

# Ventajas sociales de la Probabilidad y la Estadística

Semana de la Ciencia. Facultad de Ciencias  
Norberto Corral Blanco – Carlos Carleos Artime

Facultad de Ciencias

Oviedo, 11 de noviembre de 2022



# Todo es incertidumbre



- Según el Principio de Indeterminación de Heisenberg no se pueden determinar, simultáneamente y con precisión arbitraria, la posición y la cantidad de movimiento de un objeto dado.
- Albert Einstein mostró su desacuerdo con este principio con la famosa frase: *Dios no juega a los dados con el Universo.*

- 1 El azar
  - Cuota de mercado
  - El problema de las tres puertas
  - Cómo engañar sin mentir
- 2 Las estadísticas
  - Investigadores en España
  - El problema de la inmigración
  - Víctimas mortales por violencia de género
- 3 Predecir, ¿ciencia o brujería?
  - Asociación no es lo mismo que causa
  - Potencia de una lente intra-ocular

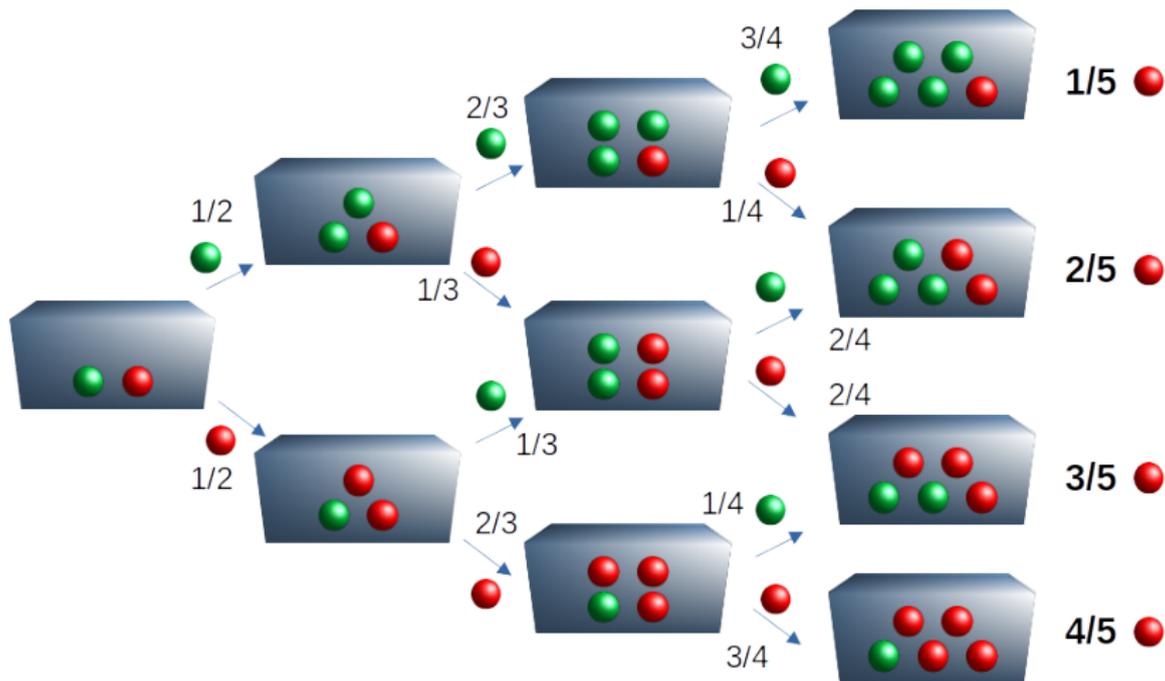
# Cuota de mercado

- Dos distribuidores ● y ● venden un nuevo dispositivo electrónico de pantalla enrollable.
- No se hace publicidad del dispositivo.
- Los aparatos funcionan muy bien.
- Cada distribuidor sortea uno de sus equipos.
- Nuevo comprador:
  - Encuentra al azar un poseedor del dispositivo.
  - Le pregunta dónde lo adquirió.
  - Compra al mismo distribuidor.

Después de alcanzar la venta número 1000,

¿cuál puede ser la cuota de mercado de cada distribuidor?

# Cuota de mercado (urna de Pólya)

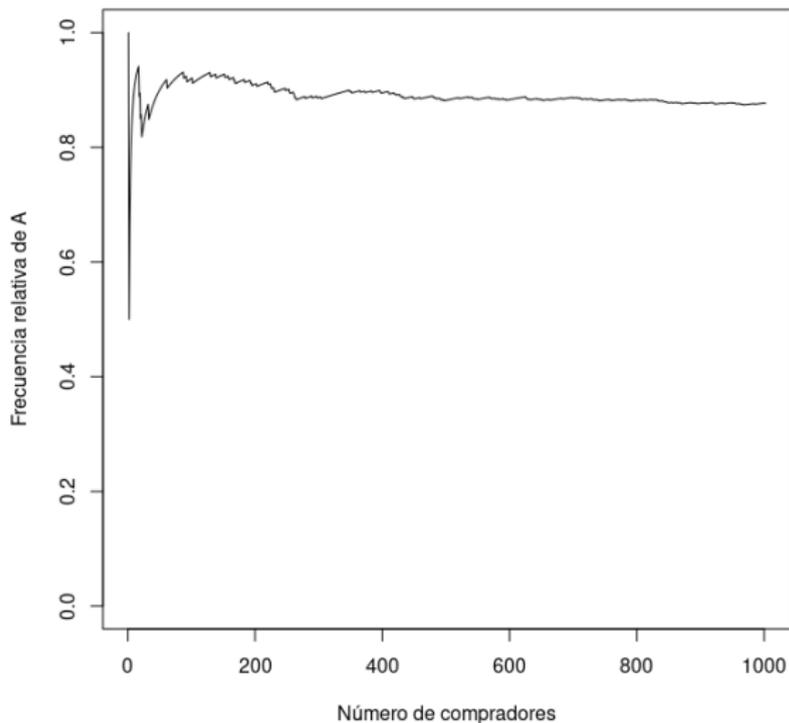


# Cuota de mercado (urna de Pólya)

```
### cuota de mercado
mercado <- c("A","B") # una persona con A y una con B ;
n0      <- length(mercado)
n       <- 1000      # número de compras ;
for (i in 1:n) {    # nuevo cliente compra lo mismo que otro cliente
  nueva.compra <- sample (mercado, 1) # encontrado al azar ;
  mercado <- c (mercado, nueva.compra)
}
## evolución de la cuota de A :
plot (cumsum(mercado=="A") / (1:(n+n0)), type="l", ylim=0:1)
mean (mercado=="A") # cuota de mercado de A al final ;
```

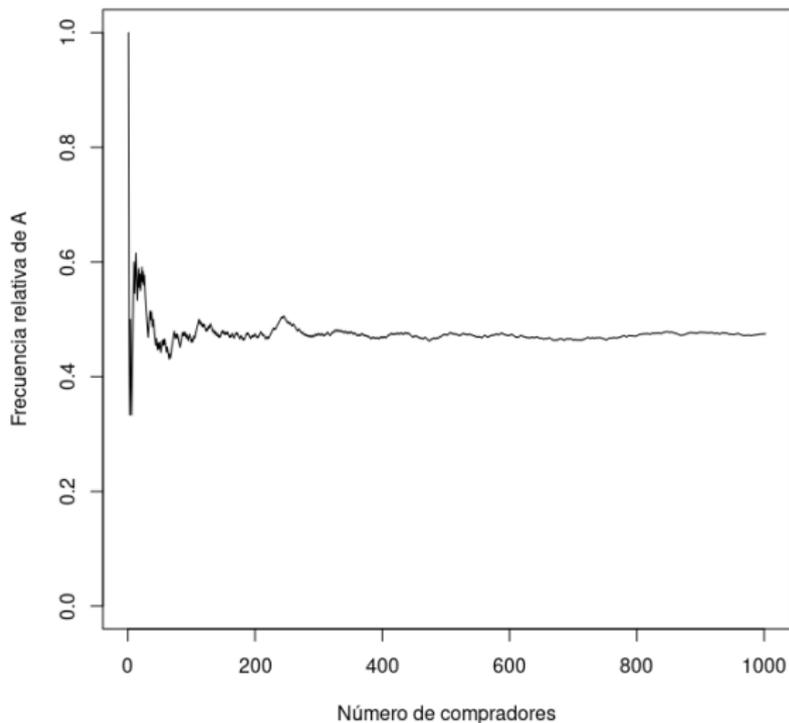
# Cuota de mercado (urna de Pólya)

88 % de A al final



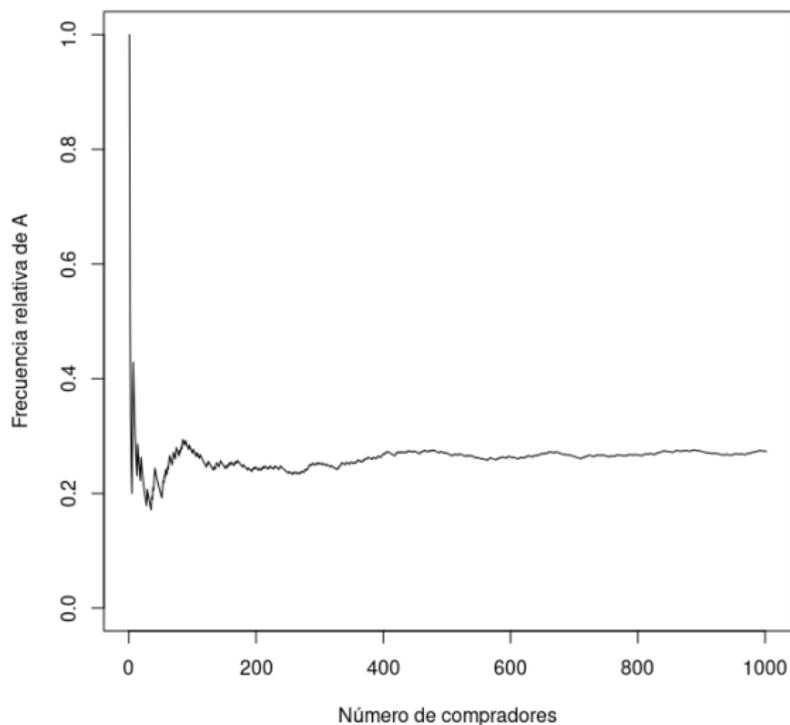
# Cuota de mercado (urna de Pólya)

48 % de A al final



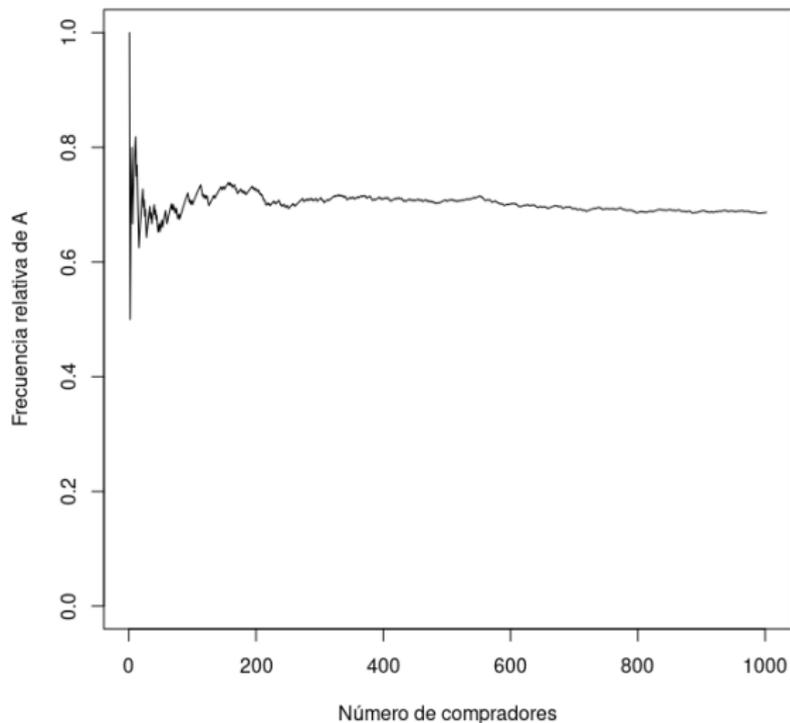
# Cuota de mercado (urna de Pólya)

27 % de A al final



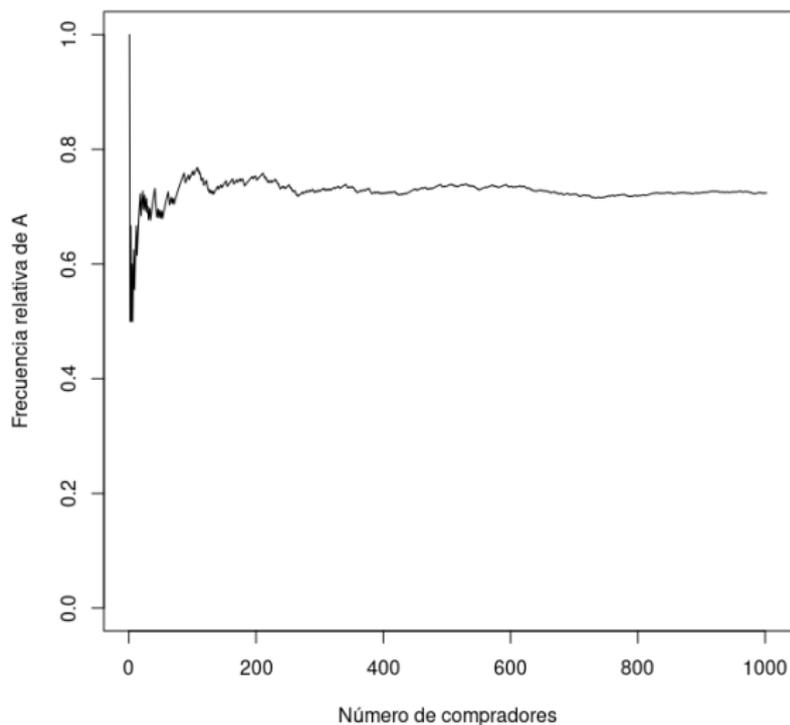
# Cuota de mercado (urna de Pólya)

69 % de A al final



# Cuota de mercado (urna de Pólya)

72 % de A al final



# Probabilidad. El problema de las tres puertas

En un concurso hay tres puertas idénticas; detrás de una hay un gran premio y en las otras dos una cabra.

La mecánica del concurso es la siguiente:

- El presentador sabe lo que hay detrás de cada puerta.
- El concursante elige libremente una de esas puertas (que continúa cerrada).
- Después el presentador abre una de las otras dos puertas, mostrando una cabra, y ofrece al concursante la posibilidad de cambiar su elección inicial.
- El concursante gana lo que hay detrás de la puerta que elija.

La cuestión es: ¿debe cambiar de puerta?

# Juego de las tres puertas



Fuente: Wikipedia

# El juego de las tres puertas

Posibilidades de ganar:

**Estrategia 1:** No cambiar de puerta.

Ganar  $\iff$  Elegir la única puerta buena la primera vez

$$P(\text{ganar}) = \frac{1}{3}$$

**Estrategia 2:** Cambiar de puerta.

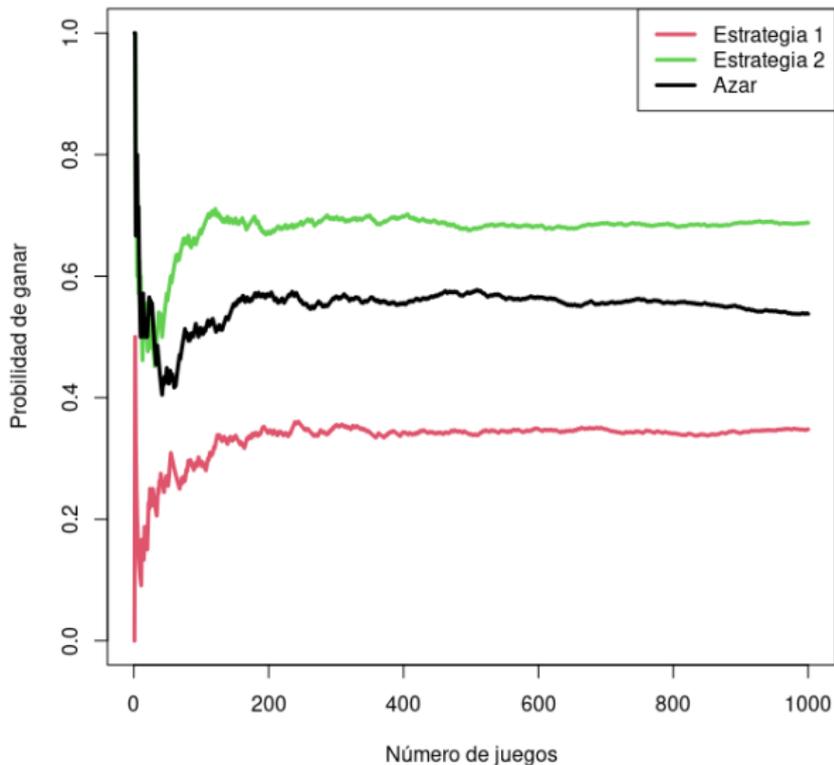
Ganar  $\iff$  Elegir cualquiera de las dos puertas “malas”.

$$P(\text{ganar}) = \frac{2}{3}$$

# El juego de las tres puertas

```
% ### tres puertas
puertas <- c("A","B","C")
unavez <- function (p) # p=prob.cambiar 0=mantengo 1=cambio 0.5=moneda
{
  premio <- sample (puertas, 1) # puerta del premio
  elegida <- sample (puertas, 1) # al azar
  mostrada <- sample (setdiff (puertas, c(premio,elegida)), 1)
  restante <- setdiff (puertas, c(elegida,mostrada))
  decision <- sample (c(elegida,restante), 1, prob=c(1-p,p))
  premio == decision # TRUE si gano, FALSE si pierdo
}
## promedios tras muchas jugadas
mean (replicate (10000, unavez(0))) # 1/3
mean (replicate (10000, unavez(1))) # 2/3
mean (replicate (10000, unavez(0.5))) # 1/2
```

# El juego de las tres puertas



# Juegos de azar *online* en España

	2015	2019
Margen neto	310 M€	750 M€
Apuestas	8.500 M€	19.000 M€
Gasto en publicidad	75 M€	180 M€
Número de cuentas activas	1.300.000	2.750.000
Número de jugadores activos	850.000	1.476.385

## Edades de los jugadores (2018)

Años	18–25	26–35	36–45	46–55	56–65	+65
%	26,9	34,4	22,1	10,5	4,4	1,7

Fuente:

<https://www.epdata.es/datos/juego-online-espana-datos-estadisticas/161>

# ¿Cómo engañar sin mentir?

Una empresa farmacéutica afirma que un nuevo método para tratar los cálculos de riñón resulta más eficaz que el tradicional y presiona para que se aplique de manera inmediata.

Para justificar su afirmación el portavoz de la empresa presentó los resultados obtenidos por ambos métodos:

- Porcentajes de éxitos:
  - Método tradicional **78**% (273/350)
  - Nuevo método **83**% (289/350)

# ¿Cómo engañar sin mentir?

A pesar de la evidencia, el responsable de la Sanidad Pública señala que el método tradicional es mejor.

## ¿Quién tiene razón?

Antes de decidir esta cuestión deberíamos plantearnos dos cuestiones previas:

- 1 ¿La comparación de resultados de dos métodos es siempre válida?
- 2 ¿Qué condiciones se deben cumplir para que lo sea?

## ¿Cómo engañar sin mentir?

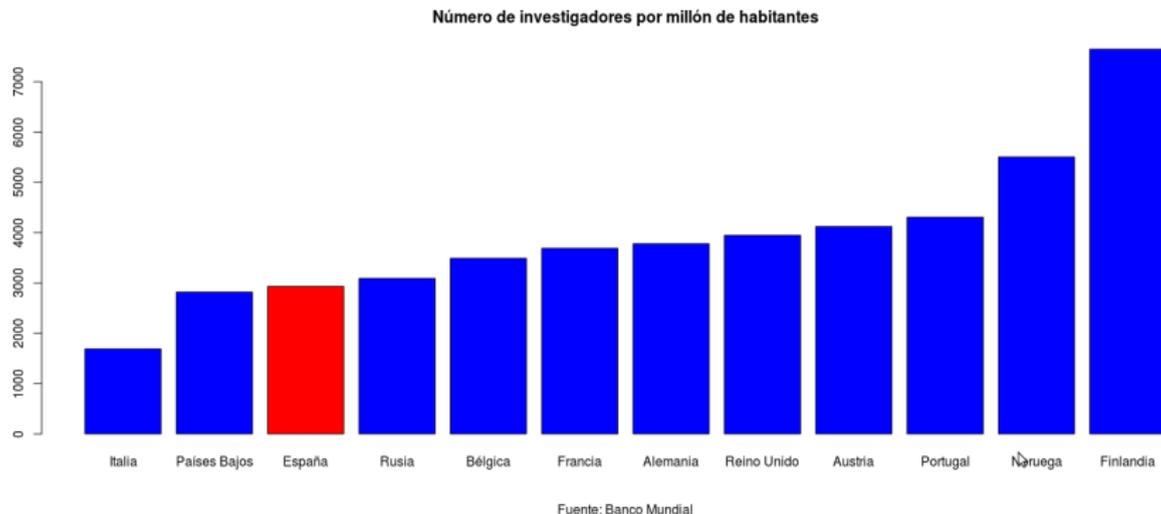
El responsable de la Sanidad Pública mostró los datos desagregados según el tamaño del cálculo renal:

	Habitual	Nuevo
C. Pequeño	(81/87) <b>93 %</b>	(234/270) <b>87 %</b>
C. Grande	(192/263) <b>73 %</b>	(55/80) <b>69 %</b>
Total	(273/350) <b>78 %</b>	(289/350) <b>83 %</b>

$$\text{fr}(\text{C. Grande} \mid \text{Nuevo}) = \frac{80}{350} = 0,23$$

$$\text{fr}(\text{C. Grande} \mid \text{Habitual}) = \frac{263}{350} = 0,75$$

¿Hay demasiados investigadores en España?

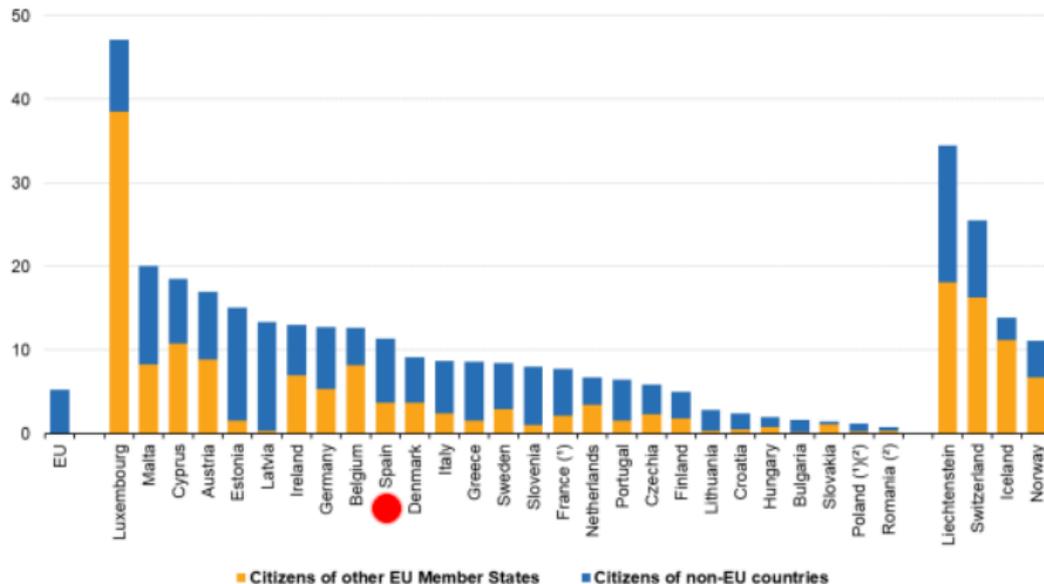


<http://datos.bancomundial.org/indicador/SP.POP.SCIE.RD.P6>

# El problema de la inmigración

## Share of non-nationals in the resident population, 1 January 2021

(%)



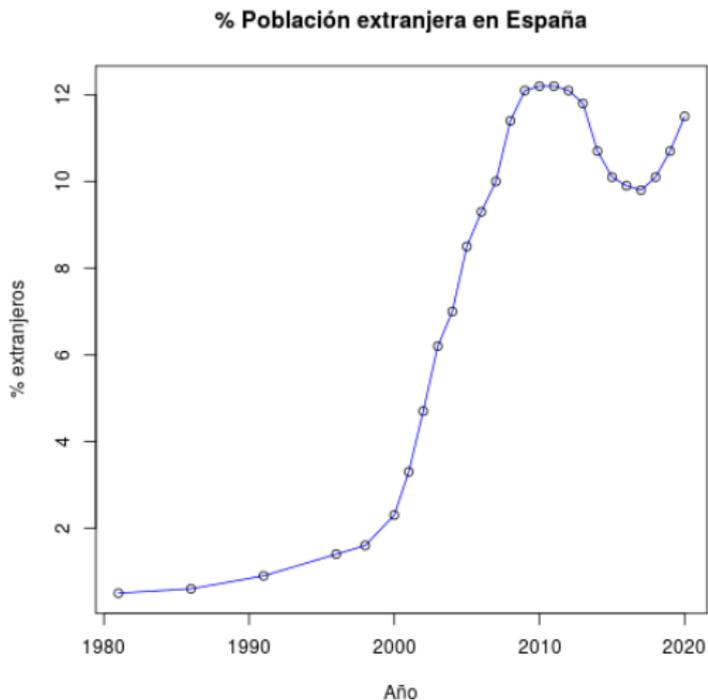
(\*) Provisional.

(\*) Estimate.

Source: Eurostat (online data code: migr\_pop1ctz)

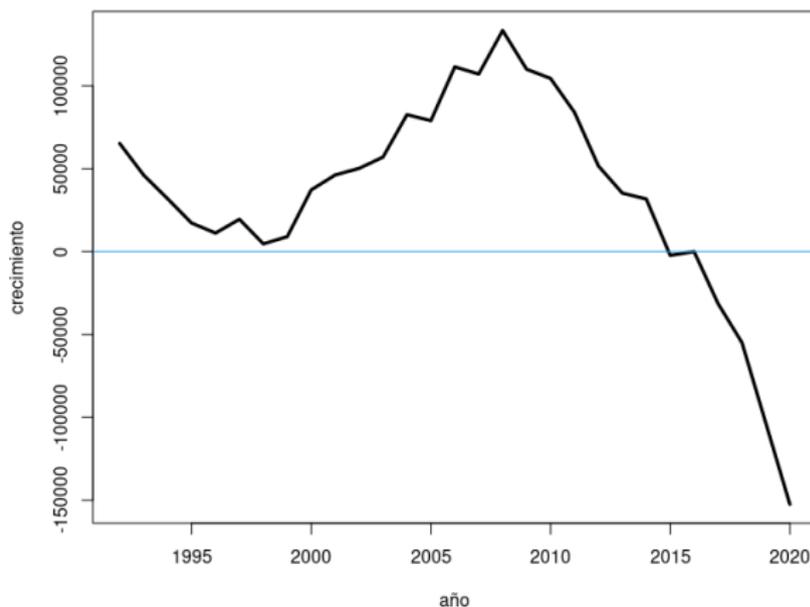
eurostat

# El problema de la inmigración



Fuente: INE

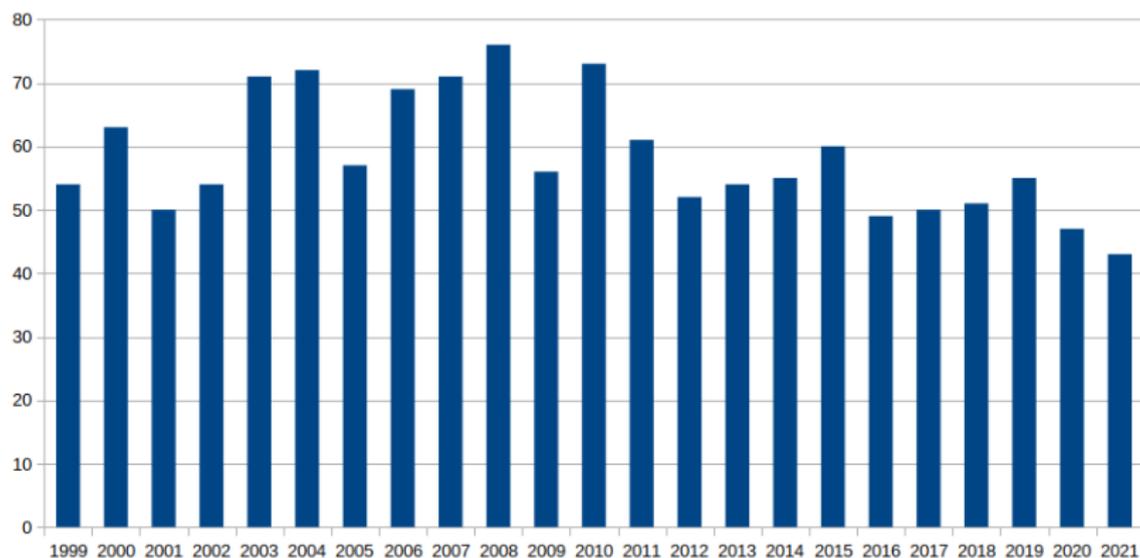
# Crecimiento de la población. El problema de la inmigración



Crecimiento vegetativo = Diferencia entre nacimientos y defunciones durante un cierto periodo de tiempo

Fuente: INE y Wikipedia

# Mujeres víctimas mortales por violencia de género



Total mujeres víctimas mortales 1999-2021

Fuentes:

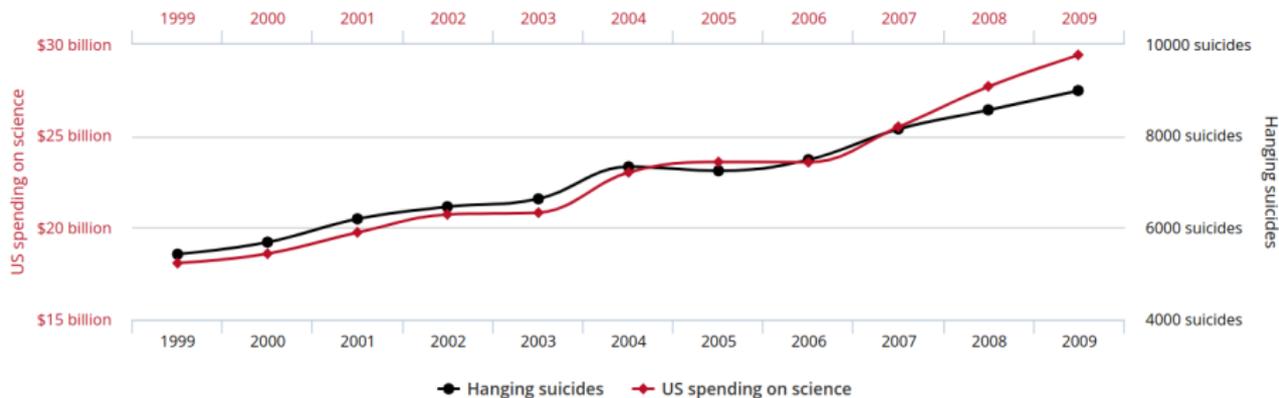
<https://www.ine.es>

<https://www.epdata.es>

# Predicción: ciencia o brujería

## US spending on science, space, and technology correlates with Suicides by hanging, strangulation and suffocation

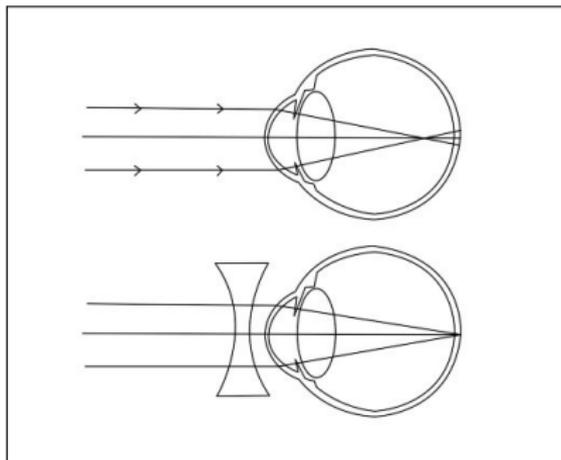
Correlation: 99.79% ( $r=0.99789126$ )



Fuente: Tylervigen.com

## Descripción del problema

El cristalino tiene forma de lente biconvexa y está situado entre el iris y el humor vítreo. Permite enfocar objetos situados a diferentes distancias. En la **miopía** el punto focal se forma delante de la retina y se corrige con una lente divergente.



## Descripción del problema

Se llama **catarata** a la opacificación total o parcial del cristalino. En la cirugía de la catarata, el cristalino se sustituye por una lente artificial denominada **lente intraocular (LIO)**.



La potencia (graduación) de la lente intraocular y se calcula, según las medidas de cada ojo, antes de colocarla.

## Objetivos

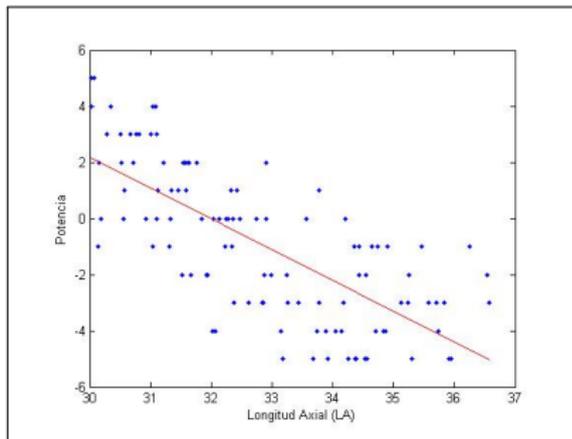
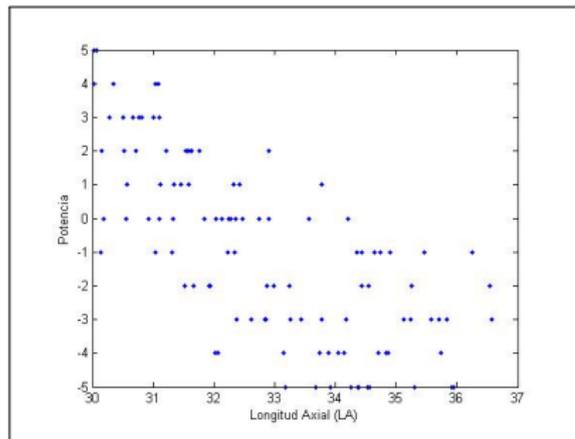
Predecir la potencia de la lente intraocular.

Etapas del problema:

- 1 Seleccionar una muestra en la que se conocen las medidas del ojo y la potencia de la lente.
- 2 Estudiar la relación entre las características del ojo con la potencia de la lente.
- 3 Formular un modelo matemático.
- 4 Estimar los parámetros del modelo.
- 5 Determinar la calidad de la predicción.

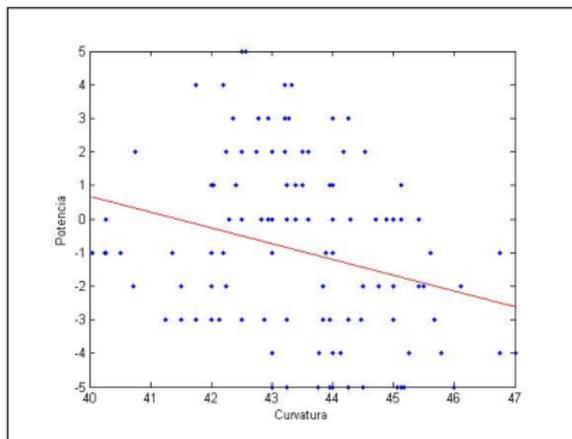
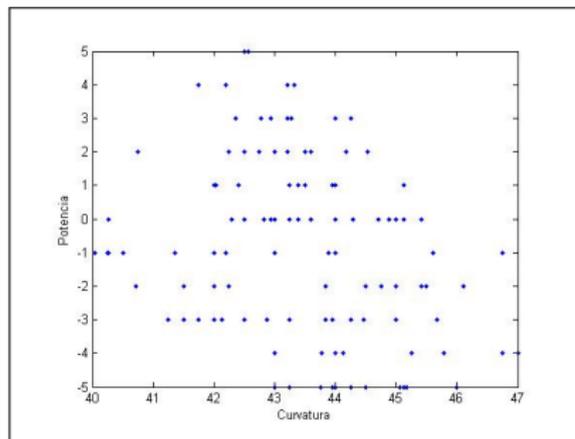
# Relaciones entre las variables

¿Que tipo de relación tiene la longitud axial del ojo con la potencia de la lente?



# Bloque 1: Determinar las relaciones entre las variables

¿Está relacionada la potencia de la lente con la curvatura?



¿Qué modelo se puede utilizar para predecir la potencia?

A la vista de las gráficas parece razonable proponer un modelo lineal (un plano), es decir,

$$\text{Potencia} = \alpha_0 + \alpha_1 LA + \alpha_2 \text{Curvatura} + \varepsilon$$

¿Cómo elegir los parámetros  $(\alpha_0, \alpha_1, \alpha_2)$  del modelo?

Las predicciones sean muy parecidas a las observaciones reales.

$$\sum_{i=1}^{108} \left( Pot_i - \widehat{Pot}_i \right)^2 = \sum_{i=1}^{108} \left( Pot_i - (\alpha_0 + \alpha_1 LA_i + \alpha_2 C_i) \right)^2$$

Se buscan  $(\alpha_0, \alpha_1$  y  $\alpha_2)$  que hagan mínima esta última expresión:

$$\min_{\alpha_0, \alpha_1, \alpha_2} \left\{ \sum_{i=1}^{108} \left( Pot_i - (\alpha_0 + \alpha_1 LA_i + \alpha_2 C_i) \right)^2 \right\}$$

Con las 108 observaciones analizadas se obtiene:

$$\alpha_0 = 97,74; \quad \alpha_1 = -1,47; \quad \alpha_2 = -1,16$$

## Calidad del modelo estimado

$$\widehat{Potencia} = 97,74 - 1,47LA - 1,16Curvatura$$

El modelo explica el **90 %** de la variabilidad de la potencia.

¿Es mejor este nuevo método que el usado habitualmente?

Comparación de los errores de predicción de ambos métodos:

- **Habitual:**  $Media(|Errores|) = 1,08$  dioptrías.
- **Nuevo método:**  $Media(|Errores|) = 0,41$  dioptrías.

Mejora relativa del nuevo método respecto de la máquina: **62,08 %**

# GRACIAS POR LA ATENCIÓN