Valor p

Intervalo de confianza
de los β

nota	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
promsecu	1.414903	.0332339	42.57	0.000	1.349531	1.480275
_cons	-6.511647	.4737255	-13.75	0.000	-7.443478	-5.579815

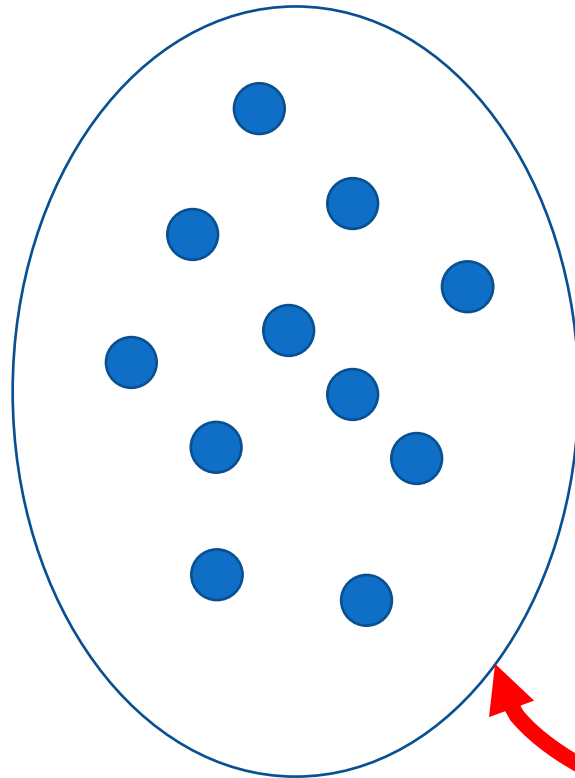
Si $\beta_1=0$: No existe asociación

Resultado:
 $\beta = 1.41$ (IC 95%: 1.35 a 1.48)
 Valor $p < 0,001$

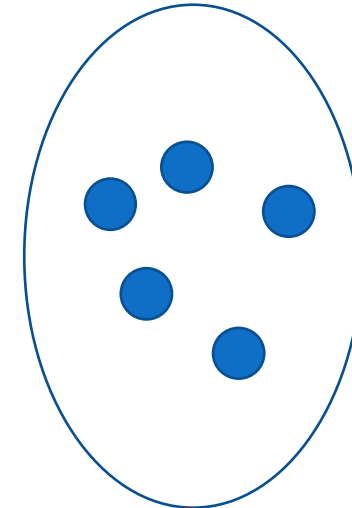
Inferencia (IC95%)

Resultado:
 $\beta = 1.41$ (IC 95%: 1.35 a 1.48)
Valor $p < 0,001$

Población



Muestra



Inferencia



Por cada punto adicional en el promedio de secundaria, la nota en el 1er año de la universidad fue en promedio entre 1.35 a 1.48 puntos mayor

Por cada punto adicional en el promedio de secundaria, la nota en el 1er año de la universidad fue en promedio 1.41 puntos mayor

Regresión lineal cuando la variable independiente es numérica

`regress nota promsecu`

	nota	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
β_1	promsecu	1.414903	.0332339	42.57	0.000	1.349531	1.480275
	_cons	-6.511647	.4737255	-13.75	0.000	-7.443478	-5.579815

Por cada punto adicional en el promedio de secundaria, la nota en el 1er año de la universidad fue en promedio 1.41 puntos mayor

Este resultado fue estadísticamente significativo

Si $\beta=0$: No existe asociación
(0 es el “valor de no efecto”)

- ¿Cuándo será estadísticamente significativo el β ?
- Cuando el IC 95% no incluya el 0
 - Cuando el valor $p < 0.05$
- (Ambos se relacionan. Basta con que uno se cumpla)

Regresión lineal cuando la variable independiente es numérica

`regress nota edad`

β_1

nota	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
edad	-.0593271	.1228929	-0.48	0.630	-.3010608	.1824066
_cons	14.70155	2.730558	5.38	0.000	9.330464	20.07263

Por cada año adicional de edad, la nota universitaria fue en promedio 0.06 puntos menor

Este resultado no fue estadísticamente significativo

Si $\beta=0$: No existe asociación

Extrapolar a las tablas de resultados

nota	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
promsecu _cons	1.414903	.0332339	42.57	0.000	1.349531	1.480275
	-6.511647	.4737255	-13.75	0.000	-7.443478	-5.579815

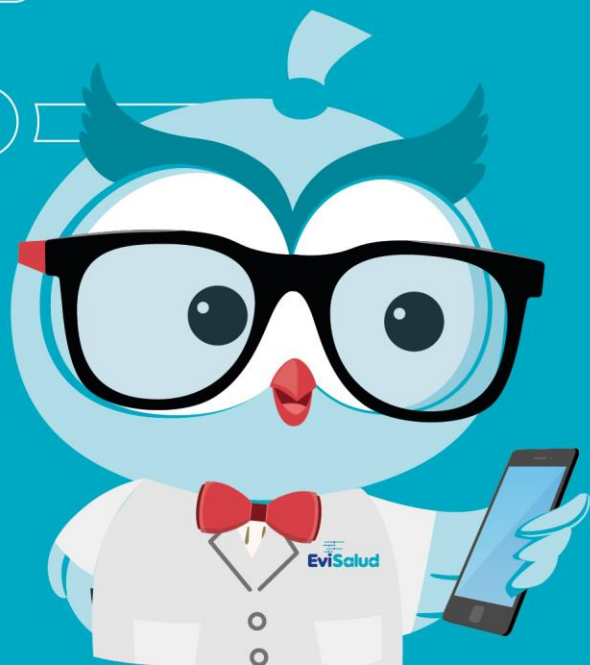
nota	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
edad _cons	-.0593271	.1228929	-0.48	0.630	-.3010608	.1824066
	14.70155	2.730558	5.38	0.000	9.330464	20.07263

Variables	Nota promedio en primer año de universidad (0 al 20)	
	B (IC95%)	
Nota promedio en secundaria	1.41	(1.35 a 1.48)
Edad	-0.06	(-0.30 a 0.18)

Muchas gracias (:



MEJORES
DECISIONES
EN SALUD



📞 950 876 703
🌐 www.evisalud.net
📘 facebook.com/evisalud/
📷 @evisaludperu



Regresión lineal con variable independiente categórica

Christopher A. Alarcon-Ruiz MD MSc(c)

Recuerda que...

$$y = \text{Cons} + \beta_1 * X$$



- Para poder realizar una **regresión lineal**, la **variable dependiente (Y)** debe ser **numérica**.
- La variable independiente (X) puede ser **numérica** o **categorica**

Regresión lineal con VI dicotómica

$$y = \text{Cons} + \beta_1 * X$$

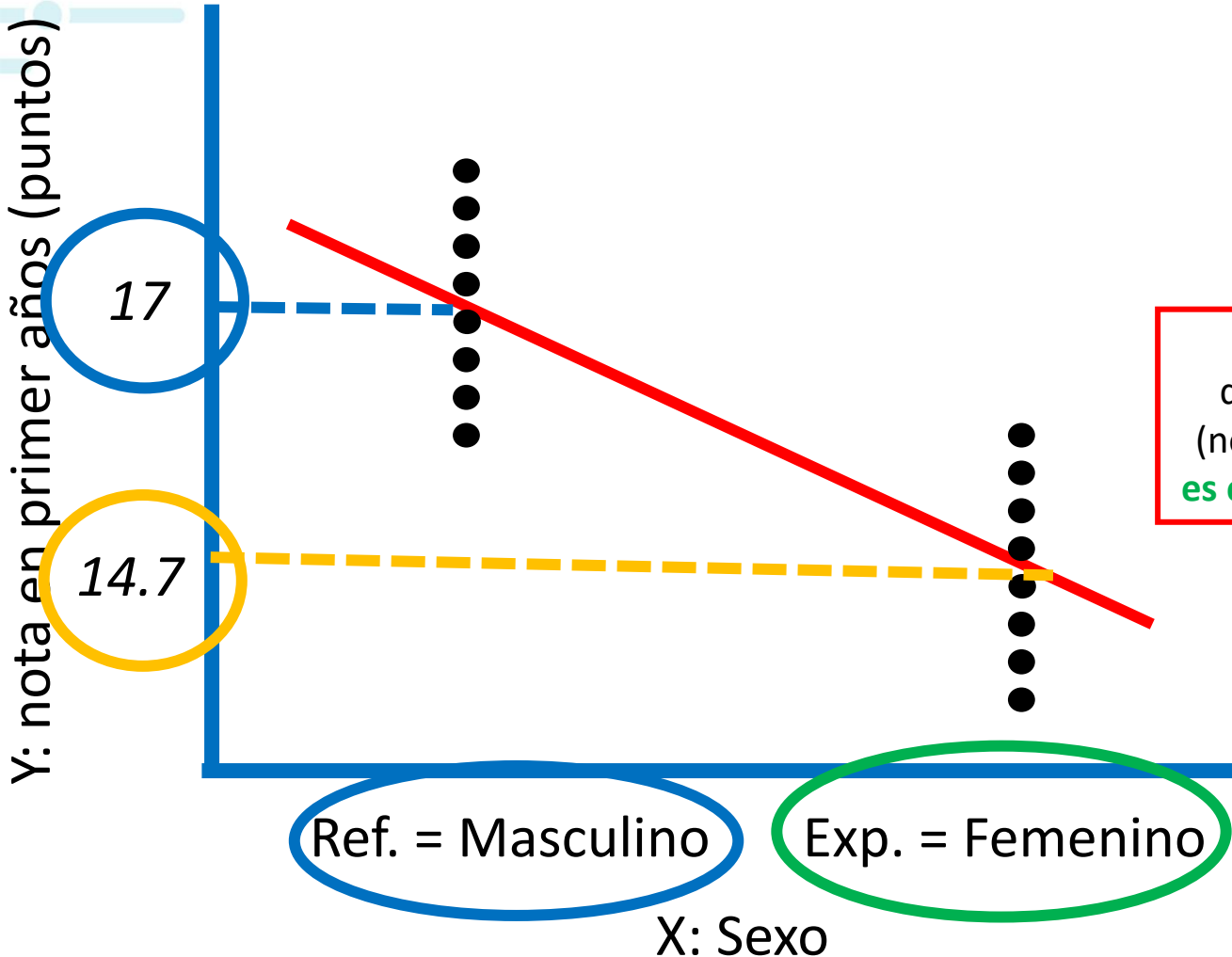
Asociación entre sexo y nota en primer año de universidad

Sexo: Variable independiente. Es dicotómica (Masculino o Femenino)

Categorías:

- **Categoría de referencia** = 0 (Masculino)
- **Categoría de exposición** = 1 (Femenino)

Regresión lineal con VI dicotómica



$$y = \text{Cons} + \beta_1 * X$$

Variable dependiente (nota) cuando **x** es cat. De interés

Intercepto (valor de y, cuando x es cat. De ref.)

Coefficiente o pendiente (β)

Categoría de interés de v. expo (Sexo = Femenino)

$$y = 17 + -2.3$$

$$14.7 = 17 + -2.3$$

Regresión lineal con VI dicotómica

Interpretación:



Comandos y resultados en STATA



Usaremos el siguiente ejemplo:

- **Población:** Universitarios
- **Variable independiente:** Sexo (0=Masculino) (1=Femenino)
- **Variable dependiente:** Nota en primer año de la universidad

Regresión lineal con VI dicotómica

Comando en STATA

```
regress nota i.sexofem
```

Comando de regresión
lineal

Variable dependiente
(desenlace)

Variable independiente
categórica
(notar el “i.” antes del
nombre de la variable)

$$\text{Nota} = \text{Cons} + \beta 1 * \text{Sexofem}$$

Regresión lineal con VI dicotómica

`regress nota i.sexofem`

Si $\beta=0$: No existe asociación

	nota	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
β_1	sexofem						
	Femenino	.313582	.4447611	0.71	0.481	-.5612756	1.18844
β_0	_cons	13.15455	.3827047	34.37	0.000	12.40175	13.90734

La media de la nota de 1er año en **mujeres** fue **0.3 puntos más** en comparación a **los varones**

La media de la nota de 1er año en los varones (Categoría de Referencia) fue **13.2 puntos**

Regresión lineal con VI dicotómica

`regress nota i.sexofem`

Valor p

Intervalo de confianza

nota

Coef.

Std. Err.

t

P>|t|

[95% Conf. Interval]

sexofem

β_1

Femenino

.313582

.4447611

0.71

0.481

- .5612756

1.18844

_cons

13.15455

.3827047

34.37

0.000

12.40175

13.90734

La media de la nota de 1er año en **mujeres** fue **0.3 puntos más** en comparación a **los varones**

Si $\beta=0$: No existe asociación
(0 es el “valor de no efecto”)

Resultado:

$\beta = 0.31$ (IC 95%: -0.56 a 1.19)

Valor p $\geq 0,05$

¿Cuándo será estadísticamente significativo el β ?

- Cuando el IC 95% no incluye el 0
- Cuando el valor p < 0.05

(Ambos se relacionan. basta con que uno se cumpla)

Regresión lineal con VI dicotómica

`regress nota i.sexofem`

Valor p

Intervalo de confianza

nota

Coef.

Std. Err.

t

P>|t|

[95% Conf. Interval]

sexofem

β_1

Femenino

.313582

.4447611

0.71

0.481

-.5612756

1.18844

_cons

13.15455

.3827047

34.37

0.000

12.40175

13.90734

La media de la nota de 1er año en **mujeres** fue **0.3 puntos más** en comparación a **los varones**

Este resultado no fue estadísticamente significativo

Si $\beta=0$: No existe asociación
(0 es el “valor de no efecto”)

¿Cuándo será estadísticamente significativo el β ?

- Cuando el IC 95% no incluye el 0
 - Cuando el valor $p < 0.05$
- (Ambos se relacionan. basta con que uno se cumpla)**

Extrapolar a las tablas de resultados

nota	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
sexofem						
Femenino	.313582	.4447611	0.71	0.481	-.5612756	1.18844
_cons	13.15455	.3827047	34.37	0.000	12.40175	13.90734

Variables	Nota promedio en primer año de universidad (0 al 20)	
	B (IC95%)	
Sexo		
Masculino	Ref.	
Femenino	0.31	-0.56 a 1.19)

Regresión lineal con VI dicotómica

$$y = \text{Cons} + \beta_1 * X$$

- y: Promedio de v. dependiente “y” en cat. exposición
- x: V. independiente dicotómica (Cat: Exposición)
- Cons: β_0 o intercepto: Valor promedio de “y” cuando “x” es la categoría de ref.
- β_1 : Promedio de “y” en cat. Exposición – Promedio de “y” en cat. Ref.

Comandos y resultados en STATA



Usaremos el siguiente ejemplo:

- **Población:** Universitarios
- **Variable independiente:** Estado nutricional (0=Normal) (1=Sobrepeso) (2=Obesidad)
- **Variable dependiente:** Nota en primer año de la universidad

Regresión lineal con VI politómica

$$y = \text{Cons} + \beta_1 * X_1 + \beta_2 * X_2$$

Asociación entre estado nutricional y nota en primer año

Variable independiente: Estado nutricional (categórica politómica)

Categorías:

- Categoría de referencia : Normal = 0
 - Categoría de exposición 1: Sobrepeso = 1
 - Categoría de exposición 2: Obesidad = 2
- β_1 (compara la cat ref con la exp 1)**

Regresión lineal con VI politómica

$$y = \text{Cons} + \beta_1 * X_1 + \beta_2 * X_2$$

Asociación entre estado nutricional y nota en primer año

Variable independiente: Estado nutricional (categórica politómica)

Categorías:

- Categoría de referencia : Normal = 0
 - Categoría de exposición 1: Sobrepeso = 1
 - Categoría de exposición 2: Obesidad = 2
- } β_2 (compara la cat ref con la exp 2)

Regresión lineal con VI politómica

`regress nota i.estadonut`

Si $\beta=0$: No existe asociación

	nota	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
	estadonut						
β_1	Sobrepeso	-1.556141	.3987109	-3.90	0.000	-2.340425	-.7718568
	Obesidad	-1.108242	.8106855	-1.37	0.173	-2.7029	.4864167
	_cons	14.07967	.2607438	54.00	0.000	13.56677	14.59257

La media de la nota universitaria en los que tuvieron sobrepeso fue **1.56 puntos menos** en comparación con **los de estado nutricional normal.**

Regresión lineal con VI politómica

`regress nota i.estadonut`

Si $\beta=0$: No existe asociación

	nota	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
	estadonut						
β_1	Sobrepeso	-1.556141	.3987109	-3.90	0.000	-2.340425	-.7718568
	Obesidad	-1.108242	.8106855	-1.37	0.173	-2.7029	.4864167
	_cons	14.07967	.2607438	54.00	0.000	13.56677	14.59257

La media de la nota universitaria en los que tuvieron sobrepeso fue **1.56 puntos menos** en comparación con **los de estado nutricional normal.**
Este resultado fue estadísticamente significativo

Resultado:
 $\beta = -1.56$ (IC 95%: -2.34 a -0.77)
Valor $p < 0,05$

Regresión lineal con VI politómica

`regress nota i.estadonut`

Si $\beta=0$: No existe asociación

nota	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
estadonut						
Sobrepeso	-1.556141	.3987109	-3.90	0.000	-2.340425	-.7718568
Obesidad	-1.108242	.8106855	-1.37	0.173	-2.7029	.4864167
_cons	14.07967	.2607438	54.00	0.000	13.56677	14.59257

La media de la nota universitaria en obesos fue **1.11 puntos menos** en comparación con **los de estado nutricional normal.**

Regresión lineal con VI politómica

`regress nota i.estadonut`

Si $\beta=0$: No existe asociación

nota	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
estadonut						
Sobrepeso	-1.556141	.3987109	-3.90	0.000	-2.340425	-.7718568
Obesidad	-1.108242	.8106855	-1.37	0.173	-2.7029	.4864167
_cons	14.07967	.2607438	54.00	0.000	13.56677	14.59257

β2

La media de la nota universitaria en obesos fue **1.11 puntos menos** en comparación con **los de estado nutricional normal**. **Este resultado no fue estadísticamente significativo**

Resultado:
 $\beta = -1.11$ (IC 95%: -2.70 a 0.49)
Valor p $\geq 0,05$

Regresión lineal con VI politómica

`regress nota i.estadonut`

Si $\beta=0$: No existe asociación

nota	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
estadonut						
Sobrepeso	-1.556141	.3987109	-3.90	0.000	-2.340425	-.7718568
Obesidad	-1.108242	.8106855	-1.37	0.173	-2.7029	.4864167
β_0 <code>_cons</code>	14.07967	.2607438	54.00	0.000	13.56677	14.59257

La media de la nota universitaria en los de **estado nutricional normal** fue **14.1 puntos** (categoría de referencia)

Extrapolar a las tablas de resultados

nota	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
estadonut						
Sobrepeso	-1.556141	.3987109	-3.90	0.000	-2.340425	-.7718568
Obesidad	-1.108242	.8106855	-1.37	0.173	-2.7029	.4864167
_cons	14.07967	.2607438	54.00	0.000	13.56677	14.59257

Variables	Nota promedio en primer año de universidad (0 al 20)	
	B (IC95%)	
Estado nutricional		
Normal	Ref.	
Sobrepeso	-1.56	(-2.34 a -0.77)
Obesidad	-1.11	(-2.70 a 0.49)

Muchas gracias (:



MEJORES
DECISIONES
EN SALUD



📞 950 876 703
🌐 www.evisalud.net
📘 facebook.com/evisalud/
📷 @evisaludperu

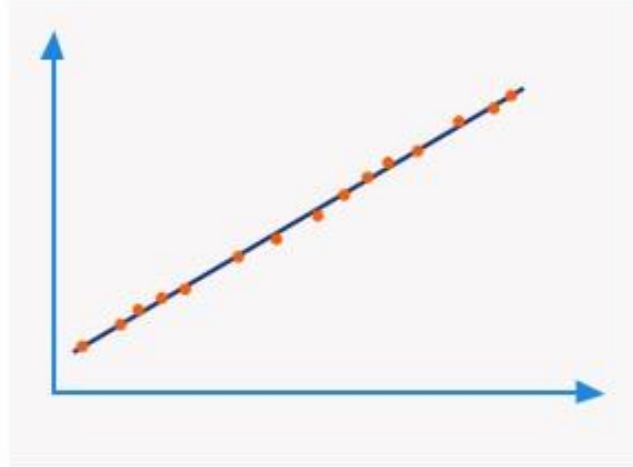


Interpretación de las salidas en Stata

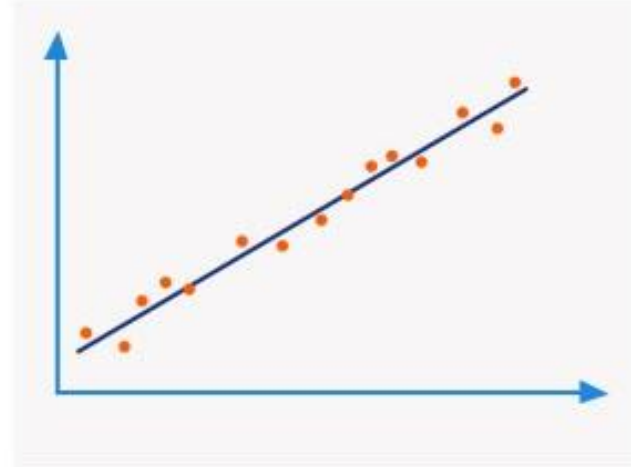
Christopher A. Alarcon-Ruiz MD MSc(c)

Correlación entre dos variables

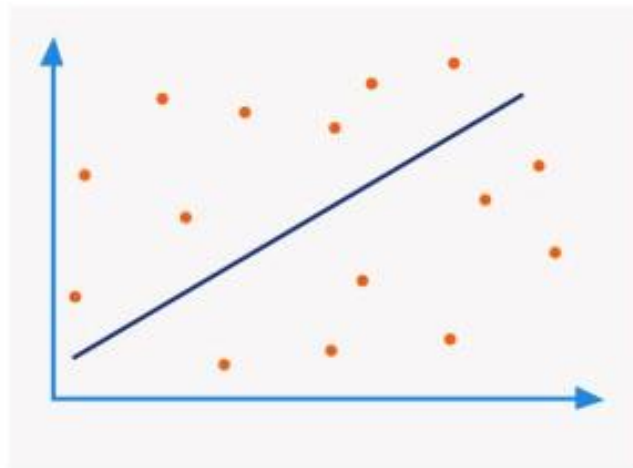
1.
Large positive
correlation



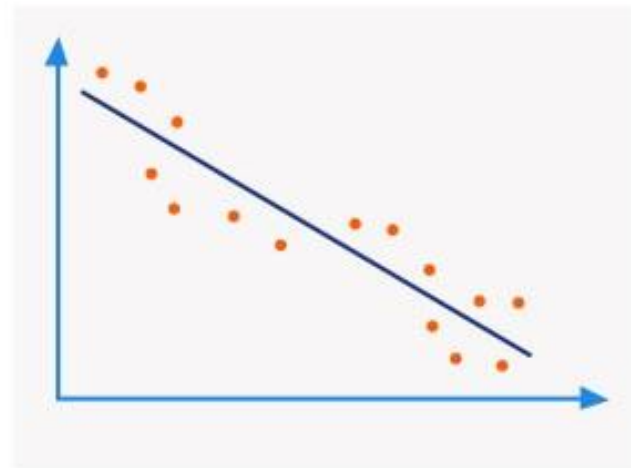
2.
Medium positive
correlation



4.
Weak / no
correlation



3.
Small negative
correlation



Regresión lineal en Stata

regress nota promsecu

Número de participantes analizados en el modelo de regresión

Valor “p” del modelo. Si hay asociación estadísticamente significativa entre las variables:
 $p < 0.05$

Source	SS	df	MS	Number of obs	=	339
Model	3667.94462	1	3667.94462	F(1, 337)	=	1812.55
Residual	681.965647	337	2.02363694	Prob > F	=	0.0000
Total	4349.91027	338	12.869557	R-squared	=	0.8432
				Adj R-squared	=	0.8428
				Root MSE	=	1.4225
nota	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
promsecu	1.414903	.0332339	42.57	0.000	1.349531	1.480275
_cons	-6.511647	.4737255	-13.75	0.000	-7.443478	-5.579815

Regresión lineal en Stata

regress nota edad

Número de participantes analizados en el modelo de regresión

Source	SS	df	MS	Number of obs = 339		
Model	3.00609997	1	3.00609997	F(1, 337)	=	0.23
Residual	4346.90417	337	12.8988254	Prob > F	=	0.6296
Total	4349.91027	338	12.869557	R-squared	=	0.0007
				Adj R-squared	=	-0.0023
				Root MSE	=	3.5915
nota	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
edad	-.0593271	.1228929	-0.48	0.630	-.3010608	.1824066
_cons	14.70155	2.730558	5.38	0.000	9.330464	20.07263

Valor “p” del modelo. Si hay asociación estadísticamente significativa entre las variables:
p<0.05

Regresión lineal en Stata

```
regress nota i.sexofem
```

Número de
participantes
analizados en
el modelo de
regresión

Valor “p” del
modelo. Si hay
asociación
estadísticamente
significativa entre
las variables:
 $p < 0.05$

Source	SS	df	MS	Number of obs	=	339
Model	6.40706375	1	6.40706375	F(1, 337)	=	0.50
Residual	4343.5032	337	12.8887335	Prob > F	=	0.4813
Total	4349.91027	338	12.869557	R-squared	=	0.0015
				Adj R-squared	=	-0.0015
				Root MSE	=	3.5901

nota	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
sexofem						
Femenino	.313582	.4447611	0.71	0.481	-.5612756	1.18844
_cons	13.15455	.3827047	34.37	0.000	12.40175	13.90734

Regresión lineal en Stata

regress nota i.estadonut

Número de participantes analizados en el modelo de regresión

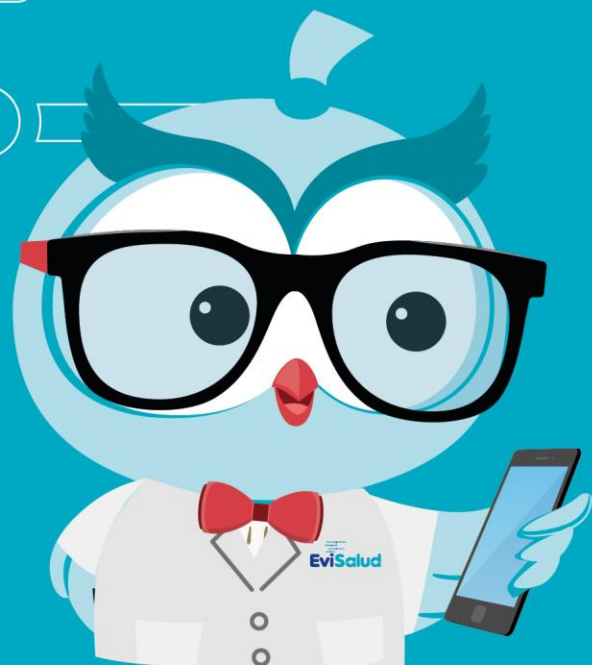
Source	SS	df	MS	Number of obs	=	339
Model	192.347922	2	96.1739611	F(2, 336)	=	7.77
Residual	4157.56234	336	12.3736975	Prob > F	=	0.0005
				R-squared	=	0.0442
				Adj R-squared	=	0.0385
Total	4349.91027	338	12.869557	Root MSE	=	3.5176
nota	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
estadonut						
Sobrepeso	-1.556141	.3987109	-3.90	0.000	-2.340425	-.7718568
Obesidad	-1.108242	.8106855	-1.37	0.173	-2.7029	.4864167
_cons	14.07967	.2607438	54.00	0.000	13.56677	14.59257

Valor “p” del modelo. Si hay asociación estadísticamente significativa entre las variables: $p < 0.05$

Muchas gracias (:



MEJORES
DECISIONES
EN SALUD



📞 950 876 703
🌐 www.evisalud.net
📘 facebook.com/evisalud/
📷 @evisaludperu



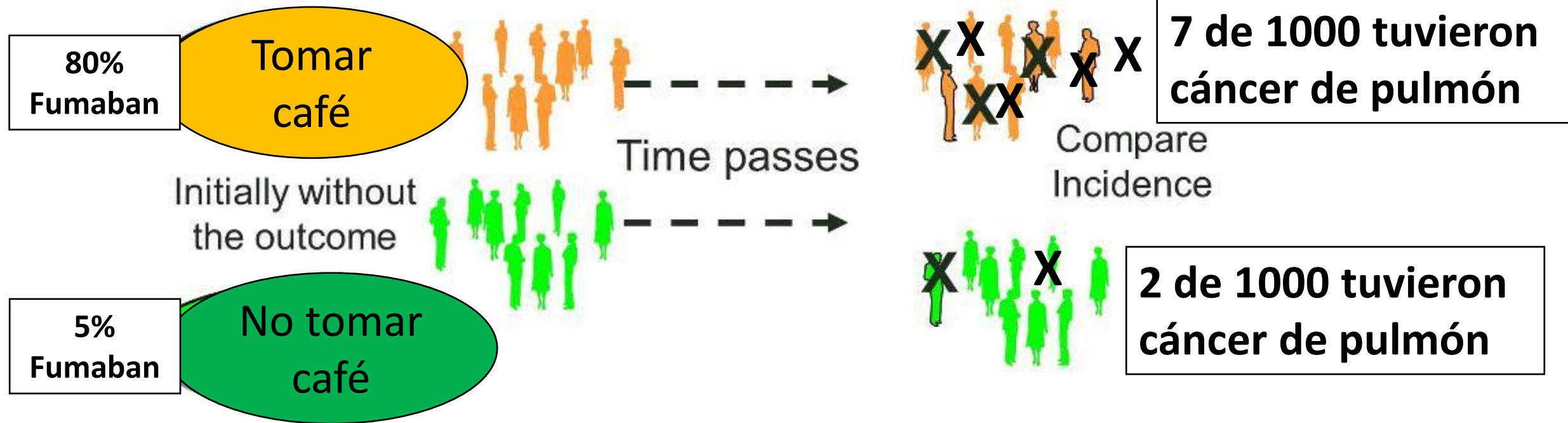
Regresión lineal múltiple

Christopher A. Alarcon-Ruiz MD MSc(c)

Confusión

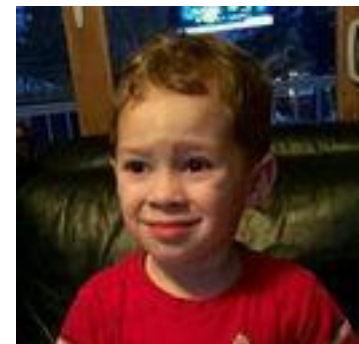
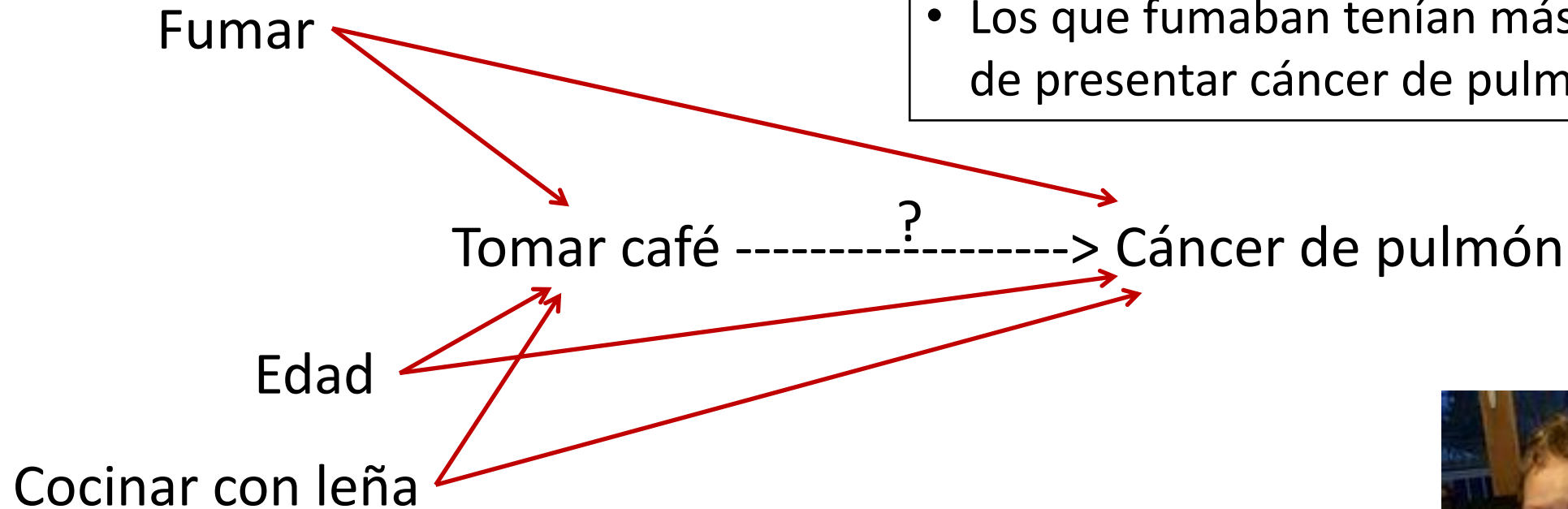
Queremos evaluar si existe una relación de causa-efecto (causalidad) entre:

- **Variable independiente:** tomar café
- **Desenlace:** cáncer de pulmón



Fumar puede ser un “confusor” si:

- Los que fumaban tenían más probabilidad de tomar café, y
- Los que fumaban tenían más probabilidad de presentar cáncer de pulmón



Variable confusora



- Es una variable que distorsiona la asociación entre otras dos variables.
- Debe estar asociada a la variable dependiente y a la variable independiente.
- Una forma de “limpiar” el efecto de los confusores es usando un **modelo de regresión lineal ajustado**.

Regresión lineal ajustada

Tabla 3

Resultados de la Asociación entre la duración del sueño y el IMC

Horas de sueño	Crudo	Modelo Ajustado 1
	Coef (IC 95%)	Coef (IC 95%)
Bajo	0	0
Normal	-1,95(-3,26;-0,64)	-1,77(-2,98;-0,57)
Alto	-2,84(-6,14;0,46)	-2,99(-6,05;0,07)

- Crudo = bivariado
- Ajustado = multivariable

Mendoza Cáceres AE, Zumaeta Gonzales NV. Asociación entre la duración del sueño y el IMC en pacientes adultos que asistieron a un Centro Universitario de Salud de Lima , Perú

Regresión lineal ajustada

- Determinar el número máximo de variables:
 - 1 var dep + 1 var indep → necesito al menos 20 participantes
 - 1 var dep + 2 var indep → necesito al menos 40 participantes
 - 1 var dep + 3 var indep → necesito al menos 60 participantes
 - ...
- ¿Qué variables incluir en tu modelo?
 - Confusores teóricos → De la literatura previa
 - Confusores estadísticos → Ejm: Incluir todas las variables indep que, al asociarse con la var dep (modelo crudo), hayan tenido $p < 0.20$

Resultados en modelos crudos

Para evaluar qué factores se asociarían con “nota”, evaluaremos las regresiones **crudas** entre cada variable independiente con “nota” y seleccionaremos a aquellas que tuvieron **$p < 0.20$ (en Prob>F o Prob>Chi2)**

. regress nota i.sexofem					
Source	SS	df	MS	Number of obs = 339	
Model	6.40706375	1	6.40706375	F(1, 337)	= 0.50
Residual	4343.5032	337	12.8887335	Prob > F	= 0.4813
				R-squared	= 0.0015
				Adj R-squared	= -0.0015
				Root MSE	= 3.5901
Total	4349.91027	338	12.869557		

. regress nota edad					
Source	SS	df	MS	Number of obs = 339	
Model	3.00609997	1	3.00609997	F(1, 337)	= 0.23
Residual	4346.90417	337	12.8988254	Prob > F	= 0.6296
				R-squared	= 0.0007
				Adj R-squared	= -0.0023
				Root MSE	= 3.5915
Total	4349.91027	338	12.869557		

. regress nota promsecu					
Source	SS	df	MS	Number of obs = 339	
Model	3667.94462	1	3667.94462	F(1, 337)	= 1812.55
Residual	681.965647	337	2.02363694	Prob > F	= 0.0000
				R-squared	= 0.8432
				Adj R-squared	= 0.8428
				Root MSE	= 1.4225
Total	4349.91027	338	12.869557		

. regress nota i.estadonut					
Source	SS	df	MS	Number of obs = 339	
Model	192.347922	2	96.1739611	F(2, 336)	= 7.77
Residual	4157.56234	336	12.3736975	Prob > F	= 0.0005
				R-squared	= 0.0442
				Adj R-squared	= 0.0385
				Root MSE	= 3.5176
Total	4349.91027	338	12.869557		

Interpretación de regresión lineal cruda

. regress nota promsecu						
Source	SS	df	MS	Number of obs = 339		
Model	3667.94462	1	3667.94462	F(1, 337)	=	1812.55
Residual	681.965647	337	2.02363694	Prob > F	=	0.0000
				R-squared	=	0.8432
				Adj R-squared	=	0.8428
Total	4349.91027	338	12.869557	Root MSE	=	1.4225
nota	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
promsecu	1.414903	.0332339	42.57	0.000	1.349531	1.480275
_cons	-6.511647	.4737255	-13.75	0.000	-7.443478	-5.579815

Coeficiente (β) crudo:
 Por **cada punto adicional** de la nota
 promedio en secundaria, la **nota**
universitaria fue en promedio **1.41**
puntos mayor

Este resultado fue estadísticamente
significativo

Interpretación de regresión lineal ajustada

```
. regress nota promsecu i.estadonut
```

Source	SS	df	MS	Number of obs	=	339
Model	3678.44773	3	1226.14924	F(3, 335)	=	611.74
Residual	671.462537	335	2.00436578	Prob > F	=	0.0000
Total	4349.91027	338	12.869557	R-squared	=	0.8456
				Adj R-squared	=	0.8443
				Root MSE	=	1.4158

nota	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
promsecu	1.401341	.0336018	41.70	0.000	1.335243	1.467438
estadonut						
Sobrepeso	-.3698477	.1629728	-2.27	0.024	-.6904266	-.0492688
Obesidad	-.2571712	.3269181	-0.79	0.432	-.9002422	.3858998
_cons	-6.156611	.4964503	-12.40	0.000	-7.133164	-5.180058

Coeficiente (β) ajustado:
 Por **cada punto adicional** de la nota promedio en secundaria, la **nota universitaria** fue en promedio **1.40 puntos mayor**; **al ajustar por estado nutricional**

Este resultado fue estadísticamente significativo

Extrapolar a las tablas de resultados

. regress nota promsecu i.estadonut

Source	SS	df	MS	Number of obs	=	339
Model	3678.44773	3	1226.14924	F(3, 335)	=	611.74
Residual	671.462537	335	2.00436573	Prob > F	=	0.0000
Total	4349.91027	338	12.869557	R-squared	=	0.8456
				Adj R-squared	=	0.8443
				Root MSE	=	1.4158

nota	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
promsecu	1.401341	.0336018	41.70	0.000	1.335243	1.467438
estadonut						
Sobrepeso	-.3698477	.1629728	-2.27	0.024	-.6904266	-.0492688
Obesidad	-.2571712	.3269181	-0.79	0.432	-.9002422	.3858998
_cons	-6.156611	.4964503	-12.40	0.000	-7.133164	-5.180058

Variables	Nota promedio en primer año de universidad (0 al 20)	
	Bc (IC 95%)	Ba (IC 95%)*
Promedio de secundaria (nota)	1.41 (1.35 a 1.48)	1.40 (1.33 a 1.47)
Estado nutricional		
Normal	Ref.	Ref.
Sobrepeso	-1.56 (-2.34 a -0.77)	-0.37 (-0.69 a -0.05)
Obesidad	-1.11 (-2.70 a 0.49)	-0.26 (-0.90 a 0.39)

* Ajustado por promedio de secundaria y estado nutricional

-0.37 (-0.69 a -0.05)
-0.26 (-0.90 a 0.39)

Muchas gracias (:



MEJORES
DECISIONES
EN SALUD



📞 950 876 703
🌐 www.evisalud.net
📘 facebook.com/evisalud/
📷 @evisaludperu



Evaluación de supuestos en regresión lineal

Christopher A. Alarcon-Ruiz MD MSc(c)

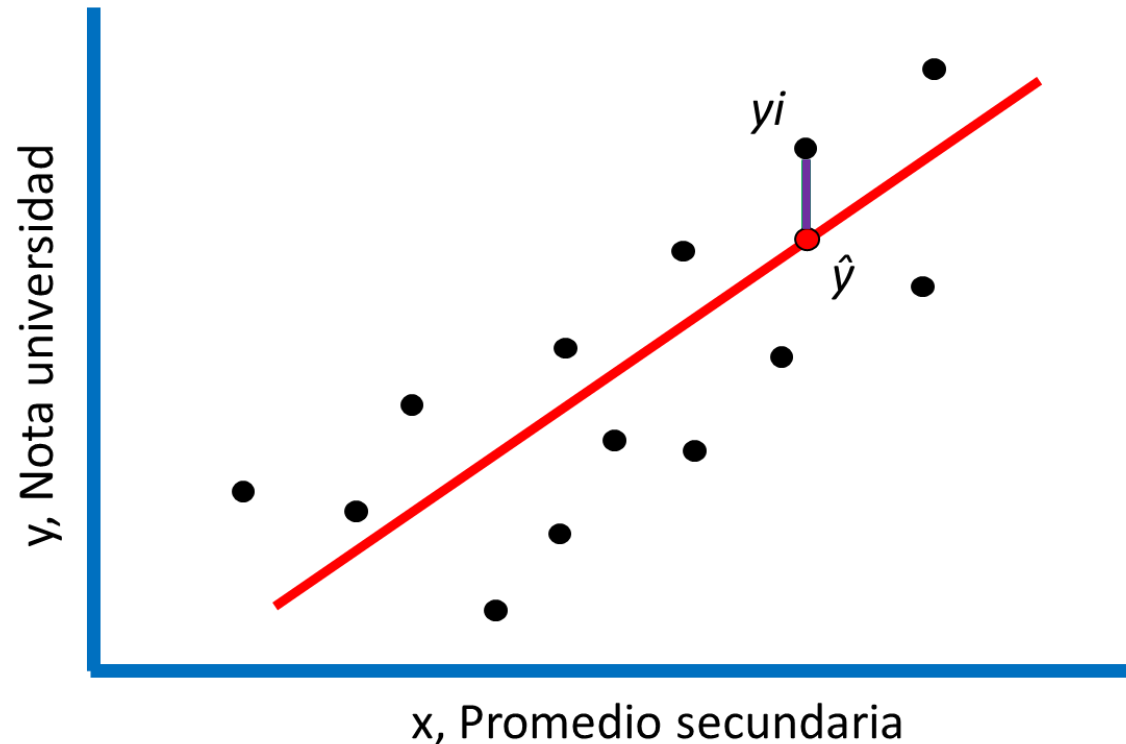
Antes de evaluar los supuestos,
conozcamos dos términos

Valores esperados de y (\hat{y}):

- Valor de y , según la pendiente β
- **Stata:** Luego de correr la regresión (*regress...*), ejecutar **predict vesp**

Residuos:

- **Resta** entre **valor observado de y (y_i)**, con su respectivo **valor esperado de y (\hat{y})**
- **Stata:** Luego de correr la regresión (*regress...*), ejecutar **predict resi**, residuals



Supuestos de regresión lineal

Regresión lineal cruda:

1. **L**= Relación **lineal**
2. **I**= Observaciones **independientes**.
3. **N**= Distribución **normal** de los residuos.
4. **E**= **Igual(equal)** varianzas (Homocedasticidad)

Regresión lineal ajustada:

1. **L**= Relación **lineal**
2. **I**= Observaciones **independientes**.
3. **N**= Distribución **normal** de los residuos.
4. **E**= **Igual(equal)** varianzas (Homocedasticidad)
5. **Multicolinealidad**

¿Qué se debe saber de los supuestos?

1. Supuesto

2. ¿Cómo
evaluar
supuesto?

3. ¿Cómo
saber si se
cumple o no?

4. ¿Qué hacer
si no se
cumple?

1. Relación lineal

Al evaluar visualmente la relación entre la VI y la VD, esta dibujará una línea.

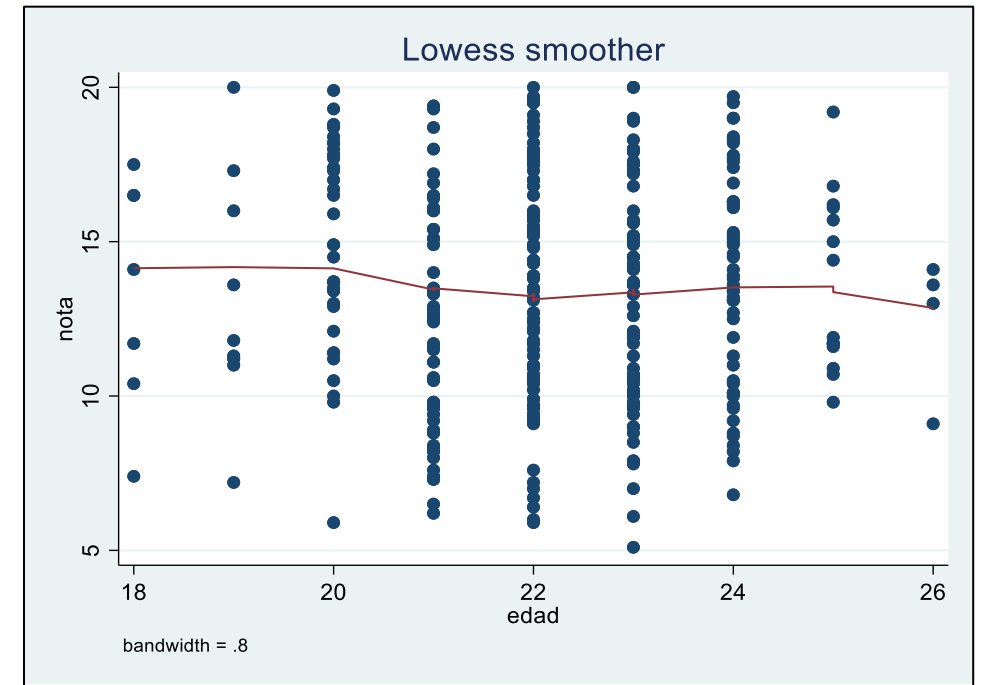
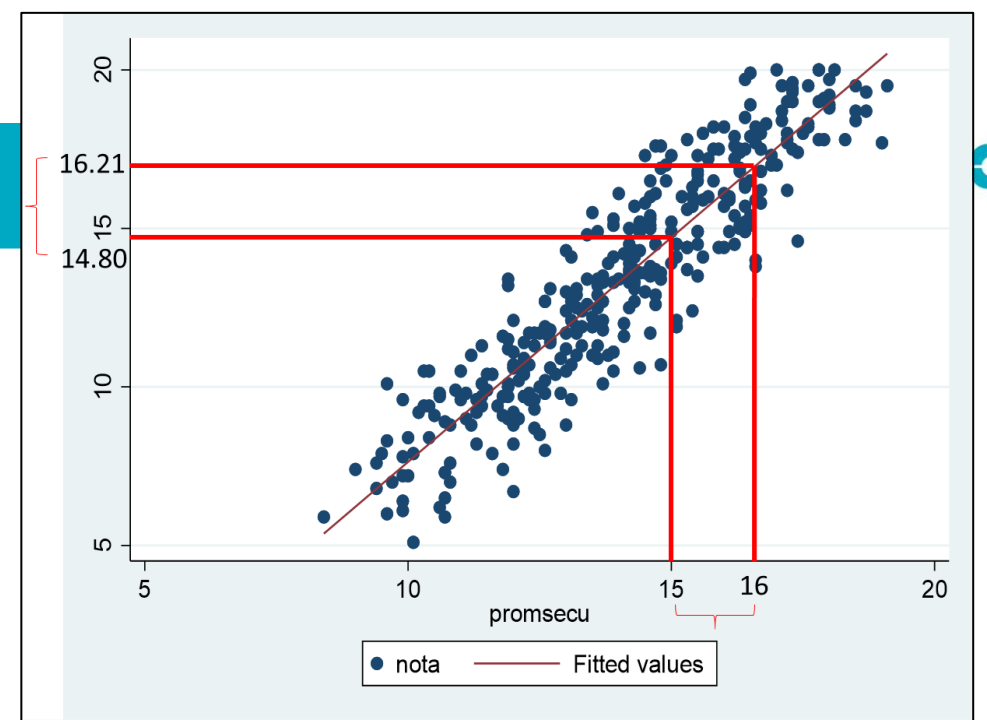
- I. Evaluación de la linealidad para VI numérica

Gráfico: VD vs VI

lowess **nota** promsecu

lowess **nota** edad

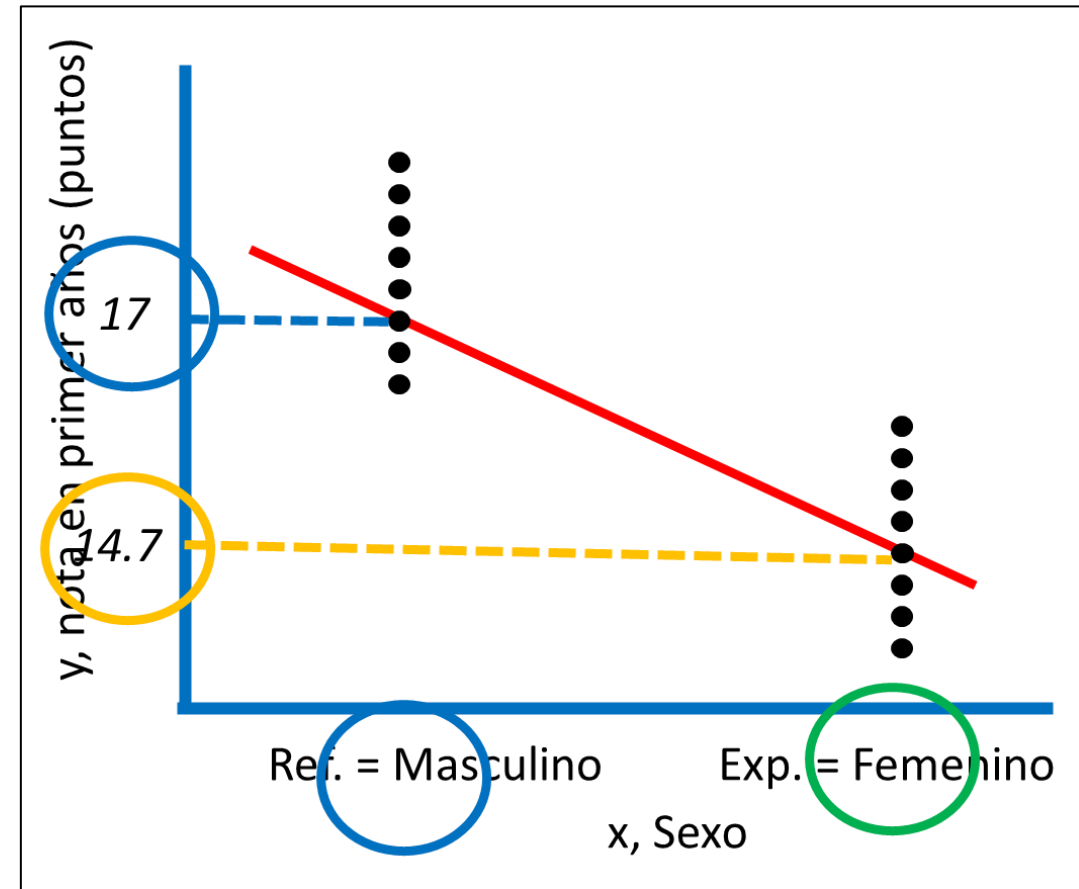
Se cumple cuando línea roja es diagonal positiva, diagonal negativa u horizontal



1. Relación lineal

II. Evaluación de la linealidad para VI categórica (dicotómica o politómica)

No es necesario evaluar linealidad. Siempre existirá linealidad



1. Relación lineal

III. Evaluación de la linealidad para regresión ajustada:

- Si ninguna VI del modelo es numérica (todas son categóricas): No es necesario evaluarla (siempre existirá linealidad)
- Si al menos una VI del modelo es numérica:

Gráfico: “Valores esperados + residuos” vs VI

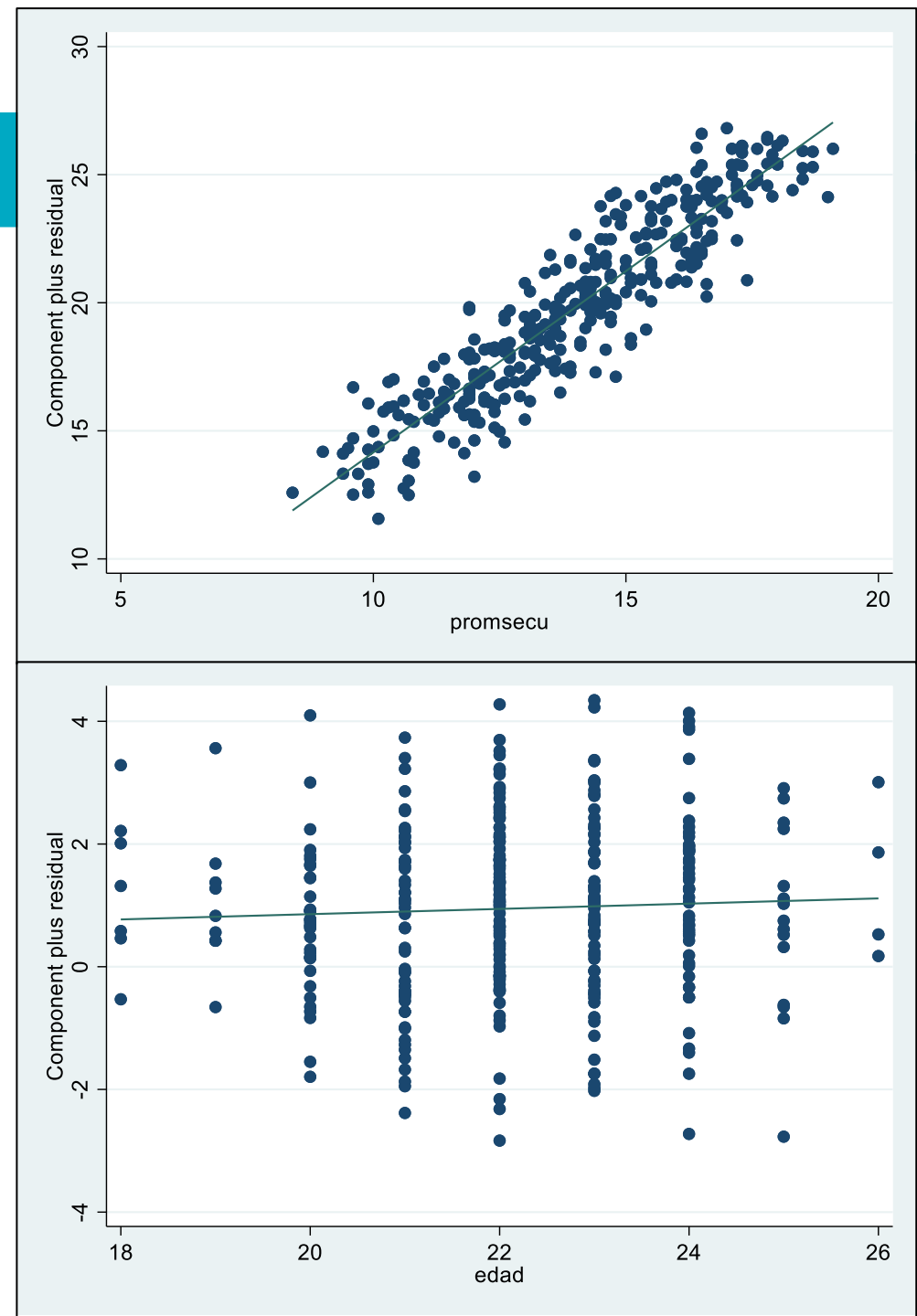
(Luego de correr la regresión ajustada)

`cprplot promsecu`

`cprplot edad`

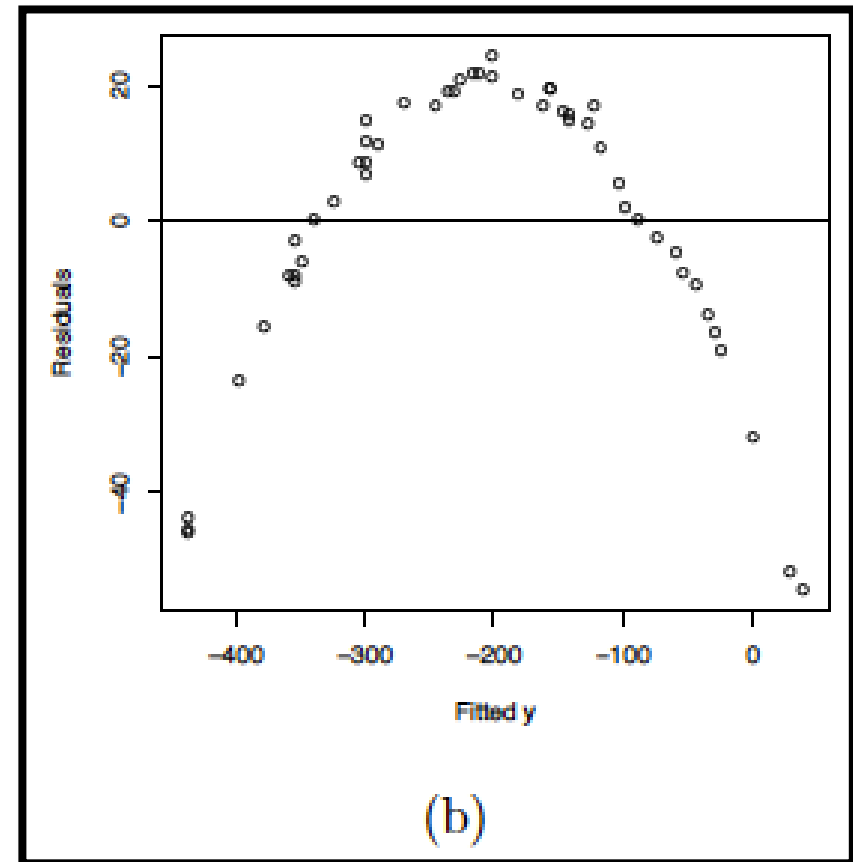
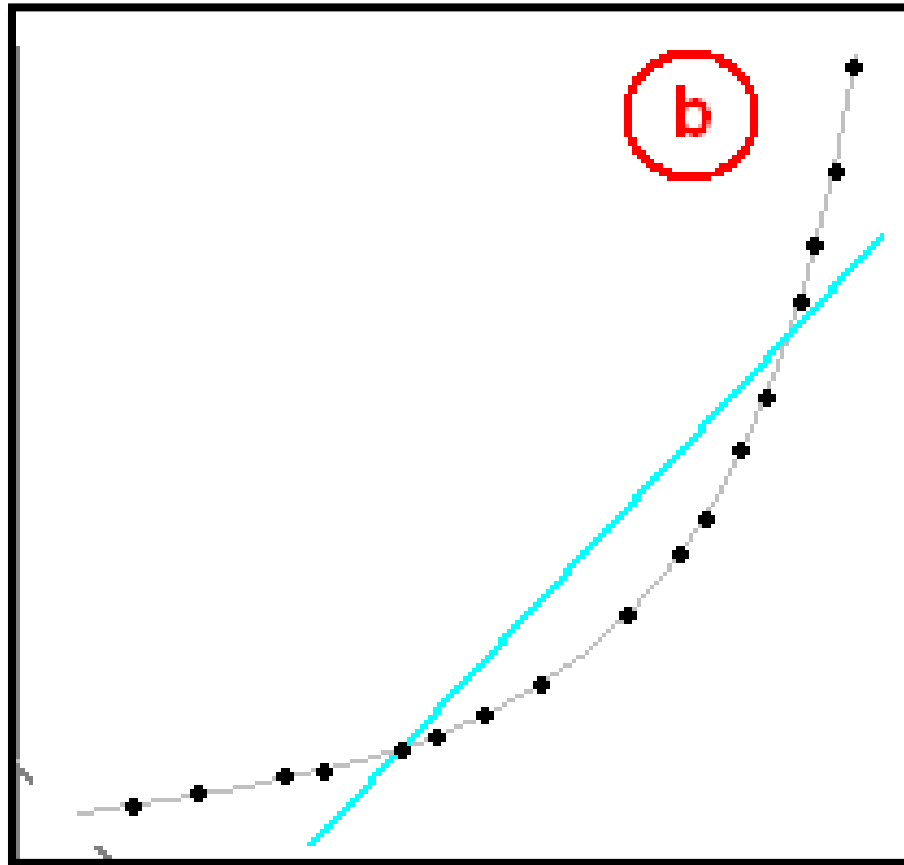
Se cumple el supuesto cuando línea azul es diagonal positiva, diagonal negativa u horizontal

No se cumple cuando dicha línea presenta altibajos pronunciados



1. Relación lineal

- Ejemplos cuando no se cumple relación no lineal



1. Relación lineal

¿Qué hacer si no se cumple con el supuesto de relación lineal?



- Hay diferentes formas de poder enfrentar el no cumplimiento de este supuesto.
- **La más sencilla es categorizar la variable independiente numérica, siguiendo una regla lógica → Automáticamente cumplirá con el supuesto**

2. Observaciones independientes

- “La medición del **desenlace** de una observación (persona / encuestado / paciente...), no debe influenciar en la medición del **desenlace** de otra observación”
- **Evaluación teórica, en base al diseño de estudio elegido.**
- **Confirmar que durante recolección de datos, esta fue independiente para cada observación.**
- **Si no lo es, no se puede usar un modelo de regresión lineal. Debe valorarse la corrección por cluster, multinivel, u otros métodos.**



3. Distribución normal de residuos

Normalidad de los **residuos**, por cada valor de VI

1. Método de evaluación para VI numérica en regresión cruda:

Evaluar la **normalidad de residuos** (Luego de correr la regresión cruda y crear sus residuos):

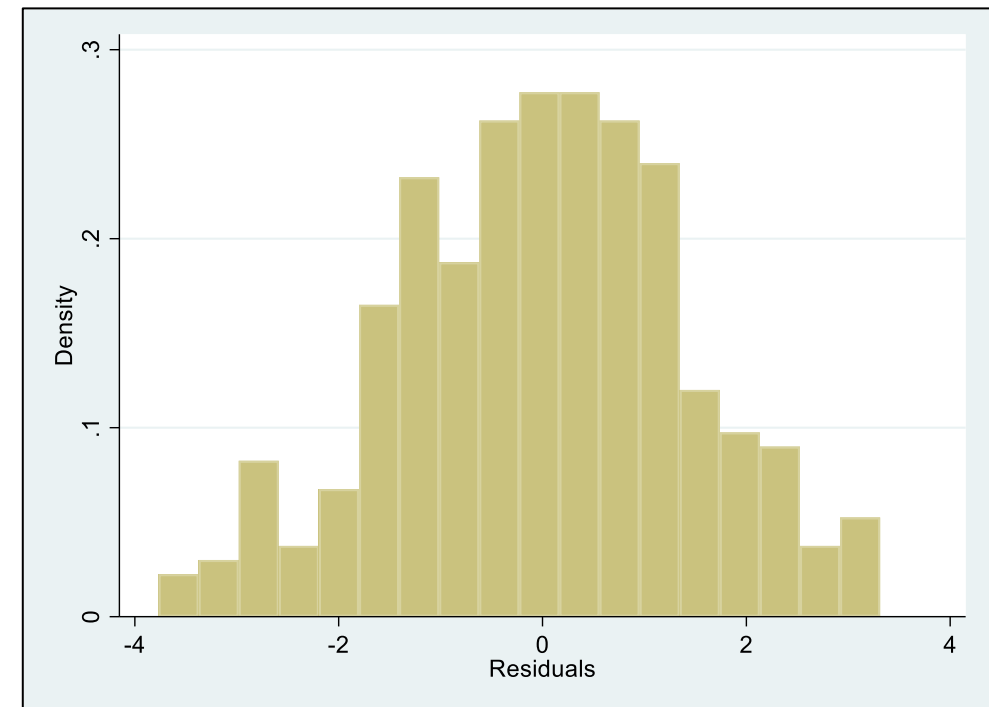
- regress **vardep varindep**
- predict **resi1**, residuals
- histogram **resi1** → Distribución normal en forma de campana

Cuando el histograma no sea concluyente:

sum **resi1**, d → Igualdad de media y mediana,

kurtosis = 3, sesgo = 0

swilk **resi1** → $p > 0.05$



3. Distribución normal de residuos

• II. Método de evaluación para VI categórica (dicotómica o politómica) en regresión cruda

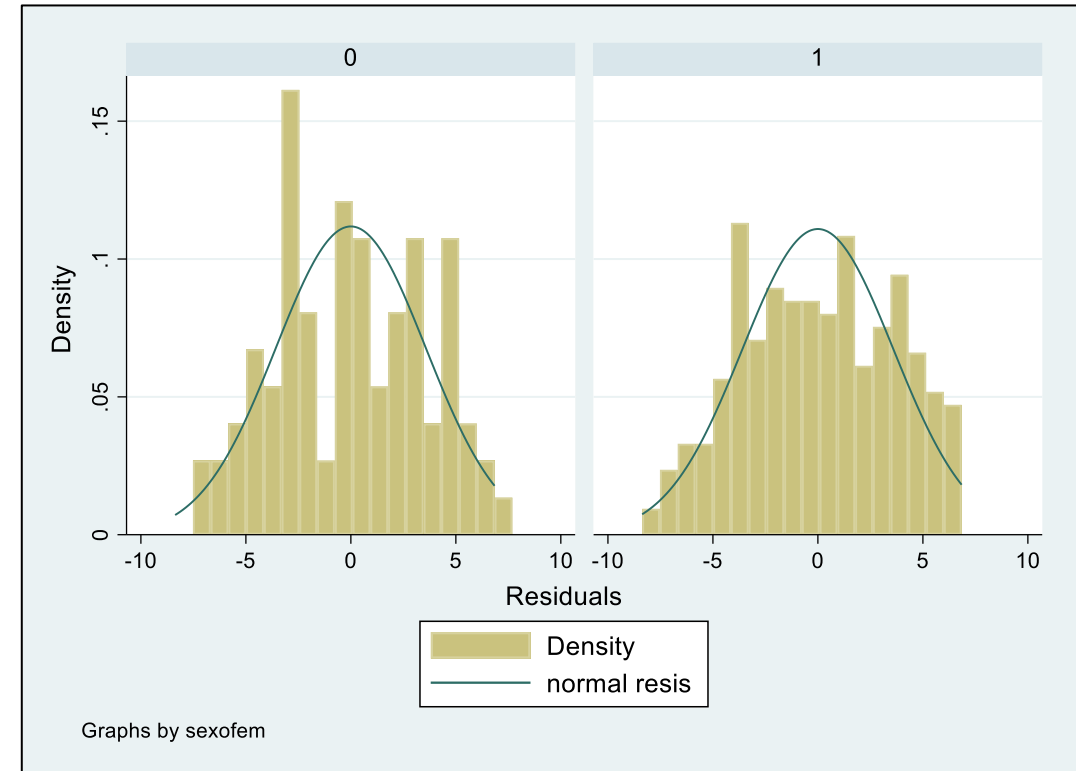
Evaluar la **Normalidad** (Luego de correr la regresión cruda y crear sus residuos)

- regress vardep i.varindep
- predict resi10, residuals
- histogram resi10, by(sexofem) normal →
Distribución normal en forma de campana

Cuando el histograma no sea concluyente:

by sexofem: sum resi10, d → Igualdad de media y mediana, kurtosis = 3, sesgo = 0

by sexofem : swilk resi10 → $p > 0.05$



3. Distribución normal de residuos

- **III. Método para regresión ajustada (con al menos una VI numérica)**

Normalidad de residuos (Luego de correr la regresión ajustada y crear residuos)

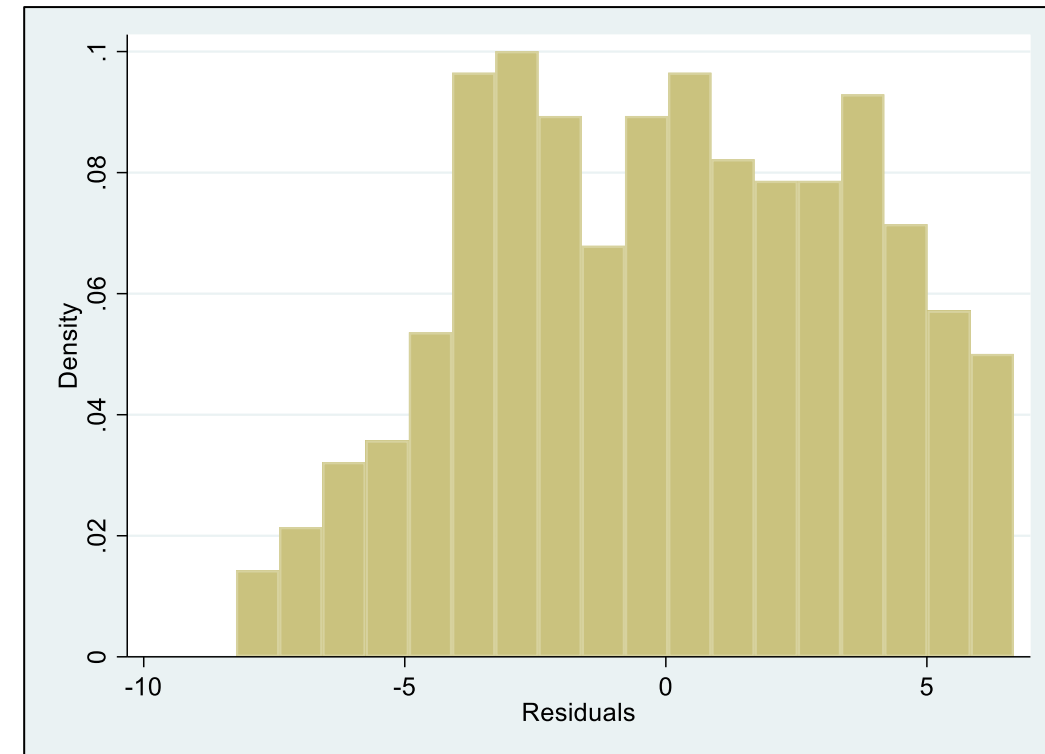
- regress **vardep** **varindep1** i.**varindep2** ...
- predict **resi5**, residuals
- histogram **resi5** → **Distribución normal en forma de campana**

Cuando el histograma no sea concluyente:

sum resi5, d → **Igualdad de media y mediana,**

kurtosis = 3, sesgo = 0

swilk resi5 → **$p > 0.05$**



3. Distribución normal de residuos

¿Cuál es el problema de no cumplir con este supuesto?

- Interpretación inferencial no es adecuada (Valor p e IC95%)

¿Qué hacer si no se cumple con supuesto?

- Hay diferentes formas de poder enfrentar el no cumplimiento de este supuesto:
 1. Considerar **no hacer regresión lineal** y hacer otro tipo de regresión con un desenlace categórico o de conteo.
 2. Usar **“Bootstrap”** en el modelo de regresión línea
 3. Por Teorema del Límite Central, **asumir que se cumple con normalidad** cuando se tiene tamaño de muestra suficiente.
 4. Usar otras correcciones estadísticas, según sea necesario

3. Distribución normal de residuos

“Bootstrap” en el modelo de regresión lineal: Asume que la muestra del estudio es representativa para toda la población. Realiza múltiples muestreos para calcular un error estándar más representativo.

Comando en Stata:

- **bootstrap, bca reps(1000): regress** vardep varindep

(Considerar que usando esta opción, podría ocasionar resultados con falsos positivos. No es perfecta.)



3. Distribución normal de residuos

Usando Bootstrap obtenemos la siguiente salida en Stata:



Prob>F ahora es Prob>Chi2
Misma utilidad para decidir si ingresa o no al modelo de regresión
ajustada

4. Homocedasticidad

Homocedasticidad = igualdad de varianzas.

Es decir, que las varianzas del **desenlace** son iguales en cada valor de la VI

Método de evaluación para cualquier tipo de variable independiente (crudo o ajustado):

Prueba estadística de Cook-Weisberg

(Correr comando luego de hacer regresión)

- `estat hettest`

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
```

```
Ho: Constant variance
```

```
Variables: fitted values of nota
```

```
chi2(1)      =      0.00
```

```
Prob > chi2   =      0.9797
```

$p < 0.05$: Se rechaza H_0 → No existe homocedasticidad (no se cumple el supuesto)

$p \geq 0.05$: No se rechaza H_0 → Existe homocedasticidad (se cumple con el supuesto)

4. Homocedasticidad

¿Cuál es el problema de no cumplir con este supuesto?

- Interpretación inferencial no es adecuada (Valor p e IC95%)

¿Qué hacer si no se cumple con supuesto?

- Hay diferentes formas de poder enfrentar el no cumplimiento de este supuesto.
- La más sencilla es usar la opción **“robust”** en el modelo de regresión lineal: Convierte los valores extremos, para homogenizar las varianzas

regress vardep varindep, **robust**

5. Evaluar multicolinealidad

Multicolinealidad (solo en modelo ajustado): Presencia de dos o más VI excesivamente correlacionadas entre sí

Método para evaluar este supuesto: Calcular el “Factor de Inflación de Varianza”
(Correr comando luego de hacer regresión)
estat hettest

Variable	VIF	1/VIF
1.sexofem	1.11	0.903149
edad	1.02	0.978545
socioeco		
2	1.76	0.567837
3	1.89	0.530046
4	1.19	0.838612
1.vivesolo	1.02	0.984655
horasredes	1.03	0.974416
estadonut		
2	1.17	0.856250
3	1.08	0.926609
promsecu	1.05	0.949292
Mean VIF	1.23	

- Algún VIF > 6: nos preocupamos
- Algún VIF > 10: nos preocupamos más

Para corregir esto: Considerar excluir una de las variables con VIF alto del modelo ajustado

Supuestos con variable independiente numérica

En resumen...		Comando para evaluar	Se cumple si...	Solución en caso de no cumplir
V. Independiente numérica	Linealidad entre vardep y varindep numérica	lowess vardep varindep	Línea diagonal lineal, u horizontal	Categorizar variable independiente
	Independencia de observaciones	Teórico	Recolección de datos fue independiente entre observaciones	Realizando regresión con cluster o regresión de multinivel
	Normalidad de residuos	Luego de correr el comando de regresión: predict resi, residuals histogram resi <i>Si hay duda con el histograma, usar:</i> sum resi, d swilk resi	Distribución normal	Usar bootstrap bootstrap, bca reps(1000): regress vardep varindep
	Homocedasticidad	Luego de correr el comando de regresión: estat hettest	Valor $p \geq 0.05$	Usar robust regress vardep varindep, robust

Supuestos con variable independiente categórica

En resumen...		Comando para evaluar	Se cumple si...	Solución en caso de no cumplir
V. Independiente categórica	Linealidad entre vardep y varindep numérica	Ya se cumple No necesidad de evaluar		
	Independencia de observaciones	Teórico	Recolección de datos fue independiente entre observaciones	Realizando regresión con cluster o regresión de multinivel
	Normalidad de residuos	Luego de correr el comando de regresión: predict resi, residuals histogram resi, by(varindep) <i>Si hay duda con el histograma, usar:</i> by varindep: sum resi, d by varindep: swilk resi	Distribución normal en cada categoría	Usar bootstrap bootstrap, bca reps(1000): regress vardep varindep
	Homocedasticidad	Luego de correr el comando de regresión: estat hettest	Valor p >= 0.05	Usar robust regress vardep varindep, robust

En resumen...

Supuestos en...		Comando para evaluar	Se cumple si...	Solución en caso de no cumplir
Regresión ajustada	Linealidad entre vardep y el resto de varindep numéricas (Evaluar sólo si hay alguna variable independiente numérica en el modelo)	Luego de correr el comando de regresión ajustada, evaluar cada VI numérica: cprplot varindepnum1 cprplot varindepnum2...	Línea diagonal lineal, u horizontal, en todas las variables independientes numéricas	Categorizar las variables independientes que no cumplan
	Independencia de observaciones	Teórico	<i>Recolección de datos fue independiente entre observaciones</i>	Realizando regresión con cluster o regresión de multinivel
	Normalidad de residuos	Luego de correr el comando de regresión ajustada: predict resi, residuals histogram resi <i>Si hay duda con el histograma, usar:</i> sum resi, d swilk resi	Distribución normal	Si se usó para alguna var indep... usar bootstrap bootstrap, bca reps(1000): regress vardep varindep
	Homocedasticidad	Luego de correr el comando de regresión ajustada: estat hettest	Valor p ≥ 0.05	Si se usó para alguna var indep... usar robust regress vardep varindep, robust
	Multicolinealidad entre varindep	Luego de correr el comando de regresión ajustada: vif	Si todos los números son menos a 6	Considerar excluir variable(s), con VIF >6, de la regresión ajustada

¿Qué hacer con una variable si...?



¿Qué hacer si tengo alguna variable cuyo modelo de regresión no cumple con los supuestos de normalidad y de homocedasticidad a la vez?

- Puedes usar Bootstrap para corregir ambos supuestos a la vez en ese modelo de regresión.

Cuando estás presentando la asociación de varias VI con una VD, en una sola tabla:

Características	Nota en el primer año de universidad	
	β crudo (IC 95%)	β ajustado (IC 95%)*
Edad (años)		
Sexo		
Masculino	Ref.	Ref.
Femenino		
Estado nutricional		
Normal	Ref.	Ref.
Sobrepeso		
Obesidad		
Nota promedio en secundaria (puntos)		

- Si todas las variables cumplen todos los supuestos → correr los comandos crudos y ajustados sin problemas.
- Si al menos una variable requirió corrección con **Bootstrap y/o Robust**, ¿Qué hacemos con el resto de las variables de la tabla que no requieren la corrección?
 - **Opción 1:** Corregir a todas las variables (que requieran o no corrección) en el modelo de regresión crudo y ajustado, con un mismo método, para que el análisis sea similar en toda la tabla → **Nosotros usaremos esta :D**
 - **Opción 2:** Dejar sin corregir a las variables que no necesiten corrección en el modelo de regresión crudo. Cada variable es corregida según su necesidad, y se tiene que especificar en la tabla).

¡Recuerda!

Si decides usar Robust o Bootstrap **debes especificarlo** en la sección de métodos de tu artículo científico / trabajo de investigación.

De preferencia, también especificarlo en la tabla de resultados.



Muchas gracias (:

Próximo video 7/7: Práctica guiada en Stata



MEJORES
DECISIONES
EN SALUD



📞 950 876 703
🌐 www.evisalud.net
📘 facebook.com/evisalud/
📷 @evisaludperu



Práctica guiada en Stata

Christopher A. Alarcon-Ruiz MD MSc(c)

¿Cómo ver este video?



- Este video (práctica guiada) es un complemento previo al taller A
- Tiene como objetivo brindar ejemplos de cómo utilizar Stata para realizar análisis de regresión lineal
- Para aprovechar mejor el taller A, te recomendamos ver este video una vez y luego **utilizar el PDF** que acompaña el video para realizar los análisis por su cuenta para el taller A.

Recordando el estudio del taller A



- **Población:** Estudiantes de medicina de primer año
- **Variables independientes:** Sexo, edad, estado nutricional, nota promedio en secundaria
- **Desenlace:** Nota en el primer año de universidad (desenlace numérico)
- **Objetivo:** Evaluar los factores asociados al desenlace de interés.

Práctica guiada

- El producto de la práctica guiada será completar esta tabla.
- En este video vamos a realizar las **regresiones lineales crudas** de las variables:
 - Edad
 - Sexo
 - Estado nutricional
 - Nota de secundaria
- También haremos la **regresión lineal ajustada** (entrarán al modelo ajustado **sólo las variables que tienen $p < 0.20$** en prueba F en el modelo de regresión lineal crudo)

Características	Nota en el primer año de universidad	
	β crudo (IC 95%)	β ajustado (IC 95%)*
Edad (años)		
Sexo		
Masculino	Ref.	Ref.
Femenino		
Estado nutricional		
Normal	Ref.	Ref.
Sobrepeso		
Obesidad		
Nota promedio en secundaria (puntos)		

*Valor β ajustado por...

Pasos para el análisis de regresión lineal crudo (variable independiente numérica)

1. Determinar tipo de variable dependiente e independiente

- **codebook** vardep
- **codebook** varindep

2. Evaluación de supuestos

Linealidad entre vardep y varindep numérica	lowess vardep varindep
Independencia de observaciones	<i>Teórico</i>
Normalidad de residuos	Luego de correr el comando de regresión: predict resi, residuals histogram resi Si hay duda con el histograma, usar: sum resi, d swilk resi
Homocedasticidad	Luego de correr el comando de regresión: estat hettest

3. Análisis de regresión

- **regress** vardep varindep
(Aplicar Bootstrap o Robust según necesidad)

4. Interpretación de resultado

5. Determinar si ingresa o no en modelo ajustado (Ver valor de prueba F o chi2)

Pasos para el análisis de regresión lineal crudo (variable independiente categórica)

1. Determinar tipo de variable dependiente e independiente

- **codebook** vardep
- **codebook** varindep

2. Evaluación de supuestos

Independencia de observaciones	<i>Teórico</i>
Normalidad de residuos	<p>Luego de correr el comando de regresión:</p> <p>predict resi, residuals</p> <p>histogram resi, by(varindep)</p> <p>Si hay duda con el histograma, usar:</p> <p>by varindep: sum resi, d</p> <p>by varindep: swilk resi</p>
Homocedasticidad	<p>Luego de correr el comando de regresión:</p> <p>estat hettest</p>

3. Análisis de regresión

- **regress** vardep i.varindep
(Aplicar Bootstrap o Robust según necesidad)

4. Interpretación de resultado

5. Determinar si ingresa o no en modelo ajustado (Ver valor de prueba F o chi2)

Pasos para el análisis de regresión ajustada

1. Elegir variables para modelo ajustado
2. Evaluación de supuestos

3. Realizar regresión ajustada

- **regress** vardep varindep1 i.varindep2 ...
(Aplicar Bootstrap o Robust según necesidad)

4. Interpretación de resultados

Linealidad entre vardep y el resto de varindep numéricas (Evaluar sólo si hay alguna variable independiente numérica en el modelo)	Luego de correr el comando de regresión ajustada, evaluar cada variable independiente numérica: cprplot varindepnum1 cprplot varindepnum2...
Independencia de observaciones	<i>Teórico</i>
Normalidad de residuos	Luego de correr el comando de regresión ajustada: predict resi, residuals histogram resi Si hay duda con el histograma, usar: sum resi, d swilk resi
Homocedasticidad	Luego de correr el comando de regresión ajustada: estat hettest
Multilinealidad entre varindep	Luego de correr el comando de regresión ajustada: vif

Vamos a Stata :D



Producto de práctica guiada

Características	Nota en el primer año de universidad	
	β crudo (IC 95%)	β ajustado (IC 95%)*
Edad (años)	-0.56 (-0.30 a 0.18)	
Sexo		
Masculino	Ref.	
Femenino	0.31 (-0.56 a 1.19)	
Estado nutricional		
Normal	Ref.	Ref.
Sobrepeso	-1.56 (-2.34 a -0.77)	-0.37 (-0.69 a -0.05)
Obesidad	-1.11 (-2.70 a 0.49)	-0.26 (-0.90 a 0.39)
Nota promedio en secundaria (puntos)	1.41 (1.35 a 1.48)	1.40 (1.34 a 1.47)

*Valor β ajustado por estado nutricional y nota promedio de secundaria

Continuar con el Taller A

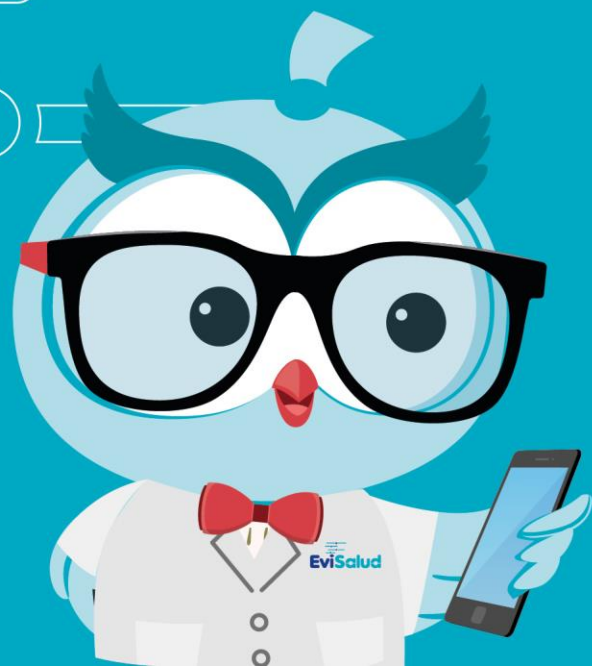
- Realizar análisis de **regresión lineal cruda** de las variables restantes:
 - Vive solo
 - Horas red
 - Nivel socioeconómico
 - Buena nota secundaria
- Realizar análisis de **regresión ajustada** (entran al modelo ajustado **sólo las variables que tienen $p < 0.20$** en prueba F o Chi2 en el modelo de regresión lineal crudo)

Características	Nota en el primer año de universidad	
	β crudo (IC 95%)	β ajustado (IC 95%)*
Edad (años)	-0.56 (-0.30 a 0.18)	
Sexo		
Masculino	Ref.	
Femenino	0.31 (-0.56 a 1.19)	
Estado nutricional		
Normal	Ref.	
Sobrepeso	-1.56 (-2.34 a -0.77)	
Obesidad	-1.11 (-2.70 a 0.49)	
Nota promedio en secundaria (puntos)	1.41 (1.35 a 1.48)	
Vive sólo		
No	Ref.	
Sí		
Horas que pasa en red (Horas)		
Nivel socioeconómico		
Muy baja	Ref.	
Baja		
Media		
Alta		
Buena nota en secundaria		
No	Ref.	
Sí		

Muchas gracias (:



MEJORES
DECISIONES
EN SALUD



📞 950 876 703
🌐 www.evisalud.net
📘 facebook.com/evisalud/
📷 @evisaludperu