

UNIVERSIDAD TORCUATO DI TELLA

# Trabajo Práctico en R

Laboratorio para el Análisis de Datos Económicos y Financieros

Marco Luciano Bunino  
Rafael Catalán

## **Docentes:**

Prof. Ian Evangelos Bounos  
Prof. Martín Escobar

Octubre 2025

## Ejercicio de Análisis de Datos

Para la realización de este ejercicio elegimos un dataset que encontramos en un repositorio de GitHub que contiene las estadísticas de todos los jugadores NBA de la temporada 2018-19. A partir de estos datos nos hicimos tres preguntas que nos parecía interesante responder:

1. ¿Cómo varía la efectividad de tiros libres (Porc\_TL) según la posición del jugador?
2. ¿Cuál es la relación entre minutos y puntos, y cómo cambia por posición?
3. ¿Cómo se relaciona la posición con el promedio de triples anotados por partido?

Para ello, primero renombramos las variables del dataset por comodidad. Después, filtramos a los jugadores por posiciones puras, siendo:

- C: pivot.
- PF: ala-pivot.
- SF: alero.
- SG: escolta.
- PG: base.

Y eliminamos las filas con N/A en las variables clave que nos interesa analizar.

### Efectividad de tiros libres por posición

Se calculó el promedio, mediana y desviación estándar del porcentaje de tiros libres (Porc\_TL) agrupando por posición. El gráfico de barras muestra que las posiciones perimetrales —bases (PG) y escoltas (SG)— presentan una mayor efectividad promedio, mientras que los pivots (C) registran los porcentajes más bajos. Esto coincide con el perfil de juego de cada posición: los jugadores interiores suelen concentrarse en la zona pintada y no destacan por su tiro libre.

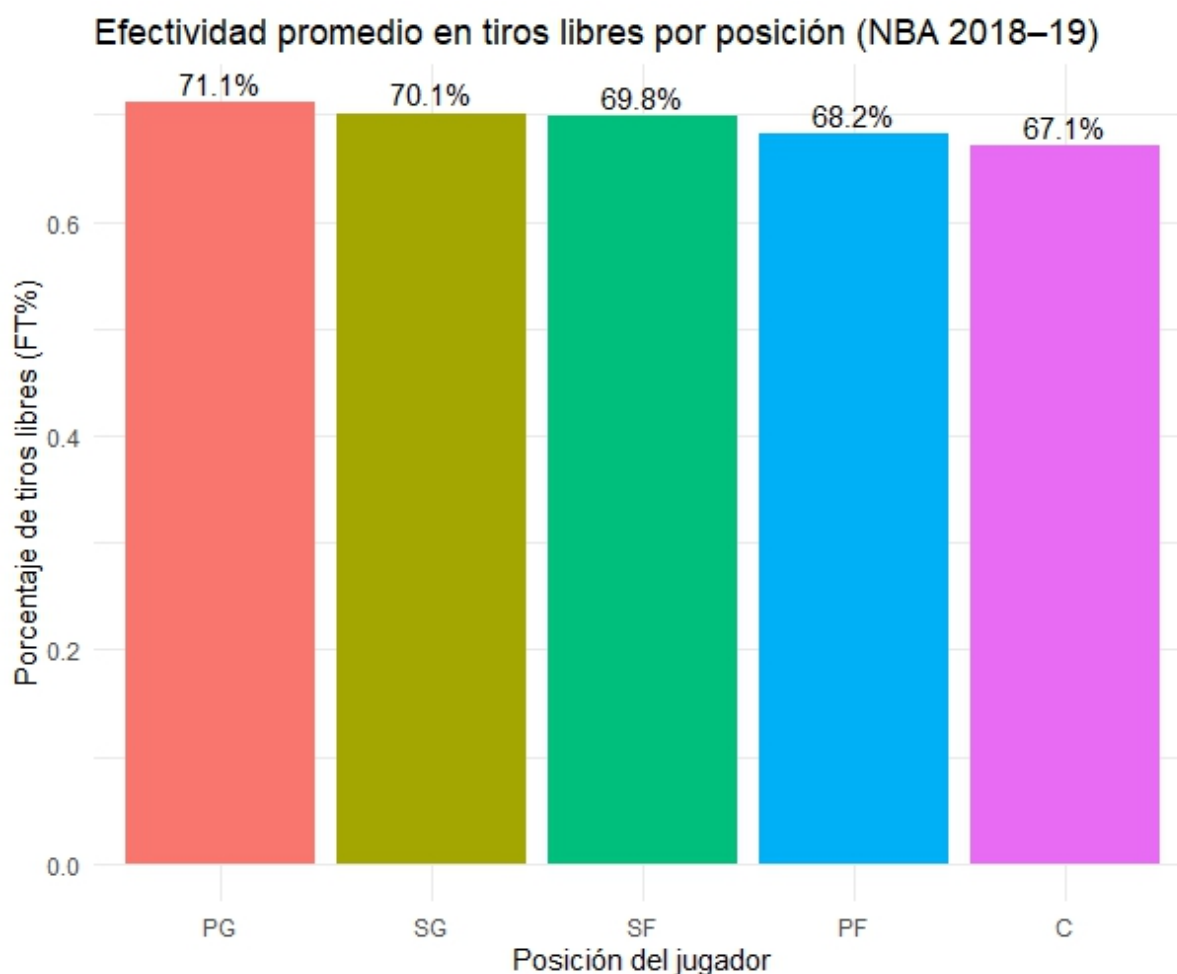


Figura 1: Gráfico: efectividad de tiros libres por posición.

## Relación entre minutos y puntos (por posición)

La correlación entre minutos jugados y puntos por partido resultó ser positiva, indicando que quienes juegan más minutos tienden a anotar más puntos. El gráfico de dispersión con ajuste lineal muestra que esta relación varía según la posición: las pendientes son más pronunciadas en PG y SG, lo que sugiere que su "productividad ofensiva", por así decirlo, crece más rápido con el tiempo en cancha que en otras posiciones. También, puede interpretarse que juegan más minutos aquellos jugadores que son más efectivos a la hora de anotar, y no que necesariamente el jugar más minutos produce que los jugadores anoten más (se puede jugar mucho y anotar pocos puntos debido a la poca efectividad, lo que provoca que estos jugadores terminen estando menos en cancha).

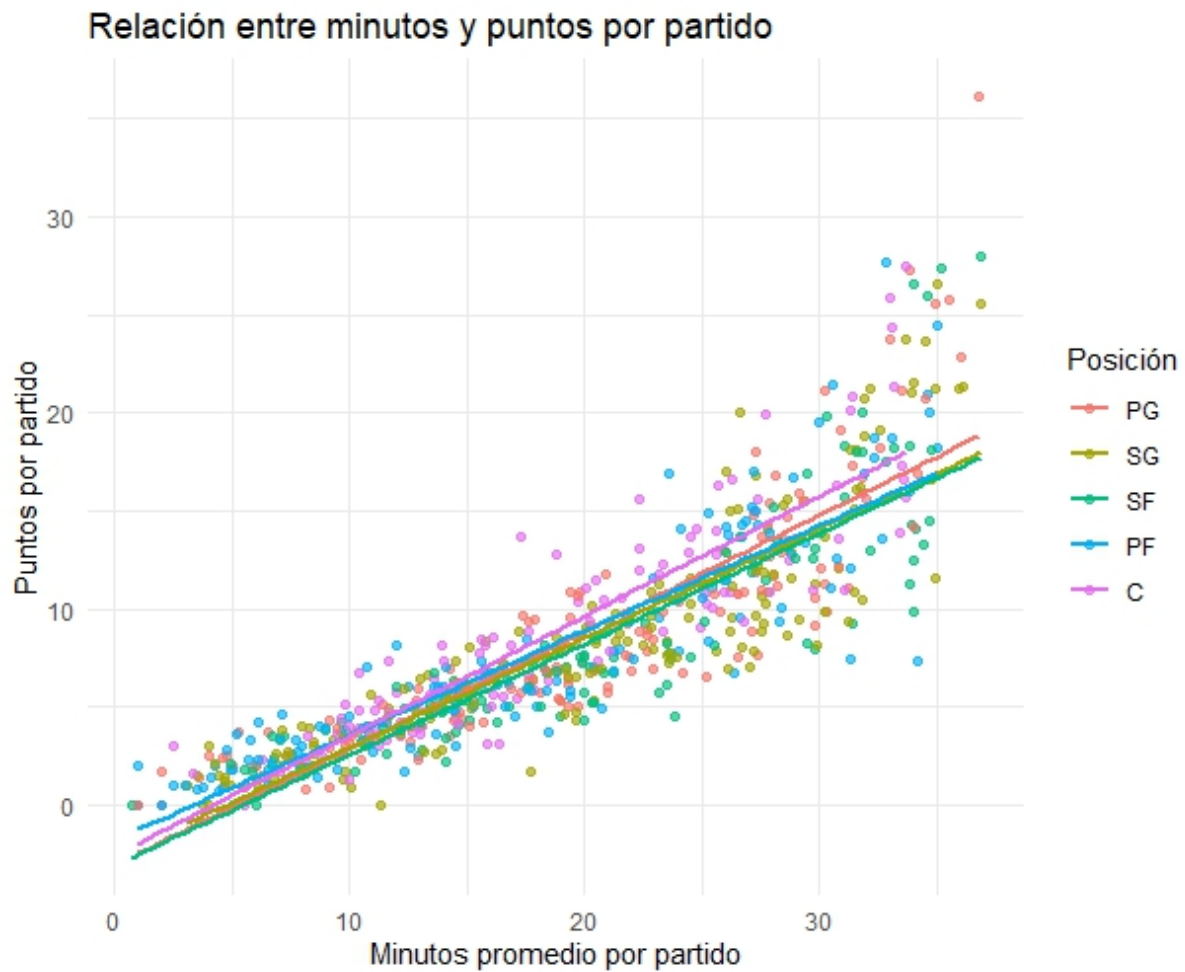


Figura 2: Gráfico: correlación entre minutos y puntos por posición.

### Promedio de triples por posición

Al agrupar por posición y calcular el promedio de triples (3P), se observa que los bases y escoltas son quienes más lanzan y aciertan desde larga distancia, seguidos por los aleros (SF). Los pivots y ala-pivots (PF, C) muestran valores considerablemente menores, reflejando su rol más interno dentro del esquema ofensivo.

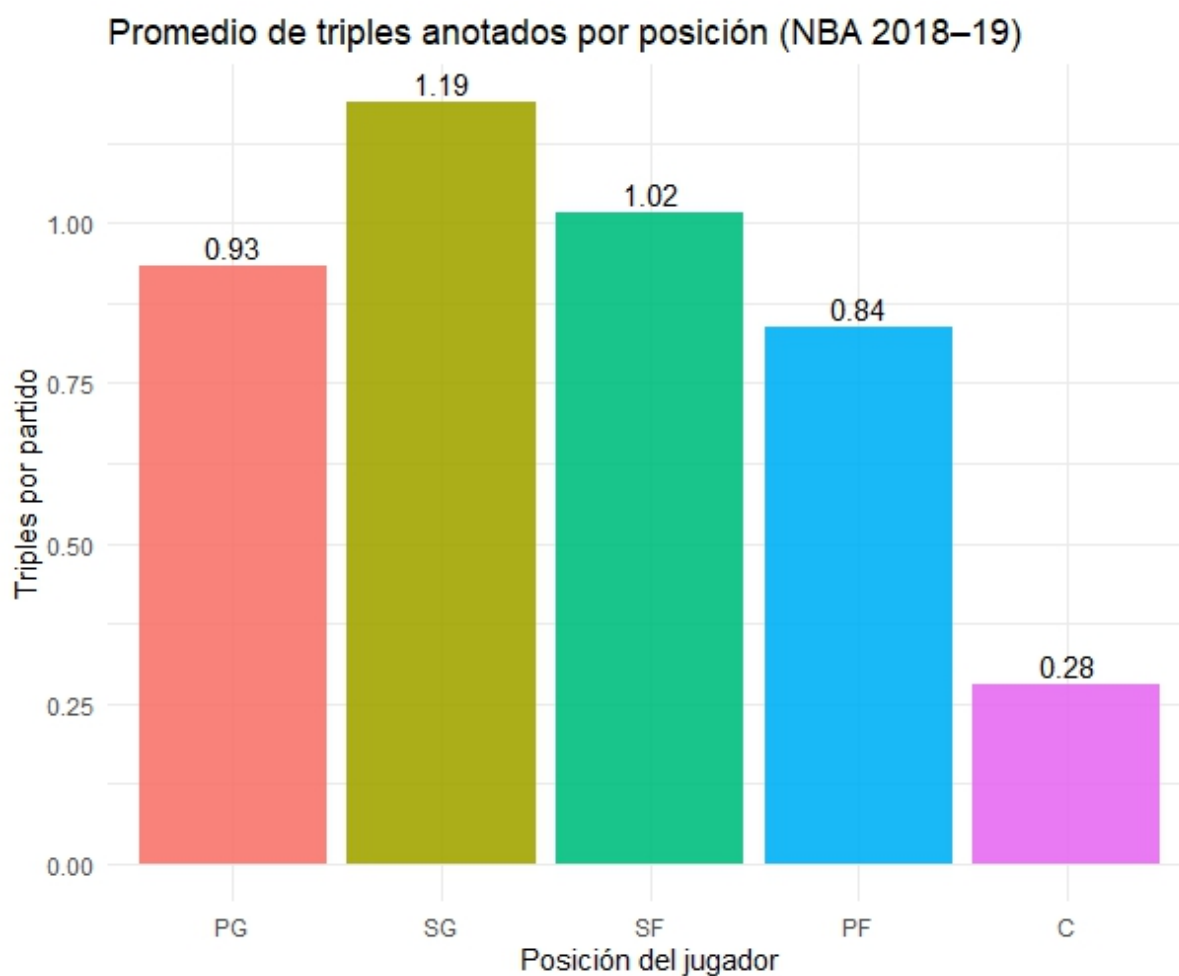


Figura 3: Gráfico: promedio de triples por posición.

## Conclusiones del ejercicio

Los resultados muestran las diferencias en los estilos de juego entre posiciones de los jugadores de baloncesto en la NBA y en general en cualquier otra liga:

- Las posiciones de perímetro (PG, SG) presentan mayor efectividad en tiros libres y triples.
- Los pivots (C) son menos eficientes en el tiro libre.
- Existe una relación positiva entre minutos y puntos, consistente con que mayor participación implica mayor producción ofensiva (o, dada la otra interpretación, que juegan más minutos aquellos jugadores que son más efectivos a la hora de anotar).

# Ejercicio: Análisis econométrico con datos de gapminder

## Parte 1: Ingreso por persona

En este ejercicio se analizó la evolución del ingreso per cápita en Argentina utilizando el dataset `gapminder`, que contiene información económica y demográfica de distintos países a lo largo del tiempo.

Primero se graficó la serie temporal de la variable `income_per_person`. El gráfico mostró una tendencia creciente de largo plazo, aunque con fluctuaciones intermedias asociadas a ciclos económicos y crisis puntuales. En términos generales, el ingreso per cápita de Argentina creció sostenidamente, pero con etapas de estancamiento o caída temporal.



Figura 4: Gráfico: ingreso por persona en Argentina.

Luego, se entrenaron tres modelos para explicar la evolución temporal del ingreso:

1. un modelo lineal simple,
2. un modelo polinómico de grado 2, y
3. un modelo polinómico de grado 10.

Los modelos se estimaron utilizando todos los años excepto los últimos diez, que se reservaron para testeo. Las predicciones mostraron que el modelo lineal describe bien la

tendencia general, mientras que el polinómico de grado 2 captura mejor la curvatura de la serie. El modelo de grado 10, en cambio, se ajusta casi perfectamente al conjunto de entrenamiento pero falla al predecir los valores recientes, evidenciando sobreajuste.

Las métricas de error (RMSE y MAE) confirmaron que los modelos lineal y de grado 2 presentan mejor capacidad de generalización, mientras que el modelo de grado 10 tiene un error significativamente mayor sobre el conjunto de testeo. Por lo que modelos simples (lineales o de bajo grado) logran representar adecuadamente la tendencia del ingreso en Argentina sin sobreajustar los datos.

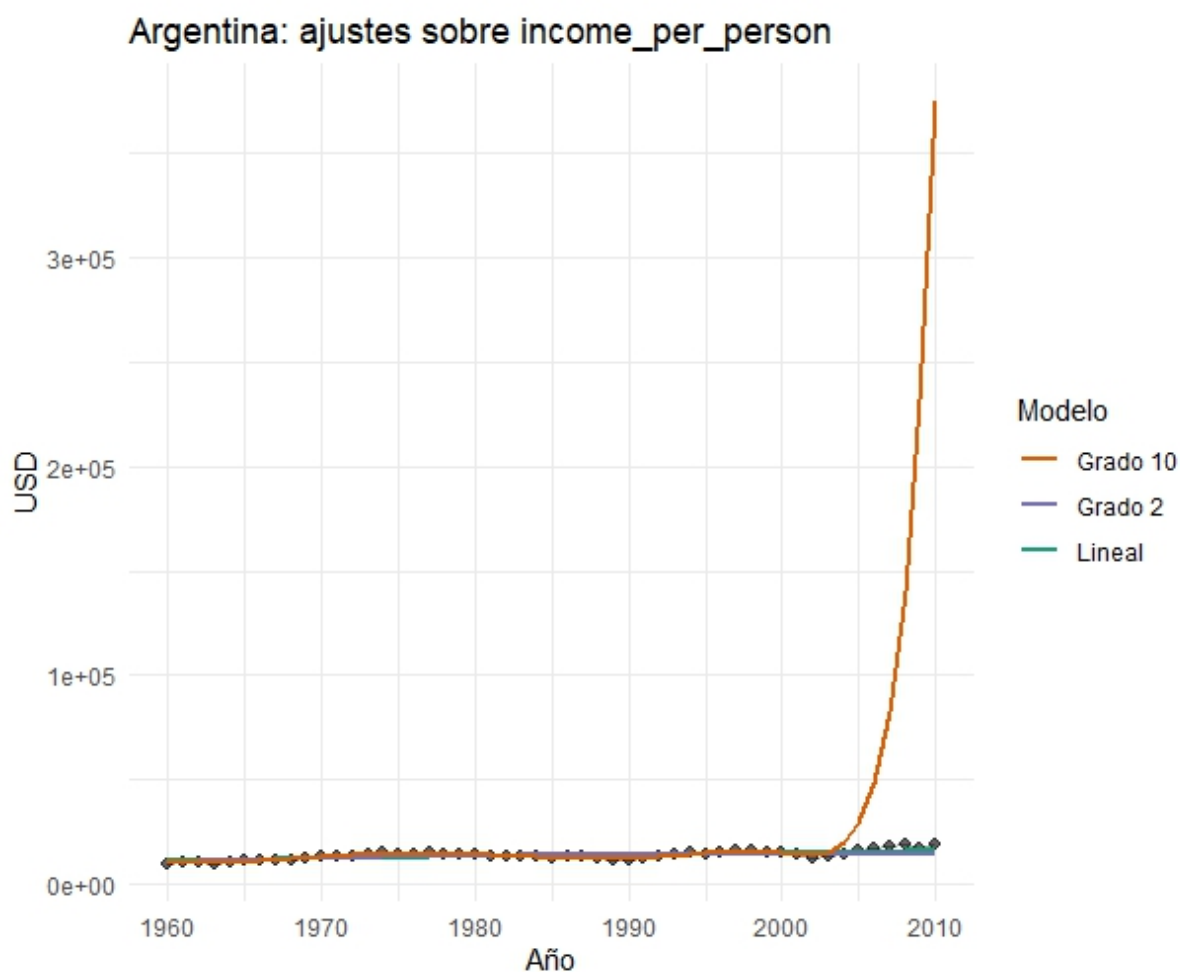


Figura 5: Gráfico: Ajustes sobre `income_per_person`.

Para extender el análisis, se seleccionaron cuatro países sudamericanos además de Argentina: Brasil, Chile, Uruguay y Perú. Con ellos se construyeron dos matrices de correlaciones:

#### 1. Correlación entre niveles de ingreso

Se calculó la matriz de correlación entre los valores de `income_per_person` de los cinco países. Los resultados mostraron correlaciones positivas y elevadas, lo que sugiere que las economías de la región tienden a evolucionar en la misma dirección

en el largo plazo.

|         | Brazil | Chile | Peru  | Uruguay |
|---------|--------|-------|-------|---------|
| Brazil  | 1,000  | 0,772 | 0,565 | 0,871   |
| Chile   | 0,772  | 1,000 | 0,549 | 0,941   |
| Peru    | 0,565  | 0,549 | 1,000 | 0,577   |
| Uruguay | 0,871  | 0,941 | 0,577 | 1,000   |

## 2. Correlación entre variaciones porcentuales anuales

Se calcularon las variaciones porcentuales anuales del ingreso ( $\Delta Y/Y$ ) y la matriz de correlación correspondiente. En este caso, las correlaciones resultaron más débiles y variables, indicando que aunque los países comparten una tendencia de crecimiento de largo plazo, sus ciclos económicos de corto plazo difieren significativamente.

|         | Brazil | Chile | Peru  | Uruguay |
|---------|--------|-------|-------|---------|
| Brazil  | 1,000  | 0,006 | 0,411 | 0,278   |
| Chile   | 0,006  | 1,000 | 0,043 | 0,366   |
| Peru    | 0,411  | 0,043 | 1,000 | 0,413   |
| Uruguay | 0,278  | 0,366 | 0,413 | 1,000   |

## Parte 2: Esperanza de vida y género

Se graficó la variable `life_expectancy` frente a `life_expectancy_female` para todos los países en el año 2004.

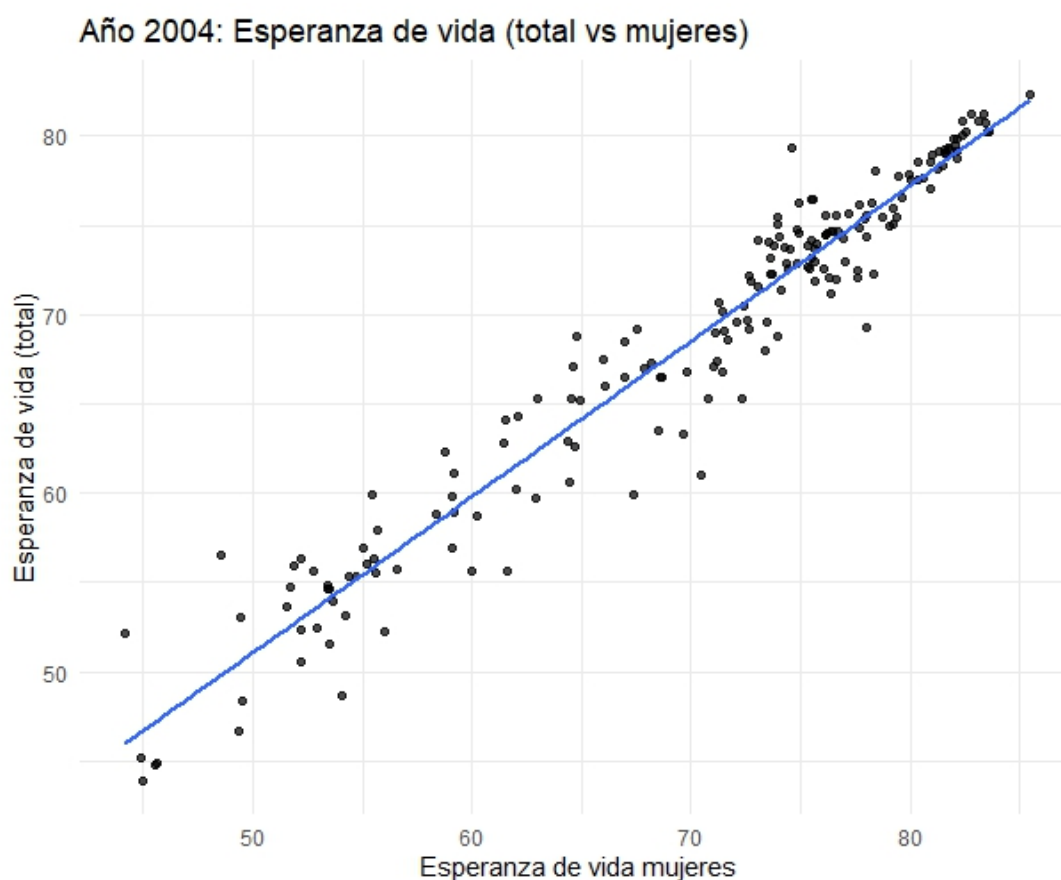


Figura 6: Gráfico: Esperanza de vida (total vs. mujeres) para el año 2004.



El gráfico muestra una relación lineal positiva muy marcada: los países en los que las mujeres viven más años presentan también una mayor esperanza de vida total. La dispersión es baja, indicando una fuerte asociación entre ambas variables. Siguiendo con el sexto inciso, se estimó una regresión lineal simple del tipo:

$$\text{life\_expectancy} = \beta_0 + \beta_1 \text{life\_expectancy\_female} + \varepsilon$$

El coeficiente  $\beta_1$  resultó cercano a 1, y el  $R^2$  fue muy alto (casi igual a 1), lo que implica que la esperanza de vida femenina explica casi toda la variación de la esperanza de vida total entre países. Esto es lógico, ya que la esperanza de vida total es una media ponderada entre hombres y mujeres. Luego, siguiendo con el contraste de hipótesis: utilizamos un test t para contrastar la hipótesis nula  $H_0$  frente a la alternativa  $H_1$ , siendo ambas:

$$H_0 : \text{life\_expectancy\_female} = \text{life\_expectancy}$$

$$H_1 : \text{life\_expectancy\_female} > \text{life\_expectancy}$$

El test t para muestras pareadas arrojó un p-valor muy pequeño, permitiendo rechazar  $H_0$ . Por tanto, la evidencia empírica muestra que la esperanza de vida de las mujeres es significativamente mayor que la esperanza de vida promedio de la población en todos los países del conjunto. Ahora, se extendió el modelo anterior incorporando el ingreso promedio por persona (en logaritmos):

$$\text{life\_expectancy} = \beta_0 + \beta_1 \text{life\_expectancy\_female} + \beta_2 \log(\text{income\_per\_person}) + \varepsilon$$

Al incluir el ingreso, el  $R^2$  aumentó levemente, indicando que esta variable agrega poder explicativo. El coeficiente de  $\log\_income$  fue positivo y significativo, lo que sugiere que a mayor nivel de ingreso, mayor esperanza de vida. Sin embargo, el efecto principal sigue siendo el de  $\text{life\_expectancy\_female}$ , lo que refuerza que la diferencia de género es un determinante clave de la longevidad. En conclusión para este inciso, si vale la pena incluir el ingreso como variable de control, aunque no altera el patrón principal.

Finalmente, se probó un modelo alternativo incorporando tres covariables adicionales:  $\log\_income$ ,  $fertility$  y  $child\_mortality$ . El modelo estimado fue:

$$\text{life\_expectancy} = \beta_0 + \beta_1 \log(\text{income\_per\_person}) + \beta_2 fertility + \beta_3 child\_mortality + \varepsilon$$

¿Por qué? Empezar con variables que tienen sentido teórico (ingreso, salud reproductiva, mortalidad infantil, carga sanitaria, tamaño poblacional, contaminación) y garantizar que haya datos para estimar el modelo. Luego, transformamos las variables con distribuciones sesgadas (por ejemplo, ingreso) y estandarizo. En el código se transformó como  $\log\_income = \log(\text{income\_per\_person})$ . Esto porque  $income$  es típicamente muy asimétrico, y usar  $\log$  lo aproxima a normalidad y hace a los coeficientes interpretables en elasticidad aproximadas. Ingreso (niveles por país) explica recursos en salud; fertilidad está relacionada con la estructura etaria y las condiciones sanitarias; mortalidad infantil resume condiciones sanitarias y acceso a cuidados. Las tres capturan distintos canales que afectan a  $\text{life\_expectancy}$ .

Los resultados mostraron que:

- El ingreso per cápita se asocia positivamente con la esperanza de vida.
- La fertilidad y la mortalidad infantil tienen efectos negativos, coherentes con la literatura demográfica.

Este modelo logra un alto  $R^2$  y resume adecuadamente las condiciones estructurales que determinan la longevidad promedio de cada país.

## Ejercicio Simulación 1: Demanda con preferencias Cobb-Douglas

Se consideraron dos bienes  $x_1$  y  $x_2$  con precios  $p_1 = 2$  y  $p_2 = 3$ . Las preferencias son Cobb-Douglas con parámetros  $\alpha_1 = 0,6, \alpha_2 = 0,4$  (tal que  $\alpha_1 + \alpha_2 = 1$ ). Se simularon  $n = 10000$  hogares. // Se modeló el ingreso individual  $Y$  como  $Y = s \cdot Z$ , con  $Z \sim \chi_k^2$ . Escogimos una escala  $s = 1000$  y probamos varios grados de libertad  $k$ . Recordemos las propiedades  $E[Y] = sk$  y  $SD(Y) = s\sqrt{2k}$ . Por tanto, la dispersión relativa  $SD/E = \sqrt{2/k}$  disminuye con  $k$ . Para documentar empíricamente esta elección se compararon varios  $k$  (ej. 1, 2, 4, 6, 10 y 20) y se corroboró que:

- $k$  aumenta la media y la varianza absoluta de  $Y$  (lineal con  $k$  para la media, proporcional a  $k$  para la varianza).
- La dispersión relativa cae con  $k$ .

En el ejercicio se adoptó  $k = 6$  porque produce una media  $E(Y) \approx 6$  (para  $s = 1000$ ) y una dispersión relativa moderada ( $SD/E \approx 0,58$ ).

### 2) Demanda óptima Cobb-Douglas

La demanda óptima, resolviendo el problema del hogar, es:

$$x_1^* = \frac{\alpha_1 Y}{p_1}, \quad x_2^* = \frac{\alpha_2 Y}{p_2}$$

y la utilidad indirecta:

$$U^* = (x_1^*)^{\alpha_1} (x_2^*)^{\alpha_2}.$$

La función `demanda_cd()` implementa exactamente estas expresiones.

### 3) Simulación base (n= 10 000)

Se simularon 10 000 ingresos con  $k = 6$  y  $s = 1000$  y se calculó la demanda óptima para cada hogar. Se presentan histogramas de  $x_1^*, x_2^*$  y  $U^*$ , y un resumen con medias y cuartiles.

- $x_1^*$  y  $x_2^*$  reflejan la heterogeneidad en ingresos.
- $U^*$  está fuertemente correlacionada con  $Y$ , por la estructura multiplicativa de la utilidad.

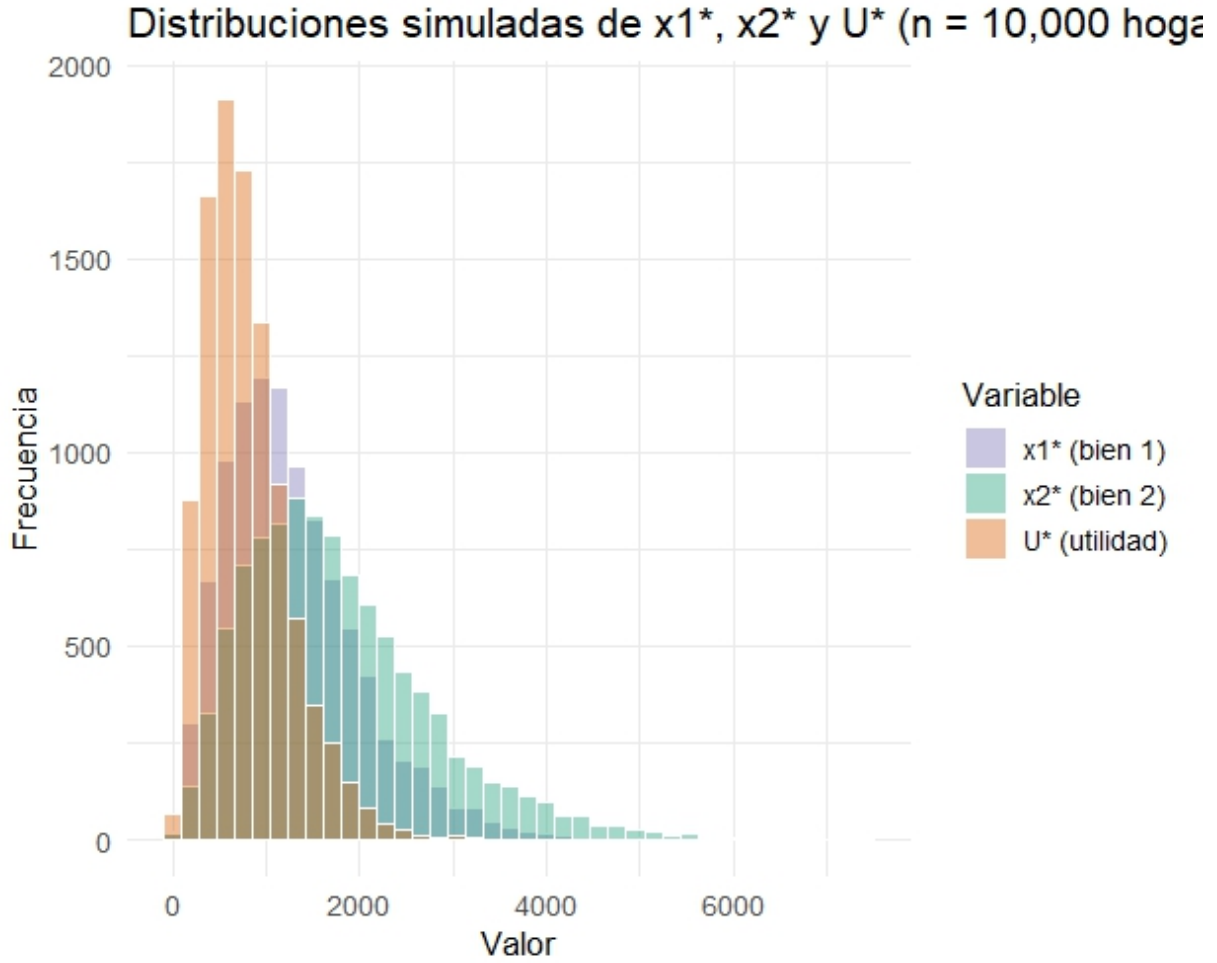


Figura 7: Gráfico: distribuciones simuladas con  $n = 10.000$  hogares.

Cuadro 1: Resumen de medias y cuartiles simulados ( $n = 10,000$  hogares)

| Variable | Media   | Q1      | Mediana | Q3      |
|----------|---------|---------|---------|---------|
| $x_1^*$  | 1802.00 | 1030.00 | 1601.00 | 2352.00 |
| $x_2^*$  | 801.00  | 458.00  | 712.00  | 1046.00 |
| $U^*$    | 1303.00 | 745.00  | 1158.00 | 1701.00 |

#### 4) Probabilidad de bajo consumo

Se definió un umbral  $c$  (en el código  $c = 50$ ) y se calculó la probabilidad empírica de que  $x_j^* < c$  para  $j = 1, 2$ . En el script se muestran los valores `pb_x1_base` y `pb_x2_base`. Antes del shock, la probabilidad de consumir menos de 50 unidades del bien 1 es 0.0001, y del bien 2 es 0.0009.

#### 5) Shock de precios

Se aumentó  $p_1$  en 20 % ( $p'_1 = 1,2p_1$ ) y se recalculó la demanda manteniendo  $Y$  y  $p_2$  constantes. Como resultado:

- $x_1$  disminuye en promedio (porque  $x_1^* \propto \frac{1}{p_1}$ ),

- la masa de hogares con  $x_1^* < c$  aumenta (se reporta `pb_x1_post;pb_x1_base`),
- la utilidad promedio disminuye.

Los histogramas pre/post muestran el desplazamiento de la masa hacia la izquierda para  $x_1^*$ .

## 6) Visualización comparada

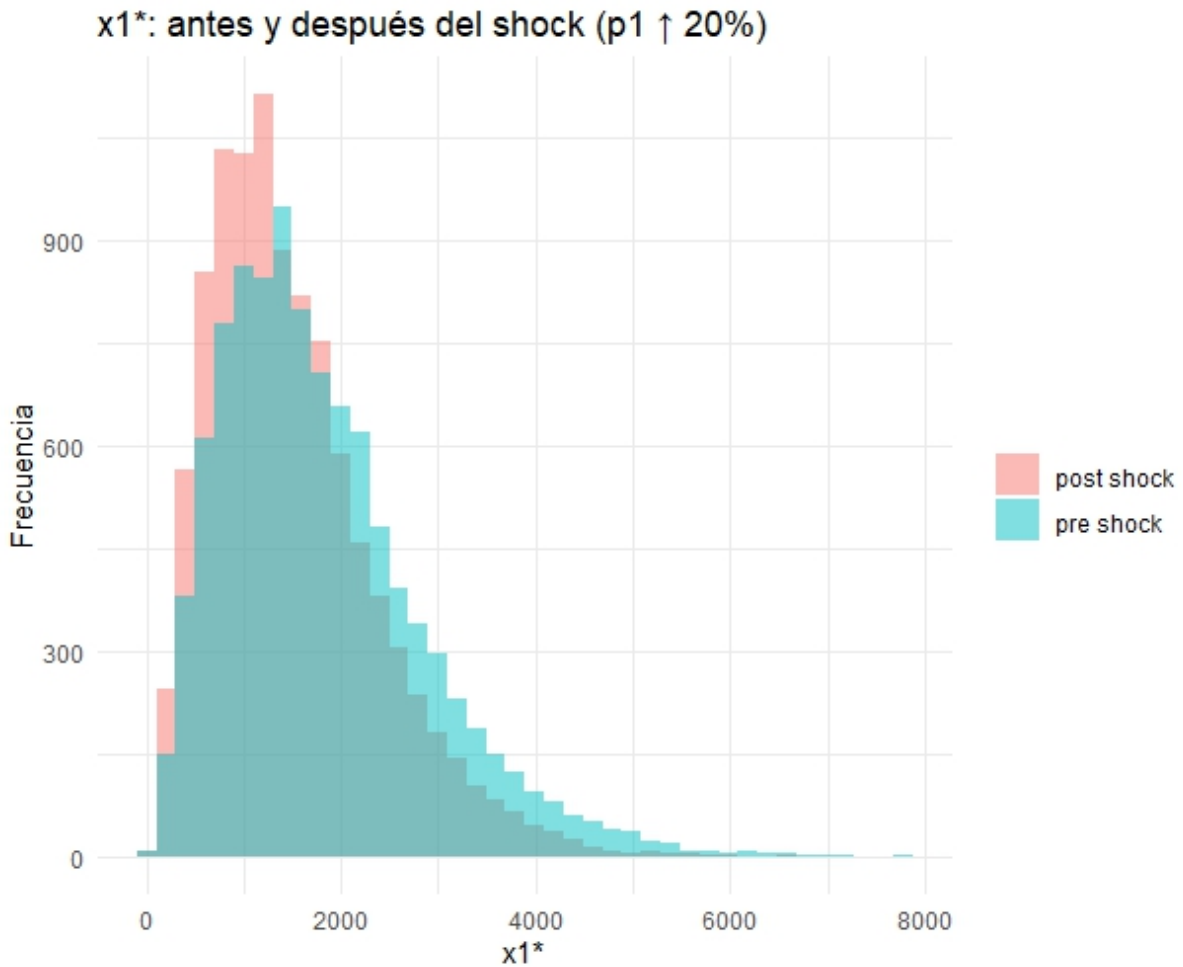


Figura 8: Gráfico:  $x_1^*$  pre y post shock.

## 7) Heterogeneidad en preferencias

Se introdujo heterogeneidad en  $\alpha_1$  tomando  $\alpha_1 \sim \text{Beta}(a = 8, b = 6)$  y  $\alpha_2 = 1 - \alpha_1$ . Con esto:

- la dispersión de  $x_1^*$  aumenta (más variabilidad entre hogares),
- la sensibilidad al shock depende de  $\alpha_1$ : hogares con  $\alpha_1$  alto pierden más consumo absoluto de  $x_1$  y ven una mayor caída en su utilidad,
- la probabilidad de bajo consumo post-shock se incrementa especialmente para los hogares con  $\alpha_1$  alto.

Se comparan `pb_x1_base_het` vs `pb_x1_post_het` y `delta_U_het` para cuantificar este efecto.