

Práctica 4: probabilidad.

Grado en Biología - Universidad de Alcalá

1. Hallar la probabilidad de que al lanzar una moneda dos veces se obtenga al menos una vez una cruz. Para ello, calcula los casos posibles y los casos favorables para usar la regla de Laplace. Solución en la página 3.
2. Las siguientes afirmaciones son necesariamente falsas. Explica por qué.
 - (a) En un hospital, la probabilidad de que un paciente permanezca ingresado durante más de dos días es de 0.5. La probabilidad de que un paciente permanezca hospitalizado durante más de un día es de 0.3.
 - (b) La probabilidad de que llueva el sábado es del 50% y de que llueva el domingo es del 50%. Por tanto, durante el fin de semana es seguro que lloverá.

Solución en la página 3.

3. Una población consta de 10 hombres y 20 mujeres. La mitad de los hombres y la mitad de las mujeres tienen los ojos castaños. Hallar la probabilidad de que una persona elegida al azar sea un hombre o tenga los ojos castaños. Solución en la página 3.
4. En cierta facultad, se sabe que (1) un 25% de los estudiantes suspendió Estadística, (2) un 15% suspendió Química y (c) un 10% suspendió ambas. Se selecciona un estudiante al azar:
 - (a) si suspendió Química, ¿cuál es la probabilidad de que también suspendiera Estadística?
 - (b) si suspendió Estadística, ¿cuál es la probabilidad de que también suspendiera Química?
 - (c) ¿Cuál es la probabilidad de que suspendiera al menos una de las dos?

Solución en la página 3.

Los ejercicios que vienen a continuación (del 6 al 10) están relacionados con el **diseño de experimentos** y, en concreto, con las **técnicas de muestreo**. Es decir, con la forma en la que recogen los datos de una población para obtener una muestra.

Se trata de pensar sobre las diferentes formas de "tomar" datos: extraer todos los individuos a la vez, hacerlo de uno en uno, con o sin reemplazamiento.

Más adelante veremos que es fundamental que los datos se recojan de forma independiente, hasta el punto de que no cumplir con esa premisa invalida toda la estadística que se pueda hacer con dichos datos.

6. En un estudio clínico se anonimizan seis muestras de tejido numerándolas del 1 al 6. Se analizan una a una (al azar y sin reemplazarlas) todas las muestras. ¿Cuál es la probabilidad de que salgan en el orden natural? (Es decir, primero la muestra etiquetada con un número uno, luego la dos, etc.) Solución en la página 4.
7. En un paquete hay 20 tarjetas numeradas del 1 al 20. Se escogen al azar dos tarjetas. Calcula la probabilidad de que las dos que se han elegido sean la número 7 y la número 11 en cada una de las siguientes situaciones:
 - (a) Sacando las dos tarjetas a la vez.

- (b) Sacando las dos tarjetas de una en una sin reemplazamiento.
 (c) Sacando las dos tarjetas de una en una con reemplazamiento.

Solución en la página 4.

8. Se escogen al azar tres lámparas (a la vez, o de una en una sin reemplazamiento) de entre 15, y sabemos que de esas 15, cinco son defectuosas. ¿Cuál es la probabilidad de que al menos una de las tres elegidas sea defectuosa? Solución en la página 6.
9. Hallar la probabilidad de que al tirar tres dados aparezca el seis en uno de los dados (no importa cual), pero sólo en uno de ellos. Solución en la página 6.
10. Elegimos al azar cinco números del 1 al 10, con reemplazamiento. Puedes pensarlo así: en una caja hay 10 bolas numeradas del 1 al 10. Sacamos una bola, anotamos el número, devolvemos la bola a la caja, y la agitamos bien. ¿Cuál es la probabilidad de que no haya repeticiones y, por tanto, obtengamos cinco números distintos? Solución en la página 7.
11. Un equipo de investigación está preparando un nuevo kit con un test rápido para detectar el contenido de nitritos en muestras de agua. El test se ha probado en 950 muestras de agua tomadas en las inmediaciones de macrogranjas porcinas del centro y levante de la península ibérica. Las muestras se han analizado también en laboratorio para determinar si el nivel de nitritos sobrepasaba realmente el umbral de contaminación. La siguiente tabla resume los resultados:

		<u>Contaminado</u>		
		Sí	No	Total
<u>Resultado del test</u>	Positivo	436	5	441
	Negativo	14	495	509
Total		450	500	950

Con estos datos, responder a las siguientes preguntas:

- (a) ¿Cuál es la probabilidad de que una muestra no contaminada haya dado positivo en el test?
- (b) ¿Cuál es la probabilidad de que una muestra contaminada haya dado negativo en el test?
- (c) Sabiendo que una muestra ha dado positivo en el test, ¿cuál es la probabilidad de que esté contaminada?
- (d) Sabiendo que una muestra ha dado negativo en el test, ¿cuál es la probabilidad de que no esté contaminada?
- (e) ¿Cuál es la sensibilidad del test ($P(+|E)$)?
- (f) ¿Y su especificidad ($P(-|S)$)?

Solución en la página 7.

12. Una empresa produce anillas para identificación de tortugas marinas en tres fábricas. El volumen de producción diario es de 500, 1000 y 2000 unidades respectivamente. Se sabe que la fracción de producción defectuosa de las tres fábricas es de 0.005, 0.008, 0.010 respectivamente. Si se selecciona una anilla de forma aleatoria del total de producción de un día y se descubre que es defectuosa,
- Si eligen una anilla al azar, ¿qué probabilidad hay de que sea defectuosa?
 - ¿de qué fábrica es más probable que provenga esa anilla? Solución en la página 8.

Soluciones

• **Ejercicio 1, pág. 1** Si escribes el espacio muestral, $(C, C), (C, X), (X, C), (X, X)$ verás que hay 4 casos posibles y 3 favorables.

• **Ejercicio 2, pág. 1**

1. El evento "permanecer ingresado durante más de dos días" está contenido (estrictamente) en el evento "permanecer hospitalizado durante más de un día", en el sentido de que si alguien está hospitalizado 2 días o más ha tenido que estar (forzosamente) hospitalizado un día o más. Por tanto, el primer evento no puede tener una probabilidad (estrictamente) mayor que el segundo.
2. Los eventos $S =$ "que llueva el sábado" y $D =$ "que llueva el domingo" son independientes y ambos tienen probabilidad $1/2$. Por las reglas de la probabilidad

$$P(\text{que llueva el fin de semana}) = P(S \cup D) = P(S) + P(D) - P(S \cap D) =$$

$$P(S) + P(D) - P(S)P(D) = \frac{1}{2} + \frac{1}{2} - \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{4} \neq 1$$

• **Ejercicio 3, pág. 1** Hay varios enfoques posibles. Uno de ellos, el más sistemático, consiste en definir los eventos

H =ser hombre

C =tener los ojos marrones

Con esa notación, la probabilidad pedida es $P(H \cup C)$. Un recuento y la regla de Laplace permite determinar que

$$P(H) = 10/30 \quad P(C) = 15/30 \quad P(H \cap C) = 5/30$$

de donde, usando las reglas de cálculo de probabilidades, se tiene

$$P(H \cup C) = P(H) + P(C) - P(H \cap C) = 10/30 + 15/30 - 5/30 = 2/3$$

• **Ejercicio 4, pág. 1** Si definimos los eventos $Q =$ "suspender química" y $E =$ "suspender Estadística", el enunciado nos da la siguiente información

$$P(Q) = 0.15 \quad P(E) = 0.25 \quad P(Q \cap E) = 0.1$$

Lo primero es traducir cada pregunta al lenguaje de la probabilidad, y a continuación, hacer el cálculo correspondiente

$$1. P(E|Q) = \frac{P(E \cap Q)}{P(Q)} = \frac{0.1}{0.15} = 0.667$$

$$2. P(Q|E) = \frac{P(E \cap Q)}{P(E)} = \frac{0.1}{0.25} = 0.4$$

$$3. P(E \cup Q) = P(E) + P(Q) - P(E \cap Q) = 0.25 + 0.15 - 0.1 = 0.3$$

• **Ejercicio 6, pág. 1** Si se extraen de una en una y sin reemplazamiento, el resultado de cada extracción influye en los siguientes (los casos posibles ya no son los mismos, porque hay una muestra menos). Llamamos E_i al evento "sale la muestra i en la extracción i -ésima. La probabilidad pedida es

$$P(E_1 \cap E_2 \cap \dots \cap E_6) =$$

Como los eventos no son independientes (porque cada extracción condiciona las demás) hay que usar la regla de la multiplicación

$$\begin{aligned} &= P(E_1)P(E_2|E_1)P(E_3|E_1 \cap E_2) \dots P(E_6|E_1 \cap E_2 \cap \dots \cap E_5) = \\ &= \frac{1}{6} \frac{1}{5} \frac{1}{4} \frac{1}{3} \frac{1}{2} \frac{1}{1} = \frac{1}{6!} \end{aligned}$$

donde cada fracción representa casos favorables (1) entre casos posibles (número de elementos que quedan por extraer).

Un enfoque alternativo es el siguiente: supón que extraes las 6 de golpe y les asignas un *orden* (al azar) para ser analizadas. En este caso hay 1 caso favorable (que el orden sea el natural) y $6!$ casos posibles (número de formas de ordenar 6 elementos) por lo que la probabilidad es $\frac{1}{6!}$, donde puedes calcular $6!$ con la orden

```
factorial(6)
```

```
[1] 720
```

• **Ejercicio 7, pág. 1**

1. Empezamos suponiendo que sacamos dos fichas a la vez. Usaremos la regla de Laplace para calcular la probabilidad. Hay un caso favorable, que es sacar 7, 11. Observa que el orden NO es importante (aunque escribamos los dos números en cierto orden...7,11 y 11,7 se refieren al mismo evento, porque se extraen dos elementos de una vez). El número de casos posibles es el número de subconjuntos de dos tarjetas que podemos hacer a partir de 20 tarjetas, sin que importe el orden. Es decir, combinaciones de 20 elementos tomados de 2 en 2. Podemos usar R para hacer el cálculo con la función `choose()`. Si no recuerdas su significado, usa el tabulador

```
choose(20,2)
```

```
[1] 190
```

```
1/choose(20,2)
```

```
[1] 0.005263158
```

2. Supongamos ahora que sacamos las dos fichas una tras otra SIN reemplazamiento. En este caso debemos tener en cuenta el orden, ya que el experimento tendría resultado positivo tanto si primero sale un 7 y después el 11 como si sucediera al revés. Si denotamos por

$A = \text{sacar-primer-un-7-y-despues-un-11}$ $B = \text{sacar-primer-un-11-y-despues-un-7}$

y queremos calcular $P(A \cup B)$. Como ambos eventos son incompatibles (no puedo sacar primero un 11 y a la vez un 11) resulta que $P(A \cup B) = P(A) + P(B)$. Entonces

$$\begin{aligned} P(A = \text{sacar} - \text{primer} - \text{un} - 7 - \text{y} - \text{despues} - \text{un} - 11) &= \\ &= P(\text{saca} - \text{un} - 7)P(\text{sacar} - \text{un} - 11 | \text{ha} - \text{salido} - \text{un} - 7) \\ &= \frac{1}{20} \frac{1}{19} \end{aligned}$$

De la misma manera,

$$P(B = \text{sacar} - \text{primer} - \text{un} - 11 - \text{y} - \text{despues} - \text{un} - 7) = \frac{1}{20} \frac{1}{19}$$

Por tanto, hay Y la probabilidad total es

$$(1/20) * (1/19) + (1/20) * (1/19)$$

$$[1] \quad 0.005263158$$

3. Para este último caso podemos reutilizar el razonamiento anterior suponiendo que hay reemplazamiento, de modo que la probabilidad de la segunda extracción es idéntica a la primera: hay 1 caso favorable y 20 posibles (porque hay reemplazamiento). Así

$$\begin{aligned} P(A = \text{sacar} - \text{primer} - \text{un} - 7 - \text{y} - \text{despues} - \text{un} - 11) &= \\ &= P(\text{saca} - \text{un} - 7)P(\text{sacar} - \text{un} - 11 | \text{ha} - \text{salido} - \text{un} - 7) \\ &= \frac{1}{20} \frac{1}{20} \end{aligned}$$

De la misma manera,

$$P(B = \text{sacar} - \text{primer} - \text{un} - 11 - \text{y} - \text{despues} - \text{un} - 7) = \frac{1}{20} \frac{1}{20}$$

Y la probabilidad total es

$$(1/20) * (1/20) + (1/20) * (1/20)$$

$$[1] \quad 0.005$$

Por tanto, la **conclusión** es que extraer k elementos de uno en uno sin reemplazamiento es equivalente a extraer los k elementos de una vez, y que ambos procedimientos son diferentes de extraer k elementos de uno en uno y con reemplazamiento.

• **Ejercicio 8, pág. 2** En el ejercicio 7 hemos visto que es importante tener en cuenta si en un experimento se hacen las extracciones con o sin reemplazamiento, y que este último caso equivalía a hacer extracciones simultáneas.

En este caso supondremos que las lámparas se extraen simultáneamente, lo que equivale a extraerlas secuencialmente y sin reemplazamiento.

Si llamamos $X =$ "número de lámpara defectuosas", se pide calcular $P(X \geq 1)$. Es decir, que haya o una, o dos o las tres lámparas defectuosas. Esto es equivalente a calcular la probabilidad del evento complementario $P(X = 0)$ para obtener $P(X \geq 1) = 1 - P(X = 0)$. Llamamos A_i al evento "la lámpara extraída en i -ésimo lugar no es defectuosa". Y el cálculo que resuelve el ejercicio es

$$P(X \geq 1) = 1 - P(X = 0) = 1 - P(A_1 \cap A_2 \cap A_3)$$

Los tres eventos no son independientes, ya que el hecho de elegir (extraer) una lámpara, cambia tanto el número de casos favorables como el de casos posibles. Por tanto, hay que usar la regla de la multiplicación:

$$1 - P(A_1 \cap A_2 \cap A_3) = 1 - P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)$$

para obtener

$$1 - (10/15) * (9/14) * (8/13)$$

$$[1] \quad 0.7362637$$

Ten en cuenta que si hubiera habido reemplazamiento cada extracción no condicionaría la extracción anterior, es decir, los eventos A_i serían independientes, y tendrías

$$1 - P(A_1 \cap A_2 \cap A_3) = 1 - P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) = P(A_1)P(A_2)P(A_3)$$

$$1 - (10/15) * (10/15) * (10/15)$$

$$[1] \quad 0.7037037$$

En el próximo tema (variable aleatoria) veremos que esa independencia permite modelizar este experimento con algo que llamaremos variable aleatoria binomial.

• **Ejercicio 9, pág. 2** Claramente hay 6^3 casos posibles (piensa en el árbol que usamos en el libro, en la Sección 3.6, para calcular el número de permutaciones de orden n). Por otro lado, los casos favorables son $3 * 5^2$, ya que el 6 puede salir sólo en uno de los tres lanzamientos y hay 5^2 posibles resultados (diferentes de 6) para los otros dos dados. La probabilidad pedida es $\frac{3*5^2}{6^3}$, es decir,

$$3*5^2/6^3$$

$$[1] \quad 0.3472222$$

También se puede resolver con la variable binomial $X =$ "número de seises al lanzar 3 dados", con probabilidad de éxito en cada lanzamiento $1/6$ y de fracaso $5/6$.

$$P(X = 1) = \binom{3}{1} \left(\frac{1}{6}\right) \left(\frac{5}{6}\right)^2$$

```
dbinom(1, size = 3, prob = 1/6)
```

```
[1] 0.3472222
```

• **Ejercicio 10, pág. 2** Como hay reemplazamiento (recuerda el ejercicio 7), si llamamos E_i a la extracción i , se calcula

$P(\text{salgan diferentes}) = P(\text{cualquier número}) * P(\text{un núm diferente de } E_1) * P(\text{un núm diferente de } E_1 \text{ y } E_2) * \dots * P(\text{núm diferente de las otras extracciones})$

Al contar casos posibles y casos favorables en cada factor se obtiene

```
(10/10)*(9/10)*(8/10)*(7/10)*(6/10)
```

```
[1] 0.3024
```

Como hay reemplazamiento, otro enfoque consiste en pensar casos favorables entre casos posibles para las 5 extracciones. Los casos favorables son las variaciones de 10 elementos tomados de 5 en 5, y los casos posibles son las variaciones con repetición de 10 elementos tomados de 5 en 5. El cociente es

$$\frac{10 * 9 * 8 * 7 * 6}{10^5}$$

• **Ejercicio 11, pág. 2** En todos los casos se trata de probabilidades condicionadas

1. $P(+|No) = \frac{P(+ \cap No)}{P(No)} = \frac{5/950}{500/950}$. A la vista de este resultado, podemos usar directamente las frecuencias absolutas de cada evento (y no las relativas). Así procedemos en los siguientes apartados
2. $P(-|Si) = \frac{P(- \cap Si)}{P(Si)} = \frac{14}{450}$.
3. $P(Si|+) = \frac{P(+ \cap Si)}{P(+)} = \frac{436}{441}$.
4. $P(No|-) = \frac{P(- \cap No)}{P(-)} = \frac{495}{509}$.
5. $P(+|Si) = \frac{P(+ \cap Si)}{P(Si)} = \frac{436}{450}$.
6. $P(-|No) = \frac{P(- \cap No)}{P(No)} = \frac{495}{500}$.

• **Ejercicio ??, pág. ??** La información que da el enunciado es (denotaremos E=enfermo y S=sano)

- $p(+|S) = 0.04 \Rightarrow p(-|S) = 0.96$
- $p(-|E) = 0.05 \Rightarrow p(+|E) = 0.95$
- $p(E) = 0.07 \Rightarrow p(S) = 0.93$

Se pide, primero,

$$P(E|+) = \frac{P(+ \cup S)}{P(S)} = \frac{0.95 * 0.07}{0.95 * 0.07 + 0.04 * 0.93} = 0.641$$

y, por otro lado,

$$P(S|-) = \frac{P(S \cup -)}{P(-)} = \frac{0.96 * 0.93}{0.05 * 0.07 + 0.96 * 0.93} = 0.996$$

- **Ejercicio 12, pág. 2** Llamaremos F_k al evento "la anilla viene de la fábrica k " y D al evento "la anilla es defectuosa".

El enunciado proporciona los siguientes datos:

$$P(F_1) = 500/3500, \quad P(F_2) = 1000/3500, \quad P(F_3) = 2000/3500$$

$$P(D|F_1) = 0.005, \quad P(D|F_2) = 0.008, \quad P(D|F_3) = 0.01$$

- En este caso hay que calcular $P(D)$. Teniendo en cuenta que la anilla puede venir de tres factorías diferentes, podemos aplicar el teorema de la probabilidad total:

$$P(D) = P(F_1)P(D|F_1) + P(F_2)P(D|F_2) + P(F_3)P(D|F_3)$$

Sustituyendo, se tiene

```
# usar notaci\ 'on vectorial para reaprovechar
# las expresiones
Prob_fabr = c(500/3500, 1000/3500, 2000/3500)
Prob_defecto_x_fabr = c(0.005, 0.008, 0.01)
sum(Prob_fabr * Prob_defecto_x_fabr)

[1] 0.008714286
```

- Para determinar de qué fábrica es más probable que provenga esa anilla hay que calcular

$$P(F_1|D), \quad P(F_2|D), \quad P(F_3|D)$$

Es decir, hay que usar tres veces el teorema de Bayes que, para la fábrica i , dice así

$$P(F_i|D) = \frac{P(F_i)P(D|F_i)}{P(F_1)P(D|F_1) + P(F_2)P(D|F_2) + P(F_3)P(D|F_3)}$$

```
# usar notaci\ 'on vectorial para reaprovechar
# las expresiones
Prob_fabr = c(500/3500, 1000/3500, 2000/3500)
Prob_defecto_x_fabr = c(0.005, 0.008, 0.01)
# P(F_1|D)
Prob_fabr[1]*Prob_defecto_x_fabr[1]/(sum(Prob_fabr*Prob_defecto_x_fabr))

[1] 0.08196721
```

```
#  $P(F_2/D)$ 
Prob_fabr[2]*Prob_defecto_x_fabr[2]/(sum(Prob_fabr*Prob_defecto_x_fabr))

[1] 0.2622951

#  $P(F_3/D)$ 
Prob_fabr[3]*Prob_defecto_x_fabr[3]/(sum(Prob_fabr*Prob_defecto_x_fabr))

[1] 0.6557377
```