

# 14.310x – Data Analysis for Social Scientists

MicroMasters en Statistics and Data Science (SDS) – MITx

**BREIT** (Brescia Institute of Technology) · Aporta — Grupo Breca

---

## Clase 1

### Fundamentos de Probabilidad y Conteo

*Teoría*

---

#### Módulos oficiales del MicroMaster:

M1 – Introduction (setup de R); M2 – Fundamentals of Probability (axiomas, conteo)

B.Sc. Cristian Lazo Quispe

✉ [clazoq@uni.pe](mailto:clazoq@uni.pe)

[in linkedin.com/in/cristian-lazo-quispe](https://www.linkedin.com/in/cristian-lazo-quispe) · [github.com/CristianLazoQuispe](https://github.com/CristianLazoQuispe)

Material del *Advanced Program in Data Science & Global Skills*, diseñado y desarrollado por **BREIT (Brescia Institute of Technology)**, iniciativa de **Aporta** —plataforma de innovación e impacto social del **Grupo Breca**— en alianza con el **MIT IDSS** (Institute for Data, Systems, and Society).

Basado en MIT 14.310x (E. Duflo & S. Ellison) y en diversas fuentes abiertas (ver bibliografía al final), bajo licencia CC BY-NC-SA 4.0.

11 de agosto de 2026

# Índice

|                                                         |   |
|---------------------------------------------------------|---|
| 1. De qué trata esta clase                              | 2 |
| 2. Espacio muestral y eventos                           | 2 |
| 3. Probabilidad: los axiomas                            | 3 |
| 4. Espacios muestrales simples: probabilidad por conteo | 3 |
| 5. Conteo: combinatoria                                 | 4 |
| 5.1. Con y sin reemplazo . . . . .                      | 5 |
| 6. Apéndice: entorno de trabajo en R                    | 5 |
| Resumen de la clase                                     | 6 |

# 1. De qué trata esta clase

En esta primera clase ponemos los cimientos de todo el curso. La pregunta de fondo es muy simple de enunciar y muy poderosa: **¿cómo le asignamos un número a la incertidumbre?** Ese número es la *probabilidad*.

Como científicos sociales trabajamos casi siempre con datos que provienen de un proceso incierto: a quién cae en una encuesta, qué hogar resulta pobre, qué estudiante aprueba. Antes de analizar esos datos necesitamos un lenguaje preciso para hablar de azar. Ese lenguaje tiene tres piezas: **espacio muestral (sample space)**, **eventos (events)** y **probabilidad (probability)**. Cuando todos los resultados son igualmente probables, calcular probabilidades se reduce a **contar** — por eso cerramos con **combinatoria (counting / combinatorics)**.

Probabilidad no es “adivinar el futuro”. Es *contabilidad de posibilidades*: medir qué fracción de los escenarios posibles coincide con lo que nos interesa.

## 2. Espacio muestral y eventos

**Definición 2.1** (Espacio muestral). El **espacio muestral (sample space)**  $S$  de un experimento es la colección de *todos* los resultados posibles. Cada resultado individual se llama *punto muestral (sample point)*.

**Definición 2.2** (Evento). Un **evento (event)**  $A$  es cualquier subconjunto de  $S$  (puede ser un solo resultado, todo  $S$ , o el conjunto vacío  $\emptyset$ ). Si el resultado observado pertenece a  $A$ , decimos que el evento *ocurrió*.

Como los eventos son conjuntos, usamos el lenguaje de la teoría de conjuntos: unión ( $A \cup B$ , “ $A$  o  $B$ ”), intersección ( $A \cap B$  o  $AB$ , “ $A$  y  $B$ ”) y complemento ( $A^c$ , “no  $A$ ”). Valen todas las propiedades usuales (asociativa, conmutativa, distributiva) y algunos resultados útiles:

$$A \subset B \Rightarrow A \cup B = B, \quad A \subset B \Rightarrow A \cap B = A, \quad A \cup A^c = S.$$

**Definición 2.3** (Mutuamente excluyentes y exhaustivos).  $A$  y  $B$  son **mutuamente excluyentes (mutually exclusive / disjoint)** si no comparten resultados, es decir  $A \cap B = \emptyset$ . Son **exhaustivos (exhaustive / complementary)** si su unión es todo  $S$ .

**Ejemplo 2.1** (Una encuesta de hogares). Encuestamos a un hogar peruano y registramos su condición de pobreza. Un espacio muestral posible es

$$S = \{\text{pobre extremo, pobre no extremo, no pobre}\}.$$

El evento “el hogar es pobre” es  $A = \{\text{pobre extremo, pobre no extremo}\}$ . Su complemento  $A^c = \{\text{no pobre}\}$ . Los tres puntos muestrales son mutuamente excluyentes (un hogar cae en una sola categoría) y exhaustivos (cubren todo  $S$ ).

**Ejemplo 2.2** (Dos lanzamientos de una moneda). Lanzamos una moneda dos veces y anotamos cara (C) o sello (S). El espacio muestral es

$$S = \{CC, CS, SC, SS\}, \quad n(S) = 4.$$

El evento “sale exactamente una cara” es  $A = \{CS, SC\}$ .

### 3. Probabilidad: los axiomas

A cada evento  $A$  le asignamos un número  $\mathbb{P}(A)$ . No cualquier asignación sirve: debe respetar tres reglas (los **axiomas de Kolmogorov**).

#### Axiomas de probabilidad (Kolmogorov)

Una **probabilidad** sobre  $S$  es una función  $\mathbb{P}$  que a cada evento le asigna un número y cumple:

**A1.**  $\mathbb{P}(A) \geq 0$  para todo evento  $A$ . (no negatividad)

**A2.**  $\mathbb{P}(S) = 1$ . (normalización)

**A3.** Si  $A_1, A_2, \dots$  son disjuntos dos a dos, entonces  $\mathbb{P}\left(\bigcup_i A_i\right) = \sum_i \mathbb{P}(A_i)$ . (aditividad)

De estos tres axiomas se deduce todo lo demás. Las consecuencias que más usaremos:

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A), \quad 0 \leq \mathbb{P}(A) \leq 1, \quad A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B),$$

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \quad (\text{regla de la suma / inclusión-exclusión}).$$

El axioma 3 dice que si los pedazos no se solapan, las probabilidades se *suman*. La regla  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB)$  corrige el doble conteo cuando *sí* se solapan: restamos una vez lo que contamos dos veces.

**Ejemplo 3.1** (Usando la regla de la suma). En una población,  $\mathbb{P}(\text{tiene empleo formal}) = 0.45$  y  $\mathbb{P}(\text{tiene seguro de salud}) = 0.60$ . Si  $\mathbb{P}(\text{ambos}) = 0.35$ , ¿cuál es la probabilidad de tener al menos uno de los dos?

$$\mathbb{P}(\text{formal} \cup \text{seguro}) = 0.45 + 0.60 - 0.35 = 0.70.$$

Y la de no tener ninguno:  $1 - 0.70 = 0.30$ .

### 4. Espacios muestrales simples: probabilidad por conteo

**Definición 4.1** (Espacio muestral simple). Un **espacio muestral simple** (simple sample space) es uno con un número finito de resultados *igualmente probables*. Si  $n(\cdot)$  cuenta

elementos, entonces

$$\mathbb{P}(A) = \frac{n(A)}{n(S)} \quad (\text{casos favorables sobre casos totales}).$$

Esta es la idea más poderosa de la clase: **si todos los resultados son igualmente probables, calcular una probabilidad es *solo contar***. Todo el esfuerzo se traslada a contar bien — de ahí la combinatoria.

**Ejemplo 4.1** (Dos dados). Lanzamos dos dados justos. ¿Probabilidad de que la suma sea 4?  $n(S) = 6 \times 6 = 36$ . El evento “suma = 4” es  $\{(1, 3), (2, 2), (3, 1)\}$ , así  $n(A) = 3$ . Luego  $\mathbb{P}(A) = 3/36 = 1/12$ .

## 5. Conteo: combinatoria

Para usar  $\mathbb{P}(A) = n(A)/n(S)$  necesitamos contar sin enumerar a mano. Tres reglas bastan para casi todo.

### Reglas de conteo

**R1. Regla del producto.** Si una tarea tiene dos partes, la primera con  $m$  opciones y la segunda con  $n$  opciones (sin importar la primera), hay  $m \times n$  resultados.

**R2. Permutaciones (permutations).** Un arreglo *ordenado* de objetos. El número de permutaciones de  $n$  objetos tomados de  $N$  es

$$P_n^N = \frac{N!}{(N - n)!}.$$

**R3. Combinaciones (combinations).** Un arreglo *sin orden*. El número de combinaciones de  $n$  tomados de  $N$  es

$$\binom{N}{n} = \frac{N!}{(N - n)! n!} \quad (“N \text{ en } n”, N \text{ choose } n).$$

donde  $N! = N \times (N - 1) \times \cdots \times 2 \times 1$  y  $0! = 1$ .

La pregunta clave antes de contar es: **¿el orden importa?** Si sí  $\rightarrow$  permutación. Si no  $\rightarrow$  combinación. Confundirlos es el error más común de la clase. Recuerda:  $\binom{N}{n}$  es  $P_n^N$  dividido entre  $n!$ , justamente para “borrar” los órdenes que ya no distinguimos.

## 5.1. Con y sin reemplazo

**Ejemplo 5.1** (Muestreo con y sin reemplazo). Una placa tiene 6 caracteres, cada uno una de 36 opciones (26 letras + 10 dígitos).

- **Con reemplazo (with replacement):** se permite repetir.  $n(S) = 36^6$ . Placas de puros dígitos:  $10^6$ , así  $\mathbb{P} = \frac{10^6}{36^6} \approx 0.0005$ .
- **Sin reemplazo (without replacement):** no se repite.  $n(S) = 36 \cdot 35 \cdots 31 = \frac{36!}{5!}$ ; favorables  $10 \cdot 9 \cdots 5 = \frac{10!}{4!}$ , así  $\mathbb{P} \approx 0.0001$ .

**Ejemplo 5.2** (Combinaciones: el problema de las canicas). Una urna tiene 5 canicas azules, 3 amarillas y 4 grises (12 en total). Si sacamos una al azar, ¿probabilidad de que sea azul? Como una sola extracción es un espacio simple,  $\mathbb{P}(\text{azul}) = \frac{5}{12}$ .

Ahora una pregunta de combinaciones: sacamos 4 canicas a la vez (sin orden). ¿Cuántos grupos contienen *al menos una azul*? Total de grupos:  $\binom{12}{4} = 495$ . Grupos *sin* ninguna azul (eligiendo entre las 7 no azules):  $\binom{7}{4} = 35$ . Por complemento, con al menos una azul:  $495 - 35 = 460$ .

“Al menos uno” casi siempre es más fácil de contar por el complemento:

$$\mathbb{P}(\text{al menos uno}) = 1 - \mathbb{P}(\text{ninguno}).$$

Contar “ninguno” suele ser un solo cálculo; contar “al menos uno” directo obliga a sumar muchos casos.

**Ejemplo 5.3** (Saludos de mano). Si 9 personas se saludan todas entre sí una vez, ¿cuántos apretones hay? Un apretón es un par *sin orden*:  $\binom{9}{2} = \frac{9 \cdot 8}{2} = 36$ .

## 6. Apéndice: entorno de trabajo en R

El curso usa **R** con **RStudio**. Instalación:

1. Instala **R** desde [cran.rstudio.com](https://cran.rstudio.com).
2. Instala **RStudio Desktop** desde [posit.co](https://posit.co).

Para practicar de forma interactiva usamos el paquete **swirl** con el curso introductorio de MITx:

**Consola de R**

```
install.packages("swirl")
library(swirl)
install_course_url(
  "https://courses.mitonline.mit.edu/asset-v1:MITxT+
  14.310x+2T2024+type@asset+block@14_310x_Intro_to_R.zip")
swirl()
```

**Probabilidad por conteo en R**

```
# Canicas: P(azul) con espacio simple
favorables <- 5; totales <- 12
favorables / totales # 0.4167

# Combinaciones: "al menos una azul" entre 4 de 12
choose(12, 4) - choose(7, 4) # 460

# Factorial y permutaciones
factorial(9) / factorial(9 - 2) # P(9,2) = 72
```

*Nota.* `choose(N, n)` calcula  $\binom{N}{n}$  y `factorial(n)` calcula  $n!$ . Con estas dos funciones resuelves casi cualquier conteo de la clase.

## Resumen de la clase

**Lo esencial de la Clase 1**

- **Espacio muestral**  $S$ : todos los resultados. **Evento**: un subconjunto de  $S$ .
- **Axiomas**:  $\mathbb{P}(A) \geq 0$ ,  $\mathbb{P}(S) = 1$ , aditividad en disjuntos. De ahí:  $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$  y  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB)$ .
- **Espacio simple**: si todo es equiprobable,  $\mathbb{P}(A) = n(A)/n(S)$  — ¡solo hay que contar!
- **Conteo**: producto, permutaciones  $\frac{N!}{(N-n)!}$  (orden importa), combinaciones  $\binom{N}{n}$  (orden no importa).
- **Trucos**: complemento para “al menos uno”; distinguir con/sin reemplazo.

## Fuentes y atribución

Material de repaso **basado en MIT 14.310x – Data Analysis for Social Scientists** (Esther Duflo & Sara Fisher Ellison), MIT OpenCourseWare, licencia **CC BY-NC-SA 4.0** (uso no comercial, con atribución, compartir bajo la misma licencia), y en las fuentes de la bibliografía. Los enunciados se basan en el material del curso y en diversas fuentes abiertas; cuando un problema se adapta de una fuente específica, se cita en el enunciado.

### Bibliografía.

- Duflo, E. & Ellison, S. F. *14.310x: Data Analysis for Social Scientists*. MIT OpenCourseWare (slides, lecture notes y problem sets). [ocw.mit.edu](https://ocw.mit.edu).
- Larsen, R. J. & Marx, M. L. *An Introduction to Mathematical Statistics and Its Applications*, 6.<sup>a</sup> ed., Pearson, 2018.
- DeGroot, M. H. & Schervish, M. J. *Probability and Statistics*, 4.<sup>a</sup> ed., Pearson, 2012.
- Blitzstein, J. & Hwang, J. *Introduction to Probability* (Harvard Stat 110); W. Chen, *Probability Cheatsheet*.
- Diez, D., Çetinkaya-Rundel, M. & Barr, C. *OpenIntro Statistics*. [openintro.org](https://openintro.org).
- Materiales abiertos de la comunidad del curso en GitHub: [gevorg-hunanyan/probability](https://github.com/gevorg-hunanyan/probability); [hanmochen/Probability-Theory-2020](https://github.com/hanmochen/Probability-Theory-2020); [mynameisjanus/14310xDataAnalysis](https://github.com/mynameisjanus/14310xDataAnalysis).

[creativecommons.org/licenses/by-nc-sa/4.0](https://creativecommons.org/licenses/by-nc-sa/4.0)