



Practica 1

Ej. 1 En una ciudad hay 20 familias, 4 de ellas tienen un hijo, 8 tienen 2 hijos, 5 familias tienen 3 hijos, hay 2 familias con 4 hijos y una con 5 hijos.

- a) Se selecciona al azar una familia, cual es la probabilidad de que tenga i hijos con $i = 1, 2, 3, 4, 5$.
- b) Se selecciona al azar un niño, cual es la probabilidad de que pertenezca a una familia de i hijos con $i = 1, 2, 3, 4, 5$.

Ej. 2 En una urna hay 3 bolitas rojas y 4 bolitas blancas. Si se seleccionan dos bolitas sin reposición, ¿Cuál es la probabilidad de que la primera sea roja y la segunda blanca? Cual es la probabilidad de que la primera sea roja? Cual es la probabilidad de que la segunda sea roja?

Ej. 3 El Problema Del Cumpleaños:

- a) Calcular la probabilidad de que entre los primeros 50 invitados en llegar a la fiesta hayan al menos dos que cumplan años en la misma fecha.
- b) Utilizando la función `sample()` simular las fechas de cumpleaños (con enteros de 1 a 365) de los primeros 50 invitados que ingresan a la fiesta. Usar la función `uplicated()` para chequear si se produce alguna repetición en las fechas de cumpleaños. Observación: usando `sum(duplicated())` podés contar cuántas repeticiones se producen.
- c) Utilizar la función `sample()` para realizar 10 simulaciones de 50 invitados ingresando a la fiesta y calcular la frecuencia relativa con la que en ese grupo de 50 personas hayan al menos dos que cumplan años en la misma fecha. Repetir para 1000 y 100000 simulaciones. Comparar cada una de las 3 frecuencias relativas con la probabilidad calculada en el inciso (a).
- d) Identificar la cantidad mínima de invitados que deben ingresar para que la probabilidad de que se produzca al menos una repetición en las fechas de cumpleaños sea superior a 0,6.

Practica 2

Ej. 1 Se lanzan dos dados *justos*. Consideramos el evento

$$S = \{\text{la suma de los dos resultados obtenidos es } 7\}$$

Calcular la probabilidad de S condicional a (dado que):

- a) La suma sea 7.
- b) La suma es mayor que 9.
- c) El 2do dado salió par.
- d) Al menos un dado salió impar.
- e) Ambos dados salieron iguales.
- f) Ambos dados salieron distintos.

Ej. 2 Un banco tiene una cartera de clientes de crédito de consumo con riesgo alto, medio y bajo. Dichos clientes tienen probabilidades de impago (no devolver el dinero prestado) de 0,02, 0,01 y 0,0025 respectivamente.

- a) Si el porcentaje de clientes de alto riesgo es 10 %, de riesgo medio es 20 % y de bajo riesgo es 70 %, entonces ¿cuál es la probabilidad de que un cliente seleccionado al azar de esta cartera haga un impago?
- b) ¿Qué ocurre con la probabilidad calculada en (a) si la cartera tiene una distribución diferente: alto riesgo 80 %, riesgo medio 15 % y bajo riesgo 5 %? ¿Tienen sentido estos resultados? ¿Qué conclusiones puedes sacar de este problema?

Ej. 3 Si A y B son dos eventos independientes, probar que A^c y B , A y B^c y A^c y B^c también son independientes.

Ej. 4 En una urna hay 3 bolitas rojas y 4 bolitas blancas. Si se seleccionan dos bolitas sin reposición, Cual es la probabilidad de que la segunda sea roja?

Ej. 5 En un bolillero hay 20 bolillas rojas y 8 verdes, de las cuales 12 están marcadas (6 de cada color). Considerando los eventos:

- A = “La bolilla extraída es roja”.
- B = “La bolilla extraída está marcada”.

Realizar los siguientes cálculos utilizando R:

- a) Calcular la frecuencia relativa (f_A) de bolillas rojas en 1000 extracciones.
- b) Usando las mismas 1000 extracciones, calcular la frecuencia relativa ($f_{A|B}$) de bolillas rojas entre las extracciones en que la bolilla obtenida está marcada.
- c) Comparar f_A y $f_{A|B}$ respecto a cuál se aproxima mejor a $P(A)$ y $P(A | B)$, respectivamente.
- d) Calcular 100 frecuencias relativas de cada tipo e identificar cuál toma un valor más cercano a la probabilidad que aproxima. Contar en cuántos casos se verifica la respuesta del inciso anterior.

Ej. 6 Una enfermedad afecta al 0.1 % de la población. Un test tiene una sensibilidad de 0.95 y una especificidad de 0.99. Es decir la probabilidad de que el test resulte positivo si la persona esta enferma es 0.95. Mientras que la probabilidad de que el test de negativo si la pesona es sana es 0.99.

- a) Calcular la probabilidad de que una persona esté realmente enferma.
- b) Si la persona le dio su examen positivo cual es la probabilidad de que la persona este enferma.
- c) Suponiendo que la enfermedad afecta al p % de la población, graficar usando R la probabilidad de estar enfermo dado un examen positivo vs p .

Ej. 7 Un Método de Sondeo Un investigador conducirá un sondeo con el objetivo de estimar el porcentaje de gente en la población que consume drogas ilícitas.

Llamemos p a la probabilidad de que una persona elegida al azar consuma drogas. Elegimos una persona al azar y le preguntamos si consumió drogas ilícitas.

- a) Supongamos que la persona no miente. Cuál es la probabilidad de que responda que sí consumió.
- b) Supongamos que si la persona no consumo no miente, pero si consumió la probabilidad de que conteste que no consumo es r . Cuál es la probabilidad de que responda que sí consumió.

- c) Preocupado de que mucha gente mentirá si se le pregunta directamente Ha consumido drogas ilícitas alguna vez? el investigador usará un método de sondeo conocido como de respuesta aleatoria. El método consiste en lo siguiente, se elige una persona al azar y se le da un dado y una moneda. Se le dice que tire un dado, si sale 1 ó 2, a continuación se le pide que tire una moneda. Si sale cara entonces la persona deberá responder que consume drogas y si sale ceca la persona deberá responder que no consume drogas. En cambio si el resultado del dado es 3, 4, 5, ó 6 se le pide a la persona que conteste a la verdad. (Obviamente que el encuestador no mirará el resultado del dado ni la moneda). Calcular la probabilidad que la persona responda que si consumió.

Practica 3

Ej. 1 Sobreventa: Supongamos que un avión tiene 100 asientos. La aerolínea vende 110 pasajes. La probabilidad de que un pasajero se presente el día del vuelo a tomar el avión es 0,9. Y supongamos que cada pasajero se presenta o no a volar independientemente del resto. S

- Calcule la probabilidad de que todos los pasajeros que se presentan a volar puedan hacerlo.
- Si la aerolínea está evaluando vender un número M pasajes en lugar de 110. Grafique la probabilidad anterior en función de M .
- Sea Y = “Número de pasajeros que se presentan a volar”. Cuál es la distribución de Y .
- Sea Z = “Número de pasajeros que no pueden volar”. Que distribución tiene $Z = \max\{0, Y - 100\}$.
- Calcule la $E(Y)$ y $E(Z)$.
- Supongamos que cada pasaje vendido le significa 500 dólares de ganancia a la aerolínea, cada pasajero que no puede volar una pérdida de 600 dólares y cada vuelo un costo operativo de 4000 dólares. Sea la variable aleatoria X = “Ganancia Neta”, definida como: $X = 500 \times 110 - 600 \times \max\{0, Y - 100\} - 40000$, calcular $E(X)$.
- Simula 100 vuelos del avión (es decir considera 100 instancias distintas donde la compañía vende 110 asientos teniendo a disposición para los pasajeros solo 100) y grafica un histograma con las realizaciones de X (la ganancia neta en cada uno de los 100 vuelos).
- Ejecuta varias veces tus líneas de código y actualiza la grafica con la distribución empírica (simulada) de la ganancia neta en cada actualización. Que puedes notar? A que se debe el comportamiento cambiante del histograma de la ganancia neta
- Repita el ejercicio anterior, pero ahora simula $r = 10,000$ vuelos del avión. Que diferencias notas con respecto al inciso anterior.
- Como podrías aproximar la media y la varianza de la variable aleatoria X utilizando tus simulaciones numéricas?
- Que ocurre con la distribución de la ganancia neta si mantenemos la política de sobre-vender 10 asientos del avión pero la probabilidad de que un pasajero no se presente a volar pasa del 0,1 al 0,05?
- Mediante simulaciones numéricas optimiza la estrategia de sobre-venta de pasajes (elige el valor óptimo de M) de tal forma que la probabilidad de obtener beneficios positivos sea de como mínimo el 90 %.

Ej. 2 Clasificación: 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, ya que remueve del nido sólo el 30 % de los huevos parásitos, mientras que la otra especie es rechazadora 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.

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.

- 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.
- 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 $0 \leq x \leq 8$.
- 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.
- 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 $0 \leq x \leq 8$.
- Calcule la probabilidad de que 5 de los huevos parásitos sean removidos del nido.
- Calcule la probabilidad de que x de los huevos parásitos sean removidos del nido, para $0 \leq x \leq 8$.

Ej. 3 La Catadora de Té : Una señora afirma que al probar una taza de té con leche puede distinguir qué fue lo primero que se echó en la taza: el té o la leche. Cómo podríamos hacer un experimento para determinar si la señora sabe distinguir?

Procuramos realizar un experimento que indique evidencia a favor de la señora. Para ello, procedemos de la siguiente manera: le damos $n = 12$ tazas preparadas de diferentes maneras y contamos el número de veces que la mujer acierta el orden en el que fue preparado el té. Sea X la variable aleatoria que cuenta el número de aciertos en los $n = 12$ ensayos.

- ¿Qué distribución propone para la variable aleatoria X ? ¿Cuáles son los parámetros de esta distribución?
- ¿Cuál es la probabilidad de que acierte las 12 tazas si responde al azar?

¿Le daremos la oportunidad de equivocarse? ¿Cuántas tazas serán suficientes para que confiemos en que la señora tiene la habilidad que ella afirma?

Propongamos diferentes reglas que podemos utilizar para decidir que la señora es una entendida en el tema.

- Criterio 0: Determinamos que sabe si acierta las 12 veces.
- Criterio 1: Determinamos que sabe si acierta 11 o más veces.
- Criterio 2: Determinamos que sabe si acierta 10 o más veces.
- Criterio 3: Determinamos que sabe si acierta 9 o más veces.

- Supongamos que la señora responde al azar. Para cada uno de los criterios propuestos, calcule la probabilidad de que consideremos que la señora sabe.

- b) Que criterio debemos utilizar si queremos que la probabilidad de equivocarnos al decidir que sabe cuando en realidad responde al azar no sea superior el valor 0,02? Es decir, que número mínimo de aciertos serían suficientes para que confiaras en que la señora tiene la habilidad si queremos que la probabilidad de equivocarnos cuando en realidad responde al azar no sea superior el valor 0,02?
- c) Supongamos que la señora responde bien en 10 ocasiones. ¿Cuál es el criterio más exigente que nos permite decir que la señora sabe? ¿Cuál es la probabilidad de error de dicho criterio cuando en realidad responde al azar? ¿Cuál es la probabilidad de observar un valor tan o más grande que el observado, asumiendo que responde al azar?
- d) Si en realidad la señora tiene un verdadero don y acierta al orden en que se pone la leche y el té con probabilidad igual a 0,8. Para cada uno de los criterios propuestos, calcule la probabilidad de que consideremos que la señora sabe. Y si la probabilidad de acertar con el orden es 0,9, ¿como cambia la probabilidad de decidir que la señora sabe con cada uno de los criterios?
- e) Si en realidad la señora tiene un verdadero don y acierta al orden en que se pone la leche y el té con probabilidad igual a 0,8. Para cada uno de los criterios propuestos, calcule la probabilidad de que consideremos que la señora NO sabe. Y si la probabilidad de acertar con el orden es 0,9, ¿como cambia la probabilidad de decidir que la señora NO sabe con cada uno de los criterios?

Practica 4

Ej. 1 Una empresa vende dos variedades de soja. La variedad I tiene un rendimiento por ha. que puede considerarse una variable aleatoria con distribución $\mathcal{N}(37, 25)$; y la variedad II tiene un rendimiento por ha. que puede considerarse $\mathcal{N}(40, 25)$. Un cliente realizó una compra de semillas de la variedad II y antes de continuar comprando a esta empresa, quiere asegurarse de que las semillas que le enviaron realmente pertenecen a esa variedad y no a la variedad I.

Sea X la variable aleatoria que mide el rendimiento de soja en una hectárea cultivada.

- a) ¿Qué distribución tiene X si se utiliza la variedad I?
- b) ¿Qué distribución tiene X si se utiliza la variedad II?
- c) ¿Cuán alto debe ser el valor de X para considerar que, efectivamente, le enviaron la variedad II?

Propongamos diferentes reglas que podemos utilizar para determinar que las semillas plantadas son de la variedad II

- Criterio 0: Determinamos que son de alto rendimiento si obtenemos 30 o más tn./ha.
- Criterio 1: Determinamos que son de alto rendimiento si obtenemos 33 o más tn./ha.
- Criterio 2: Determinamos que son de alto rendimiento si obtenemos 35 o más tn./ha.
- Criterio 3: Determinamos que son de alto rendimiento si obtenemos 38 o más tn./ha.

- 1) Supongamos que las semillas son de la variedad I. Para cada uno de los criterios propuestos, calcule la probabilidad de que consideremos que las semillas son de alto rendimiento.

- 2) Que criterio debemos utilizar si queremos que la probabilidad de equivocarnos al decidir que las semillas son de alto rendimiento cuando en realidad son de la variedad I no sea superior el valor 0,05?
- 3) Supongamos al probar las semillas obtenemos 31 toneladas en una hectárea. ¿Cuál es el criterio más exigente que nos permite decir que se utilizó la variedad II? ¿Cuál es la probabilidad de error de dicho criterio cuando en realidad se utiliza la variedad I? ¿Cuál es la probabilidad de observar un valor tan o más grande que el observado, asumiendo que se utiliza la variedad I?
- 4) Supongamos que las semillas son de la variedad II, para cada uno de los criterios propuestos, calcule la probabilidad de que consideremos que las semillas son de la variedad I.

Practica 5

Ej. 1 Clasificación: 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 hospedadoras 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, ya que remueve del nido sólo el 30 % de los huevos parásitos, mientras que la otra especie es rechazadora 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. 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. Y sea Y la variable que vale 1 si la especie es aceptadora y 0 si es rechazadora

- a) Haga la tabla de la función de probabilidad conjunta del vector (X, Y) .
- b) Considere la regla de clasificación que determina que un nido corresponde a la especie rechazadora 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. Calcule el error de medio de clasificación, es decir, calcule la probabilidad de que la regla de clasificación, clasifique mal.
- c) Considere la regla de clasificación que determina que un nido corresponde a la especie rechazadora si la cantidad de huevos removidos es par. Caso contrario, determina que el nido es de un hospedador aceptador. Calcule el error de medio de clasificación, es decir, calcule la probabilidad de que la regla de clasificación, clasifique mal.
- d)Cuál de las reglas consideradas en los items anteriores prefiere?
- e) Para cada $t \in \{0, \dots, 8\}$, considere la regla de clasificación que determina que un nido corresponde a la variedad rechazadora si la cantidad de huevos removidos es mayor o igual a t . Caso contrario, determina que el nido es del hospedador aceptador. Calcule el error de medio de clasificación correspondiente a cada regla con punto de corte en t para $0 \leq t \leq 8$.
- f) Con qué punto de corte obtiene la regla con menor error de clasificación?

Practica 6

Ej. 1

Practica 7

Ej. 1 Cuando se realiza una encuesta con preguntas delicadas, donde la gente tiene cierta resistencia a responder, se suele usar un método indirecto para hacer la pregunta. Supongamos que estamos interesados en conocer (o estimar) la proporción de mujeres que se realizaron un aborto. Un modo de realizar la encuesta sería el siguiente: a cada mujer le damos un dado y una moneda y le decimos que tire el dado. Si el resultado del dado es 3, 4, 5, o 6 la mujer responde la verdad y si sale 1 o 2 tira la moneda; si sale cara responde SI y si sale ceca responde NO.

Obviamente, el encuestador no ve el resultado del dado ni el de la moneda y por lo tanto no sabría si la respuesta del encuestado es la verdad o es producto del azar.

A partir de este mecanismo quisieramos poder estimar la verdadera proporción de mujeres que se hicieron un aborto que llamaremos p , es decir $p = \text{Probabilidad de que una mujer elegida al azar se haya realizado un aborto}$.

- a) Calcular (en función de p) la probabilidad de que una persona elegida al azar responda afirmativamente a la pregunta ($P(SI)$). Ayuda: condicione al resultado del dado.
- b) A partir de encuestar a n personas con este mecanismo, como estimaría la probabilidad de que alguien responda que SI, es decir $P(SI)$? Mas precisamente, si pensamos en n variables i.i.d W_1, \dots, W_n donde $W_i = 1$ si la mujer responde SI y 0 en caso contrario, que cuenta deberá hacer para estimar la proporción de gente que responde SI, $P(SI)$.
- c) Usando a) y b), proponga una cuenta para estimar p , la probabilidad de que una mujer elegida al azar se haya realizado un aborto. Llamemos \hat{p} a esta cuenta.
- d) Para convencernos que este mecanismo resulta efectivo, realice la siguiente simulación. Supongamos que el verdadero valor de $p = 0,1$ es decir la verdadera proporción de mujeres que se realizó un aborto. Simule $n = 100$ respuestas, W_1, \dots, W_n según el mecanismo de respuesta propuesto y compare con la verdadera proporción p . Repita esto $Nrep = 1000$ veces.
- e) Si $p = 0,1$, ¿ qué distribución tiene W_i ? Usando el Teorema Central del Límite, aproxime la probabilidad $P(0 < \hat{p} < 0,2)$.
- f) Supongamos como en el ejercicio anterior que $p = 0,1$ pero hacemos la encuesta directa y nadie miente. En ese caso, estimaríamos a p simplemente contando el porcentaje de gente que contesta si. Es decir, tendríamos X_1, \dots, X_n variables aleatorias i.i.d $Bi(1, p)$ y estimaríamos p con $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. Usando el Teorema Central del Límite, aproxime la probabilidad $P(0 < \bar{X}_n < 0,2)$.
- g) Explique la relación o las diferencias entre los dos item anteriores.