

Las aves parásitas de cría ponen huevos en nidos de otras especies (hospedador), las cuales incuban los huevos y crían al pichón parásito. En un bosque de talas de la provincia de Buenos Aires hay dos especies hospederas que son indistinguibles a simple vista. Una de las principales diferencias entre estas especies radica en el grado de discriminación y remoción de huevos parásitos de sus nidos. Una de las especies es “aceptadora” de huevos parásitos ($Y = 0$), ya que remueve del nido sólo el 30% de los huevos parásitos, mientras que la otra especie es “rechazadora” ($Y = 1$) ya que remueve el 80% de los huevos parásitos presentes en su nido. Además, se sabe que el 90% de los nidos del bosque corresponden a la especie “aceptadora”, mientras que apenas el 10% restante son nidos de la especie “rechazadora”.

Cálculo de probabilidades: Se elige al azar un nido del bosque y se colocan $n = 8$ huevos parásitos. Denotemos con X a la variable aleatoria que indica el número de huevos removidos del nido. Asuma que, en cada nido, la remoción (o no) de los diferentes huevos se realiza de manera independiente.

1. Calcule la probabilidad de que 5 de los huevos parásitos sean removidos del nido si se sabe que el nido parasitado es de la especie aceptadora.
2. Calcule la probabilidad de que x de los huevos ajenos sean removidos del nido si se sabe que el nido parasitado es de la especie aceptadora, para $0 \leq x \leq 8$, y represente las probabilidades calculadas en un diagrama de barras. Incluya un título pertinente en el gráfico.
3. Calcule la probabilidad de que 5 de los huevos parásitos sean removidos del nido si se sabe que el nido parasitado es de la especie rechazadora.
4. Calcule la probabilidad de que x de los huevos parásitos sean removidos del nido si se sabe que el nido parasitado es de la especie rechazadora, para $0 \leq x \leq 8$, y represente las probabilidades calculadas en un diagrama de barras. Incluya un título pertinente en el gráfico.
5. Calcule la probabilidad de que 5 de los huevos parásitos sean removidos del nido.
6. Calcule la probabilidad de que x de los huevos parásitos sean removidos del nido, para $0 \leq x \leq 8$, y represente las probabilidades calculadas en un diagrama de barras. Incluya un título pertinente en el gráfico.
7. Haga la tabla de la función de probabilidad conjunta del vector (X, Y) .

Algunos datos: El archivo `depredadosclasificadas.txt` contiene mediciones correspondientes a nidos elegidos al azar en el bosque, registrándose en cada uno de ellos el número de huevos parásitos removidos (primera columna) y la especie del hospedador (segunda columna), determinada mediante un análisis genético. Tenemos entonces observaciones de (X, Y) donde X indica la cantidad de huevos removidos del nido, mientras que $Y = 0$ si el hospedador es aceptador, $Y = 1$ indica que el hospedador es rechazador.

8. Calcule la proporción de hospedadores aceptadores observados en la muestra. ¿Qué probabilidad está estimando esta proporción?
9. Calcule la proporción de nidos donde se removieron $x = 5$ huevos parásitos entre los hospedadores aceptadores. ¿Qué probabilidad está estimando esta proporción?
10. Compare la función de probabilidad puntual condicional de X dado $Y = 0$ con la estimada, presentando ambas en un mismo gráfico de barras.
11. Compare la función de probabilidad puntual condicional de X dado $Y = 1$ con la estimada, presentando ambas en un mismo gráfico de barras.
12. Hacer la tabla (con dos decimales) de las frecuencias relativas observadas de los pares (x, y) , para $0 \leq x \leq 8$, $0 \leq y \leq 1$. Indique qué valores está estimando con esta tabla.

Y / X	0	1	2	3	4	5	6	7	8
0									
1									

Clasificación: Toy Example

13. Considere la regla de clasificación que determina que un nido corresponde a la especie rechazadora ($Y = 1$) si la cantidad de huevos removidos es mayor o igual a $t = 5$. Caso contrario, determina que el nido es de la especie hospedadora aceptadora ($Y = 0$). Implemente una función **regla5** que dado x , el número de huevos agredidos, devuelva el valor 1 si $x \geq 5$, y devuelva el valor 0, caso contrario. Observe que la regla de clasificación es una función g que para cada x determina un valor en $\{0, 1\}$. Calcule \hat{R}_5 , la tasa de error empírica correspondiente a esta regla, utilizando los datos de `depredadosclasificadas.txt`. Es decir, calcule la proporción de datos mal clasificados según esta regla.

14. Considere la regla de clasificación que determina que un nido corresponde a la especie rechazadora ($Y = 1$) si la cantidad de huevos removidos es par. Caso contrario, determina que el nido es de un hospedador aceptador ($Y = 0$). Implemente una función **rechazadorpar** que dado x , el número de huevos agredidos, devuelva el valor 1 si x es par, y devuelva el valor 0, caso contrario. Calcule la tasa de error empírica correspondiente a esta regla, utilizando los datos depredadosclasificadas. Es decir, calcule la proporción de datos mal clasificados según esta regla.
15. ¿Cuál de las reglas consideradas en los items anteriores (**regla5** o **rechazadorpar**) prefiere?
16. Para cada $t \in \{0, \dots, 8\}$, considere la regla de clasificación que determina que un nido corresponde a la variedad rechazadora ($Y = 1$) si la cantidad de huevos removidos es mayor o igual a t . Caso contrario, determina que el nido es del hospedador aceptador ($Y = 0$). Implemente una función **reglacorte** que dados x y t , el número de huevos removidos y el punto de corte, respectivamente, devuelva el valor 1 si $x \geq t$, y devuelva el valor 0, caso contrario. Calcule \hat{R}_t , la tasa de error empírica correspondiente a regla con punto de corte en t utilizando los datos depredadosclasificadas. Es decir, calcule la proporción de datos mal clasificados según cada regla, para $0 \leq t \leq 8$.
17. ¿Con qué punto de corte obtiene la regla con menor error de clasificación empírico?

(*) **Bonus Track**

18. Una regla de clasificación h puede expresarse mediante una tira de ceros y unos que indica cuánto vale h para los diferentes valores de x . Por ejemplo, la tira $(0, 0, 0, 1, 1, 1, 0, 1, 1)$ indica que $h(0) = 0$, $h(1) = 0$, $h(2) = 0$, $h(3) = 1$, $h(4) = 1$, etc. Implemente la función **ErrorClassTRUE** que tenga por input un vector de ceros y unos y de longitud 9 y devuelva el error de clasificación medio de la regla h , dado por $P(h(X) \neq Y)$.
19. Calcule el error de clasificación medio correspondiente a la regla con punto de corte en t , para cada $t \in \{0, \dots, 8\}$. Recuerde que si $h(X)$ es una regla de clasificación, su error de clasificación medio está dado por

$$\mathbb{P}(h(X) \neq Y) .$$

Grafique el error de clasificación medio como función de t .

20. Implemente una función **rHuevosNido** que tenga por input un número natural n y genere n pares (X_i, Y_i) , $0 \leq i \leq n$, según la distribución descrita en el enunciado de la lista.