present

Métodos econométricos en las Relaciones Internacionales

¿Cómo representamos la incertidumbre?[1]

- Usamos distribuciones de probabilidad para cuantificar la incertidumbre.
- En problemas que involucran incertidumbre, es esencial poder comparar la verosimilitud de diferentes afirmaciones
- Por ejemplo, nos gustaría ser capaces de representar que la proposición A es más plausible que la proposición B.
- Podemos representar la plausibilidad de proposiciones por medio de una función P valuada en los números reales.

$$P(A) > P(B)$$

$$P(A) = P(B)$$

[1]: Kochenderfer, M. J., Wheeler, T. A., & Wray, K. H. (2022). Algorithms for decision making. MIT press.

¿Cómo representamos la incertidumbre?

- Con supuestos adicionales de la forma de la función P, esta cumple con los *axiomas de probabilidad*.
- Si tenemos completa certeza de la ocurrencia de A, entonces $P(A) = 1$.
- Si creemos que A es imposible, entonces $P(A) = 0$.
- La incertidumbre en la ocurrencia de A está representada por valores entre estos dos extremos :

$$0 \leq P(A) \leq 1$$

Distribuciones de probabilidad

- Una *distribución de probabilidad* asigna probabilidades a diferentes salidas.
- La forma como representamos las distribuciones de probabilidad depende de si estas involucran eventos **discretos** o **continuos**.

Distribuciones de probabilidad discretas

- Una distribución de probabilidad discreta es una distribución definida sobre un conjunto discreto de valores (eventos).
- Podemos representar tal distribución como una *función de probabilidad* la cual asigna una probabilidad a cada evento posible. Esta función de probabilidad toma como entrada una variable aleatoria X que puede tomar uno de los n eventos posibles.
- Una restricción de esta función es que la suma de las probabilidades de los eventos debe ser igual a 1

$$\sum_{i=1}^{n} P(X = i) = 1$$

y $0 \leq P(X = i) \leq 1$ para todo i

Consejo

La variable X representa alguna cantidad **desconocida** de interés. Si el valor de X es desconocido o puede cambiar, llamamos a X una variable aleatoria.

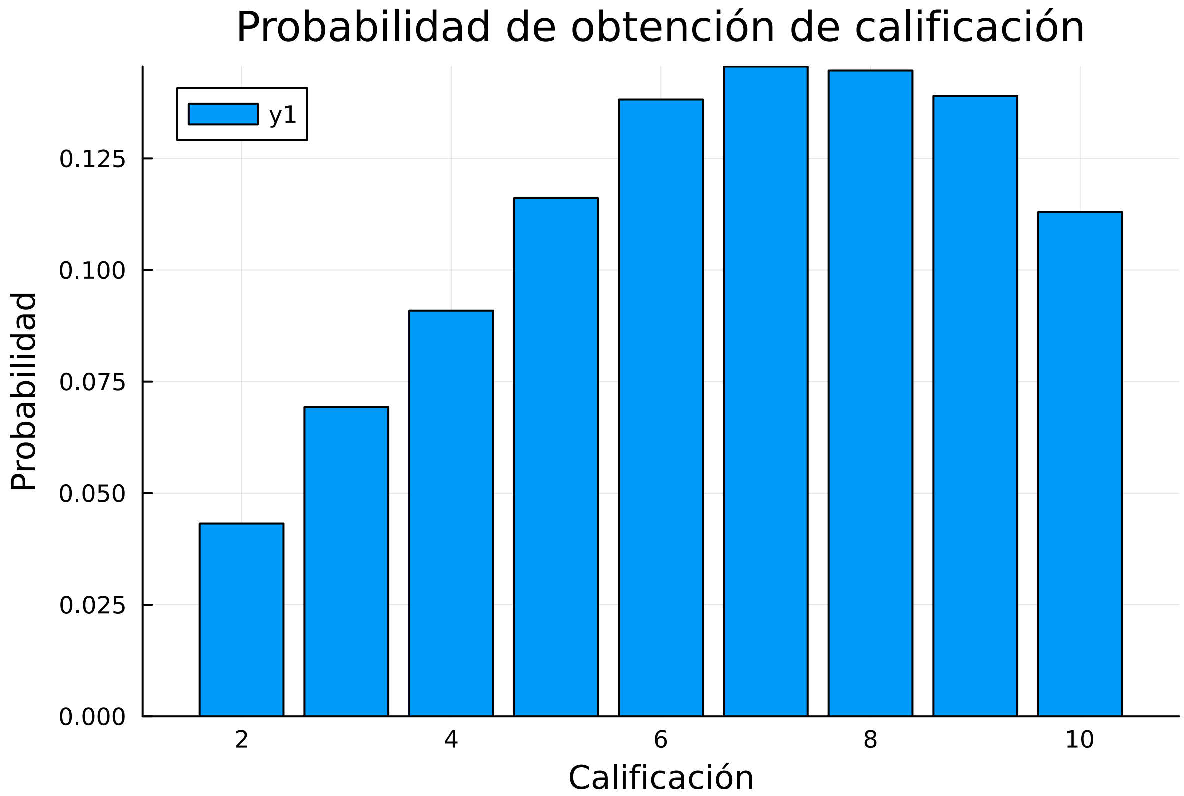
El conjunto de valores posibles lo denotamos como $\mathcal{X}$, que se conoce como el **espacio muestral** o **espacio de estados**. Un evento es un conjunto de realizaciones de un espacio muestral espacio muestral dado.

Por ejemplo, si X representa las caras de un dado que es lanzado, $X = \{1, 2, 3, 4, 5, 6\}$, el evento de *ver un 1* se denota $X = 1$, de *ver un numero impar* es $X = \{1, 3, 5\}$

Distribuciones de probabilidad continuas

- Una distribución de probabilidad continua es una distribución sobre un conjunto de valores continuos.
- Representamos una distribución de probabilidad continua mediante una *función de densidad de probabilidad*, la cual la definimos con letras minúsculas, $p(x)$.
- Similar a la función de probabilidad del caso discreto que las probabilidades deben sumar 1, una función de densidad de probabilidad debe sumar $p(x)$ debe integrar a 1

$$\int_{-\infty}^{\infty} p(x)\, dx = 1$$



Distribución de Probabilidad Uniforme

- La distribución uniforme $U(a, b)$ asigna una densidad de probabilidad igual o uniforme entre los límites a y b.
- Por lo tanto, la función de densidad es igual a

$$p(x) = \frac{1}{(b-a)} \quad \text{para } x \text{ en el intervalo } [a, b]$$

- Reescribimos la densidad de la siguiente manera:

$$U(x|a, b)$$

Los **parámetros** a y b definen a la densidad. Por su parte, x es una **variable aleatoria** (*i.e* un valor que proviene de una densidad de probabilidad).

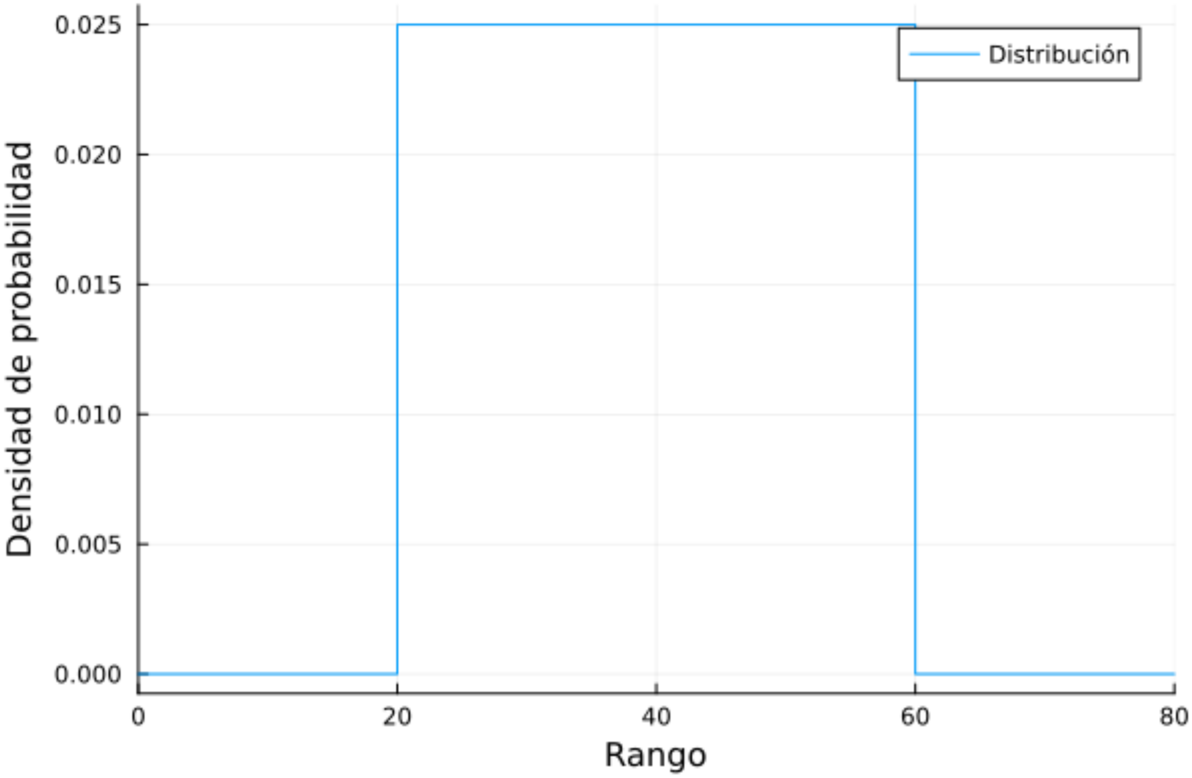
Distribución de Probabilidad Uniforme

Selecciona valor de a

20

Selecciona valor de b

60



Distribución Normal (Gausiana)

La *Distribución Normal* está parametrizada por la media μ y varianza σ^2:

$$p(x) = N(x|\mu, \sigma^2)$$

el parámetro σ es la *desviación estándar*, que es igual a la raíz cuadrada de la varianza.

La densidad tiene la siguiente forma funcional y parametrización:

$$N(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x-\mu}{2\sigma^2}\right)$$

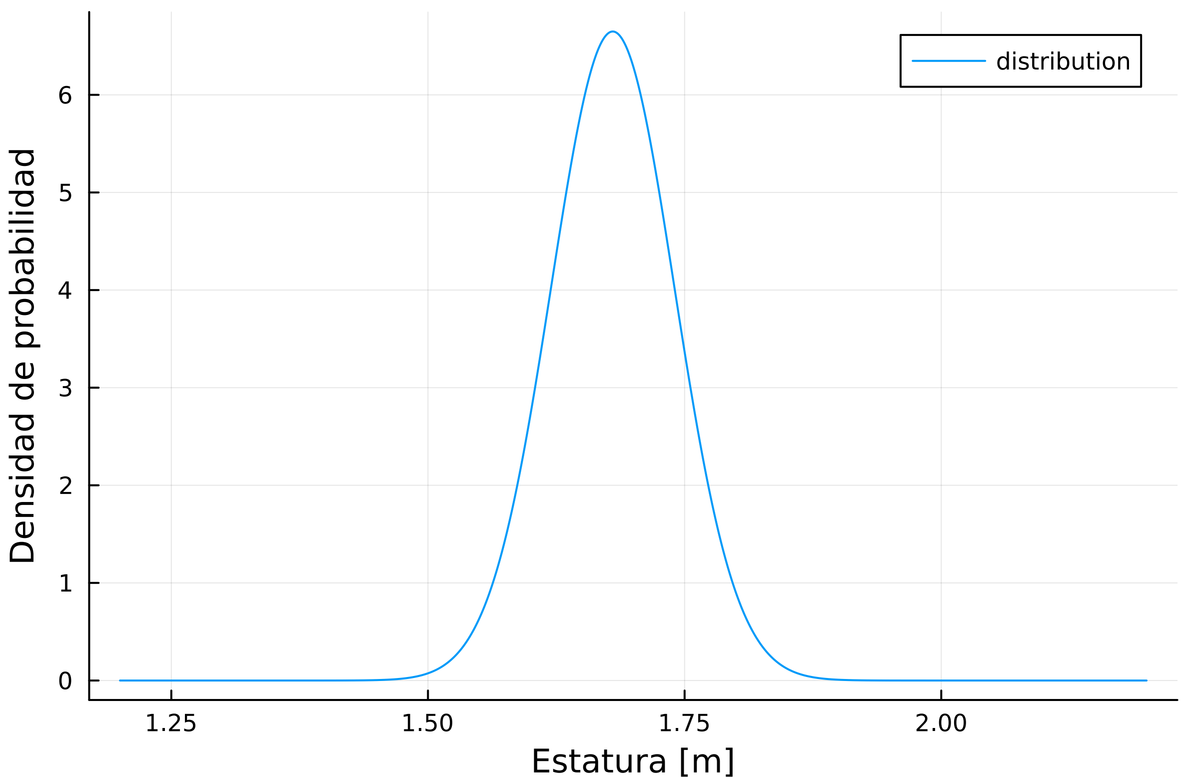
Distribución Normal (Gausiana)

Selecciona valor de μ = 1.68

Consejo

Recuerda: la media μ es un parámetro de localización y la varianza σ^2 es un parámetro de dispersión.

Selecciona valor de σ^2 = 0.06



Distribuciones Conjuntas

- Una *Distribución Conjunta* es una distribución de probabilidad sobre múltiples variables.
- Una distribución sobre una sola variable es una *distribución univariada*.
- Una distribución sobre varias variables es llamada *distribución multivariada*.
- Si tenemos una distribución conjunta sobre dos variables discretas X y Y, entonces $P(X, Y)$ denota la probabilidad que se de una realización $X = x$ y $Y = y$.

Distribuciones Conjuntas Discretas

Si las variables son discretas, la distribución conjunta puede ser representada por una tabla como la siguiente, la cual representa la distribución conjunta de tres variables binarias X, Y y Z.

X	Y	Z	$P(X, Y, Z)$
0	0	0	0.08
0	0	1	0.31
0	1	0	0.09
0	1	1	0.37
1	0	0	0.01
1	0	1	0.05
1	1	0	0.02
1	1	1	0.07

- En algunos casos podemos asumir que nuestras variables son *independientes*, lo que significa que la realización de una no afecta la disribución de probabilidad de la otra.
- Si X y Y son independientes ($X \perp Y$), entonces tenemos que $P(X, Y) = P(X)P(Y)$ para toda x y y.

Distribuciones Conjuntas Continuas

Una de las distribuciones conjuntas continuas más comunes es la *Distribución Normal Multivariada*, la cual se representa como

$$N(\mathbf{x}|\mu, \mathbf{\Sigma})$$

donde μ es un vector de tamaño n y $\mathbf{\Sigma}$ es una matriz $n \times n$. Suele llamarse a μ como el vector de medias, cuyos elementos individuales denotan los valores esperados correspondientes a cada variable aleatoria, mientras que $\mathbf{\Sigma}$ describe la varianza de cada variable así como la correlación entre ellas:

Mediante la matriz de covarianza, la distribución normal multivariada describe no sólo el comportamiento de aleatorias independientes sino también la correlación entre esas variables.

Covarianza

La covarianza entre dos variables aleatorias X y Y mide el grado en que estas están relacionadas (linealmente). La covarianza se define como:

$$Cov(X, Y) = E[(X - E[X])(Y - E[Y])]$$

Si $\mathbf{x}$ es un vector de tamaño n, su matriz de covarianza está definida como

$$\mathbf{\Sigma} = \begin{pmatrix} Var(X_1) & Cov(X_1, X_2) & \dots & Cov(X_1, X_n) \\ Cov(X_2, X_1) & Var(X_2) & \dots & Cov(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_n, X_1) & Cov(X_n, X_2) & \dots & Var(X_n) \end{pmatrix}$$

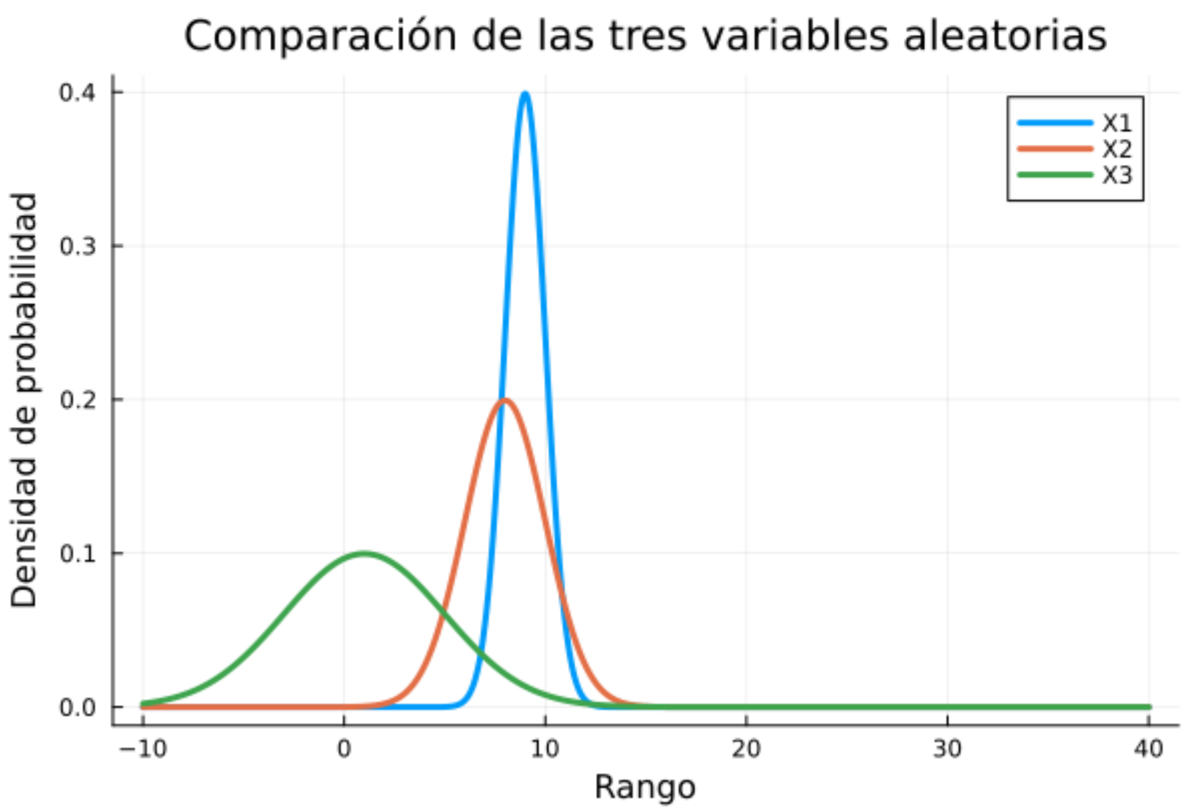
Ejemplo : Distribución Normal Multivariada

Supongamos tres variables aleatorias que siguen una distribución normal univariada y que describen los precios de vivienda de tres colonias de la Ciudad de México:

- La variable aleatoria X_1 modela la distribución de los precios de vivienda en la colonia Napoles la cual tiene parámetros $N(9,1)$.
- La variable aleatoria X_2 modela la distribución de los precios de vivienda en la colonia Del Valle la cual tiene parámetros $N(8,2)$.
- La variable aleatoria X_3 modela la distribución de los precios de vivienda en la colonia Morelos la cual tiene parámetros $N(1,4)$.

$$\mathbf{x} \sim N(\mu, \mathbf{\Sigma}) = N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}, \begin{bmatrix} Var(X_1) & Cov(X_1,X_2) & Cov(X_1,X_3) \\ Cov(X_2,X_1) & Var(X_2) & Cov(X_2,X_3) \\ Cov(X_3,X_1) & Cov(X_3,X_2) & Var(X_3) \end{bmatrix} \right)$$

$$\mathbf{x} \sim N(\mu, \mathbf{\Sigma}) = N\left(\begin{bmatrix} 9 \\ 8 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 & 0.9 & 0.1 \\ 0.9 & 2 & 0.1 \\ 0.1 & 0.1 & 4 \end{bmatrix} \right)$$



Distribuciones Condicionales

Una Distribución Condicional representa la distribución de una variable **dado** el valor de una o más variables aleatorias.

La definición de *Propabilidad Condicional* es

$$P(y|x)$$

donde $P(y|x)$ se lee como **la probabilidad de y dado x**. En algunos contextos se refieren a x como la **evidencia**.

Modelos Lineales Gaussianos(Normales)

Un **Modelo Lineal Gaussiano** para $P(Y|X)$ representa la distribución sobre una variable continua Y como una distribución Normal donde la media es una función lineal del valor de la variable continua X.

La función de densidad está representada por

$$p(y \mid x; \theta) = N(y \mid b + mx \ , \ \sigma^2)$$

con parámetros $\theta = [m, b, \sigma]$. La media es una función lineal de x definida por parámetros m y b. La varianza σ es constante.

Regresión Lineal. Caso Univariado

En el contexto de econometría, los parámetros b, m corresponden a los parámetros β_0 y β_1. De manera que el Modelo Lineal del caso univariado (*i.e* donde sólo hay una **variable independiente o exógena**) de la regresión lineal es igual a

$$p(y \mid x; \theta) = N(y \mid \beta_0 + \beta_1 x \ , \ \sigma^2)$$

con parámetros $\theta = [\beta_0, \beta_1, \sigma]$.

En el contexto de regresión, se asume que el parámetro σ es constante. Este supuesto se le conoce como **homoscedasticidad**.

Consejo

A la ecuación

$$\beta_0 + \beta_1 x$$

Se le concce como **ecuación de regresión**.

A los parámetros β_0 y β_1 se les conoce como **parámetros poblacionales**, en el sentido que generan las observaciones del modelo de regresión.

Ilustración del Modelo de Regresión Lineal Univariado (Varianza constante u Homocedástica)

Consejo

La linea roja se le denomina la **linea de regresión** y es obtenida a partir de la **ecuación de regresión**

$$\beta_0 + \beta_1 x$$

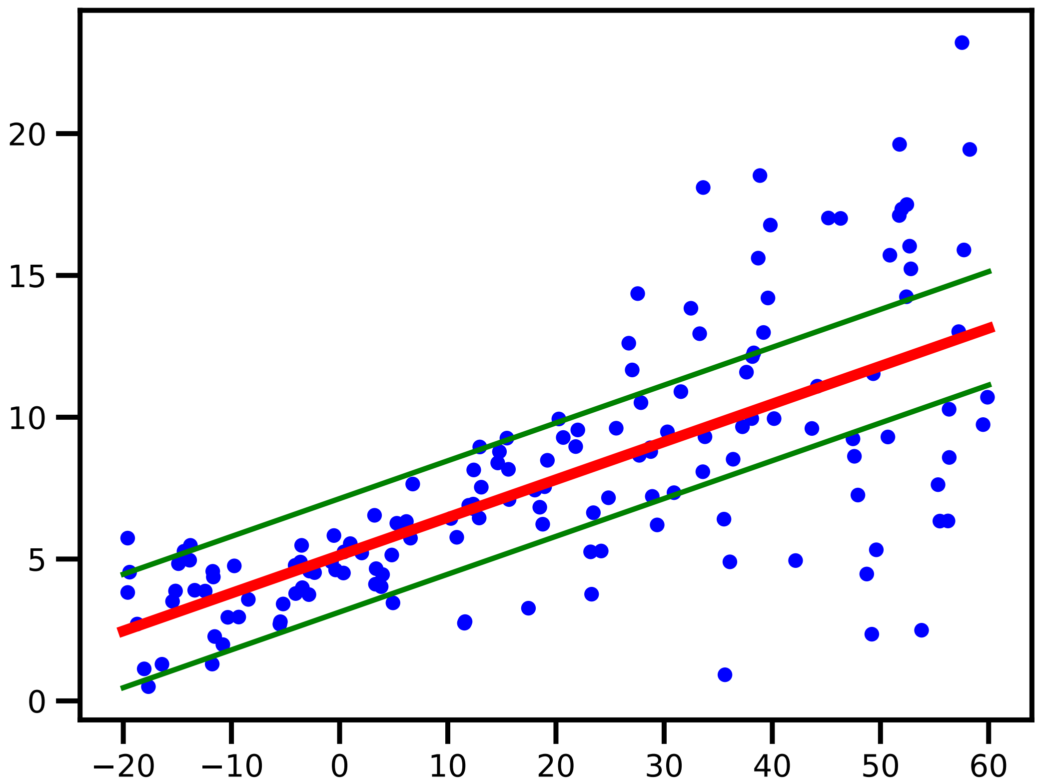
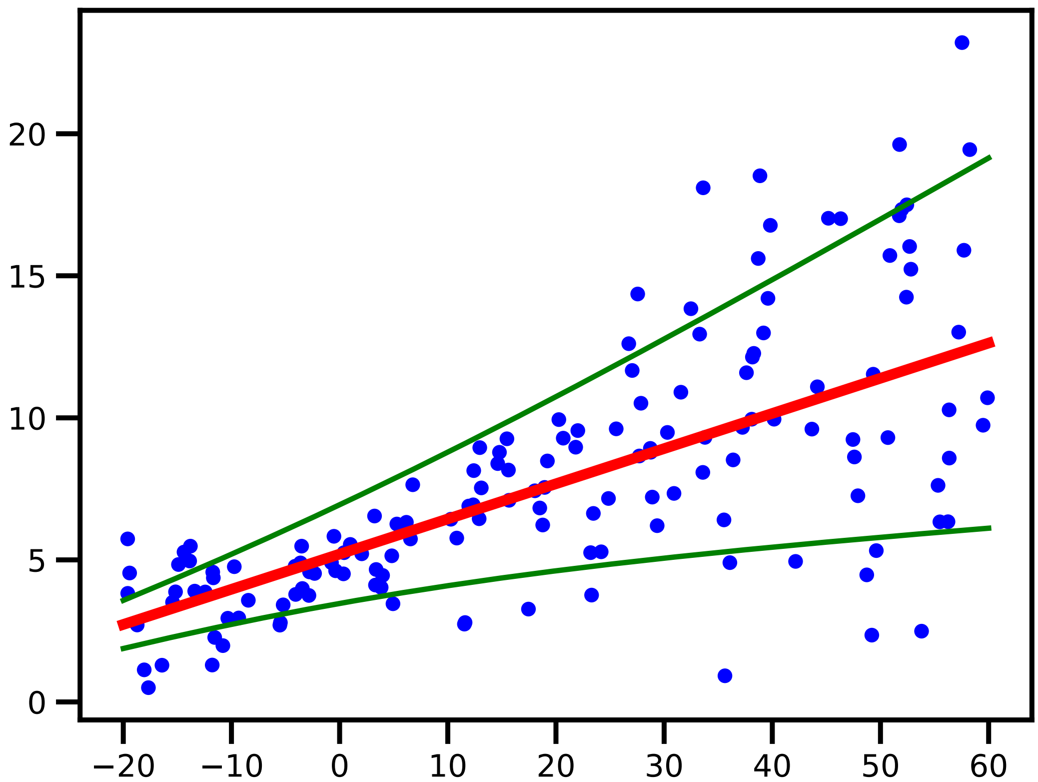


Ilustración del Modelo de Regresión Lineal Univariado (Varianza variable u Heterocedástica)

Consejo

La linea roja se le denomina la **linea de regresión** y es obtenida a partir de la **ecuación de regresión**

$$\beta_0 + \beta_1 x$$



Consideraciones del modelo de Regresión Lineal Univariado

- Podemos usar los valores de una variable (x) para **predecir** los valores de otra variable (y). Decimos que explicamos a y usando x.

Aviso:

Esta es una **explicación estadística**. Esta **no** es una explicación del **por qué** pasa y (a menos que encontremos el efecto de x sobre y). El modelo se limita a la capacidad de predecir y mediante x.

Consideraciones del modelo de Regresión Lineal Univariado [2]

- Fácil interpretación de los coeficientes. Por ejemplo, el incremento en una unidad en x está asociado con un incremento en β_1 veces en y.
- Si usamos observaciones de las variables predictoras en la regresión estimada, obtenemos una predicción $\hat{y}$. Esta es la parte de y que es explicada por la regresión. La diferencia entre y y $\hat{y}$ corresponde a la parte no explicada por la regresión, y se le denomina el **residual** ε.

$$\varepsilon = y - \hat{y}$$

- Si agregamos más variables a la ecuación de regresión, como $y = \beta_0 + \beta_1 x + \beta_2 z$, el parámetro β_1 se interpreta como *la parte de y que no es explicada por z* y *la parte de x que no es explicada por z*.
- Si consideramos que la relación entre y y x no está bien explicada por una línea recta, podemos usar una línea curva en su lugar. Para esto, podemos reformular la ecuación de regresión de la siguiente manera

$$y = \beta_0 + \beta_1 x + \beta_2 x^2$$

[1]: Huntington-Klein, N. (2021). The effect: An introduction to research design and causality. Chapman and Hall/CRC.

¿Cómo ajustamos la línea de regresión?

- Queremos que nuestro modelo de regresión tenga una capacidad predictoria alta.
- Nos gustaría que la parte que **no** es explicada por la regresión, el **residual** ε sea **mínimo**, idealmente igual a 0.
- O de forma equivalente, nos gustaría que la diferencia entre y y $\hat{y}$ sea la mínima o, idealmente, 0.
- Necesitamos resolver un problema de **optimización** donde buscamos decidir los valores de β_0 y β_1 que **minimizan** el residual.
- Cuando el problema de optimización se restringe a minimizar la diferencia **al cuadrado** entre y y $\hat{y}$:

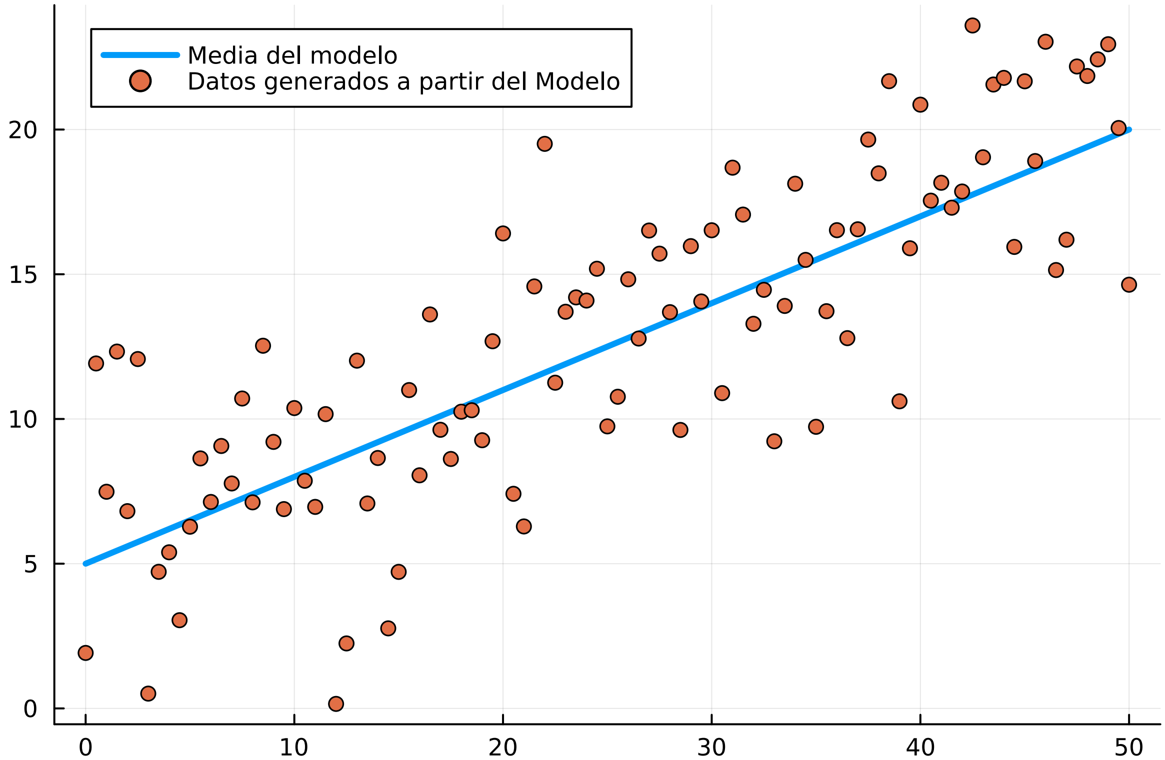
$$\min_{\beta_0,\beta_1} \varepsilon = (y - \hat{y})^2$$

- estamos utilizando el método de Mínimos Cuadrados Ordinarios (OLS en sus siglas en inglés *Ordinary least squares*)
- Los valores de β_0 y β_1 estimados por este método se les conoce como **coeficientes** y se definen como $\hat{\beta}_0$ y $\hat{\beta}_1$.

¿Cómo ajustamos la línea de regresión?

- Supongamos que el **modelo verdadero** o **proceso generador de datos** está dado por algún modelo condicional:

$$p(y \mid x; \theta) = N(y \mid \beta_0 + \beta_1 x \ , \ \sigma^2)$$



Actividad: Selecciona los valores de β_0 y β_1 que hagan 0 (minimicen) el residual

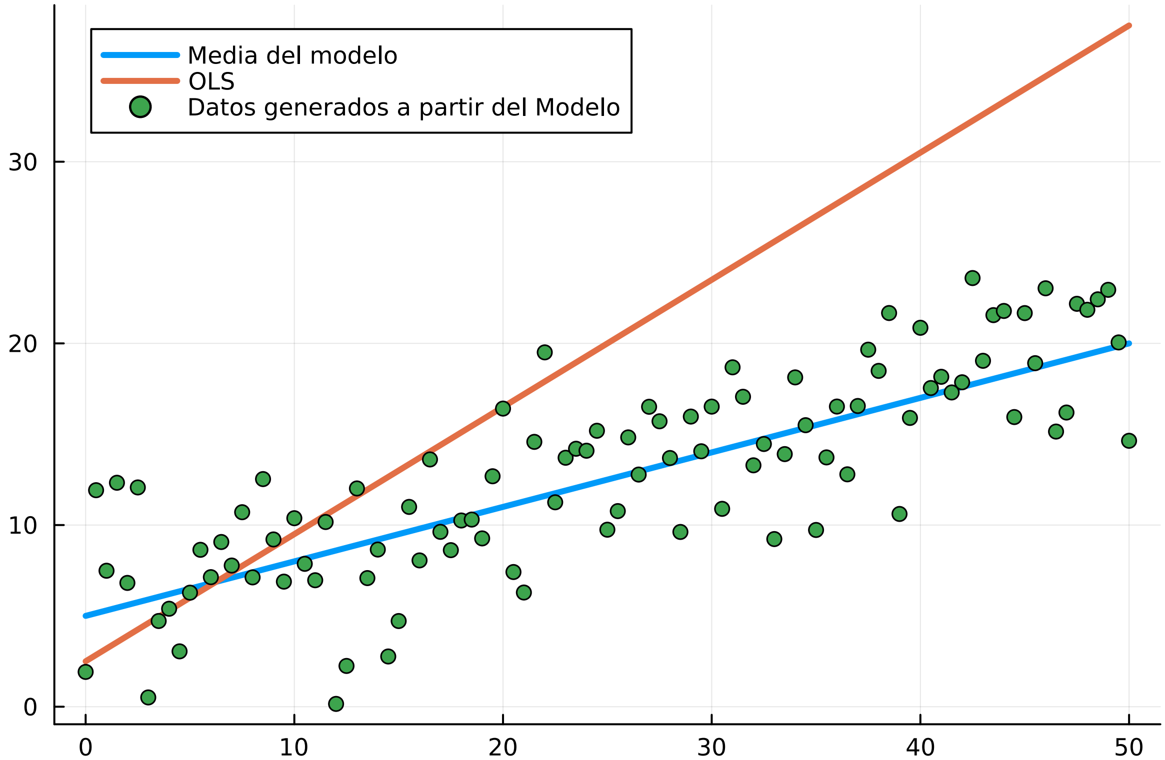
Selecciona valor de $\hat{\beta}_0$ = 2.5

2.5

Selecciona valor de $\hat{\beta}_1$ = 0.7

0.7

Valor del residual ε = 9115.249999999998



Prueba de hipótesis

- Los coeficientes β_0 y β_1 **también siguen una distribución de probabilidad**.
- Bajo la estimación por OLS, el coeficiente $\hat{\beta}_1$ sigue una distribución normal

$$\hat{\beta}_1 \sim N\left(\beta_1, \sqrt{\frac{\sigma^2}{\mathrm{var}(X)n}}\right)$$

con media β_1 (*i.e* media en el parámetro poblacional) y desviación estándar $\sqrt{\frac{\sigma^2}{\mathrm{var}(X)n}}$, donde n es la cantidad de observaciones en los datos, σ es la desviación estándar del residual ε y $\mathrm{var}(X)$ es la varianza de los datos.

- **¿Por qué queremos conocer la distribución de los coeficientes?** Para realizar **Pruebas de Hipótesis**.

Pruebas de Hipótesis[3]

- Cuando decidimos si un modelo es un buen descriptor de nuestros datos o no lo es, siempre es útil preguntarnos **¿con respecto a qué?**
- Para formalizar esta pregunta, supongamos que tenemos 2 hipótesis: una **hipótesis nula** H_0 y una **hipótesis alternativa** H_1.
- Queremos elegir aquella hipótesis que pensamos es más probable.
- En el contexto de regresión, las hipótesis a contrastar para β_1 son las siguientes:

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

- Si consideramos que la H_1 es más probable que H_0, es decir, **rechazamos** la H_0, podemos decir que *hay evidencia para pensar que es poco probable que el valor del coeficiente* sea 0.
- La poca probabilidad que el valor del coeficiente sea 0, significa que hay **alguna** relación entre la variables explicativa y la variable de respuesta.

[3]: Murphy, K. P. (2022). Probabilistic machine learning: an introduction. MIT press.

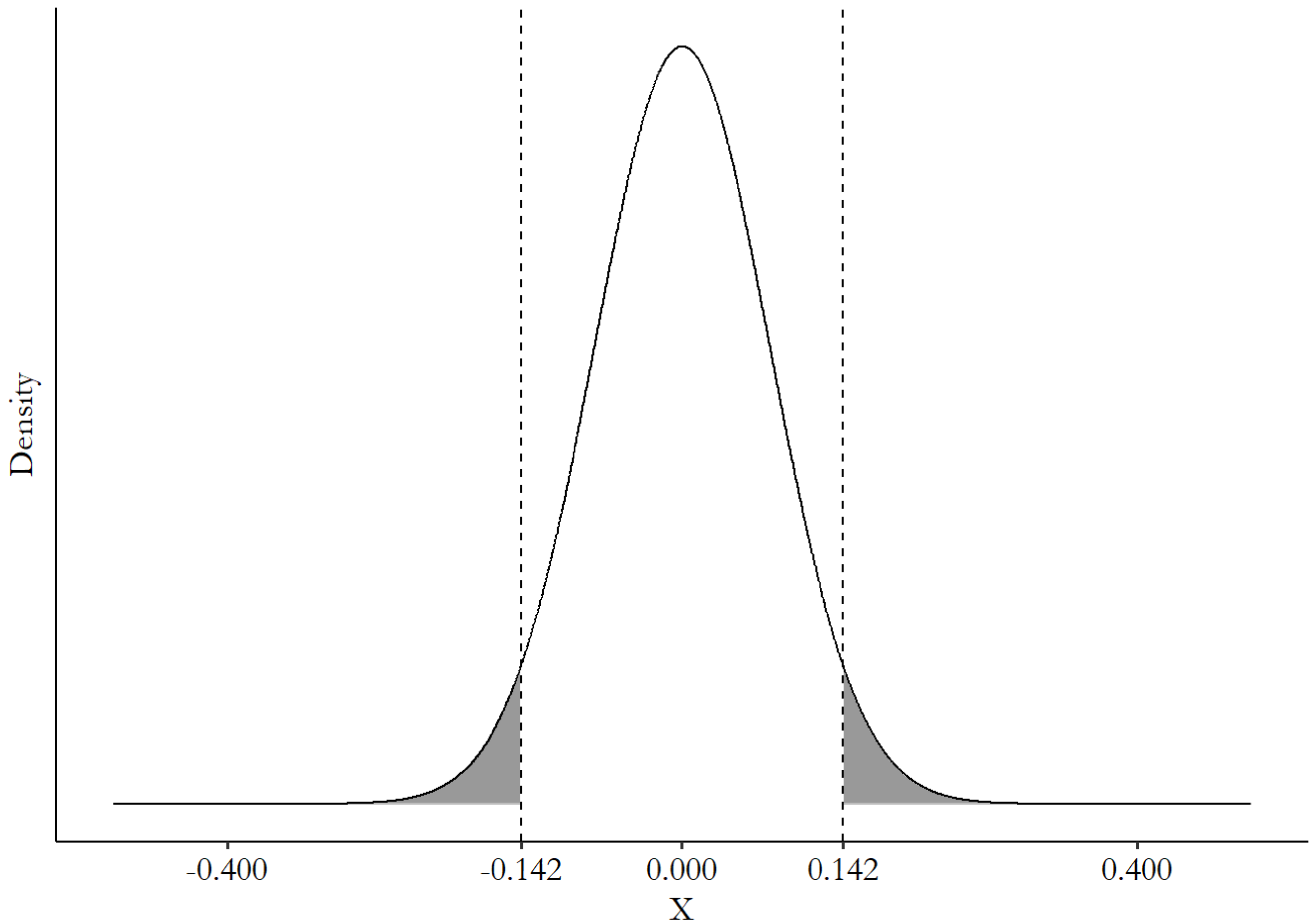
Pruebas de Hipótesis

- Si el valor estimado de $\hat{\beta}_1$ *es inesperadamente más grande* al valor propuesto en H_0 ($\beta_1 = 0$), rechazamos H_0.
- Es decir, la probabilidad de observar el valor estimado de $\hat{\beta}_1$ es muy baja dado el valor de β_1 en H_0
- El **p-value** es la probabilidad, bajo la hipótesis nula H_0 de observar un **estadístico de prueba** que es tan grande o mayor que el valor de H_0.
- Valores pequeños del p-value corresponden a evidencia fuerte en contra de H_0.
- Tradicionalmente, se rechaza la hipótesis nula H_0 si el p-value es más chico que $\alpha = 0.05$, el cual es llamado **nivel de significancia** de la prueba.
- Cuando al aplicar la prueba estadística a nuestro estimador $\hat{\beta}_1$ se rechaza la H_0, decimos que el estimador tiene **significancia estadística**.

Consejo

- Podemos pensar el *inesperadamente más grande* como que el valor estimado de β_1 se encuentra *muy alejado* al valor propuesto en H_0 ($\beta_1 = 0$)
- La significancia estadística estadística **solo** proporciona información acerca si el valor propuesto en la hipótesis nula H_0 es **improbable**. No dice nada acerca si el valor estimado **importa**.

Pruebas de Hipótesis



Regresión lineal en la práctica

¿Cadenas de restaurantes con obtienen mejores resultados en inspecciones sanitarias que los restaurantes con pocas sucursales?

- Para responder a la pregunta, usaremos los datos de Louis-Ashley Camus sobre inspecciones en restaurantes en USA.
- Los datos tienen información del puntaje de inspección (máximo puntaje es 100), el año de la inspección, y el número de sucursales de la cadena de restaurantes.
- La siguiente tabla presenta los datos:

27178×5 DataFrame *27153 rows omitted*

Row	business_name String	inspection_score Int64	Year Int64	NumberofLocations Int64	Weekend Bool
1	MCGINLEYS PUB	94	2017	9	false
2	VILLAGE INN #1	86	2015	66	false
3	RONNIE SUSHI 2	80	2016	79	false
4	FRED MEYER - RETAIL FISH	96	2003	86	false
5	PHO GRILL	83	2017	53	false
6	TACO KING #2	95	2008	89	false
7	AJ'S RIB-A-GO-GO	94	2017	28	false
8	MT SPURR SCHOOL	100	2005	37	false
9	RAY'S PLACE LLC	92	2007	109	false
10	SOUTHSIDE BISTRO	91	2005	59	false
11	TACO BELL #18131	98	2015	45	false
12	NEW ASIA RESTAURANT	94	2015	32	false
13	JOJO'S HYPER BEAN LLC	96	2017	11	false
⋮	⋮	⋮	⋮	⋮	⋮
27167	HULA HANDS RESTAURANT	89	2018	181	false
27168	HULA HANDS RESTAURANT	89	2018	181	false
27169	HULA HANDS RESTAURANT	89	2018	181	false
27170	BRIDGE SEAFOOD	100	2018	7	false
27171	HULA HANDS RESTAURANT	89	2018	181	false
27172	TWIN DRAGON	90	2018	127	false
27173	TWIN DRAGON	90	2018	127	false
27174	TWIN DRAGON	90	2018	127	false
27175	TWIN DRAGON	90	2018	127	false
27176	TWIN DRAGON	90	2018	127	false
27177	TWIN DRAGON	90	2018	127	false
27178	TWIN DRAGON	90	2018	127	false

Regresión lineal en la práctica

- La siguiente tabla presenta algunos estadísticos descriptivos:

	variable	mean	std	min	q25	median	q75	max
1	:inspection_score	93.6435	6.25689	66	90.0	95.0	100.0	100
2	:NumberofLocations	64.7662	84.2669	1	27.0	41.0	71.0	646

Regresión lineal en la práctica

- Para la primera regresión definimos como **variable de respuesta o dependiente** al *puntaje de inspección* y como **variable independiente-explicativa-regresora** la cantidad de sucursales:

$$\text{InspectionScore} = \beta_0 + \beta_1 \text{NumberofLocations} + \varepsilon$$

Realizaremos una segunda regresión a la cual agregaremos el año de inspección:

$$\text{InspectionScore} = \beta_0 + \beta_1 \text{NumberofLocations} + \beta_2 \text{YearofInspection} + \varepsilon$$

Regresión lineal en la práctica

```
---------------------------------------------------------
                                 inspection_score
                    -------------------------------------
                                   (1)                (2)
---------------------------------------------------------
(Intercept)                   94.866***         225.333***
                    (94.775, 94.956)   (201.006, 249.659)
NumberofLocations             -0.019***          -0.019***
                    (-0.020, -0.018)     (-0.020, -0.018)
Year                                             -0.065***
                                         (-0.077, -0.053)
---------------------------------------------------------
Obs.                            27,178             27,178
R2                               0.065              0.068
F                            1,876.705            997.386
Adjusted R2                      0.065              0.068
---------------------------------------------------------
```

Consejo

- Los estadísticos que se encuentran debajo de la información de los coeficientes de regresión nos da indicios de la *calidad del modelo*.
- El estadístico R^2 indica el porcentaje de variación explicado por nuestro modelo, mientras que R^2 Adj es R^2 pero penalizando por la cantidad de regresoras.
- El estadístico F es una prueba de significancia conjunta para contrastar si, en conjunto, todos los coeficientes de regresión estimados son distintos de 0.

Subíndices en las ecuaciones de regresión

Cuando se escriben ecuaciones de regresión, normalmente se utilizan **subíndices** en las variables que indican las unidades de variación en tiempo e individuos.

Una regresión puede ser expresada como:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

donde i nos dice en qué índice varían los datos. Es decir, ¿qué son nuestras observaciones? Generalmente i hace referencia a la variación en **individuos** (personas, municipios, países, empresas, etc). En este tipo de especificación de la ecuación de regresión, Y y X difieren entre individuos.

Subíndices en las ecuaciones de regresión

Podemos tener variación **temporal** con la siguiente especificación:

$$Y_t = \beta_0 + \beta_1 X_{t-1} + \varepsilon_t$$

donde el subíndice t corresponde al tiempo. Esta especificación describe una regresión donde cada observación corresponde a un periodo de tiempo distinto. La variable X_{t-1} indica que estamos relacionando Y con un periodo de tiempo rezagado en una unidad temporal.

Subíndices en las ecuaciones de regresión

Cuando utilizamos tanto variación temporal como entre individuo, la especificación de regresión se le conoce como **panel** o **longitudinal**:

$$Y_{it} = \beta_g + \beta_t + \beta_1 X_{it} + \beta_2 W_i + \varepsilon_{it}$$

esta especificación describe una regresión en que Y y X varían tanto entre individuos i como en el tiempo t. Hay un intercepto para cada periodo de tiempo β_t así como también para cada agrupación de individuos g. La variable W_i indica que sus observaciones sólo varían entre individuos y no cambian con el tiempo.

Regresoras discretas

- Una variable discreta es aquella que tiene un conjunto finito de posibles valores.
- Una variable binaria es un ejemplo (Mujer-Hombre, Casado-Soltero).
- Estas variables pueden ser incluidas en la regresión.
- Agregamos la variable binaria Weekend a nuestra regresión

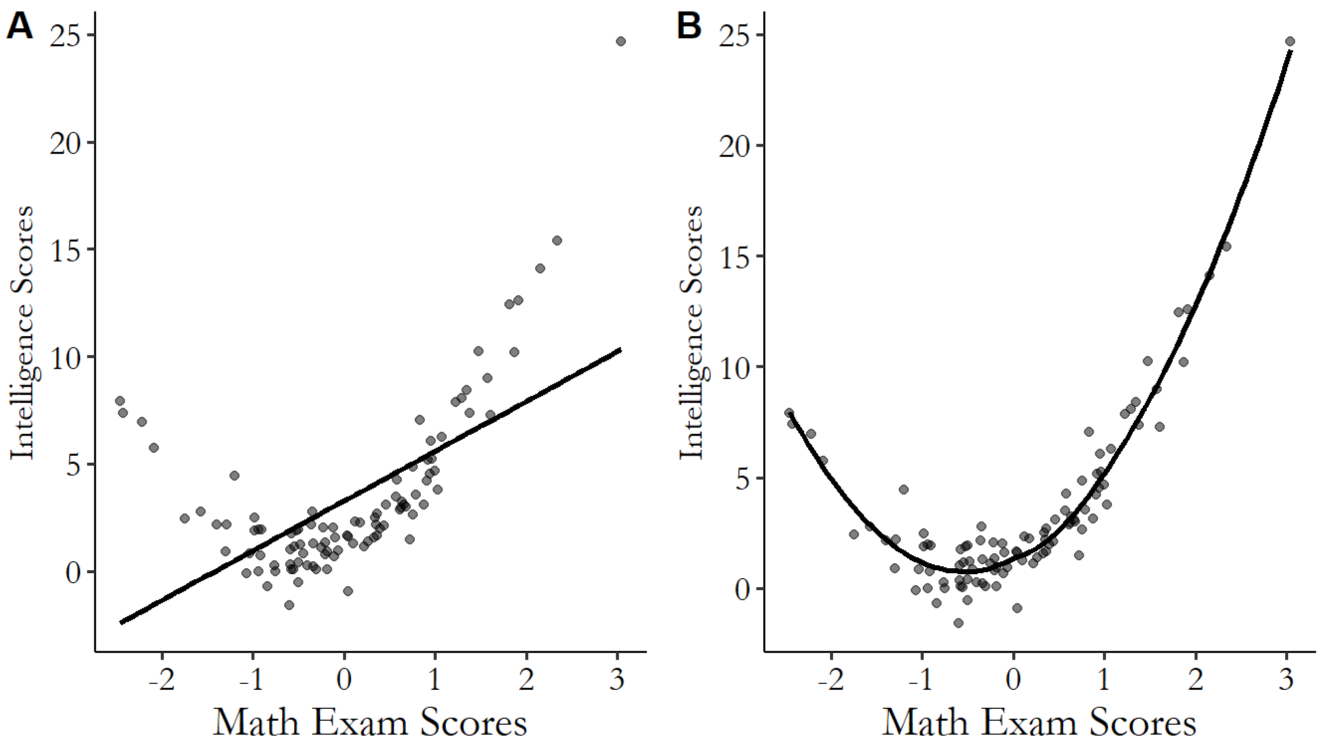
$$\text{InspectionScore} = \beta_0 + \beta_1 \text{NumberofLocations} + \beta_2 \text{Weekend} + \varepsilon$$

$$\text{InspectionScore} = \beta_0 + \beta_1 \text{NumberofLocations} + \beta_2 \text{YearofInspection} + \beta_3 \text{Weekend} +$$

- Para las regresiones del puntaje de inspección, para la regresión 3 y 4, en promedio, el puntaje es 1.49 y 1.43 más alto cuando la inspección se hace en fin de semana que cuando se hace entre semana.

	inspection_score			
	(1)	(2)	(3)	
(Intercept)	94.866***	225.333***	94.851***	22
	(94.775, 94.956)	(201.006, 249.659)	(94.760, 94.942)	(200.355,
NumberofLocations	-0.019***	-0.019***	-0.019***	-
	(-0.020, -0.018)	(-0.020, -0.018)	(-0.020, -0.018)	(-0.020,
Year		-0.065***		-
		(-0.077, -0.053)		(-0.077,
Weekend			1.498***	
			(0.676, 2.320)	(0.611
Obs.	27,178	27,178	27,178	
R2	0.065	0.068	0.065	
F	1,876.705	997.386	945.135	
Adjusted R2	0.065	0.068	0.065	

Polinomios : ¿Qué ocurre si una línea recta no es suficiente?



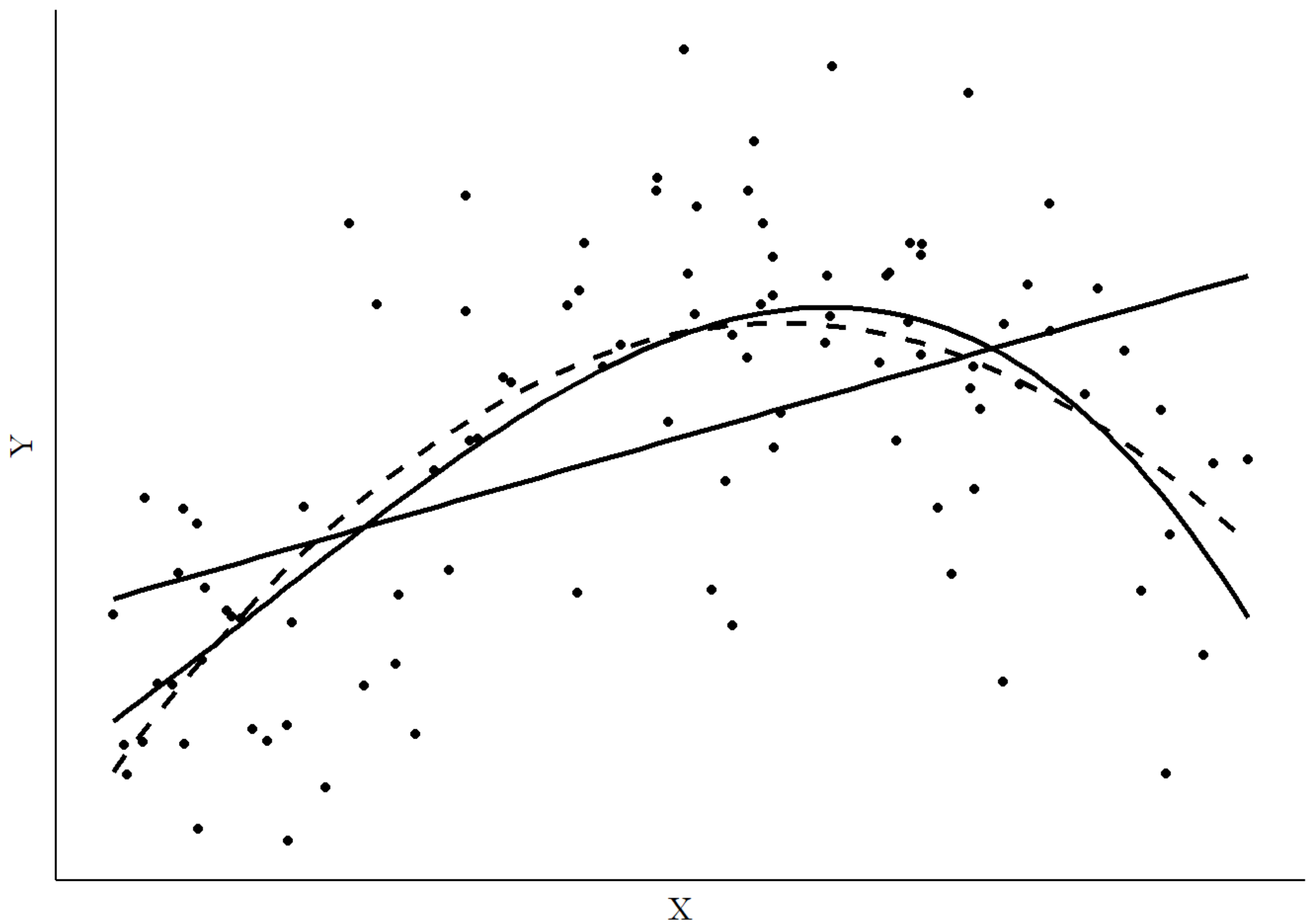
Polinomios : ¿Qué ocurre si una línea recta no es suficiente?

- Dos opciones : agregar **términos polinomiales** o **transformar variables**.
- Un **polinomio** es cuando agregamos a una misma variable a la ecuación de regresión pero elevada a alguna potencia.
- La siguiente ecuación corresponde a un polinomio de orden 3

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3$$

Polinomios : ¿Qué ocurre si una línea recta no es suficiente?

- La siguiente figura presenta el ajuste para regresiones de polinomios de orden 1,2 y 3.
- La línea sólida recta corresponde a $Y = \beta_0 + \beta_1 X$.
- La línea punteada corresponde a $Y = \beta_0 + \beta_1 X + \beta_2 X^2$
- La linea sólida curveada corresponde a $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3$



Polinomios : ¿Por qué no sólo usamos polinimos de orden infinito?

- Problema : **Sobre ajuste (overfitting)**. El modelo no es capaz de generalizar.
- Peor: No mejora el desempeño del modelo y sólo lo hace menos interpretable.

Polinomios : Interpretación de coeficientes

- Para interpretar los coeficientes necesitamos hacer un poco de cálculo.
- Para mantener la interpretación "manteniendo el resto constante, un incremento en una unidad de X está asociado con un incremento Δ in Y".
- Para calcular Δ, derivamos la ecuación de regresión $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3$ con respecto a X:

$$\frac{\partial Y}{\partial X} = \beta_1 + 2\beta_2 X + 3\beta_3 X^2$$

- Un cambio en una unidad de X está asociado con un cambio $\beta_1 + 2\beta_2 X + 3\beta_3 X^2$ en Y.
- Para nuestro modelo de puntaje de inspección, el incremento en una unidad de sucursales está asociado con un incremento en

$$\beta_1 + \beta_2 \text{NumberofLocations}$$

$$-0.0801894 + 0.0001168 \text{NumberofLocations}$$

```
----------------------------------------
                        inspection_score
----------------------------------------
(Intercept)                  97.5178574***
                        (97.403, 97.633)
NumberofLocations            -0.0801894***
                        (-0.082, -0.078)
NumberofLocations ^ 2         0.0001168***
                          (0.000, 0.000)
----------------------------------------
Obs.                              27,178
R2                                 0.194
F                              3,279.008
Adjusted R2                        0.194
----------------------------------------
```

Transformación de Variables

- Además de los polinomios, podemos aplicar transformaciones a las variables para ajustar formas no lineales.
- Generalmente la transformación de variables consiste en aplicar una función a las variables antes de usarlas.
- El logaritmo natural es una función que se utiliza con frecuencia para transformar variables.
- Por ejemplo, en vez de utilizar la regresión

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

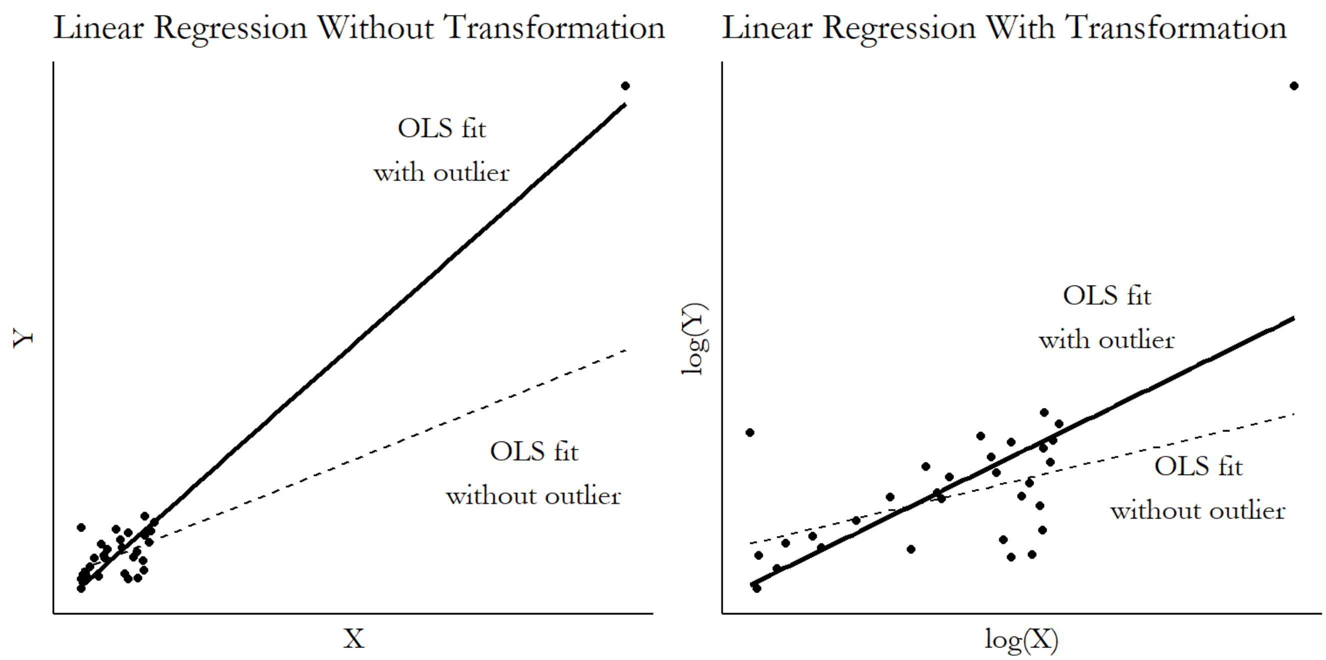
- Podemos correr:

$$Y = \beta_0 + \beta_1 \ln X + \varepsilon$$

Consejo

- La transformación de variables reduce la varianza de los datos, además que resuelve problemas de datos atípicos.

Transformación de Variables



Transformación de Variables

- Podemos aplicar transformaciones a la variable dependiente Y.
- Por ejemplo, si tenemos la siguiente especificación:

$$\ln Y = \beta_0 + \beta_1 \ln X + \varepsilon$$

Podemos interpretar el coeficiente β_1 como una **elasticidad** : el incremento en 1% en X representa, en promedio y manteniendo el resto constante, un incremento el β_1%.

Términos de interacción

- Los polinimios y transformaciones generan relaciones más flexibles entre Y y X.
- Pero puede que la relación entre Y y X difiera no sólo en base al valor de X, sino también en base a un valor distinto de una variable Z.
- Por ejemplo, ¿cuál es la relación entre el precio de la gasolina y cuánto elige conducir un individuo?
- Para las personas que poseen un automóvil, esa relación puede ser bastante fuerte y negativa.
- Para las personas que no poseen un automóvil, esa relación probablemente sea cercana a cero.
- La relación entre Y y X puede diferir entre grupos.
- Para modelar este cambio de relación entre grupos usamos **términos de interacción**.

Términos de interacción

- Un término de interacción es cuando multiplicas dos variables independientes e incluyes su producto en la regresión, por ejemplo

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ + \epsilon$$

- ¿Cómo interpretamos los términos de interacción? Dos preguntas:
- 1.- ¿Cuál es el efecto de una variable X cuándo hay una interacción entre X y otra variable del modelo?
- 2.- ¿Cómo se interpreta el término de interacción?

¿Cuál es el efecto de una variable X cuándo hay una interacción entre X y otra variable del modelo?

Si tenemos la regresión

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ + \epsilon$$

El cambio en Y por un cambio en una unidad de X es

$$Y = \beta_1 + \beta_3 Z$$

Es decir, no hay un valor único del cambio en Y por X, depende de Z. De manera que nos preguntamos cuál es el efecto de X **dado el valor** de Z.

¿Cómo se interpreta el término de interacción?

- El coeficiente β_3 nos dice **cuánto más fuerte es el efecto de X en Y cuando Z se incrementa en una unidad**.
- Supongamos una variable de interacción binaria. Tenemos la ecuación de regresión

$$Y = \beta_0 + \beta_1 X + \beta_2 \text{Child} + \beta_3 X \times \text{Child}$$

- Donde Child es una variable binaria igual a 1 si hay un infante y 0 en caso contrario.
- Sabemos que el efecto de X sobre Y es igual a

$$\beta_1 + \beta_3 \text{Child}$$

- Lo que indica que el efecto de X en Y, cuando $\text{Child} = 0$, es igual a β_1, es decir, el efecto de X en Y, cuando no hay infantes, es igual a β_1.
- El efecto de X en Y, cuando hay infantes ($\text{Child} = 1$), es igual a $\beta_1 + \beta_3$.
- Cuando encontramos significancia estadística e interpretabilidad del coeficiente de interacción β_3, podemos decir que **el efecto de X en Y difiere entre grupos que tienen infantes y los que no los tienen**.

Volvamos al ejemplo de puntaje de inspección

Agregamos dos interacciones a nuestra regresión de puntaje de inspección:

$$\begin{aligned}\text{InspectionScore} = \beta_0 + \beta_1 \text{NumberofLocations} + \\ \beta_2 \text{YearofInspection} + \beta_3 \text{NumberofLocations} \times \text{YearofInspection} + \varepsilon\end{aligned}$$

$$\begin{aligned}\text{InspectionScore} = \beta_0 + \beta_1 \text{NumberofLocations} + \\ \beta_2 \text{YearofInspection} + \beta_3 \text{NumberofLocations} \times \text{Weekend} + \varepsilon\end{aligned}$$

El coeficiente β_3 del término de interacción $\text{NumberofLocations} \times \text{Weekend}$ resultó estadísticamente significativo. Podemos decir que cada año, la relación entre un incremento unitario en la cantidad de sucursales y el puntaje de inspección se vuelve más positivo en $\beta_3 = 0.001$

	inspection_score		
	(1)	(2)	(3
(Intercept)	224.680***	311.891***	225.126**
	(200.355, 249.005)	(281.625, 342.157)	(200.793, 249.460
NumberofLocations	-0.019***	-1.462***	-0.019**
	(-0.020, -0.018)	(-1.764, -1.161)	(-0.020, -0.018
Year	-0.065***	-0.108***	-0.065**
	(-0.077, -0.052)	(-0.123, -0.093)	(-0.077, -0.053
Weekend	1.432***		1.759**
	(0.611, 2.252)		(0.803, 2.715
NumberofLocations & Year		0.001***	
		(0.001, 0.001)	
NumberofLocations & Weekend			-0.01
			(-0.025, 0.005
Obs.	27,178	27,178	27,17
R2	0.069	0.071	0.06
F	669.086	696.477	502.25
Adjusted R2	0.069	0.071	0.06

Términos de interacción

- Siempre mantén en mente : piensa cuidadosamente el **por qué** vas a incluir términos de interacción.
- ¿Por qué hincapié en pensar cuidadosamente qué términos de interacción agregar?
- Hay muchos términos de interacción diferentes que son tentadores.
- ¿El efecto de la capacitación laboral difiere entre géneros? ¿Entre carreras? ¿Entre diferentes edades?
- Acotar la especificación téorica al problema que queremos modelar.
- ¿Hay una razón teórica sólida como para esperar efectos distintos entre subgrupos?

Consideraciones del modelo de regresión lineal

- Supuestos del modelo de regresión lineal (comportamiento del término de error).
  - El término de error ε se distribuye como una distribución normal con media 0 y varinza constante $\varepsilon \sim N(0, \sigma_\varepsilon^2)$.
  - El valor esperado del error es independiente de las variables independientes $E[\varepsilon|x] = E[\varepsilon] = 0$. Supuesto de exogeneidad.
- Variables instrumentales para resolver endogeneidad.
- Estimación robusta de los errores.
- Ponderación muestral del conjunto de datos.
- Colinearidad.
- Regresión con términos de penalización

Consideraciones del modelo de regresión lineal : Causalidad

- ¿X causa a Y?
- ¿La política o programa X causa un efecto en la variable de interés Y?
- Sí lo hay, ¿Qué tan grande es?
- Para poder encontrar un efecto causal tenemos que seguir una **estrategia de identificación**.
- **Inferencia causal**.
- Enfoque experimental : RCT (Randomized Controlled Trial) $\rightarrow$ Construir un contrafactual. Grupos de control y grupos de tratamiento.
- Efecto causal medio : *Average treatment effect*.

Consideraciones del modelo de regresión lineal : Métodos no experimentales- Experimentos naturales

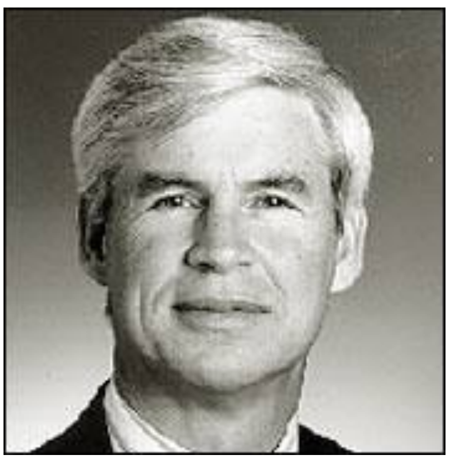
- Variables Instrumentales
- Efectos fijos y diferencias en diferencias
- Diseño de regresión discontinua
- Emparejamiento (Matching)
- Simulación

Revolución de credibilidad: Card, Angrist e Imbens. Nobel 2021.

Consideraciones del modelo de regresión lineal : Series de tiempo

- Correlaciones espurias.
- Series estacionarias vs Series no estacionarias.
- Pruebas de raíz unitaria.
- Cointegración.

Cointegración: Engle y Granger. Nobel 2003.




Regresiones para Crecimiento del RCA, Apariciones y Desapariciones

Con el objetivo de ofrecer evidencia empírica sobre la trayectoria de dependencia de la evolución industrial para el conjunto de Zonas Metropolitanas y El Salvador, se estimaron regresiones para tres variables explicaticas:

- **Crecimiento del RCA** : variable continua igual a la inversa hiperbólica de la tasa de crecimiento del RCA entre 2015 y 2020.
- **Aparición de industrias**: variable dicotómica que toma la etiqueta 1 si la industria tenía un $\mathrm{RCA} < 0.05$ en 2015 y un $\mathrm{RCA} > 0.2$ en 2020. En caso contrario, se asigna la etiqueta 0. La variable indica si la actividad se especializó en el periodo.
- **Desaparición de industrias** : variable dicotómica que toma la etiqueta 1 si la industria tenía un $\mathrm{RCA} < 0.2$ en 2015 y un $\mathrm{RCA} > 0.05$ en 2020. En caso contrario, se asigna la etiqueta 0. La variable indica si la actividad perdió especialización en el periodo.

Las regresiones toman la siguiente forma funcional:

$$y_{j,k,t_1} = \alpha_0 + \alpha_1 Densidad_{j,k,t_0} + \alpha_2 RCA_{j,k,t_0} + \delta_1 ZM + \delta_2 CIIU + \varepsilon_{j,k,t_1}$$

Donde el subíndice j corresponde a la actividad CIIU, k corresponde a la zona metropolitana, t_0 corresponde al año de inicio del periodo de estudio(2015), t_1 corresponde al año final del periodo de estudio (2020). La estimación controla por efectos fijos para zonas metropolitanas (ZM) y actividades CIIU (CIIU).

Para examinar el efecto de las métricas de viabilidad en la trayectoria de evolución industrial, a la forma funcional anterior agregamos como regresores las metricas de viabilidad:

- **Input Presence** : mide la presencia de actividades que producen insumos necesarios para las actividades.
- **Output Presence** : mide la presencia de actividades que compran productos de las actividades.
- **Input Presence Similarity** : mide la presencia de actividades que son similares respecto a la demanda de mismos insumos.
- **Output Presence Similarity** : mide la presencia de actividades similares respecto a si le venden sus producción a las mismas industrias.
- **Co-empleo** : mide la presencia de actividades que demandan los mismos tipos de ocupaciones.

Crecimiento del RCA

	Crecimiento RCA				
	(1)	(2)	(3)	(4)	
Density	0.057	0.074	0.023	0.012	-0.
	(0.152)	(0.182)	(0.189)	(0.172)	(0.1
RCA	-0.071***	-0.069***	-0.073***	-0.076***	-0.076
	(0.021)	(0.019)	(0.020)	(0.021)	(0.0
Input Presence		-0.035			
		(0.106)			
Input Presence Similarity			0.126		
			(0.334)		
Output Presence				0.086	
				(0.067)	
Output Presence Similarity					0.
					(0.2
Co-empleo					
Zona Metropolitana Fixed Effects	Yes	Yes	Yes	Yes	
Actividad CIIU Fixed Effects	Yes	Yes	Yes	Yes	
Obs.	1,488	1,488	1,488	1,488	1,
R2	0.167	0.167	0.167	0.169	0.
Within-R2	0.021	0.021	0.021	0.023	0.
Pseudo-R2	0.182	0.182	0.182	0.183	0.

```
estimated covariance matrix of moment conditions not positive-semidefinite.
                adjustment from Cameron, Gelbach & Miller (2011, JBES) applied.
                 model tests should be interpreted with caution.
```

Aparición de industrias

	Apariciones					
	(1)	(2)	(3)	(4)	(5)	
Density	0.056**	0.060**	0.062**	0.057**	0.050*	0.0
	(0.027)	(0.027)	(0.027)	(0.027)	(0.028)	(0.0
RCA	0.090	0.092	0.088	0.090	0.091	0.
	(0.230)	(0.230)	(0.230)	(0.230)	(0.230)	(0.2
Input Presence		-0.014				
		(0.010)				
Input Presence Similarity			-0.047			
			(0.033)			
Output Presence				-0.002		
				(0.006)		
Output Presence Similarity					0.035	
					(0.033)	
Co-empleo						0.10
						(0.0
Co-empleo Calificado						
Zona Metropolitana Fixed Effects	Yes	Yes	Yes	Yes	Yes	
Actividad CIIU Fixed Effects	Yes	Yes	Yes	Yes	Yes	
Obs.	3,640	3,640	3,640	3,640	3,640	3,
R2	0.092	0.092	0.092	0.092	0.092	0.
Within-R2	0.001	0.002	0.002	0.001	0.002	0.
Pseudo-R2	-0.033	-0.034	-0.034	-0.033	-0.034	-0.

Desaparición de industrias

	Desapariciones				
	(1)	(2)	(3)	(4)	(
Density	-0.128***	-0.122***	-0.087**	-0.133***	-0.122*
	(0.031)	(0.033)	(0.036)	(0.032)	(0.03
RCA	-0.004	-0.003	-0.002	-0.004	-0.0
	(0.004)	(0.004)	(0.004)	(0.004)	(0.00
Input Presence		-0.013			
		(0.018)			
Input Presence Similarity			-0.151**		
			(0.069)		
Output Presence				0.011	
				(0.016)	
Output Presence Similarity					-0.0
					(0.06
Co-empleo					
Co-empleo Calificado					
Zona Metropolitana Fixed Effects	Yes	Yes	Yes	Yes	Y
Actividad CIIU Fixed Effects	Yes	Yes	Yes	Yes	Y
Obs.	1,236	1,236	1,236	1,236	1,2
R2	0.257	0.257	0.260	0.257	0.2
Within-R2	0.026	0.027	0.031	0.027	0.0
Pseudo-R2	-0.164	-0.165	-0.167	-0.165	-0.1