

Ejercicios

1. Estadística descriptiva

1.1 (1 punto) Se identifica las reservas de petróleo, en miles de millones de barriles, de cuatro regiones de América :

Estados Unidos	38,7
Sudamérica	22,6
Canadá	8,8
México	60,0

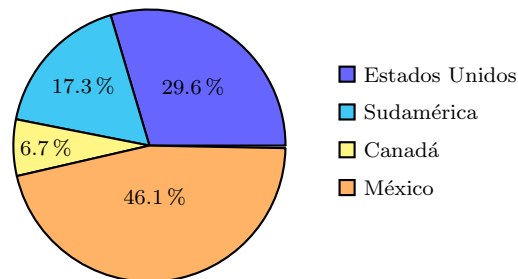
Representen estos datos con un gráfico circular (sean precisos).

Solution: Tenemos

$$38,7 + 22,6 + 8,8 + 60,0 = 130,1.$$

Entonces, las reservas en por ciento son

$$29,7\%, \quad 17,3\%, \quad 6,7\% \quad \text{y} \quad 46,1\%$$



1.2 ($1\frac{1}{2}$ puntos) Sea una muestra de tamaño 5 que satisface

$$\bar{x} = 104 \quad \text{y} \quad s^2 = 16.$$

Supongan que conocen 3 valores de la muestra: 102, 100 y 105. Calculen los dos otros valores.

Solution: Sean a y b los dos valores que faltan. Entonces

$$\frac{102 + 100 + 105 + a + b}{5} = \bar{x} = 104,$$

de tal manera que

$$a + b = 213.$$

También, tenemos

$$\frac{(102 - 104)^2 + (100 - 104)^2 + (105 - 104)^2 + (a - 104)^2 + (b - 104)^2}{4} = 16,$$

que se simplifica en

$$(a - 104)^2 + (b - 104)^2 = 43.$$

Podemos substituir por $b = 213 - a$ y obtenemos

$$(a - 104)^2 + (213 - a - 104)^2 = 43.$$

Se simplifica en

$$2a^2 - 426a + 22654 = 0.$$

Los raizes son 102,59 y 110,41. Entonces, hay dos soluciones posibles:

$$(a, b) = (102,59; 110,41) \quad \text{y} \quad (a, b) = (110,41, 102,59).$$

- 1.3 ($1\frac{1}{2}$ puntos) Un estudiante en química mide el número de pulgadas de agua evaporada durante 55 días del mes de julio divididos entre 4 años. Obtiene las observaciones siguientes:

0.0		2,6
0.1		1,4
0.2		1,1,1,3,3,4,5,5,5,6,9
0.3		0,0,2,2,2,3,3,3,3,4,4,5,5,5,6,6,7,8,9
0.4		0,1,2,2,2,3,4,4,4,5,5,5,7,8,8,8,9,9
0.5		2,5,6

Entonces, el valor mínimo es 0,02 y el valor máximo es 0,56. Calculen (a) la media, (b) la mediana, (c) la moda, (d) la varianza y (e) la desviación estándar de esta muestra.

Solution: La media es 0,3476, la mediana es 0,35, la moda es 0,33, la varianza es 0,0138 y la desviación estándar es 0,1176.

- 1.4 (1 punto) Los resultados obtenidos por los campeones del Masters en Estados Unidos entre 1982 y 1991 son

284, 280, 277, 282, 279, 285, 281, 283, 278, 277.

(a) Calculen los cuartiles y (b) dibujen el diagrama de caja correspondiente.

Solution: Primero, debemos ordenar los datos. Obtenemos

277, 277, 278, 279, 280, 281, 282, 283, 284, 285.

Hay 10 datos. Entonces, los tres cuartiles son en posiciones

$$0,25 \cdot 10 = 2,5 \rightarrow \mathbf{3}$$

$$0,50 \cdot 10 = 5 \rightarrow \mathbf{5 - 6}$$

$$0,75 \cdot 10 = 7,5 \rightarrow \mathbf{8}.$$

Los valores son 277, 278, $(280 + 281)/2 = 280,5$, 283 y 285.

2. Probabilidad

2.1 (1 punto) Simplifiquen las siguientes expresiones.

- (a) $E \cup \bar{E}$
- (b) $E \cap \bar{E}$
- (c) $(E \cup F) \cap (E \cup \bar{F})$
- (d) $(E \cup F) \cap (\bar{E} \cup F) \cap (E \cup \bar{F})$
- (e) $(E \cup F) \cap (F \cup G)$.

Solution: (a) $E \cup \bar{E} = S$.

(b) $E \cap \bar{E} = \emptyset$.

(c) $(E \cup F) \cap (E \cup \bar{F}) = E \cup (F \cap \bar{F}) = E \cup \emptyset = E$.

(d) Tenemos

$$\begin{aligned}
 (E \cup F) \cap (\bar{E} \cup F) \cap (E \cup \bar{F}) &= [(E \cap \bar{E}) \cup F] \cap (E \cup \bar{F}) \\
 &= [\emptyset \cup F] \cap (E \cup \bar{F}) \\
 &= F \cap (E \cup \bar{F}) \\
 &= (F \cap E) \cup (F \cap \bar{F}) \\
 &= (F \cap E) \cup \emptyset \\
 &= F \cap E.
 \end{aligned}$$

(e) No se simplifica mucho. Es posible escribir

$$(E \cup F) \cap (F \cup G) = F \cup (E \cap G).$$

2.2 (1 punto) Lanzamos 7 dados regulares de 6 caras. ¿Cuál es la probabilidad que todos los números aparecen?

Solution: Hay 6^7 resultados posibles si lanzamos 7 dados (6 resultados posibles por dado).

Ahora contamos cuantas tuplas tienen todos los números. En particular, hay un único número que aparece 2 veces. Tenemos **6** para el número que aparece dos veces. También, tenemos **C(7, 2)** opciones donde los dos números idénticos son colocados. Finalmente, los otros números son colocados en cualesquiera posiciones que quedan. Hay **5!** maneras de elegir estas posiciones. La probabilidad buscada es

$$\frac{6 \cdot C(7, 2) \cdot 5!}{6^7} \approx 0,0540.$$

2.3 (1 punto) Un cajón contiene 4 pares de calcetines. Se selecciona al azar 4 calcetines.

(a) ¿Cuál es la probabilidad de que no hay ningún par completo?

Solution: En total, hay $8 \cdot 7 \cdot 6 \cdot 5$ combinaciones (con orden) posibles. También, tenemos

- 8 opciones para el primer calcetín;
- 6 para el segundo (para no completar el primer par);
- 4 para el tercero (para no completar los dos primeros pares) y
- 2 para el cuarto (para no completar los tres primeros pares).

Entonces, la probabilidad buscada es

$$\frac{8 \cdot 6 \cdot 4 \cdot 2}{8 \cdot 7 \cdot 6 \cdot 5} = \frac{384}{1680} = \frac{8}{35} \approx 0,2286.$$

(b) ¿Exactamente un par completo?

Solution: Hay $C(4, 1)$ maneras de elegir el tipo de par que es completo. También, hay $C(3, 2)$ maneras de elegir los dos tipos de par que son presentes y hay 2 opciones de calcine en cada caso. Finalmente, hay $C(8, 4)$ maneras de elegir 4 calcetines entre las 8. La probabilidad buscada es

$$\frac{C(4, 1) \cdot C(3, 2) \cdot 2 \cdot 2}{C(8, 4)} = \frac{24}{35} \approx 0,6857.$$

2.4 (1 punto) Dos plantas diferentes fabrican radios (se llaman Planta A y Planta B). Suponemos que cada radio producida por A tiene una probabilidad 0,05 que sea defectuosa mientras que la probabilidad que sea defectuosa en B es 0,01.

Elegimos una de las plantas con igual probabilidad y compramos dos radios de esta planta. Si la primera radio es defectuosa, ¿cuál es la probabilidad que la segunda es también defectuosa?

Solution: Sean los eventos siguientes

- A : «Los radios resultan de la planta A»
- B : «Los radios resultan de la planta B»
- D_i : «La radio i es defectuosa»

La probabilidad buscada es

$$P(D_2|D_1) = \frac{P(D_1 \cap D_2)}{P(D_1)}.$$

Primero, por $i = 1, 2$,

$$\begin{aligned} P(D_i) &= P(A \cap D_i) + P(B \cap D_i) \\ &= P(D_i|A)P(A) + P(D_i|B)P(B) \\ &= 0,05 \cdot 0,5 + 0,01 \cdot 0,5 \\ &= 0,03. \end{aligned}$$

Entonces,

$$\begin{aligned} P(D_1 \cap D_2) &= P(D_1 \cap D_2 \cap A) + P(D_1 \cap D_2 \cap B) \\ &= P(D_1 \cap D_2|A)P(A) + P(D_1 \cap D_2|B)P(B) \\ &= P(D_1|A)P(D_2|A)P(A) + P(D_1|B)P(D_2|B)P(B) \\ &= 0,05 \cdot 0,05 \cdot 0,5 + 0,01 \cdot 0,01 \cdot 0,5 \\ &= 0,0013, \end{aligned}$$

donde la tercera igualdad viene del hecho que las radios de la misma planta son defectuosas con probabilidad independiente. Finalmente,

$$P(D_2|D_1) = \frac{0,0013}{0,03} = 0,0433.$$

2.5 (1 punto) Lanzamos dos dados. Sean los eventos siguientes:

- E : «La suma es 7»;
- C : «El primer dado es 4» y
- T : «El segundo dado es 3».

Demostren que

- (a) E y C son independientes;
- (b) E y T son independientes.

Solution: Tenemos

$$P(E) = \frac{6}{36} = \frac{1}{6}$$

$$P(C) = \frac{1}{6}$$

$$P(T) = \frac{1}{6}$$

$$P(E \cap C) = \frac{1}{36}$$

$$P(E \cap T) = \frac{1}{36}.$$

Entonces,

$$P(E)P(C) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} = P(E \cap C)$$

y

$$P(E)P(T) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} = P(E \cap T),$$

de tal manera que los eventos son independientes.

3. Variables aleatorias

- 3.1 (1 punto) Sea X una variable aleatoria que puede tomar solo los valores 1, 2, 3 y 4 con probabilidad uniforme. Calculen $E[X]$ y $\text{Var}(X)$.

Solution: Obtenemos

$$\begin{aligned}
 E[X] &= \sum_{i=0}^4 i \cdot P(X = i) \\
 &= \sum_{i=0}^4 i \cdot \frac{1}{4} \\
 &= \frac{1 + 2 + 3 + 4}{4} = \frac{10}{4} = 2,5 \\
 E[X^2] &= \sum_{i=0}^4 i^2 \cdot P(X = i) \\
 &= \sum_{i=0}^4 i^2 \cdot \frac{1}{4} \\
 &= \frac{1 + 4 + 9 + 16}{4} = \frac{30}{4} = 7,5. \\
 \text{Var}(X) &= E[X^2] - E[X]^2 \\
 &= 7,5 - (2,5)^2 = 1,25.
 \end{aligned}$$

- 3.2 (1 punto) En un grupo, hay 2 hombres et 2 mujeres que toman un examen. Suponemos que todas las rankings son equiprobables. Sea X la variable aleatoria que corresponde al mejor resultado de un hombre (por ejemplo, $X = 2$ si el mejor de los hombres es segundo). Calculen $P(X = i)$ para $i = 1, 2, 3, 4$.

Solution: En total, hay $4!$ rankings posibles.

$X = 1$ si un hombre es primero. Entonces, hay 2 opciones para el primero hombre y entonces, $3!$ para los tres personas que quedan. La probabilidad es $2 \cdot 3!/4! = 1/2$.

$X = 2$ si una mujer es primera y un hombre es siguiente. Hay 2 opciones para la mujer y 2 opciones para el hombre. Luego, hay 2 opciones para las dos últimas personas, entonces la probabilidad es $2 \cdot 2 \cdot 2/4! = 1/3$.

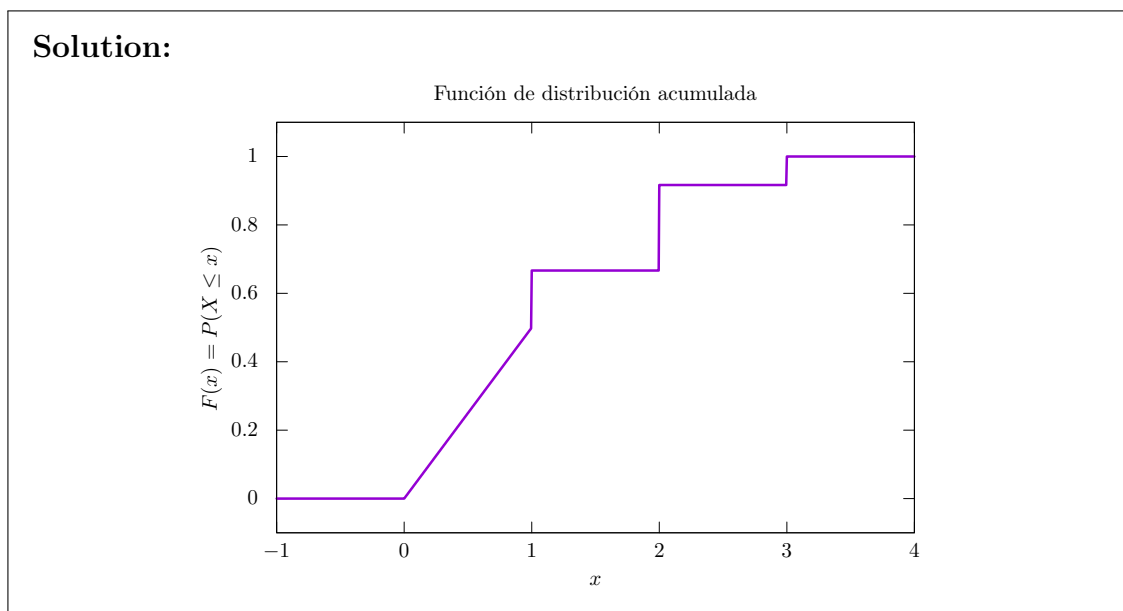
$X = 3$ si los hombres son los dos últimos y las mujeres son primeras. Hay 2 ordenaciones para los hombres y 2 para las mujeres, entonces la probabilidad es $2 \cdot 2/4! = 1/6$.

$X = 4$ no es posible, entonces $P(X = 4) = 0$.

3.3 (2 puntos) La FDA de una variable aleatoria X es

$$F(x) = \begin{cases} 0, & \text{si } x < 0; \\ x/2, & \text{si } 0 \leq x < 1; \\ 2/3, & \text{si } 1 \leq x < 2; \\ 11/12, & \text{si } 2 \leq x < 3; \\ 1, & \text{si } x \geq 3. \end{cases}$$

(a) Dibujen $F(x)$.



(b) Calculen las probabilidades siguientes:

- I. $P(X > 1/2)$,
- II. $P(2 < X \leq 4)$,
- III. $P(X < 3)$,
- IV. $P(X = 1)$.

Solution: Obtenemos

$$\begin{aligned} P(X > 1/2) &= 1 - P(X \leq 1/2) \\ &= 1 - F(1/2) \\ &= 1 - (1/2)/2 = 3/4 \\ P(2 < X \leq 4) &= F(4) - F(2) \\ &= 1 - 11/12 = 1/12 \\ P(X < 3) &= P(X \leq 2) \\ &= F(2) \end{aligned}$$

$$P(X = 1) = 0.$$

3.4 (2 puntos) Cuatro autobuses que contienen respectivamente 40, 33, 25 y 50 estudiantes llegan al estadio. Se elige al azar un estudiante x . Sea X el número de estudiantes que han utilizado el mismo autobús que x . También, se elige al azar uno de los 4 conductores de autobús y . Sea Y el número de estudiantes que ha conducido y .

(a) ¿Cuál valor entre $E[X]$ y $E[Y]$ es mayor? ¿Por qué?

Solution: $E[X]$ est mayor que $E[Y]$ porque el peso de cada de los 4 valores 25, 33, 40 y 50 aumenta la esperanza en el caso de X . En el caso de Y , los pesos son uniformes.

(b) Calculen $E[X]$ y $E[Y]$.

Solution: Primero, notamos que los valores posibles para X y Y son 25, 33, 40 y 50.

Entonces

$$\begin{aligned} E[X] &= \sum_{i \in \{25, 33, 40, 50\}} iP(X = i) \\ &= 25P(X = 25) + 33P(X = 33) + 40P(X = 40) + 50P(X = 50) \\ &= 25 \cdot \frac{25}{148} + 33 \cdot \frac{33}{148} + 40 \cdot \frac{40}{148} + 50 \cdot \frac{50}{148} \\ &= \frac{25^2 + 33^2 + 40^2 + 50^2}{148} = \frac{2907}{74} \approx 39,28. \end{aligned}$$

También,

$$\begin{aligned} E[Y] &= \sum_{i \in \{25, 33, 40, 50\}} iP(Y = i) \\ &= 25P(X = 25) + 33P(X = 33) + 40P(X = 40) + 50P(X = 50) \\ &= 25 \cdot \frac{1}{4} + 33 \cdot \frac{1}{4} + 40 \cdot \frac{1}{4} + 50 \cdot \frac{1}{4} \\ &= \frac{25 + 33 + 40 + 50}{4} = \frac{148}{4} = 37. \end{aligned}$$

4. Distribuciones conocidas

4.1 (1 punto) Sea X una variable normal de parámetros $\mu = 10$ y $\sigma^2 = 36$. Calculen

- (a) $P(X > 5)$
- (b) $P(4 < X < 16)$
- (c) $P(X < 20)$

Solution: Sea $Z \sim \mathcal{N}(0, 1)$. Entonces

$$\begin{aligned}P(X > 5) &= P\left(\frac{X - \mu}{\sigma} > \frac{5 - 10}{6}\right) \\&= P(Z > -0,8333) \\&= P(Z < 0,8333) \\&\approx 0,7967.\end{aligned}$$

$$\begin{aligned}P(4 < X < 16) &= P\left(\frac{4 - 10}{6} < \frac{X - \mu}{\sigma} < \frac{16 - 10}{6}\right) \\&= P(-1 < Z < 1) \\&= P(Z < 1) - P(Z < -1) \\&= P(Z < 1) - P(Z > 1) \\&= P(Z < 1) - (1 - P(Z < 1)) \\&= 2P(Z < 1) - 1 \\&\approx 2 \cdot 0,8413 - 1 \\&= 0,6826.\end{aligned}$$

$$\begin{aligned}P(X < 20) &= P\left(\frac{X - \mu}{\sigma} < \frac{20 - 10}{6}\right) \\&= P(Z < 2,6667) \\&\approx 0,6064.\end{aligned}$$

4.2 (1 punto) El tiempo necesario para reparar una máquina tiene una distribución exponencial de parámetro $\lambda = 1$.

- (a) ¿Cuál es la probabilidad que el tiempo de reparación exceda 2 horas?
- (b) ¿Cuál es la probabilidad condicional que la reparación exceda 3 horas dado que la duración sea al menos 2 horas? *Ayuda:* La distribución exponencial no tiene memoria.

Solution: Sea X el tiempo necesario para reparar la máquina. Entonces $X \sim \text{Exp}(1)$. También, sabemos que la FDA de X es

$$F(x) = \begin{cases} 0, & \text{si } x < 0; \\ 1 - e^{-x}, & \text{si } x \geq 0. \end{cases}$$

(a) Obtenemos

$$\begin{aligned} P(X > 2) &= 1 - P(X \leq 2) \\ &= 1 - F(2) \\ &= 1 - (1 - e^{-2}) \\ &= e^{-2} \\ &= 0,1353. \end{aligned}$$

(b) Sabemos que la distribución exponencial no tiene memoria. Entonces

$$\begin{aligned} P(X > 3 \mid X > 2) &= P(X > 1) \\ &= 1 - P(X \leq 1) \\ &= 1 - F(1) \\ &= 1 - (1 - e^{-1}) \\ &= e^{-1} \\ &= 0,3679. \end{aligned}$$

4.3 (1 punto) Un avión funciona correctamente si al menos la mitad de sus motores funcionan. Si cada motor funciona correctamente con probabilidad p , ¿para cuál valor de p un avión de 4 motores tiene más suerte funcionar que un avión de 2 motores?

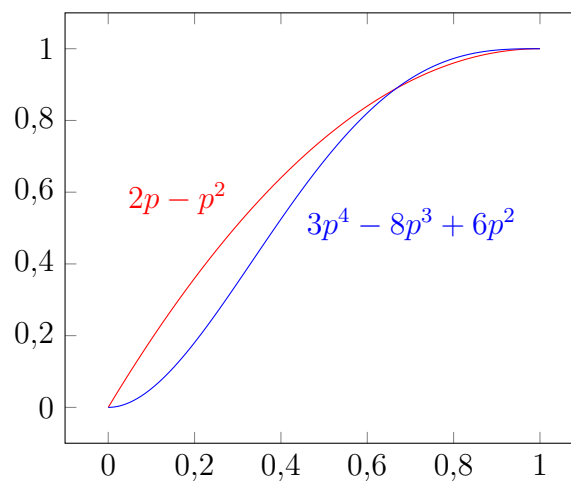
Solution: Sea X el número de motores funcionales en el caso de 4 motores y Y el número de motores funcionales en el caso de 2 motores.

Entonces $X \sim \mathcal{B}(4, p)$ y $Y \sim \mathcal{B}(2, p)$. También, tenemos que

$$\begin{aligned} P(Y \geq 1) &= P(Y = 1) + P(Y = 2) \\ &= \binom{2}{1} p^1 (1-p)^{2-1} + \binom{2}{2} p^2 (1-p)^{2-2} \\ &= 2p(1-p) + p^2 \\ &= 2p - 2p^2 + p^2 \\ &= 2p - p^2 \end{aligned}$$

$$\begin{aligned}
P(X \geq 2) &= P(X = 2) + P(X = 3) + P(X = 4) \\
&= \binom{4}{2} p^2 (1-p)^{4-2} + \binom{4}{3} p^3 (1-p)^{4-3} + \binom{4}{4} p^4 (1-p)^{4-4} \\
&= 6p^2(1-p)^2 + 4p^3(1-p) + p^4 \\
&= 6p^2(1-2p+p^2) + 4p^3(1-p) + p^4 \\
&= 6p^2 - 12p^3 + 6p^4 + 4p^3 - 4p^4 + p^4 \\
&= 3p^4 - 8p^3 + 6p^2.
\end{aligned}$$

Tenemos que identificar cuando las curvas inducidas por los polinomiales se encuentran.



Los puntos de intersección son $(0,0)$, $(2/3, 8/9)$ y $(1,1)$. Entonces, un avión de 4 motores tiene más suerte funcionando si $2/3 \leq p \leq 1$ (la curva azul está encima de la curva roja).

- 4.4 (1 punto) Una persona compra un conjunto de 100 transistores. Prueba 10 de estos transistores al azar. Decide que guarda el conjunto solamente si al menos 9 de los 10 transistores funcionan. Dado que hay 20 transistores defectuosos entre los 100, ¿cuál es la probabilidad que guarde los transistores?

Solution: Sea X el número de transistores funcionales. Entonces $X \sim \text{Hiper}(N, M, n)$ donde $N = 80$, $M = 20$ y $n = 10$. La probabilidad buscada es

$$\begin{aligned}
P(X \geq 9) &= P(X = 9) + P(X = 10) \\
&= \frac{\binom{80}{9} \binom{20}{1}}{\binom{100}{10}} + \frac{\binom{80}{10} \binom{20}{0}}{\binom{100}{10}}
\end{aligned}$$

$$\approx 0,3630.$$

- 4.5 (2 puntos) El número de gripes que una persona tiene cada año tiene una distribución de Poisson de parámetro $\lambda = 3$. Suponemos que una nueva medicación disminua el parámetro a $\lambda = 2$ para 75% de la población, pero para los otros 25%, es sin efecto. Si alguien toma la medicación durante un año y no tiene ningun gripe, ¿cuál es la probabilidad que la medicación sea sin efecto?

Solution: Sea X el número de gripes que la persona tiene. Entonces $X \sim \text{Poisson}(\lambda)$. También, sea E el evento «La medicación afecta el resultado». Buscamos $P(X = 0)$. Claramente, tenemos que

$$\begin{aligned} P(X = 0) &= P(X = 0 \cap E) + P(X = 0 \cap \bar{E}). \\ &= P(X = 0 \mid E)P(E) + P(X = 0 \mid \bar{E})P(\bar{E}). \\ &= P(X = 0 \mid \lambda = 2)P(E) + P(X = 0 \mid \lambda = 3)P(\bar{E}). \end{aligned}$$

Como

$$\begin{aligned} P(E) &= 0,75, \\ P(\bar{E}) &= 0,25, \\ P(X = 0 \mid \lambda = 2) &= \frac{e^{-2}2^0}{0!} = e^{-2}, \\ P(X = 0 \mid \lambda = 3) &= \frac{e^{-3}3^0}{0!} = e^{-3} \end{aligned}$$

deducimos que

$$P(X = 0) = 0,75e^{-2} + 0,25e^{-3} \approx 0,1139.$$

5. Distribuciones muestrales

- 5.1 (1 punto) Sea X una variable binomial de parámetros $n = 150$ y $p = 0,6$. Con un ordenador se obtiene $P(X \leq 80) \approx 0,05745956$. Comparen este valor con la aproximación normal (a) con corrección de continuidad (b) sin corrección de continuidad.

Solution: Utilizando la corrección de continuidad, obtenemos

$$\begin{aligned}
 P(X \leq 80) &\approx P\left(\frac{X - np}{\sqrt{np(1 - np)}} \leq \frac{80,5 - 150 \cdot 0,6}{\sqrt{150 \cdot 0,6 \cdot 0,4}}\right) \\
 &\approx P(Z \leq -1,5833) \\
 &= P(Z \geq 1,5833) \\
 &= 1 - P(Z \leq 1,5833) \\
 &\approx 1 - 0,9429 \\
 &= 0,0571.
 \end{aligned}$$

Sin la corrección, es

$$\begin{aligned}
 P(X \leq 80) &\approx P\left(\frac{X - np}{\sqrt{np(1 - np)}} \leq \frac{80 - 150 \cdot 0,6}{\sqrt{150 \cdot 0,6 \cdot 0,4}}\right) \\
 &\approx P(Z \leq -1,6667) \\
 &= P(Z \geq 1,6667) \\
 &= 1 - P(Z \leq 1,6667) \\
 &\approx 1 - 0,9525 \\
 &= 0,0475.
 \end{aligned}$$

- 5.2 (1 punto) En la población, estimamos que 12% de las personas son zurdas. Calculen la probabilidad que haya entre 10 y 14 personas zurdas en una muestra de 100 personas.

Solution: Sea X el número de personas zurdas en la muestra. Entonces $X \sim \mathcal{B}(100; 0,12)$, pero se puede aproximar con una distribución normal, así que

$$\begin{aligned}
 P(10 \leq X \leq 14) &\approx P\left(\frac{9,5 - 100 \cdot 0,12}{\sqrt{100 \cdot 0,12 \cdot 0,88}} \leq \frac{X - np}{\sqrt{np(1 - p)}} \leq \frac{14,5 - 100 \cdot 0,12}{\sqrt{100 \cdot 0,12 \cdot 0,88}}\right) \\
 &\approx P(-0,7693 \leq Z \leq 0,7693) \\
 &\approx 2P(Z \leq 0,7693) - 1 \\
 &\approx 2 \cdot 0,7794 - 1
 \end{aligned}$$

$$= 0,5588.$$

5.3 (2 puntos) Un dado regular se lanza hasta que el total sea más que 400 (es decir que se suman los resultados hasta que la suma sea más que 400).

¿Cuál es la probabilidad que necesitemos más de 140 tiradas? *Nota:* pueden utilizar una aproximación.

Solution: El problema se puede formular de una otra manera. De hecho, busquemos la probabilidad que la suma de 140 tiradas sea menor o igual a 400.

Sea X_i el resultado obtenido en la tirada i , donde $i = 1, 2, 3, \dots, 140$, y $X = \sum_{i=1}^n X_i$. Sabemos que X_i sigue una distribución uniforme discreta en el conjunto $\{1, 2, 3, 4, 5, 6\}$, cuya media es $7/2$ y varianza es $35/12$. Entonces, busquemos $P(X \leq 400)$.

Por el teorema del límite central, X sigue aproximadamente una distribución normal de media $n\mu = 140 \cdot 7/2 = 490$ y de varianza $n\sigma^2 = 140 \cdot 35/12 = 1295/3$. Luego,

$$\begin{aligned} P(X \leq 400) &\approx P\left(\frac{X - \mu}{\sigma/\sqrt{n}} \leq \frac{400,5 - 490}{\sqrt{1295/3}}\right) \\ &\approx P(Z \leq -4,3077) \\ &\approx 1 - P(Z \leq 4,3077) \\ &\approx 0. \end{aligned}$$

5.4 (2 puntos) La media y la desviación estándar de una muestra son respectivamente 517 y 120. Aproximen la probabilidad que una muestra aleatoria de tamaño 144 tenga una media mayor que (a) 507 (b) 517 (c) 537 y (d) 550.

Solution: Sea \bar{X} la media muestral. Sabemos que \bar{X} sigue aproximadamente una distribución normal de media $\mu = 517$ y de desviación estándar $\sigma/\sqrt{n} = 120/\sqrt{144} = 10$. Entonces

$$\begin{aligned} P(\bar{X} > 507) &\approx P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > \frac{507 - 517}{10}\right) \\ &= P(Z > -1) \\ &= P(Z < 1) \\ &\approx 0,8413, \end{aligned}$$

$$\begin{aligned}P(\bar{X} > 517) &\approx P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > \frac{517 - 517}{10}\right) \\&= P(Z > 0) \\&= 0,5, \\P(\bar{X} > 537) &\approx P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > \frac{537 - 517}{10}\right) \\&= P(Z > 2) \\&= 1 - P(Z < 2) \\&\approx 1 - 0,9772 \\&= 0,0228, \\P(\bar{X} > 550) &\approx P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > \frac{550 - 517}{10}\right) \\&= P(Z > 3,3) \\&= 1 - P(Z < 3,3) \\&\approx 1 - 0,9995 \\&= 0,0005.\end{aligned}$$

6. Estimación de parámetros

6.1 (2 puntos) Calculen el estimador de máxima verosimilitud del parámetro θ para una muestra de tamaño n si la función de densidad de las variables es

$$f(x) = \begin{cases} e^{-(x-\theta)}, & \text{si } x \geq \theta; \\ 0, & \text{de otra manera.} \end{cases}$$

Solution: Para $x_1, x_2, \dots, x_n \geq \theta$, la función de verosimilitud es

$$\begin{aligned} L(\theta) &= e^{-(x_1-\theta)} e^{-(x_2-\theta)} \dots e^{-(x_n-\theta)} \\ &= \exp\left(-\sum_{i=1}^n (x_i - \theta)\right) \end{aligned}$$

Entonces

$$\ln L(\theta) = -\sum_{i=1}^n (x_i - \theta).$$

Como $x_i \geq \theta$ for $i = 1, 2, \dots, n$, esta suma no puede ser positiva, y entonces es máxima cuando

$$0 = -\sum_{i=1}^n (x_i - \theta) = -\sum_{i=1}^n x_i + n\theta,$$

es decir $\hat{\theta} = \bar{X}$.

6.2 (2 puntos) Los resultados de los estudiantes de un examen tienen una desviación estándar de 11,3. Un año dado, sabemos que 81 estudiantes que pasan el examen obtienen una media de 74,6. Calculen un intervalo de confianza a 90 % para estimar la media que se obtiene en general a este examen.

Solution: Aquí, la desviación estándar de la población es conocida ($\sigma = 11,3$). También, $\bar{x} = 74,6$, $\alpha = 0,1$, $n = 81$ y

$$z_{\alpha/2} = z_{0,05} \approx 1,645.$$

Entonces, un intervalo de confianza a 90 % para la media es

$$\left(\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) \approx (72,5346; 76,6654)$$

6.3 (2 puntos) La vida de 10 baterías eligidas al azar es

140, 136, 150, 144, 148, 152, 138, 141, 143, 151.

- (a) Den una estimación puntual de la varianza σ^2 .
 (b) Calculen un intervalo de confianza bilateral para σ^2 .

Solution: (a) Como $E[S^2] = \sigma^2$, una estimación puntual de la varianza es la varianza muestral. Con Python, obtenemos $\bar{X} = 144,3$ y $s^2 = 32,23$:

```
>>> X = [140, 136, 150, 144, 148, 152, 138, 141, 143, 151]
>>> Xb = sum(X) * 1.0 / len(X)
>>> s2 = sum((x - Xb)**2 for x in X) / (len(X) - 1)
>>> Xb
144.30000000000000
>>> s2
32.23333333333333
```

(b) Utilizamos $\alpha = 0,05$. También, tenemos $n = 10$ y

$$\begin{aligned}\chi_{\alpha/2, n-1}^2 &= \chi_{0,025;9}^2 = 19,023 \\ \chi_{1-\alpha/2, n-1}^2 &= \chi_{0,975;9}^2 = 2,700.\end{aligned}$$

Entonces, un intervalo de confianza a 95 % es

$$\left(\frac{(n-1)s^2}{\chi_{\alpha/2, n-1}^2}, \frac{(n-1)s^2}{\chi_{1-\alpha/2, n-1}^2} \right) = (15,2500; 107,4444).$$

7. Pruebas de hipótesis

- 7.1 (2 puntos) El peso de los salmones que se venden en una pescadería siguen una distribución normal de desviación estándar 1,2. La pescadería dice que el peso medio este año es 7,6. Suponemos que una muestra de tamaño 16 tiene una media de 7,2. ¿Se puede rechazar la afirmación de la pescadería al nivel de significado 5%? ¿1%? ¿Cuál es el valor p ?

Solution: Aquí, tenemos $n = 16$, $\sigma = 1,2$ (la varianza es conocida), $\mu_0 = 7,6$ y $\bar{x} = 7,2$. Las hipótesis son

$$H_0 : \mu = 7,6 \quad \text{y} \quad H_1 : \mu \neq 7,6.$$

Entonces,

$$Z = \frac{\sqrt{n}}{\sigma} |\bar{x} - \mu_0| = \frac{\sqrt{16}}{1,2} |7,2 - 7,6| = \frac{4}{3}$$

y el valor p es

$$P(|Z| > 4/3) = 2P(Z > 1,33) = 2(1 - P(Z < 1,33)) \approx 2(1 - 0,9082) = 0,1836.$$

Como el valor p es bastante grande, la hipótesis H_0 no se rechaza.

- 7.2 (2 puntos) Hace 20 años, los estudiantes de una escuela secundaria podían hacer 24 flexiones en 60 segundos. Para verificar si esta media es la misma hoy, se pasa la misma prueba a 36 estudiantes elegidos al azar. Si obtienen una media de 22,5 y una desviación estándar de 3,1, ¿podemos concluir que la media es diferente de 24? Utilicen un nivel de significado de 5%.

Solution: En este caso, $n = 36$, $\bar{x} = 22,5$, $s = 3,1$ (la varianza no es conocida), $\mu_0 = 24$, $\alpha = 0,05$. Las hipótesis son

$$H_0 : \mu = 24 \quad \text{y} \quad H_1 : \mu \neq 24.$$

Entonces,

$$T = \frac{\sqrt{n}(\bar{x} - \mu_0)}{s} = \frac{\sqrt{36}(22,5 - 24)}{3,1} \approx -2,9032.$$

El valor p es

$$P(t_{35} < -2,9032) \approx 0,0032,$$

de tal manera que se rechaza H_0 y entonces se acepta H_1 , es decir que la media es diferente de 24.

7.3 (2 puntos) Se lanza un dado 100 veces para verificar si es trucado. Obtenemos los valores siguientes

Résultats	1	2	3	4	5	6
Effectifs	7	18	26	15	18	16

¿Cuál es la conclusión?

Solution: Debemos utilizar un bondad de ajuste. Sea X el número de dado obtenido. Si el dado no es trucado, deberíamos obtener cada resultado 16,67 veces. Entonces, la hipótesis H_0 es

$$H_0 : P(X = i) = \frac{1}{6}, \quad \text{para } i = 1, 2, 3, 4, 5, 6.$$

Entonces, $n = 100$, $x = (7, 18, 26, 15, 18, 16)$, $p_i = 1/6$ para $i = 1, 2, 3, 4, 5, 6$ y

$$\begin{aligned} T &= \sum_{i=1}^6 \frac{x_i^2}{np_i} - n \\ &= \frac{7^2 + 18^2 + 26^2 + 15^2 + 18^2 + 16^2}{100/6} - 100 \\ &= 11,24. \end{aligned}$$

El valor p es

$$P(T \geq 11,24) \approx P(\chi_5^2 \geq 11,24) \approx 0,0468.$$

Entonces, H_0 se acepta si $\alpha \leq 0,468$, es decir que la prueba no es conclusiva.