

A Minimal Book Example

John Doe

2025-02-16

Contents

1	About	5
1.1	Usage	5
1.2	Render book	5
1.3	Preview book	6
2	Introducción	7
3	Análisis de regresión	9
3.1	Metodología clásica	9
3.2	Probabilidad condicional	11
3.3	Función de regresión poblacional	12
3.4	Función de regresión muestral	13
3.5	Objetivo del análisis de regresión	13
3.6	Conceptos estadísticos	13

Chapter 1

About

This is a *sample* book written in **Markdown**. You can use anything that Pandoc’s Markdown supports; for example, a math equation $a^2 + b^2 = c^2$.

1.1 Usage

Each **bookdown** chapter is an .Rmd file, and each .Rmd file can contain one (and only one) chapter. A chapter *must* start with a first-level heading: **# A good chapter**, and can contain one (and only one) first-level heading.

Use second-level and higher headings within chapters like: **## A short section** or **### An even shorter section**.

The `index.Rmd` file is required, and is also your first book chapter. It will be the homepage when you render the book.

1.2 Render book

You can render the HTML version of this example book without changing anything:

1. Find the **Build** pane in the RStudio IDE, and
2. Click on **Build Book**, then select your output format, or select “All formats” if you’d like to use multiple formats from the same book source files.

Or build the book from the R console:

```
bookdown::render_book()
```

To render this example to PDF as a `bookdown::pdf_book`, you'll need to install XeLaTeX. You are recommended to install TinyTeX (which includes XeLaTeX): <https://yihui.org/tinytex/>.

1.3 Preview book

As you work, you may start a local server to live preview this HTML book. This preview will update as you edit the book when you save individual .Rmd files. You can start the server in a work session by using the RStudio add-in “Preview book”, or from the R console:

```
bookdown::serve_book()
```

Chapter 2

Introducción

Chapter 3

Análisis de regresión

Econometría \implies Medición económica.

3.1 Metodología clásica

1. Planteamiento de teoría (hipótesis)
2. Especificación del modelo matemático
3. Especificación del modelo econométrico
4. Obtención de datos
5. Estimación de parámetros del modelo
6. Pruebas de hipótesis
7. Pronóstico (predicción)
8. Modelo para fines de control/política

Ej. Función consumo keynesiana:

$$c = \alpha + \beta y \quad \forall 0 < \beta < 1$$

Las relaciones entre variables económicas son **inexactas**, dada la injerencia de otras variables:

$$c = \alpha + \beta y + u$$

Modelo econométrico

$$u = \text{Error (perturbación estocástica)}$$

Variable aleatoria con propiedades probabilísticas

u incluye todos los factores que afectan **consumo** pero no están en la ecuación: tamaño de familia, edades, etc.

El modelo requiere ser estimado: obtener valores α y β a partir de **datos**.

Ej. Gasto en consumo personal

- Regresión: técnica estadística para el estudio de una **variable dependiente** que está en función de una o más **variables independientes**.
- Usando los datos de consumo y PIB de BUA se obtiene:

$$\hat{c} = -231.8 + 0.7194y$$

Donde $\alpha = -231.8$, $\beta = 0.7194$, $\hat{c} = \text{Consumo (estimado)}$, $y = \text{PIB}$.

La interpretación consiste en que un incremento de tasa en el ingreso incrementa (en promedio) el consumo en 0.72 USD.

Se debe probar si los valores estimados:

1. Son estadísticamente significativos ($\alpha, \beta \neq 0$)
2. Confirman la teoría (hipótesis) que está siendo probada ($0 < \beta < 1$)

Si el modelo confirma la teoría (hipótesis), se pueden **pronosticar** valores futuros de la variable dependiente.

Ej. Suponer un PIB esperado para 1995 de 6,000 mmd, ¿cuál es el pronóstico de consumo?

$$\hat{c} = -231.8 + 0.7194(6,000) = 4,084.6$$

Suponer que el gobierno considera que un gasto de \$4,000 mmd mantendrá la tasa de desempleo en 6.5%. ¿Cuál nivel de ingreso garantizará esta meta de consumo?

$$\begin{aligned}
\hat{c} &= -231.8 + 0.7194y \\
4,000 &= -231.8 + 0.7194y \\
0.7194y &= 4,000 + 231.8 \\
y &= \frac{4,231.8}{0.7194} \\
y^* &= 5,882.40
\end{aligned}$$

Un modelo estimado puede ser usado para fines de control o de política económica (fiscal y monetaria).

3.2 Probabilidad condicional

Suponer un país con una población de 60 familias. Se estudia el gasto en consumo familiar semanal (y) y el ingreso familiar semanal (x).

Se presenta la distribución del gasto en consumo (y) correspondiente a un ingreso fijo (x): la distribución condicional de y dada x .

Para $P(y|x = 80)$:

$$\begin{aligned}
P(y = 55|x = 80) &= \frac{1}{5} \\
P(y = 60|x = 80) &= \frac{1}{5} \\
P(y = 65|x = 80) &= \frac{1}{5} \\
P(y = 70|x = 80) &= \frac{1}{5} \\
P(y = 75|x = 80) &= \frac{1}{5}
\end{aligned}$$

Para cada distribución de probabilidad condicional de y , calculamos su **media (media condicional)**:

$$E(y|x = 80) = 55 \left(\frac{1}{5}\right) + 60 \left(\frac{1}{5}\right) + 65 \left(\frac{1}{5}\right) + 70 \left(\frac{1}{5}\right) + 75 \left(\frac{1}{5}\right) = 65$$

$$\mu_{y|x=100} = \frac{\sum_y y_i}{n} = \frac{462}{6} = 77$$

3.3 Función de regresión poblacional

Lugar geométrico de las medias condicionales de la variable dependiente para valores fijos de la variable independiente.

Se puede deducir que: $E(y|x_i) = f(x_i)$.

$E(y|x_i) \rightarrow$ Función de Regresión Poblacional (FRP)

Forma funcional de la FRP:

$$E(y|x_i) = \alpha + \beta x_i \rightarrow \text{Ecuación de recta}$$

$y = ax + b$

$$\begin{array}{ll} \alpha, \beta & \rightarrow \text{Coeficientes de regresión} \\ \alpha & \rightarrow \text{Intercepto} \\ \beta & \rightarrow \text{Coeficiente de la pendiente} \end{array}$$

Objetivo: Estimar α y β con base en observaciones de x y y .

Esta desviación de un y_i alrededor de su valor esperado se expresa como:

$$\begin{aligned} u_i &= y_i - E(y|x_i) \\ y_i &= E(y|x_i) + u_i \end{aligned}$$

Donde u_i es el término de error estocástico, que representa todas las variables omitidas que puedan afectar y , pero no están incluidas en el modelo de regresión.

$$\begin{aligned} y_i &= \alpha + \beta x_i + u_i \quad \therefore E(y|x_i) = \alpha + \beta x_i \\ \therefore y_i &= \alpha + \beta x_i + u_i \rightarrow \text{Ecuación de regresión} \end{aligned}$$

Adicionalmente, si se toma $E(\cdot)$ a y_i , tenemos que:

$$\begin{aligned} y_i &= E(y|x_i) + u_i \\ E(y|x_i) &= E[E(y|x_i)] + E(u_i|x_i) \\ &= E(y|x_i) + E(u_i|x_i) \quad \therefore \\ E(u_i|x_i) &= E(y|x_i) - E(y|x_i) \quad \therefore \\ E(u_i|x_i) &= 0 \end{aligned}$$

3.4 Función de regresión muestral

A diferencia del caso anterior, en realidad se trabaja con muestras. Se estima la FRP con base en información **muestral**. Suponer que se extrae una muestra de la población de 60 familias donde se estudia el gasto en consumo y el ingreso familiar. Muestra aleatoria de 10 observaciones.

[IMAGEN]

La función de regresión muestral (FRM) puede escribirse como:

$$\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i$$

Donde: - $\hat{y}_i \equiv y$ estimada - $\hat{\alpha}_i \equiv$ estimador de α - $\hat{\beta}_i \equiv$ estimador de β

Se estiman los parámetros poblacionales a partir de información muestral.

Estimador	→	Fórmula
Estimado	→	Valor numérico

FRM (forma estocástica): $y_i = \hat{\alpha} + \hat{\beta}x_i + \hat{u}_i$. Por lo tanto, \hat{u}_i es el residual muestral (estimado de u_i).

3.5 Objetivo del análisis de regresión

Estimar FRP, $y_i = \alpha + \beta x_i + u_i$, a partir de estimar FRM, $y_i = \hat{\alpha} + \hat{\beta}x_i + \hat{u}_i$.

[IMAGEN]

3.6 Conceptos estadísticos

3.6.1 Sumatoria y multiplicatoria

$$\sum_{i=1}^n x_i = x_1 + x_2 + \dots + x_n$$

Propiedades del operador de sumatoria (\sum):

1. $\sum_{i=1}^n k = nk \quad \therefore \quad k \equiv \text{Constante.}$
2. $\sum_{i=1}^n kx_i = k \sum_{i=1}^n x_i \quad \therefore \quad k \equiv \text{Constante.}$

3. $\sum_{i=1}^n (a + bx_i) = \sum_{i=1}^n a + \sum_{i=1}^n bx_i = na + b \sum_{i=1}^n x_i \quad \forall \quad a, b \equiv \text{Constante.}$
4. $\sum_{i=1}^n (x_i + y_i) = \sum_{i=1}^n x_i + \sum_{i=1}^n y_i.$

Se pueden tener sumatorias múltiples:

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^m &= \sum_{i=1}^n (x_{i1} + x_{i2} + \dots + x_{im}) \\ &= (x_{11} + x_{12} + \dots + x_{1m}) + (x_{21} + x_{22} + \dots + x_{2m}) + \dots + (x_{n1} + x_{n2} + \dots + x_{nm}) \end{aligned}$$

Propiedades del operador de sumatoria ($\sum \sum$):

1. $\sum_{i=1}^n \sum_{j=1}^m x_{ij} = \sum_{j=1}^m \sum_{i=1}^n x_{ij} \rightarrow \text{Intercambiable}$
2. $\sum_{i=1}^n \sum_{j=1}^m x_i y_j = \sum_{i=1}^n x_i \sum_{j=1}^m y_j$
3. $\sum_{i=1}^n \sum_{j=1}^m (x_{ij} + y_{ij}) = \sum_{i=1}^n \sum_{j=1}^m x_{ij} + \sum_{i=1}^n \sum_{j=1}^m y_{ij}$

Adicionalmente, el operador multiplicatoria:

$$\prod_{i=1}^n x_i = x_1 \cdot x_2 \cdot \dots \cdot x_n$$

3.6.2 Valor esperado

El valor esperado de una variable aleatoria discreta se define como:

$$E(x) = \sum_x x f(x)$$

$$\begin{aligned} \therefore x &\rightarrow \text{Valores de la variable aleatoria discreta} \\ f(x) &\rightarrow \text{FDP (discreta) de } x \end{aligned}$$

En términos poblacionales:

$$E(x) = \mu_x \quad \text{La media de la variable aleatoria discreta}$$

Propiedades del valor esperado $E(\cdot)$:

1. $E(b) = b \quad \therefore \quad b \equiv \text{Constante}$
2. $E(ax + b) = E(ax) + E(b) = aE(x) + b \quad \forall \quad a, b \text{ Constantes}$
3. $E(xy) = E(x)E(y) \quad \forall \quad x, y \rightarrow \text{Variables aleatorias independientes}$

3.6.3 Varianza