

14.310x – Data Analysis for Social Scientists

MicroMasters en Statistics and Data Science (SDS) – MITx

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

Clase 2

Probabilidad Condicional, Independencia y Bayes

Ejercicios (5 niveles)

Módulos oficiales del MicroMaster:

M2 – Fundamentals of Probability (prob. condicional, independencia, Bayes); M3 – Joint/Marginal/Conditional (entrada)

B.Sc. Cristian Lazo Quispe

✉ clazoq@uni.pe

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

Niveles de dificultad (*difficulty levels*)

Cada problema está rotulado por color según su nivel, de lo más accesible (Nivel 1) a lo más difícil (Nivel 5). Empieza por donde te sientas cómodo y avanza a tu ritmo; los niveles 4–5 son extensiones para profundizar, no un requisito.

- **Nivel 1 – Fundamentos:** calentamiento muy guiado; para quien parte de casi cero.
- **Nivel 2 – Núcleo:** el nivel objetivo del curso/examen.
- **Nivel 3 – Aplicación:** combina varias ideas en contexto realista.
- **Nivel 4 – Reto:** difícil, integra varios temas.
- **Nivel 5 – Reto+:** muy difícil / fuera del scope típico.

Indicaciones

Resuelve los ejercicios de tu nivel; si terminas, sube al siguiente (ver la leyenda de arriba). Justifica cada paso; al condicionar, indica siempre cuál es el “nuevo” espacio muestral.

1. Problemas

Nivel 1 – Fundamentos (warm-up)

Para quien parte de casi cero: cálculos directos y guiados.

Ejercicio 1.1 (Calentamiento: probabilidad condicional con una tabla pequeña). En un salón de 40 estudiantes, 24 aprobaron matemática y, de *esos* aprobados, 18 también aprobaron lectura. Se elige un estudiante al azar.

- (a) ¿Cuál es la probabilidad de que haya aprobado matemática?
- (b) Dado que aprobó matemática, ¿cuál es la probabilidad de que haya aprobado lectura?

Ejercicio 1.2 (Calentamiento: regla del producto con números redondos). En una comunidad, la probabilidad de que una familia tenga acceso a internet es $\mathbb{P}(I) = 0.50$. Dado que tiene internet, la probabilidad de que un hijo estudie en línea es $\mathbb{P}(E | I) = 0.80$. ¿Cuál es la probabilidad de que una familia *tenga internet y además* un hijo estudie en línea?

Ejercicio 1.3 (Calentamiento: ¿independientes? compara dos números). Se sabe que $\mathbb{P}(A) = 0.4$, $\mathbb{P}(B) = 0.5$ y $\mathbb{P}(A \cap B) = 0.20$. ¿Son A y B independientes?

Ejercicio 1.4 (Calentamiento: Bayes con dos cajas y números redondos). Hay dos cajas. La caja 1 tiene 1 bola roja y 1 azul; la caja 2 tiene 3 rojas y 1 azul. Eliges una caja al azar (cada una con probabilidad $1/2$) y sacas una bola.

- (a) ¿Cuál es la probabilidad de sacar una bola roja?

(b) Si sacaste una roja, ¿cuál es la probabilidad de que viniera de la caja 2?

Ejercicio 1.5 (Definición de condicional). En un distrito, $\mathbb{P}(\text{un hogar es pobre y rural}) = 0.18$ y $\mathbb{P}(\text{un hogar es rural}) = 0.30$. Entre los hogares rurales, ¿qué fracción es pobre?

Ejercicio 1.6 (Regla del producto). La probabilidad de que un estudiante de una escuela rural termine la secundaria es $\mathbb{P}(S) = 0.40$. Dado que terminó la secundaria, la probabilidad de que acceda a la universidad es $\mathbb{P}(U|S) = 0.25$. ¿Cuál es la probabilidad de que un estudiante *termine la secundaria y además* acceda a la universidad?

Ejercicio 1.7 (¿Independientes? — caso sí). Se elige al azar a una persona de un padrón. Sea $A = \text{“nacido en un mes par”}$ con $\mathbb{P}(A) = 0.5$ y $B = \text{“su DNI termina en dígito par”}$ con $\mathbb{P}(B) = 0.5$. Se sabe que $\mathbb{P}(A \cap B) = 0.25$. ¿Son A y B independientes?

Ejercicio 1.8 (Complemento condicional). En cierta clínica, dado que un paciente tiene síntomas de gripe, la probabilidad de que sea realmente influenza es $\mathbb{P}(I|S) = 0.35$. ¿Cuál es la probabilidad de que *no* sea influenza, dado que tiene síntomas?

Ejercicio 1.9 (Tabla de contingencia 2x2 — lectura). Se encuestó a 200 personas sobre si participan en un programa social y si conocen sus derechos. Los conteos son:

	Conoce derechos	No conoce	Total
Participa	70	30	100
No participa	40	60	100
Total	110	90	200

Calcula $\mathbb{P}(\text{conoce derechos})$ y $\mathbb{P}(\text{conoce} | \text{participa})$.

Nivel 2 – Núcleo (core)

El nivel objetivo del curso/examen; todos deberían llegar aquí.

Ejercicio 1.10 (Operaciones de conjuntos y leyes). Sea el espacio muestral $S = \{1, 2, 3, 4, 5, 6\}$ y los eventos $A = \{1, 2\}$, $B = \{2, 4, 5\}$, $C = \{1, 5, 6\}$. Halla:

(a) $A \cup B$; (b) $A \cap B$; (c) A^c ; (d) B^c .

(e) Verifica una ley de De Morgan: $(A \cup B)^c = A^c \cap B^c$.

(f) Verifica la ley distributiva: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

Ejercicio 1.11 (Independencia con un número aleatorio). Se elige N al azar de manera uniforme en $\{1, 2, \dots, 10\}$. Sean los eventos $A = \{N < 7\}$ y $B = \{N \text{ es par}\}$. ¿Son A y B independientes, aunque ambos hablen del *mismo* número N ?

Ejercicio 1.12 (¿Independientes? — caso no). En una encuesta, $\mathbb{P}(\text{tiene seguro de salud}) = 0.60$, $\mathbb{P}(\text{tiene empleo formal}) = 0.50$ y $\mathbb{P}(\text{seguro y formal}) = 0.45$. ¿Son independientes “tener seguro” y “tener empleo formal”?

Ejercicio 1.13 (Tabla 2x2 — ¿independencia?). Con la misma tabla del nivel anterior (200 personas; participar vs. conocer derechos: celdas 70, 30, 40, 60), ¿son independientes “participar” y “conocer derechos”?

Ejercicio 1.14 (Regla del producto en dos pasos). Una urna de un sorteo comunal tiene 5 balotas premiadas y 15 sin premio (20 en total). Se extraen 2 sin reemplazo. ¿Probabilidad de que *ambas* sean premiadas?

Ejercicio 1.15 (Dos encuestadores: al menos uno, exactamente uno). Dos encuestadores trabajan de forma independiente un mismo barrio. El encuestador 1 logra completar una entrevista en una vivienda con probabilidad 0.7 y el encuestador 2 con probabilidad 0.8. Para una vivienda dada:

- (a) ¿Probabilidad de que la vivienda sea entrevistada *al menos una vez*?
- (b) ¿Probabilidad de que sea entrevistada *exactamente una vez*?

Ejercicio 1.16 (Condicional con dos dados sociales). En un juego de feria comunal se lanzan dos dados justos. Dado que la suma de las caras es 8, ¿cuál es la probabilidad de que *al menos uno* de los dados muestre un 6?

Ejercicio 1.17 (Moneda hasta la primera cruz). Se lanza una moneda justa repetidamente hasta obtener la primera cruz. Sea X el número total de lanzamientos realizados. Halla $\mathbb{P}(X = 5)$.

Ejercicio 1.18 (Tres monedas (una de dos caras)). Una caja tiene 3 monedas: dos normales (con cara y cruz) y una falsa con *dos caras* (siempre sale cara). Eliges una moneda al azar y la lanzas una vez.

- (a) ¿Cuál es la probabilidad de obtener cara?
- (b) Dado que saliste cara, ¿cuál es la probabilidad de que hayas tomado la moneda de dos caras?

Ejercicio 1.19 (Ley de probabilidad total con tres regiones). Un programa nacional opera en tres regiones. La región A atiende al 50 por ciento de los beneficiarios, la B al 30 por ciento y la C al 20 por ciento. La tasa de error administrativo es 4 por ciento en A, 6 por ciento en B y 9 por ciento en C. ¿Cuál es la probabilidad de que un beneficiario al azar tenga un error en su expediente?

Ejercicio 1.20 (Bayes con partición de tres). Con los datos del ejercicio anterior, un expediente tiene error. ¿Cuál es la probabilidad de que provenga de la región C?

Nivel 3 – Aplicación

Combina dos o más ideas en contexto realista; consolida lo aprendido.

Ejercicio 1.21 (Conjuntos contables o incontables). Clasifica cada conjunto como *contable* (finito o numerable) o *incontable*:

- (a) $A = \{x \in \mathbb{Q} \mid -100 \leq x \leq 100\}$.
- (b) $B = \{(x, y) \mid x \in \mathbb{N}, y \in \mathbb{Z}\}$.
- (c) $C = (0, 0.1]$ (intervalo de números reales).
- (d) $D = \{1/n \mid n \in \mathbb{N}\}$.

Ejercicio 1.22 (Diagrama de árbol: dos proveedores de vacunas). Un centro de salud recibe vacunas de dos proveedores: el 70 por ciento del lote viene del proveedor 1 y el 30 por ciento del proveedor 2. El 2 por ciento de las dosis del proveedor 1 sale defectuosa y el 5 por ciento de las del proveedor 2. Dibuja el árbol y calcula la probabilidad de que una dosis al azar sea defectuosa.

Ejercicio 1.23 (Bayes sobre el árbol anterior). Con el árbol del ejercicio previo, una dosis resultó defectuosa. ¿Cuál es la probabilidad de que provenga del proveedor 2?

Ejercicio 1.24 (Falsos positivos y tasa base). Una prueba rápida de anemia tiene sensibilidad 95 por ciento (detecta al enfermo) y da falso positivo en el 10 por ciento de los sanos. En una población donde el 5 por ciento tiene anemia, una persona da positivo. ¿Cuál es la probabilidad de que realmente tenga anemia?

Ejercicio 1.25 (Dos detecciones independientes). Un sistema de auditoría usa dos revisores independientes. El revisor 1 detecta un fraude con probabilidad 0.7 y el revisor 2 con probabilidad 0.6, de forma independiente.

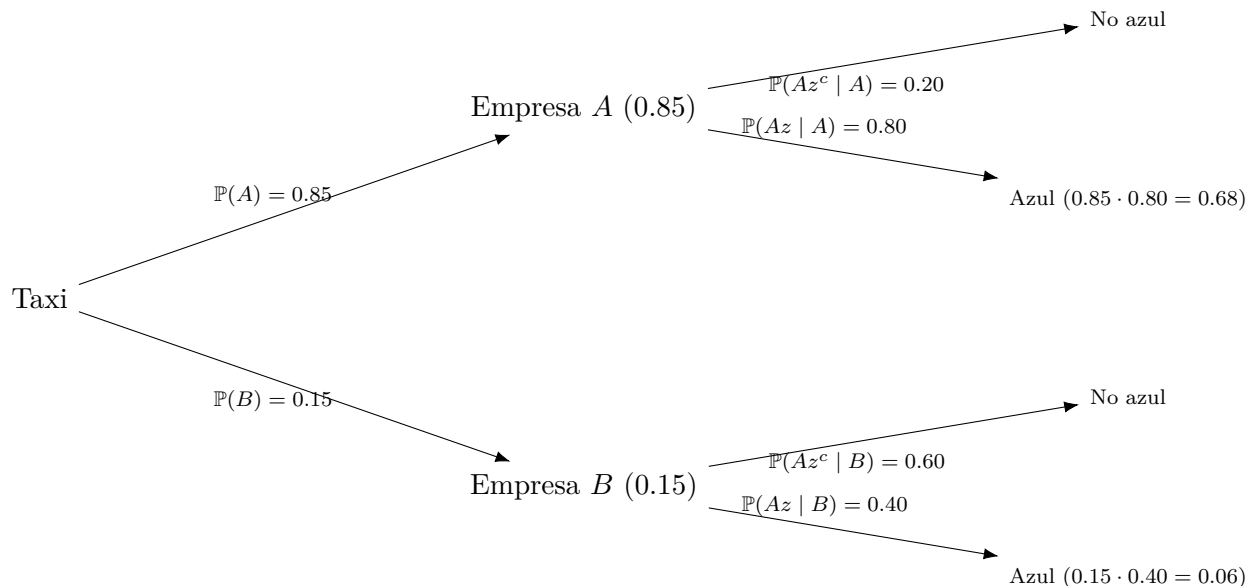
- (a) ¿Probabilidad de que *al menos uno* detecte el fraude?
- (b) ¿Probabilidad de que *ambos* lo detecten?

Ejercicio 1.26 (Dos urnas y traslado de información). Hay dos cajas de expedientes. La caja I tiene 3 aprobados y 7 rechazados; la caja II tiene 6 aprobados y 4 rechazados. Se elige una caja al azar (mismas probabilidades) y de ella se extrae un expediente.

- (a) ¿Probabilidad de extraer un aprobado?
- (b) Si salió aprobado, ¿probabilidad de que viniera de la caja II?

Ejercicio 1.27 (Dos padrones, mismo perfil). Una ONG tiene dos padrones de beneficiarios. El padrón I contiene 5 urbanos, 11 rurales y 8 periurbanos; el padrón II contiene 10 urbanos, 8 rurales y 6 periurbanos. Se extrae al azar una ficha de *cada* padrón (de forma independiente). ¿Cuál es la probabilidad de que ambas fichas correspondan al *mismo* tipo de zona?

Ejercicio 1.28 (Árbol y Bayes: el testigo del taxi). En una ciudad, el 85 por ciento de los taxis pertenece a la empresa *A* (de los cuales el 80 por ciento es azul) y el 15 por ciento a la empresa *B* (de los cuales el 40 por ciento es azul). Un testigo de un accidente reporta haber visto un taxi *azul*. Suponiendo que el reporte del color es fiable, construye el árbol y halla la probabilidad de que el taxi involucrado fuera de la empresa *A*.



Cada rama es una probabilidad condicional: depende del nodo del que sale; la probabilidad de un camino es el producto de sus ramas.

Ejercicio 1.29 (Diagrama de árbol de dos etapas con Bayes). Un tamizaje de desnutrición tiene dos etapas. En la etapa 1, el 25 por ciento de los niños es marcado “en riesgo”. De los marcados en riesgo, el 60 por ciento resulta desnutrido en la etapa 2; de los *no* marcados, solo el 5 por ciento resulta desnutrido. Construye el árbol y halla: (a) la probabilidad de estar desnutrido; (b) dado que un niño está desnutrido, la probabilidad de que hubiera sido marcado “en riesgo”.

Ejercicio 1.30 (Bayes secuencial: dos observaciones). Una ONG cree que el 30 por ciento de los hogares de una zona es elegible para un bono ($\mathbb{P}(E) = 0.3$). Aplica un cuestionario de tamizaje que, en hogares elegibles, da “apto” con probabilidad 0.9, y en no elegibles da “apto” con probabilidad 0.2. El mismo cuestionario se aplica *dos veces* de forma independiente (condicional al verdadero estado del hogar) y ambas veces sale “apto”. ¿Cuál es la probabilidad de que el hogar sea elegible tras las dos observaciones?

Ejercicio 1.31 (Bayes con tres hipótesis y dos señales). Una agencia clasifica solicitudes en tres tipos: legítimas (L , 60 por ciento), dudosas (D , 30 por ciento) y fraudulentas (F , 10 por ciento). Un detector emite “alerta” con probabilidad 0.05 en legítimas, 0.40 en dudosas y 0.90 en fraudulentas. Una solicitud genera alerta. ¿Cuál es la probabilidad de que sea fraudulenta?

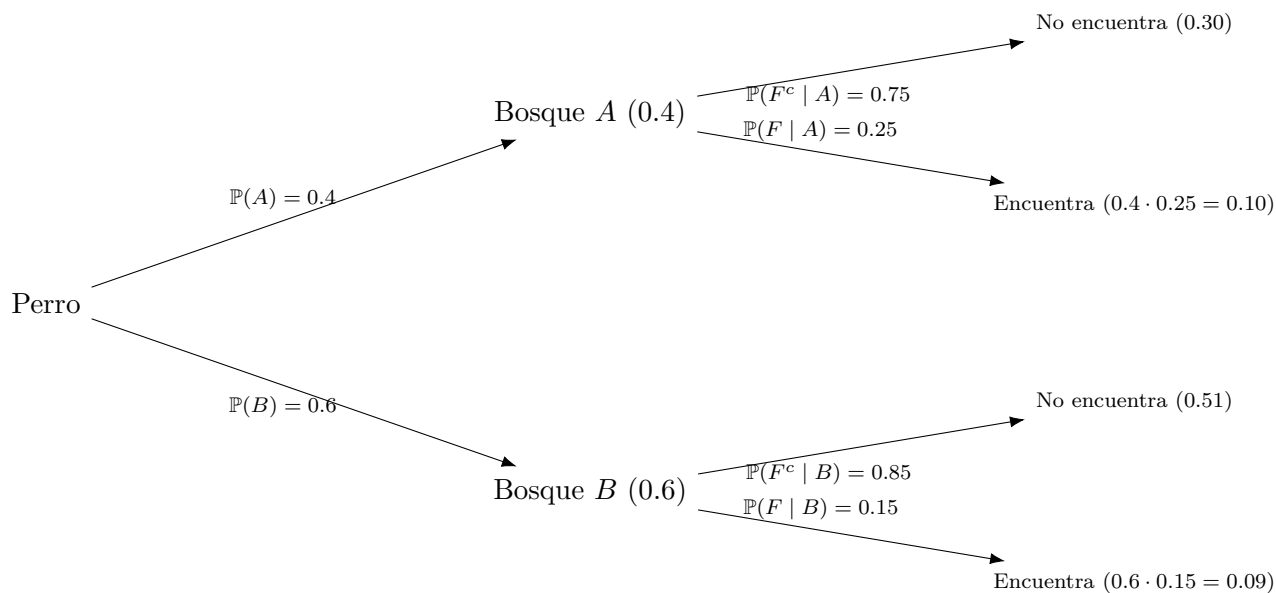
Nivel 4 – Reto

Difícil: integra varios temas en un mismo problema.

Ejercicio 1.32 (Vida útil de un producto (condicional)). El tiempo de vida T (en años) de un producto cumple $\mathbb{P}(T \geq t) = e^{-t/5}$ para $t \geq 0$ (por ejemplo $\mathbb{P}(T \geq 2) = e^{-2/5} \approx 0.6703$). El producto ya funcionó 2 años sin fallar. ¿Cuál es la probabilidad de que falle *durante el tercer año*?

Ejercicio 1.33 (El perro perdido de Oscar). El perro de Oscar se perdió: está en el bosque A con probabilidad 0.4 o en el bosque B con probabilidad 0.6. Si Oscar busca *un día* en el bosque correcto, lo encuentra con probabilidad 0.25 si está en A y 0.15 si está en B ; si busca en el bosque equivocado, la probabilidad de encontrarlo ese día es 0.

- ¿En qué bosque debe buscar el primer día para maximizar la probabilidad de encontrarlo?
- Buscó en A el primer día y *no* lo encontró. ¿Cuál es ahora la probabilidad de que el perro esté en A ?



Cada rama es una probabilidad condicional: depende del nodo del que sale; la probabilidad de un camino es el producto de sus ramas.

Ejercicio 1.34 (Tamizaje masivo con tasa base baja: ¿sirve un positivo?). Un programa de salud pública aplica una prueba rápida de tuberculosis a *toda* la población de un distrito. La prevalencia real de la enfermedad es del 0.5 por ciento. La prueba tiene **sensibilidad** 99 por ciento ($\mathbb{P}(+ | \text{enfermo}) = 0.99$) y **tasa de falsos positivos** 5 por ciento ($\mathbb{P}(+ | \text{sano}) = 0.05$).

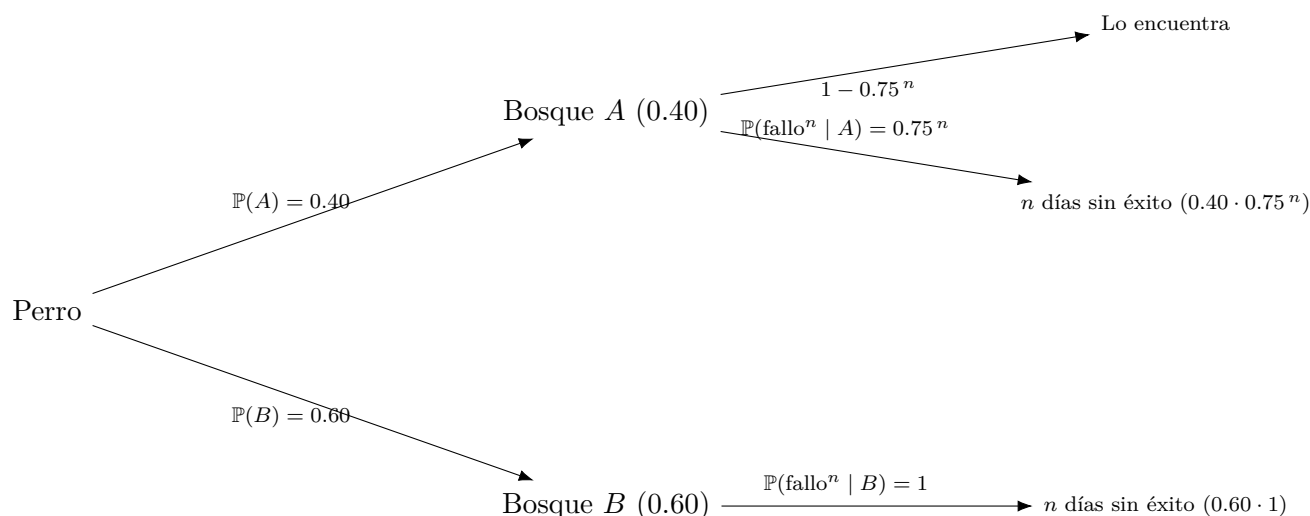
- Imagina 100,000 personas tamizadas. Llena la tabla de cuatro celdas (enfermo/sano $\times$ positivo/negativo).
- Una persona da positivo. ¿Cuál es el *valor predictivo positivo* $\mathbb{P}(\text{enfermo} | +)$? Interpreta por qué es tan bajo pese a una prueba “buena”.

- (c) A esa misma persona se le repite la prueba (independiente, condicional al estado) y vuelve a salir positiva. ¿Cuál es ahora $\mathbb{P}(\text{enfermo} \mid +, +)$?

(adaptado de MIT 18.05, Class 3 — the base rate fallacy)

Ejercicio 1.35 (El perro de Oscar, versión extendida (varios días)). El perro de Oscar se perdió: está en el bosque A con probabilidad 0.40 o en el bosque B con probabilidad 0.60. Oscar decide buscar *cada día en el bosque A*. Si el perro está en A , cada búsqueda diaria lo encuentra con probabilidad 0.25, de forma independiente día a día; si está en B , nunca lo encuentra buscando en A .

- (a) Tras *un* día sin éxito en A , ¿cuál es la probabilidad de que el perro esté en A ?
- (b) Tras *tres* días sin éxito en A , ¿cuál es esa probabilidad?
- (c) Halla la fórmula general $\mathbb{P}(A \mid n \text{ días sin éxito})$ y describe su límite cuando $n \rightarrow \infty$.



Cada rama es una probabilidad condicional: la probabilidad de un camino es el producto de sus ramas; “ n días sin éxito en A ” tiene probabilidad 0.75^n si el perro está en A y probabilidad 1 si está en B .

Ejercicio 1.36 (Paradoja de Simpson con datos concretos). Se comparan dos modalidades de un programa de capacitación laboral, A (intensiva) y B (estándar), según su tasa de inserción (insertados / capacitados). Los datos, desglosados por tipo de participante (urbano = perfil “fácil”, rural = perfil “difícil”), son:

	Urbano (fácil)	Rural (difícil)	Total
Modalidad A	81/87 = 0.931	192/263 = 0.730	273/350 = 0.780
Modalidad B	234/270 = 0.867	55/80 = 0.688	289/350 = 0.826

- (a) ¿Qué modalidad inserta mejor en cada subgrupo por separado?
- (b) ¿Qué modalidad tiene mayor tasa *global*? Explica la aparente contradicción identificando la variable de confusión.

(adaptado de MIT 18.05, Class 3 — conditional probability; estructura clásica de Simpson)

Ejercicio 1.37 (Monty Hall en clave social: tres sobres). Un funcionario te ofrece tres sobres cerrados; uno contiene una beca y dos están vacíos. Eliges el sobre 1 (sin abrirlo). El funcionario, que *sabe* dónde está la beca, abre uno de los otros dos sobres que *sabe* que está vacío (el sobre 3) y te ofrece cambiar tu elección al sobre 2. ¿Te conviene cambiar? Justifica con probabilidades.

Nivel 5 – Reto+

Muy difícil / fuera del scope típico: demostraciones y casos límite.

Ejercicio 1.38 (Reto+: independencia por pares pero no mutua). Se lanzan dos dados justos. Define los eventos

$$A_1 = \{\text{el primer dado es par}\}, \quad A_2 = \{\text{el segundo dado es impar}\}, \quad A_3 = \{\text{la suma es impar}\}.$$

Demuestra que A_1, A_2, A_3 son independientes *por pares* pero *no* mutuamente independientes. (Moraleja: la independencia mutua es estrictamente más fuerte que la independencia por pares.)

Ejercicio 1.39 (Reto+: independencia marginal sin independencia condicional). Se lanzan dos monedas justas. Define

$$A = \{\text{la primera moneda es cara}\}, \quad B = \{\text{la segunda moneda es cara}\}, \quad C = \{\text{ambas monedas salen iguales}\}.$$

- (a) Muestra que A y B son *independientes marginalmente*.
- (b) Muestra que A y B **no** son independientes *condicionalmente dado* C . Da una explicación intuitiva.

(adaptado de MIT 18.05, Class 3 — independence)

Ejercicio 1.40 (Reto+: el problema de los dos hijos con condicionamiento sutil). Una funcionaria del registro civil revisa familias con *exactamente dos hijos*. Supone que cada hijo es niño o niña con probabilidad $1/2$, de forma independiente, y que cada hijo nació en uno de los 7 días de la semana con igual probabilidad e independiente del sexo.

- (a) Sabiendo que la familia tiene *al menos un niño*, ¿cuál es la probabilidad de que *ambos* hijos sean niños?
- (b) Sabiendo que el hijo *mayor* es niño, ¿cuál es esa probabilidad? Compara con (a).
- (c) Sabiendo que la familia tiene *al menos un niño nacido un martes*, ¿cuál es la probabilidad de que ambos sean niños? (Resultado contraintuitivo.)

(adaptado de MIT 18.05, Class 3 — conditional probability; “the two children problem”)

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.
- Larsen, R. J. & Marx, M. L. *An Introduction to Mathematical Statistics and Its Applications*, 6.^a ed., Pearson, 2018.
- DeGroot, M. H. & Schervish, M. J. *Probability and Statistics*, 4.^a 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.
- 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