1.1. PROBLEMAS RESUELTOS CAPÍTULO 3

1. Sean A, B y C tres sucesos de un espacio muestral. Encontrar expresiones para los siguientes sucesos en función de A, B y C. Ilustrarlo mediante diagramas de Venn.

   (1) Ocurre A ó B.

   (2) Ocurren A y B.

   (3) Sólo ocurre A.

   (4) Ocurren tanto A como B pero no C.

   (5) Los tres ocurren.

   (6) Ocurre por lo menos uno de los tres.

   (7) Por lo menos dos ocurren.

   (8) Ocurre uno y no más de uno.

   (9) Ocurren dos y no más de dos.

   (10) No ocurre ninguno.

   (11) No ocurren más de dos.

   Solución.

   (1) $A \cup B$

   (2) $A \cap B$

   (3) $A \cap \overline{B} \cap \overline{C}$

   (4) $A \cap B \cap \overline{C}$

   (5) $A \cap B \cap C$

   (6) $A \cup B \cup C$

   (7) $(A \cap B) \cup (A \cap C) \cup (B \cap C)$

   (8) $\left(A \cap \overline{B} \cap \overline{C}\right) \cup \left(\overline{A} \cap B \cap \overline{C}\right) \cup \left(\overline{A} \cap \overline{B} \cap C\right)$

   (9) $\left(A \cap B \cap \overline{C}\right) \cup \left(A \cap \overline{B} \cap C\right) \cup \left(\overline{A} \cap B \cap C\right)$

   (10) $\overline{A} \cap \overline{B} \cap \overline{C} = \overline{A \cup B \cup C}$

   (11) $\overline{A \cap B \cap C} = \overline{A} \cup \overline{B} \cup \overline{C}$

2. Una moneda se lanza tres veces. Se pide:

   (a) Construir el espacio muestral asociado al experimento.

   (b) Expresar en función de los sucesos elementales los siguientes sucesos:

      (1) Los tres lanzamientos producen el mismo resultado.

      (2) El mismo resultado aparece exactamente dos veces.

      (3) Sale cara al menos dos veces

      (4) La cara aparece en el primero y en el segundo lanzamientos.

(5) Sale cara exactamente dos veces.

Solución.

(a) $\Omega = \{CCC, CCX, CXC, CXX, XCC, XCX, XXC, XXX\}$

(b) (1) $\{CCC, XXX\}$

(2) $\{CCX, CXC, CXX, XCC, XCX, XXC\}$

(3) $\{CCC, CCX, CXC, XCC\}$

(4) $\{CCC, CCX\}$

(5) $\{CCX, CXC, XCC\}$

3. Un número es seleccionado al azar entre los números 1 al 10. Se A el suceso "el número elegido es par", B el suceso "el número elegido es primo" y C el suceso "el número elegido es múltiplo de 3". Expresar verbalmente los siguientes sucesos y encontrar los elementos que los forman:

   $A \cap B \qquad A - B \qquad A \cup B \qquad \bar{A} \cap \bar{B} \qquad A \cap C \qquad A - C \qquad B \cap C \qquad \bar{A} \cap C \qquad (A \cup B) \cap \bar{C}$

   Solución.

   $A \cap B$={número par y primo} $= \{2\}$

   $A - B =$ {número par no primo} $= \{4, 6, 8, 10\}$

   $A \cup B =$ {número par ó primo} $= \{1, 2, 3, 4, 5, 6, 7, 8, 10\}$

   $\bar{A} \cap \bar{B} =$ {número no par y no primo} $= \{9\}$

   $A \cap C =$ {número par y múltiplo de 3} $= \{6\}$

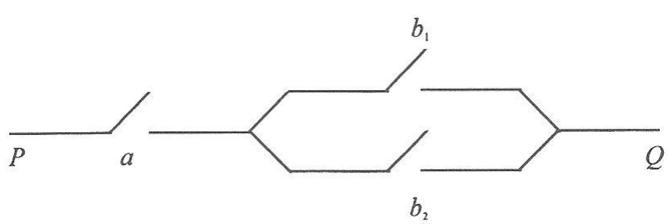
   $A - C =$ {número par y no múltiplo de 3} $= \{2, 4, 8, 10\}$

   $B \cap C =$ {número primo y múltiplo de 3} $= \{3\}$

   $\bar{A} \cap C =$ {número impar y múltiplo de 3} $= \{3, 9\}$

   $(A \cup B) \cap \bar{C} =$ {número par o primo no múltiplo de 3} $= \{1, 2, 4, 5, 7, 8, 10\}$

4. En la figura se muestra el esquema de un circuito eléctrico entre dos puntos P y Q. Sea A el suceso "el interruptor a está cerrado" y B_i "el interruptor b_i está cerrado" $(i = 1, 2)$. Expresar el suceso de que el circuito esté cerrado entre P y Q.



   Solución. El suceso de que el circuito esté cerrado se expresaría como $A \cap (B_1 \cup B_2)$

5. Un sistema está construido con elementos de dos tipos. Sea A_k el suceso que la unidad k-ésima del primer tipo está útil y B_j que la unidad j-ésima del segundo tipo está útil. El sistema funciona si

al menos están útiles una unidad del primer tipo y una del segundo. Expresar este hecho mediante los sucesos A_k y B_j.

Solución. El suceso buscado será $(\bigcup A_k) \cap (\bigcup B_j)$.

6. De 25 micro computadoras disponibles en un almacén, 10 de ellas tienen tarjeta para impresora, 5 tienen tarjeta adaptadora para módem y 13 no tienen ninguna de éstas. Expresar simbólicamente los siguientes conjuntos y calcular el número de micro computadoras que hay en cada uno de ellos:

   (a) Las que tengan ambas tarjetas.

   (b) Las que no tengan ninguna tarjeta.

   (c) Las que sólo tengan tarjeta para impresora.

   (d) Las que tengan exactamente una de las tarjetas.

   Solución. Consideramos los conjuntos $I = \{\text{micro computadoras con tarjeta para impresora}\}$ y $M = \{\text{micro computadoras con tarjeta para modem}\}$.

   (a) Las que tengan ambas tarjetas serán las del conjunto $I \cap M$ y sumarán un total de $10 + 5 - (25 - 13) = 3$

   (b) Las que no tienen ninguna tarjeta serán las del conjunto $\overline{I} \cap \overline{M} = \overline{I \cup M}$ y son 13, como dice el enunciado.

   (c) Las que sólo tienen tarjeta para impresora serán las del conjunto $I \cap \overline{M}$ y sumarán un total de $10 - 3 = 7$

   (d) Las que tienen exactamente una tarjeta serán las del conjunto $\left(I \cap \overline{M}\right) + \left(\overline{I} \cap M\right)$ y sumarán un total de $(10 - 3) + (5 - 3) = 9$

7. Simplificar la siguiente expresión: $(A \cup B) \cap (A \cup \bar{B}) \cap (\bar{A} \cup B)$

   Solución. Calculamos primero $(A \cup B) \cap (A \cup \bar{B})$ utilizando las propiedades de la teoría de conjuntos:

$$(A \cup B) \cap (A \cup \bar{B}) = [(A \cup B) \cap A] \cup \left[(A \cup B) \cap \bar{B}\right] = A \cup \left(A \cap \bar{B}\right) = A$$

   Por tanto tendremos que

$$(A \cup B) \cap (A \cup \bar{B}) \cap (\bar{A} \cup B) = A \cap (\bar{A} \cup B) = \varnothing \cup (A \cap B) = A \cap B$$

8. Una caja contiene seis bombillas de las cuales 2 son defectuosas. Se prueban las bombillas hasta encontrar la primera defectuosa. Encontrar el espacio muestral asociado al experimento y la probabilidad de cada uno de los sucesos elementales

   Solución. Si en cada prueba denotamos por D_i al suceso $D_i = \{\text{la } i\text{-ésima bombilla probada es defectuosa}\}$, el espacio muestral sería $\Omega = \left\{D_1, \overline{D_1}D_2, \overline{D_1}\ \overline{D_2}D_3, \overline{D_1}\ \overline{D_2}\ \overline{D_3}D_4, \overline{D_1}\ \overline{D_2}\ \overline{D_3}\ \overline{D_4}D_5\right\}$ y las probabilidades serían:

$$p(D_1) = \frac{2}{6} = \frac{1}{3}$$
$$p\left(\overline{D_1}D_2\right) = p\left(\overline{D_1}\right)p\left(D_2/\overline{D_1}\right) = \left(\frac{4}{6}\right)\left(\frac{2}{5}\right) = \frac{4}{15}$$
$$p\left(\overline{D_1}\ \overline{D_2}D_3\right) = p\left(\overline{D_1}\right)p\left(\overline{D_2}/\overline{D_1}\right)p\left(D_3/\overline{D_1}\ \overline{D_2}\right) = \left(\frac{4}{6}\right)\left(\frac{3}{5}\right)\left(\frac{2}{4}\right) = \frac{1}{5}$$
$$p\left(\overline{D_1}\ \overline{D_2}\ \overline{D_3}D_4\right) = \left(\frac{4}{6}\right)\left(\frac{3}{5}\right)\left(\frac{2}{4}\right)\left(\frac{2}{3}\right) = \frac{2}{15}$$
$$p\left(\overline{D_1}\ \overline{D_2}\ \overline{D_3}\ \overline{D_4}D_5\right) = \left(\frac{4}{6}\right)\left(\frac{3}{5}\right)\left(\frac{2}{4}\right)\left(\frac{1}{3}\right)(1) = \frac{1}{15}$$

9. Una cadena comercial tiene dos establecimientos en una ciudad. Se sabe que el 30 % de los clientes potenciales compra productos sólo en la tienda 1, el 50 % compra solamente en la tienda 2, el 10 % compra en las tiendas 1 y 2, y el 10 % de los consumidores no compra en ninguna de las dos tiendas. Sea A el suceso en el que el cliente potencial compra en la tienda 1 y B el suceso en el que compra en la tienda 2. Calcular y explicar el significado de las siguientes probabilidades. Expresarlas como porcentajes.

   (a) $p(A)$ (b) $p(A \cup B)$ (c) $p(\bar{B})$ (d) $p(A \cap B)$ (e) $p(A \cup \bar{B})$ (f) $p(\bar{A} \cap \bar{B})$

   Solución. Sabemos que $p\left(A\overline{B}\right) = 0.3$, $p\left(\overline{A}B\right) = 0.5$, $p(AB) = 0.1$ y $p\left(\overline{AB}\right) = 0.1$.

   (a) $p(A) = p(AB) + p\left(A\overline{B}\right) = 0.1 + 0.3 = 0.4$ y por tanto el 40 % de los potenciales clientes compra en la tienda 1.

   (b) $p(A \cup B) = p(A) + p(\overline{A}B) = 0.4 + 0.5 = 0.9$ y por tanto el 90 % de los potenciales clientes compra en alguna de las dos tiendas.

   (c) $p(\bar{B}) = p\left(A\overline{B}\right) + p\left(\overline{A}\ \overline{B}\right) = 0.3 + 0.1 = 0.4$ y por tanto el 40 % de los potenciales clientes no compra en la tienda 2.

   (d) $p(A \cap B) = p(AB) = 0.1$ y por tanto el 10 % de los potenciales clientes compra en las dos tiendas.

   (e) $p(A \cup \bar{B}) = p(A) + p\left(\overline{A}\ \overline{B}\right) = 0.4 + 0.1 = 0.5$ y por tanto el 50 % de los potenciales clientes compran solamente en la tienda 1, o no compra.

   (f) $p(\bar{A} \cap \bar{B}) = p\left(\overline{A}\ \overline{B}\right) = 0.1$ y por tanto el 10 % de los potenciales clientes no compra en ninguna de las dos tiendas.

10. Un experimentador desea averiguar el efecto de tres variables (temperatura, grado de abonado y tipo de semilla) sobre el rendimiento de un cultivo de invernadero. Si el investigador desea analizar cuatro niveles de temperatura, cinco dosis de diferentes de abonado y tres tipos de semilla, ¿cuántos ensayos experimentales ha de llevar a cabo si desea probar todas las combinaciones posibles de temperatura, abonado y tipo de semilla?

    Solución. Para probar todas la combinaciones deberá realizar $(4)(5)(3) = 60$ ensayos.

11. Un equipo de baloncesto cuenta con 3 bases, 6 aleros y 5 pivots. Si el quinteto inicial está formado por 1 base, 2 aleros y 2 pivots, ¿cuántos equipos titulares se pueden formar?

    Solución. Las posibles maneras de elegir el base serán $C_{3,1}$ Del mismo modo para los aleros tendremos $C_{6,2}$ y para los pivots $C_{5,2}$. Por tanto los posibles equipos titulares serán $\binom{3}{1}\binom{6}{2}\binom{5}{2} = (3)(15)(10) = 450$

12. En un banco se pueden sentar 8 personas. Si entre un grupo de 8 amigos hay una pareja de novios, ¿cuál es la probabilidad de que se sienten juntos?

    Solución. Las posibles maneras de colocar a las 8 personas serán $P_8 = 8! = 40320$. Para que la pareja de novios se sienten juntos tenemos 14 posibilidades (7 posiciones multiplicado por las 2 posibilidades de intercambiarlos entre sí). Para cada una de ellas las posibles maneras de colocar a las otras 6 personas serán $P_6 = 6! = 720$. Por tanto, utilizando la regla de Laplace, la probabilidad buscada será:

$$p = \frac{\text{casos favorables}}{\text{casos posibles}} = \frac{(14)(6!)}{8!} = \frac{14}{56} = 0.25$$

13. ¿Cuál es la probabilidad de que en una quiniela de fútbol haya exactamente cinco x's?

    Solución. Definimos el suceso $A = \{\text{obtener cinco } x\text{'s en una quiniela de fútbol}\}$ y supondremos que todas las posibles quinielas son igualmente probables. Por tanto la probabilidad se puede calcular por la regla de Laplace. Los casos posibles son $VR_{3,14} = 3^{14} = 4782969$ y los casos favorables al suceso A serán las posibles maneras de elegir las cinco casillas para colocar las cinco x's ($C_{14,5} = \binom{14}{5} = 2002$) multiplicado por las posibles maneras de rellenar las 9 casillas restantes con los resultados 1 o 2 ($VR_{2,9} = 2^9 = 512$). Por tanto, la probabilidad buscada es

$$p(A) = \frac{C_{14,5}VR_{2,9}}{VR_{3,14}} = \frac{1025024}{4782969} = 0.2143$$

14. Un jugador italiano expresó su sorpresa a Galileo por observar que al jugar con tres dados la suma 10 aparece con más frecuencia que la suma 9. Según el jugador los casos favorables serían:

Casos favorables al 9	Casos favorables al 10
1 2 6	1 3 6
1 3 5	1 4 5
1 4 4	2 2 6
2 2 5	2 3 5
2 3 4	2 4 4
3 3 3	3 3 4

    Galileo, en su libro *Considerazione sopra il giuco dei dadi*, vio que estas posibilidades no se pueden considerar igualmente probables. Explicar porqué y calcular las correspondientes probabilidades.

    Solución. El experimento consiste en lanzar tres dados indistinguibles a la vez y el espacio muestral Ω estaría por todas las combinaciones de los números naturales comprendidos entre 1 y 6 ambos incluidos, elegidos de tres en tres con repetición y sin que importe el orden (porque los dados son indistinguibles). Por tanto el número de elementos que hay en Ω sería $CR_{6,3} = \binom{6+3-1}{3} = \binom{8}{3} = 56.$

Tal y como puede comprobarse en la tabla, los casos favorables a la suma 9 son 6, al igual que los casos favorables a la suma 10, que también son 6. Por eso el jugador italiano, utilizando la regla de Laplace, sospechaba que los dos sucesos deberían ser igualmente probables (6/56) y por tanto las frecuencias relativas de ambos sucesos deberían ser iguales. Sin embargo esto no es cierto porque no todos los sucesos elementales son igualmente probables y por tanto no se puede aplicar la regla de Laplace. Veamos cual sería el calculo correcto de las probabilidades. Si denotamos los sucesos por $A = \{$suma de las puntuaciones igual a 9$\}$ y $B = \{$suma de las puntuaciones igual a 10$\}$ tendremos que sumar en cada caso las probabilidades de cada uno de los 6 sucesos elementales incluidos en los sucesos A y B. Para el caso del suceso A las probabilidades de los sucesos elementales $(1, 2, 6)$, $(1, 3, 5)$ y $(2, 3, 4)$ podrían calcularse en los tres casos como $\frac{P_3}{VR_{6,3}} = \frac{6}{6^3} = \frac{6}{216}$, para los sucesos $(1, 4, 4)$ y $(2, 2, 5)$ sería $\frac{PR_2^{1,2}}{VR_{6,3}} = \frac{3}{216}$, y para el suceso $(3, 3, 3)$ sería $\frac{1}{216}$ Por tanto tendríamos que

$$p(A) = 3\left(\frac{6}{216}\right) + 2\left(\frac{3}{216}\right) + \frac{1}{216} = \frac{25}{216} = 0.1157$$

Del mismo modo, para el suceso B tendríamos que las probabilidades para cada uno de los sucesos elementales $(1, 3, 6)$, $(1, 4, 5)$ y $(2, 3, 5)$ sería $\frac{6}{216}$, y para cada uno de los sucesos $(2, 2, 6)$, $(2, 4, 4)$ y $(3, 3, 4)$ sería $\frac{3}{216}$. Por tanto tendríamos que

$$p(B) = 3\left(\frac{6}{216}\right) + 3\left(\frac{3}{216}\right) = \frac{27}{216} = 0.125$$

Es decir $p(B) > p(A)$ y por tanto es lógico que la suma 10 aparezca con más frecuencia que la suma 9.

15. Obtener la probabilidad p de que al lanzar dos dados n veces se obtenga al menos un 6 doble. ¿Cuántas partidas habrá que jugar para que tengamos $p = 1/2$ de obtener un 6 doble? (*Problema propuesto a Pascal por el caballero de Mere*)

    Solución. Si suponemos que los dos dados son distinguibles y todos los resultados son igualmente probables, vamos a calcular primero la probabilidad de obtener un 6 doble en cada lanzamiento utilizando la regla de Laplace. Si definimos el suceso $A_1 = \{$obtener un 6 doble al lanzar dos dados$\}$, sólo habría un caso favorable al suceso A_1 y los casos posibles serían $VR_{6,2} = 6^2 = 36$. Por tanto tendríamos que $p(A_1) = \frac{1}{36}$. Si los dados fuesen indistinguibles también tendríamos que $p(A_1) = \frac{1}{36}$ porque el hecho de distinguir o no distinguir los dados no altera la aleatoriedad del experimento. Entonces, si repetimos este experimento n veces, denotamos por A_i al suceso consistente en obtener un 6 doble en el i-ésimo lanzamiento y denotamos por A al suceso consistente en obtener al menos un 6 doble en los n lanzamientos tendremos que $A = A_1 \cup A_2 ... \cup A_n$. Además, por la primera ley de De Morgan, tendremos que $\overline{A} = \overline{A_1} \cap \overline{A_1} \cap ... \cap \overline{A_n}$ y, como los resultados de cada lanzamiento son independientes entre sí, podemos calcular la probabilidad como

$$p(A) = 1 - p\left(\overline{A}\right) = 1 - p\left(\overline{A_1}\right) p\left(\overline{A_2}\right) ... p\left(\overline{A_n}\right) = 1 - \left(1 - \frac{1}{36}\right)^n = 1 - \left(\frac{35}{36}\right)^n$$

Si queremos que $p(A) \geq \frac{1}{2}$ necesitaríamos que $\left(\frac{35}{36}\right)^n \leq 0.5$, o equivalentemente, $n[\ln(35) - \ln(36)] \leq \ln(0.5)$. Despejando n tendremos entonces que $n \geq \frac{\ln(0.5)}{\ln(35) - \ln(36)} = 24.6$ (observar que $\ln(35) -$

$\ln(36) \leq 0$ y por tanto la desigualdad cambia). Luego necesitaríamos al menos 25 partidas para que la probabilidad de obtener al menos un 6 doble sea mayor que 0.5.

16. ¿Cuál es la probabilidad de torpedear un barco enemigo sabiendo que sólo pueden lanzarse 3 torpedos y que la probabilidad de hacer blanco con cada uno de ellos es 0.20?

    Solución. Si consideramos los sucesos $A_1 = \{\text{hacer blanco con el primer torpedo}\}$, $A_2 = \{\text{hacer blanco con el segund}$ y $A_3 = \{\text{hacer blanco con el tercer torpedo}\}$ tendremos que calcular $p(A_1 \cup A_2 \cup A_3)$. Utilizando la primera ley de De Morgan y la independencia de los disparos tendremos:

$$\begin{aligned} p(A_1 \cup A_2 \cup A_3) &= 1 - p\left(\overline{A_1 \cup A_2 \cup A_3}\right) = 1 - p\left(\overline{A_1}\ \overline{A_2}\ \overline{A_3}\right) \\ &= 1 - p\left(\overline{A_1}\right) p\left(\overline{A_2}\right) p\left(\overline{A_3}\right) = 1 - 0.8^3 = 0.488 \end{aligned}$$

17. Sea un dado tal que la probabilidad de las distintas caras es proporcional al número de puntos inscritos en ellas. Hallar la probabilidad de obtener con este dado un número par.

    Solución. Para este dado tendremos que $p_1 = 1/21$, $p_2 = 2/21$, $p_3 = 3/21$, $p_4 = 4/21$, $p_5 = 5/21$ y $p_6 = 6/21$. Por tanto la probabilidad de obtener un número para será $p_2 + p_4 + p_6 = 12/21 = 0.571$

18. Un examen de reválida de licenciatura consta de 14 temas. Se eligen dos al azar, y el alumno deberá escoger uno para contestarlo. (a) Calcular la probabilidad de que a un alumno que ha preparado 5 temas le toque al menos uno que sabe. (b) ¿Cuál es el número mínimo de temas que debe de preparar para que tenga una probabilidad superior a 1/2 de superar el examen?

    Solución. Las posibles propuestas de examen serán $C_{14,2} = \binom{14}{2} = 91$ y son todas igualmente probables. Podemos entonces aplicar la regla de Laplace.

    (a) Si definimos el suceso $A = \{\text{al menos uno de los temas es de los 5 que el alumno ha estudiado}\}$ y consideramos el suceso $\overline{A}$, los casos favorables a este último serán $C_{9,2} = \binom{9}{2} = 36$. Por tanto la probabilidad buscada es

$$p(A) = 1 - p\left(\overline{A}\right) = 1 - \frac{C_{9,2}}{C_{14,2}} = 1 - \frac{36}{91} = 0.6044$$

    (b) Si n es el número de temas estudiados por el alumno, la probabilidad anterior sería

$$p(A) = 1 - \frac{C_{14-n,2}}{C_{14,2}} = 1 - \frac{(14-n)(13-n)}{182}$$

    y por tanto necesitaríamos que

$$1 - \frac{(14-n)(13-n)}{182} \geq \frac{1}{2}$$

    Con $n = 5$ ya hemos visto en el apartado anterior que esto se cumple. Para $n = 4$ tendríamos

$$1 - \frac{(14-n)(13-n)}{182} = 1 - \frac{90}{182} = 0.5055 \geq \frac{1}{2}$$

    mientras que, con $n = 3$ tendríamos

$$1 - \frac{(14-n)(13-n)}{182} = 1 - \frac{110}{182} = 0.3956 < \frac{1}{2}$$

Por tanto como mínimo debería estudiar al menos 4 temas para que la probabilidad sea superior a 1/2.

19. Se hacen tres disparos simultáneos con tres cañones distintos, siendo la probabilidad de alcanzar el objetivo 0.1, 0.2 y 0.3 , respectivamente. Calcular la probabilidad de cada uno de los números posibles de blancos. Calcular la probabilidad de obtener al menos un blanco.

    Solución. Consideramos los sucesos $A_i = \{\text{el } i\text{-ésimo cañón alcanza el objetivo}\}$ con $i = 1, 2, 3$. Sabemos que $p(A_1) = 0.1$, $p(A_2) = 0.2$ y $p(A_3) = 0.3$. El espacio muestral para el número posible de blancos será $\Omega = \{0, 1, 2, 3\}$ y, teniendo en cuenta que los sucesos A_i son independientes, las probabilidades de cada uno de los sucesos elementales será:

    $p(\{0\}) = p\left(\overline{A_1}\ \overline{A_2}\ \overline{A_3}\right) = p\left(\overline{A_1}\right) p\left(\overline{A_2}\right) p\left(\overline{A_3}\right) = (0.9)(0.8)(0.7) = 0.504$

    $p(\{1\}) = p\left(A_1\overline{A_2}\ \overline{A_3}\right) + p\left(\overline{A_1}A_2\overline{A_3}\right) + p\left(\overline{A_1}\ \overline{A_2}A_3\right) = (0.1)(0.8)(0.7) + (0.9)(0.2)(0.7) + (0.9)(0.8)(0.3) = 0.398$

    $p(\{2\}) = p\left(A_1A_2\overline{A_3}\right) + p\left(A_1\overline{A_2}A_3\right) + p\left(\overline{A_1}A_2A_3\right) = (0.1)(0.2)(0.7) + (0.1)(0.8)(0.3) + (0.9)(0.2)(0.3) = 0.092$

    $p(\{3\}) = p(A_1A_2A_3) = (0.1)(0.2)(0.3) = 0.006$

    La probabilidad de obtener al menos un blanco sería $1 - p(\{0\}) = 0.496$

20. En el lejano reino de Fabulandia a los condenados a muerte se les concedía la gracia de que su vida dependiera de que sacasen una bola blanca en el siguiente sorteo: se ponían 50 bolas blancas en una urna y 50 bolas negras en otra. Se vendaban los ojos al condenado, éste elegía en primer lugar una urna y luego extraía una bola. Mas en cierta ocasión un reo pidió la gracia de que se le dejara distribuir las bolas antes de hacer el sorteo. Tras alguna discusión con los magos de la corte se le concedió la gracia y colocó 1 bola blanca en una urna y en la otra 49 blancas y las 50 negras. ¿Mejoró el reo la probabilidad de salvar su vida?

    Solución. Consideramos los sucesos $B = \{\text{la bola extraída es blanca}\}$ y $A_i = \{\text{la urna elegida es la } i\text{-ésima}\}$ con $i = 1, 2$. Suponemos que la elección de la urna es aleatoria, con $p(A_1) = p(A_2) = 1/2$.Si tenemos 50 bolas blancas en el urna 1 y 50 bolas negras en la urna 2 tendremos que:

    $$p(B) = p(A_1)\,p(B/A_1) + p(A_2)\,p(B/A_2) = (1/2)(1) + (1/2)(0) = 0.5$$

    y la probabilidad de que el reo salve su vida será 0.5. Sin embargo, si la urna 1 contiene sólo una bola blanca y la urna 2 contiene 49 blancas y 50 negras tendremos que

    $$p(B) = p(A_1)\,p(B/A_1) + p(A_2)\,p(B/A_2) = (1/2)(1) + (1/2)(49/99) = 0.7475$$

    y la probabilidad de que el reo salve su vida sería mayor que antes.

21. Una empresa tiene establecido un programa que le permite servir todos sus pedidos con un retraso inferior a una semana. De datos anteriores se conoce que esta condición es cumplida en el 95 % de

los casos, lo cual se considera como satisfactorio. Reclamaciones por parte de algunos clientes hacen sospechar a la dirección que actualmente hay un mayor número de demoras, por lo cual se plantea revisar el proceso. Para ello establece la siguiente norma: se seleccionan al azar tres pedidos y se procederá a observar cómo son servidos. A la vista de la información que se obtenga procederá de la siguiente manera: Si ninguna de la tres órdenes se retrasa, no se harán más comprobaciones; si una o más órdenes se retrasan se procederá a revisar el proceso. ¿Cuál es la probabilidad de que el proceso sea revisado sin necesidad?

Solución. Definimos los sucesos $A_1 = \{\text{el primer pedido se retrasa}\}$, $A_2 = \{\text{el segundo pedido se retrasa}\}$ y $A_3 = \{\text{el tercer pedido se retrasa}\}$. Si el proceso funciona satisfactoriamente y no es necesario revisarlo tendremos que $p(A_1) = p(A_2) = p(A_3) = 0.05$. Para que el proceso sea revisado sin necesidad deberá ocurrir al menos uno de los tres sucesos, es decir, necesitaremos calcular $p(A_1 \cup A_2 \cup A_3)$. Aplicando la primera ley de De Morgan podemos asegurar que $\overline{A_1 \cup A_2 \cup A_3} = \overline{A_1} \cap \overline{A_2} \cap \overline{A_3}$ y por tanto $p(A_1 \cup A_2 \cup A_3) = 1 - p\left(\overline{A_1} \cap \overline{A_2} \cap \overline{A_3}\right)$. Además, como los tres sucesos son independientes tendremos que:

$$p(A_1 \cup A_2 \cup A_3) = 1 - p\left(\overline{A_1} \cap \overline{A_2} \cap \overline{A_3}\right) = 1 - p\left(\overline{A_1}\right) p\left(\overline{A_2}\right) p\left(\overline{A_3}\right) = 1 - (0.95)^3 = 0.143$$

22. En una parcela con 100 árboles sabemos que 30 están enfermos. Si se eligen, al azar, 20 árboles para ser talados, ¿cuál es la probabilidad de que 8 de esos 20 árboles estén enfermos?

Solución. Los posibles resultados para la elección de los 20 árboles serán $C_{100,20} = \binom{100}{20}$ y, como la elección es aleatoria, todos son igualmente probables. Podemos aplicar entonces la regla de Laplace en este experimento. Si definimos el suceso $A = \{\text{obtener 8 árboles enfermos en la muestra de tamaño 20}\}$ los casos favorables a A serán $C_{30,8}C_{70,12}$ y por tanto la probabilidad será

$$p(A) = \frac{C_{30,8}C_{70,12}}{C_{100,20}} = \frac{\binom{30}{8}\binom{70}{12}}{\binom{100}{20}} = 0.1162$$

23. Sean A y B dos sucesos de un espacio muestral. Supuestas conocidas $p(A)$, $p(B)$ y $P(A \cap B)$, calcular:

(a) La probabilidad de que ocurran exactamente k de los sucesos A y B ($k = 0, 1, 2$).

(b) La probabilidad de que ocurran al menos k de los sucesos A y B ($k = 0, 1, 2$).

(c) La probabilidad de que ocurran a lo sumo k de los sucesos A y B ($k = 0, 1, 2$).

(d) La probabilidad de que ocurra A pero no B.

Solución.

(a) $p(\text{exactamente } 0) = p\left(\overline{A} \cap \overline{B}\right) = p\left(\overline{A \cup B}\right) = 1 - p(A \cup B) = 1 - p(A) - p(B) + p(A \cap B)$

$$p(\text{exactamente } 1) = p\left(A \cap \overline{B}\right) + p\left(\overline{A} \cap B\right) = p(A) - p(A \cap B) + p(B) - p(A \cap B)$$
$$= p(A) + p(B) - 2p(A \cap B)$$

$$p(\text{exactamente } 2) = p(A \cap B)$$

(b) $p(\text{al menos } 0) = 1$

$p\,(\text{al menos } 1) = p\,(A \cup B) = p\,(A) + p\,(B) - p\,(A \cap B)$

$p\,(\text{al menos } 2) = p\,(A \cap B)$

(c) $p\,(\text{a lo sumo } 0) = p\,(\text{exactamente } 0) = p\left(\overline{A} \cap \overline{B}\right) = 1 - p(A) - p(B) + p\,(A \cap B)$

$p\,(\text{a lo sumo } 1) = p\,(\text{exactamente } 0) + \ p\,(\text{exactamente } 1) = 1 - p\,(A \cap B)$

$p\,(\text{a lo sumo } 2) = 1$

(d) $p\left(A \cap \overline{B}\right) = p\,(A) - p\,(A \cap B)$

24. Se llama fiabilidad de un sistema a la probabilidad de que el sistema funcione con éxito durante un tiempo fijado. Normalmente un sistema está formado por varios subsistemas cuyo éxito o fallo afecta al éxito del sistema total. Un sistema en serie se caracteriza porque el sistema total opera con éxito si y sólo si todos los subsistemas operan con éxito. Un sistema en paralelo se caracteriza porque funciona si al menos uno de los subsistemas funciona.

(a) Consideremos un sistema formado por n subsistemas con fiabilidad p_i $(i = 1, .., n)$. Obtener la fiabilidad del sistema en función de los p_i: (i) si el sistema está montado en serie. (ii) si el sistema está montado en paralelo.

(b) Sea A un sistema formado por 10 subsistemas acoplados en serie. Calcular la fiabilidad del sistema si la fiabilidad de cada uno de los subsistemas es de 0.99.

(c) Sea A un sistema formado por 3 subsistemas acoplados en paralelo. Calcular la fiabilidad del sistema si la fiabilidad de cada uno de los subsistemas es de 0.80.

Solución.

(a) Si definimos los sucesos $A_i = \{\text{el sistema } i \text{ opera con éxito}\}$ tendremos que la fiabilidad de un sistema en serie será $p\left(\bigcap\limits_{i=1}^{n} A_i\right)$. Como los sistemas funcionan de forma independiente tendremos $p\left(\bigcap\limits_{i=1}^{n} A_i\right) = \prod\limits_{i=1}^{n} p\,(A_i) = \prod\limits_{i=1}^{n} p_i$. En cambio, para un sistema en paralelo la fiabilidad será $p\left(\bigcup\limits_{i=1}^{n} A_i\right) = 1 - p\left(\bigcap\limits_{i=1}^{n} \overline{A_i}\right) = 1 - \prod\limits_{i=1}^{n} (1 - p_i)$

(b) Si $n = 10$, $p_i = 0.99$ y el sistema está acoplado en serie la fiabilidad será: $p = 0.99^{10} = 0.904$

(c) Si $n = 3$, $p_i = 0.80$ y el sistema está acoplado en paralelo la fiabilidad será: $p = 1 - 0.20^3 = 0.992$

25. Una flota de nueve taxis se destina al azar a tres aeropuertos A, B y C: 2 taxis van a A, 4 van a B y 3 taxis se destinan al aeropuerto C.

(a) ¿De cuántos modos se puede hacer esta asignación?

(b) ¿Cuál es la probabilidad de que el taxi que conduce Juan sea asignado al aeropuerto C?

Solución.

(a) Las posibles asignaciones serán $VR_{3,9}^{2,4,3} = \frac{9!}{2!4!3!} \frac{362880}{288} = 1260$ y si la elección es aleatoria todas son igualmente probables.

(b) Si el taxi de Juan va al aeropuerto C, las posibles asignaciones sería $VR_{3,8}^{2,4,2} = \frac{8!}{2!4!2!} = \frac{40320}{96} = 420$ y por tanto, utilizando la regla de Laplace, la probabilidad que buscamos será $\frac{VR_{3,8}^{2,4,2}}{VR_{3,9}^{2,4,3}} = \frac{420}{1260} = \frac{1}{3}$

26. Demostrar que si $1 \leq r \leq n$ entonces $\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$

Solución. Utilizando las fórmulas de los números combinatorios tenemos:

$$\begin{aligned} \binom{n-1}{r-1} + \binom{n-1}{r} &= \frac{(n-1)...(n-r+1)}{(r-1)...1} + \frac{(n-1)...(n-r+1)(n-r)}{r(r-1)...1} \\ &= \frac{[(n-1)...(n-r+1)r] + [n(n-1)...(n-r+1)(n-r)]}{r(r-1)...1} \\ &= \frac{(n-1)...(n-r+1)\,(r+n-r)}{r(r-1)...1} = \frac{n(n-1)...(n-r+1)}{r(r-1)...1} = \binom{n}{r} \end{aligned}$$

27. Un inspector de calidad tiene a su cargo la inspección de 10 líneas de envasado. Cada mañana (de una semana de 5 días) selecciona al azar una de las líneas para inspeccionarla ese día. Calcular la probabilidad de que seleccione una línea más de una vez durante la semana.

Solución. La probabilidad de seleccionar una línea más de una vez por semana será uno menos la probabilidad no repetir ninguna línea durante la semana. Como la selección se hace de forma aleatoria podemos utilizar la regla de Laplace. Los casos posibles para la elección de las líneas durante una semana serán $VR_{10,5} = 10^5$ y los casos favorables para no repetir ninguna línea durante la semana serán $V_{10,5}$. Por tanto la probabilidad buscada será:

$$p = 1 - \frac{V_{10,5}}{VR_{10,5}} = 1 - \frac{(10)\,(9)\,(8)\,(7)\,(6)}{10^5} = 1 - 0.3024 = 0.6976$$

28. El Servicio de Información de Incendios dio la siguiente tabla acerca de los incendios que ocurrieron en Estados Unidos durante los últimos diez años. Todas las cifras de la tabla son en porcentajes y, por ejemplo, el 22 de la esquina superior izquierda señala que el 22 % de los incendios que se registraron en hogares fueron provocados por la calefacción.

Causa del incendio	Hogares	Apartamentos	Casas móviles	Hoteles y moteles	Otros lugares
Calefacción	22	6	22	8	45
Al cocinar	15	24	13	7	0
Materias inflamables	10	15	7	16	8
Por fumar	7	18	6	36	19
Instalación eléctrica	8	5	15	7	28
Otras	38	32	37	26	0
Todas las causas	73	20	3	2	2

Si en una estación de bomberos se recibe una llamada, determinar la probabilidad de que sea por:

(a) un incendio en un hogar.

(b) un incendio causado por la calefacción, dado que fue en un apartamento.

(c) un incendio provocado por cada una de las causas (es decir, la probabilidad de un incendio originado por la calefacción, la probabilidad de un incendio provocado al cocinar, ...)

(d) un incendio en un apartamento y originado por la calefacción.

(e) un incendio en un hogar, dado que fue provocado por la calefacción.

(f) en una casa móvil, dado que fue causado por fumar.

Solución. Las probabilidades expresadas en porcentaje que se dan en le interior de la tabla son las probabilidades de las causas del incendio condicionadas por el lugar donde se produce. La última fila de la tabla nos da las probabilidades, también porcentuales, de los lugares del incendio. Entonces tenemos:

(a) $p\,(\text{incendio en hogar}) = 0.73$

(b) $p\,(\text{calefacción/apartamento}) = 0.06$

(c) $p\,(\text{calefacción}) = (0.73)\,(0.22)+(0.20)\,(0.06)+(0.03)\,(0.22)+(0.02)\,(0.08)+(0.02)\,(0.45) = 0.1898$

$p\,(\text{cocinar}) = (0.73)\,(0.15) + (0.20)\,(0.24) + (0.03)\,(0.13) + (0.02)\,(0.07) + (0.02)\,(0) = 0.1628$

$p\,(\text{inflamables}) = (0.73)\,(0.10)+(0.20)\,(0.15)+(0.03)\,(0.07)+(0.02)\,(0.16)+(0.02)\,(0.08) = 0.1099$

$p\,(\text{fumar}) = (0.73)\,(0.07) + (0.20)\,(0.18) + (0.03)\,(0.06) + (0.02)\,(0.36) + (0.02)\,(0.19) = 0.0999$

$p\,(\text{electricidad}) = (0.73)\,(0.08)+(0.20)\,(0.05)+(0.03)\,(0.15)+(0.02)\,(0.07)+(0.02)\,(0.28) = 0.0799$

$p\,(\text{otras}) = (0.73)\,(0.38) + (0.20)\,(0.32) + (0.03)\,(0.37) + (0.02)\,(0.26) + (0.02)\,(0) = 0.3577$

(d) $p\,(\text{apartamento y calefacción}) = p\,(\text{apartamento})\,p\,(\text{calefacción/apartamento}) = (0.20)\,(0.06) = 0.012$

(e) $p\,(\text{hogar/calefacción}) = \frac{p(\text{hogar y calefacción})}{p(\text{calefacción})} = \frac{p(\text{hogar})p(\text{calefacción/hogar})}{p(\text{calefacción})} = \frac{(0.73)(0.22)}{0.1898} = 0.8462$

(f) $p\,(\text{casa movil/fumar}) = \frac{p(\text{casa movil y fumar})}{p(\text{fumar})} = \frac{p(\text{casa móvil})p(\text{fumar/casa móvil})}{p(\text{fumar})} = \frac{(0.03)(0.06)}{0.0999} = 0.0180$

29. Una empresa produce resistencias que vende como resistencias de 10 omhnios. Sin embargo, los omhnios reales de los resistores pueden variar. Se ha observado que el 5 % de los valores son menores que 9.5 omhnios y el 10 % mayores que 10.5 omhnios. Si en un determinado sistema se usan dos de esas resistencias, seleccionadas al azar, calcular la probabilidad de que:

(a) ambas tengan valores reales comprendidos entre 9.5 y 10.5 omhnios.

(b) al menos una tenga un valor real mayor que 10.5 omhnios.

Solución.

(a) Sean los sucesos $A_i = \{\text{la resistencia } i \text{ está entre 9.5 y 10.5 ohmios}\}$, con $i = 1, 2$. Por los datos del problema tendremos que $p\,(A_i) = 1 - 0.05 - 0.10 = 0.85$. Ahora necesitamos calcular $p\,(A_1A_2)$ y como los sucesos son independientes tendremos que $p\,(A_1A_2) = p\,(A_1)\,p\,(A_2) = 0.85^2 = 0.7225$

(b) Sena los sucesos $B_i = \{\text{la resistencia } i \text{ tiene un valor mayor que 10.5 ohmios}\}$, con $i = 1, 2$. Por los datos del problema tendremos que $p\,(B_i) = 0.10$. Ahora necesitamos calcular $p\,(B_1 \cup B_2)$ y como los sucesos son independiente tendremos:

$$p\,(B_1 \cup B_2) = p\,(B_1) + p\,(B_2) - p\,(B_1B_2) = 0.1 + 0.1 - 0.1^2 = 0.19$$

30. En una clase hay 23 estudiantes. ¿Cuál es la probabilidad de que el cumpleaños de por lo menos dos de ellos sea en la misma fecha (día y mes)? (suponer que el año tiene 365 días)

Solución. Si suponemos que todos los días son igualmente probables para nacer, los casos posibles para las 23 fechas de nacimiento de los estudiantes serán $VR_{365,23}$ y todas ellas serán igualmente probables. Ahora, si consideramos el suceso $A = \{\text{al menos dos fechas de cumpleaños coinciden}\}$ los casos favorables para el suceso $\overline{A}$ serán $V_{365,23}$ y por tanto, aplicando la regla de Laplace, tendremos que

$$p(A) = 1 - p\left(\overline{A}\right) = 1 - \frac{V_{365,23}}{VR_{365,23}} = 1 - 0.4927 = 0.5073$$

31. Supóngase que la probabilidad de exposición a la gripe durante una epidemia es 0.6. La experiencia ha demostrado que una vacuna tiene el 80 % de éxito en la prevención de la gripe en una persona que ha sido inoculada con ella, si la persona se expone al virus. Alguien que no esté vacunado se enfrenta a una probabilidad de 0.9 de contagiarse de la enfermedad, si se expone al virus.

    (a) ¿Cuál es la probabilidad de que una persona que ha sido vacunada coja la gripe?

    (b) ¿Cuál es la probabilidad de que se contagie una persona que no ha sido vacunada?

    Solución. Si consideramos los sucesos $A = \{\text{la persona se expone al virus}\}$ y $B = \{\text{la persona se contagia de gripe}\}$ tenemos que $p(A) = 0.6$. Además, si la persona ha sido vacunada tendremos que $p(B/A) = 1 - 0.8 = 0.2$ mientras que si la persona no ha sido vacunada tendremos que $p(B/A) = 0.9$.

    (a) Si la persona ha sido vacunada tendremos:

$$p(B) = p(A)\,p(B/A) + p\left(\overline{A}\right)p\left(B/\overline{A}\right) = (0.6)(0.2) + (0.4)(0) = 0.12$$

    (b) Si la persona no ha sido vacunada tendremos:

$$p(B) = p(A)\,p(B/A) + p\left(\overline{A}\right)p\left(B/\overline{A}\right) = (0.6)(0.9) + (0.4)(0) = 0.54$$

32. En un experimento epidemiológico con ratas se utilizan jaulas con una rata sana en cada jaula. Una segunda rata, infectada con uno de los microorganismos A, B y C se introduce en cada jaula. La probabilidad de que la rata lo esté con el microorganismo A, B o C es de $1/3$ en cada caso. Si se introduce en la jaula una rata con la infección A la probabilidad es de $1/2$ de que la rata sana se contagie. Estas probabilidades son de $1/3$ y $1/4$ si la rata introducida estaba infectada con los microorganismos B o C, respectivamente. La rata sana de la jaula enfermó. ¿Cuál es la probabilidad de que la rata introducida estuviera contagiada con A? ¿Y con B? ¿Y con C?

    Solución. Consideramos los sucesos $A = \{\text{rata introducida con } A\}$, $B = \{\text{rata introducida con } B\}$, $C = \{\text{rata introducida con } C\}$ y $D = \{\text{rata sana de la jaula enferma}\}$. Sabemos que $p(A) = p(B) = p(C) = 1/3$ y además $p(D/A) = 1/2$, $p(D/B) = 1/3$ y $p(D/C) = 1/4$. Sabemos que ha ocurrido el suceso D y tenemos que calcular $p(A/D)$. Aplicando la regla de Bayes tendremos

$$\begin{aligned} p(A/D) &= \frac{p(AD)}{p(D)} = \frac{p(A)\,p(D/A)}{p(A)\,p(D/A) + p(B)\,p(D/B) + p(C)\,p(D/C)} \\ &= \frac{(1/3)(1/2)}{(1/3)(1/2) + (1/3)(1/3) + (1/3)(1/4)} = \frac{1/2}{13/12} = 0.4615 \end{aligned}$$

Del mismo modo, para B y C tendremos

$$p\,(B/D) = \frac{(1/3)\,(1/3)}{(1/3)\,(1/2) + (1/3)\,(1/3) + (1/3)\,(1/4)} = \frac{1/3}{13/12} = 0.3077$$

$$p\,(C/D) = \frac{(1/3)\,(1/4)}{(1/3)\,(1/2) + (1/3)\,(1/3) + (1/3)\,(1/4)} = \frac{1/4}{13/12} = 0.2308$$

33. (a) Se considera una caja de 50 naranjas de las que el 10 % están heladas. Se extrae una naranja al azar. Sea A el suceso "la naranja está helada". Se extrae, a continuación, una segunda naranja (sin volver a reponer la primera naranja extraída). Sea B el suceso "la segunda naranja está helada". Calcular $p(A)$, $p(B)$, $p(B/A)$ y $p(B/\bar{A})$. ¿Son los sucesos A y B independientes?

    (b) Repetir los cálculos anteriores suponiendo que se considera un camión cargado con 100000 naranjas de las que el 10 % están heladas. ¿Puede considerarse que los sucesos A y B son independientes?

    Solución.

    (a) La caja contendrá 5 naranjas heladas y 45 no heladas. Si las extracciones se hacen de forma aleatoria y sin devolución tendremos

    $p\,(A) = 5/50 = 0.1$, $p\,(B/A) = 4/49 = 0.8163$, $p\left(B/\bar{A}\right) = 5/49 = 0.1020$ y, por el teorema de la probabilidad total, tendremos

$$p\,(B) = p\,(A)\,p\,(B/A) + p\left(\bar{A}\right)p\left(B/\bar{A}\right) = \left(\frac{5}{50}\right)\left(\frac{4}{49}\right) + \left(\frac{45}{50}\right)\left(\frac{5}{49}\right) = 0.1$$

    Además también tendremos que

$$p\,(AB) = p\,(A)\,p\,(B/A) = \left(\frac{5}{50}\right)\left(\frac{4}{49}\right) = 0.0082$$

    mientras que $p\,(A)\,p\,(B) = 0.01$. Por tanto $p\,(AB) \neq p\,(A)\,p\,(B)$ y los sucesos no son independientes.

    (b) Si repetimos el problema con 100000 naranjas, de las cuales 10000 estarían heladas, seguiríamos teniendo $p\,(A) = p\,(B) = 0.1$ pero ahora sería

$$p\,(AB) = p\,(A)\,p\,(B/A) = \left(\frac{10000}{100000}\right)\left(\frac{9999}{99999}\right) = 0.0099991$$

    Por tanto prácticamente tenemos $p\,(AB) = p\,(A)\,p\,(B)$ y los sucesos son prácticamente independientes.

34. Un hombre va de pesca. En un bote lleva 3 carnadas de tipo A, 7 carnadas de tipo B y 10 carnadas de tipo C. La mejor carnada es la A: la probabilidad de pescar un pez con ella es $3/5$. La probabilidad de pescar un pez con las otras carnadas es sólo de $2/7$. Mete la mano en el bote y saca una carnada al azar. Calcular:

    (a) Probabilidad de que pesque un pez.

    (b) Si el hombre tiene éxito en la pesca, ¿cuál es la probabilidad de que utilizara una carnada de tipo B?

(c) Si saca dos carnadas a la vez, ¿cuál es la probabilidad de que ninguna sea de tipo A?

Solución.

(a) Definimos los siguiente sucesos para el problema: $A = \{\text{elegir carnada tipo } A\}$, $B = \{\text{elegir carnada tipo } B\}$, $C = \{\text{elegir carnada tipo } C\}$, y $D = \{\text{el hombre tiene éxito en la pesca}\}$. Con los datos del problema sabemos que $p(A) = 3/20$, $p(B) = 7/20$, $p(C) = 10/20$, $p(D/A) = 3/5$, $p(D/B) = p(D/C) = 2/7$. Tenemos que calcular $p(D)$. Para ello aplicamos el Teorema de la Probabilidad Total utilizando los sucesos A, B y C como partición.

$$\begin{aligned} p(D) &= p(A)\,p(D/A) + p(B)\,p(D/B) + p(C)\,p(D/C) = \left(\frac{3}{20}\right)\left(\frac{3}{5}\right) + \left(\frac{7}{20}\right)\left(\frac{2}{7}\right) + \left(\frac{1}{2}\right)\left(\frac{2}{7}\right) \\ &= \frac{233}{700} = 0.333 \end{aligned}$$

(b) Ahora hay que calcular $p(B/D)$ y para ello utilizamos la Regla de Bayes:

$$p(B/D) = \frac{p(B)\,p(D/B)}{p(D)} = \frac{\left(\frac{7}{20}\right)\left(\frac{2}{7}\right)}{\frac{233}{700}} = \frac{70}{233} = 0.300$$

(c) Definimos ahora el suceso $E = \{\text{ninguna de las dos carnada elegidas es de tipo } A\}$. Como la elección se hace de forma aleatoria entre las 20 carnadas podemos aplicar la regla de Laplace, es decir casos favorables dividido por casos posibles. Los casos posibles son $C_{20,2}$ y los casos favorables son $C_{17,2}$. Por tanto la probabilidad buscada es:

$$p(E) = \frac{\binom{17}{2}}{\binom{20}{2}} = \frac{(17)(16)}{(20)(19)} = 0.716$$

35. La siguiente tabla da el número de muertes por accidente que se registraron en Estados Unidos durante 1984, para cuatro causas específicas más un apartado para otros tipos de accidente, desglosadas por edades:

	Tráfico	Atropello	Caídas	Ahogamientos	Otras	
Menos de 5 años	1132	1024	114	638	744	3652
De 5 a 14	2263	925	68	532	410	4198
De 15 a 24	19738	2546	399	1353	765	24801
De 25 a 44	15036	2396	936	1549	1581	21498
De 45 a 64	6954	3521	1624	763	1411	14273
De 65 a 74	4020	2954	1702	281	467	9424
Más de 75	3114	4381	7067	272	1231	16065
	52257	17747	11910	5388	6609	93911

(a) Obtener la distribución de probabilidades marginales según la edad.

(b) Obtener la distribución de probabilidades marginales según el tipo de accidente.

(c) Obtener la tabla de probabilidades condicionales del tipo de accidente según las distintas edades de la población accidentada.

(d) ¿Cuál es la causa más probable de accidente para una víctima de más de 75 años?

(e) ¿Cuál es la causa más probable de accidente para una víctima de edad comprendida entre los 15 y los 24 años?

(f) ¿Cuál es el tipo de accidente menos probable?

(g) ¿Cuál es la probabilidad de que la víctima tenga una edad comprendida entre los 25 y 44 años, si se sabe que el accidente fue de tráfico?

Solución.

(a) Expresando los resultados en porcentajes redondeados, la distribución de probabilidades es:

Menos de 5	De 5 a 14	De 15 a 24	De 25 a 44	De 45 a 64	De 65 a 74	Más de 75
$\frac{3652}{93911}$=4 %	$\frac{4198}{93911}$=4 %	$\frac{24801}{93911}$=26 %	$\frac{21498}{93911}$=23 %	$\frac{14273}{93911}$=15 %	$\frac{9424}{93911}$=10 %	$\frac{16065}{93911}$=17 %

(b) Expresando los resultados en porcentajes redondeados, la distribución de probabilidades es:

Tráfico	Atropello	Caídas	Ahogamientos	Otras
$\frac{52257}{93911}$=56 %	$\frac{17747}{93911}$=19 %	$\frac{11910}{93911}$=13 %	$\frac{5388}{93911}$=6 %	$\frac{6609}{93911}$=7 %

(c) Expresando los resultados en porcentajes redondeados, las distribuciones condicionales del tipo de accidente según la edad son:

	Tráfico	Atropello	Caídas	Ahogamientos	Otras
Menos de 5 años	$\frac{1132}{3652}$=31 %	$\frac{1024}{3652}$=28 %	$\frac{114}{3652}$=3 %	$\frac{638}{3652}$=17 %	$\frac{744}{3652}$=201 %
De 5 a 14	$\frac{2263}{4198}$=54 %	$\frac{925}{4198}$=22 %	$\frac{68}{4198}$=2 %	$\frac{532}{4198}$=13 %	$\frac{410}{4198}$=10 %
De 15 a 24	$\frac{19738}{24801}$=80 %	$\frac{2546}{24801}$=10 %	$\frac{399}{24801}$=2 %	$\frac{1353}{24801}$=5 %	$\frac{765}{24801}$=3 %
De 25 a 44	$\frac{15036}{21498}$=70 %	$\frac{2396}{21498}$=11 %	$\frac{936}{21498}$=4 %	$\frac{1549}{21498}$=7 %	$\frac{1581}{21498}$=7 %
De 45 a 64	$\frac{6954}{14273}$=49 %	$\frac{3521}{14273}$=25 %	$\frac{1624}{14273}$=11 %	$\frac{763}{14273}$=5 %	$\frac{1411}{14273}$=10 %
De 65 a 74	$\frac{4020}{9424}$=43 %	$\frac{2954}{9424}$=31 %	$\frac{1702}{9424}$=18 %	$\frac{281}{9424}$=3 %	$\frac{467}{9424}$=5 %
Más de 75	$\frac{3114}{16065}$=19 %	$\frac{4381}{16065}$=27 %	$\frac{7067}{16065}$=44 %	$\frac{272}{16065}$=2 %	$\frac{1231}{16065}$=8 %

(d) Caídas con un porcentaje de 44 %

(e) Tráfico con un porcentaje de 80 %

(f) Según la tabla del apartado (b) el tipo de accidente menos probable será el de ahogamiento, con un porcentaje de 6 %

(g) El total de accidentes de tráfico fue de 52257, de los cuales 15036 afectaron a una edad comprendida entre los 25 y 44 años. Por tanto la probabilidad buscada será 15036/52257 = 0.288

36. El 30 % de los enfermos de hepatitis sufre hepatitis obstructiva que exige una intervención quirúrgica mientras que el otro 70 % tiene hepatitis infecciosa que puede curarse simplemente con el reposo y la medicación. Para discernir entre ambas situaciones cuando se ingresa en un hospital se realiza una determinada prueba clínica, que puede dar positiva o negativa. Se sabe que la probabilidad de que la prueba resulte positiva es de 0.95 cuando los enfermos tienen hepatitis obstructiva y de 0.10 cuando la tienen infecciosa. Supongamos que en un hospital se producen 150 ingresos por hepatitis durante cierto año y que se opera a todos aquellos que dan positivo en la prueba.

(a) ¿Cuántas intervenciones, aproximadamente, se realizarán en dicho año? ¿Cuántas intervenciones que no deberían realizarse se practican ese año? ¿Cuántos diagnósticos erróneos se realizarán?

(b) Sabiendo que un enfermo ha dado positivo en la prueba, ¿cuál es la probabilidad de que tenga hepatitis infecciosa?

(c) Si la prueba da negativa en un paciente, ¿cuál es, pese a ello, la probabilidad de que realmente tenga una hepatitis obstructiva?

Solución. Definimos los sucesos $A = \{\text{paciente con hepatitis obstructiva}\}$ y $B = \{\text{prueba clínica positiva}\}$. Sabemos que $p(A) = 0.30$, $p(B/A) = 0.95$ y $p(B/\overline{A}) = 0.10$

(a) Necesitamos calcular $p(B)$ para saber el número de intervenciones. Por el Teorema de la Probabilidad Total tendremos que

$$p(B) = p(A)p(B/A) + p(\overline{A})p(B/\overline{A}) = (0.30)\,(0.95) + (0.70)\,(0.10) = 0.355$$

y se intervendrán el 22 % de los ingresos, es decir, $(150)\,(0.355) = 53.25 \simeq 53$ intervenciones anuales.

Para saber cuántas de ellas no deberían haberse practicado necesitamos calcular $p\left(\overline{A}B\right)$ utilizando la regla de la multiplicación:

$$p\left(\overline{A}B\right) = p(\overline{A})p(B/\overline{A}) = (0.70)\,(0.10) = 0.07$$

Por tanto el 7 % de las intervenciones no deberían haberse practicado, es decir, $(150)\,(0.07) = 10.5 \simeq 11$

Para saber cuántos diagnósticos erróneos se realizarán necesitamos calcular $p\left(A\overline{B} + \overline{A}B\right)$:

$$p\left(A\overline{B} + \overline{A}B\right) = p\left(A\overline{B}\right) + p\left(\overline{A}B\right) = p\,(A)\,p\left(\overline{B}/A\right) + 0.07 = (0.30)\,(0.05) + 0.07 = 0.085$$

Por tanto tendremos $(150)\,(0.105) = 12.75 \simeq 13$ diagnósticos erróneos.

(b) Tendremos que calcular $p\left(\overline{A}/B\right)$ utilizando la regla de Bayes:

$$p\left(\overline{A}/B\right) = \frac{p\left(\overline{A}B\right)}{p(B)} = \frac{p(\overline{A})p(B/\overline{A})}{p(A)p(B/A) + p(\overline{A})p(B/\overline{A})} = \frac{0.07}{0.355} = 0.1972$$

(c) Tendremos que calcular $p\left(A/\overline{B}\right)$ también por la regla de Bayes:

$$p\left(A/\overline{B}\right) = \frac{p\left(A\overline{B}\right)}{p\left(\overline{B}\right)} = \frac{p\,(A)\,p\left(\overline{B}/A\right)}{1 - p\,(B)} = \frac{(0.30)\,(0.05)}{1 - 0.355} = 0.0233$$

37. Un concursante debe elegir entre tres puertas detrás de una de las cuales se encuentra el premio. Hecha la elección y antes de abrir la puerta, el presentador le muestra que en una de las dos puertas no elegidas no está el premio y le da la posibilidad de reconsiderar su elección. ¿Qué debería hacer el concursante?

Solución. Consideramos los sucesos $A_i = \{\text{premio en puerta } i\}$ y $B_i = \{\text{presentador abre puerta } i\ \}$ con $i = 1, 2, 3$. Vamos a resolver el problema suponiendo que el concursante ha elegido inicialmente la puerta 1 (igual se haría si hubiese elegido la 2 o la 3) y el presentador abre la puerta 2 (igual se razonaría se el presentador abre la puerta 3). Suponemos que $p(A_i) = 1/3$ porque a priori las tres puertas con igualmente probables. Además, si el presentador actúa de forma aleatoria, tendremos que $p(B_2/A_1) = p(B_3/A_1) = 1/2$, $p(B_2/A_2) = 0$, $p(B_3/A_2) = 1$, $p(B_2/A_3) = 1$ y $p(B_3/A_3) = 0$. Para saber si el concursante debería cambiar su elección inicial necesitaríamos calcular $p(A_1/B_2)$ y $p(A_3/B_2)$. Utilizando la regla de Bayes estas probabilidades serían

$$\begin{aligned} p(A_1/B_2) &= \frac{p(A_1)\,p(B_2/A_1)}{p(A_1)\,p(B_2/A_1) + p(A_2)\,p(B_2/A_2) + p(A_3)\,p(B_2/A_3)} = \\ &= \frac{(1/3)\,(1/2)}{(1/3)\,(1/2) + (1/3)\,(0) + (1/3)\,(1)} = \frac{1/2}{3/2} = \frac{1}{3} \end{aligned}$$

$$\begin{aligned} p(A_3/B_2) &= \frac{p(A_3)\,p(B_2/A_3)}{p(A_1)\,p(B_2/A_1) + p(A_2)\,p(B_2/A_2) + p(A_3)\,p(B_2/A_3)} = \\ &= \frac{(1/3)\,(1)}{(1/3)\,(1/2) + (1/3)\,(0) + (1/3)\,(1)} = \frac{1}{3/2} = \frac{2}{3} \end{aligned}$$

Por tanto el concursante debería cambiar su elección y quedarse con la puerta 3 en vez de con la 1 que había elegido inicialmente.

38. Un banco ha estimado, por experiencias anteriores, que la probabilidad de que una persona falle en los pagos de un préstamo personal es de 0.3. También ha estimado que el 30 % de los préstamos no pagados a tiempo se han hecho para financiar viajes de vacaciones y el 60 % de los préstamos pagados a tiempo también se han hecho para financiar viajes de vacaciones.

    (a) ¿Cuál es la probabilidad de que un préstamo que se haga para financiar un viaje de vacaciones no se pague a tiempo?

    (b) ¿Cuál es la probabilidad de que, si el préstamo se ha hecho para propósitos distintos a viajes de vacaciones, sea pagado a tiempo?

    Solución. Consideramos los sucesos $A = \{\text{el préstamo es para financiar viaje de vacaciones}\}$ y $B = \{\text{la persona no paga a tiempo}\}$. Sabemos que $p(B) = 0.3$, $p(A/B) = 0.3$ y $p\left(A/\overline{B}\right) = 0.6$

    (a) Tenemos que calcular $p(B/A)$ y para ello utilizamos la regla de Bayes:

$$p(B/A) = \frac{p(B)\,p(A/B)}{p(B)\,p(A/B) + p\left(\overline{B}\right)p\left(A/\overline{B}\right)} = \frac{(0.3)\,(0.3)}{(0.3)\,(0.3) + (0.7)\,(0.6)} = \frac{0.09}{0.51} = 0.176$$

    (b) Tenemos calcular $p\left(\overline{B}/\overline{A}\right)$ y para ello utilizamos la regla de Bayes:

$$p\left(\overline{B}/\overline{A}\right) = \frac{p\left(\overline{B}\right)p\left(\overline{A}/\overline{B}\right)}{p(B)\,p\left(\overline{A}/B\right) + p\left(\overline{B}\right)p\left(\overline{A}/\overline{B}\right)} = \frac{(0.7)\,(1-0.6)}{(0.3)\,(1-0.3) + (0.7)\,(1-0.6)} = \frac{0.28}{0.49} = 0.571$$

39. En el jardinero del Sr. Rodríguez no se puede confiar: la probabilidad de que se olvide de regar al rosal favorito durante la ausencia del Sr. Rodríguez es de 2/3. El rosal está en un estado inseguro: si se le riega tiene igual probabilidad de progresar que de secarse, pero solamente una probabilidad de 0.25 de progresar si no se le riega. Después de su regreso, el Sr. Rodríguez se encuentra con que su rosal favorito se ha secado. ¿Cuál es la probabilidad de que el jardinero no lo haya regado?

Solución. Consideramos los sucesos $A = \{\text{el jardinero olvida regar el rosal}\}$ y $B = \{\text{el rosal se seca}\}$. Sabemos que $p(A) = 2/3$, $p\left(\overline{B}/A\right) = 0.25$ y $p\left(B/\overline{A}\right) = 0.5$. Tenemos que calcular $p(A/B)$ y para ello usamos la regla de Bayes:

$$p(A/B) = \frac{p(A)\,p(B/A)}{p(A)\,p(B/A) + p\left(\overline{A}\right)p\left(B/\overline{A}\right)} = \frac{(2/3)\,(1-0.25)}{(2/3)\,(1-0.25) + (1/3)\,(0.5)} = 0.75$$

40. Un parque natural está separado en dos zonas A y B por un río. Hay 10 ciervos en la zona A y otros 10 en la zona B. Un biólogo está realizando investigaciones sobre la conducta de un cierto ciervo que vive en la zona A, al que cariñosamente se le conoce como Bambi. Por un descuido de los vigilantes del parque, 9 ciervos de la zona A se pasan a la zona B. Los vigilantes lo advierten, y devuelven 9 ciervos (escogidos al azar) al territorio A. Informado el biólogo de tal contingencia y deseando proseguir sus pesquisas sobre Bambi, ¿en cuál de las dos zonas es preferible que empiece a buscar a su ciervo?

Solución. Suponemos que los 9 ciervos que pasaron a la zona B son elegidos aleatoriamente entre los 10 que hay en la zona A. Consideramos los sucesos $C = \{\text{Bambi está en la zona } A\}$ y $D = \{\text{Bambi pasó a la zona } B\}$. Aplicando el Teorema de la Probabilidad Total tendremos que

$$\begin{aligned} p(C) &= p(D)\,p(C/D) + p(\overline{D})p\left(C/\overline{D}\right) = \left(\frac{C_{9,8}}{C_{10,9}}\right)\left(\frac{C_{18,8}}{C_{19,9}}\right) + \left(\frac{C_{9,9}}{C_{10,9}}\right)(1) \\ &= \left(\frac{C_{9,1}}{C_{10,1}}\right)\left(\frac{C_{18,8}}{C_{19,9}}\right) + \left(\frac{1}{10}\right) = \left(\frac{9}{10}\right)\left(\frac{9}{19}\right) + \left(\frac{1}{10}\right) = 0.5263 \end{aligned}$$

Por tanto el biólogo debería buscar a Bambi en la zona A, porque es más probable que esté allí.

41. En un sistema de emergencia, la probabilidad de que se produzca una situación de peligro es de 0.1. Si ésta se produce, la probabilidad de que el sistema de alarma funcione es de 0.95. La probabilidad de que la alarma funcione sin haber habido situación de peligro es de 0.03. Hallar:

(a) La probabilidad de que habiendo funcionado la alarma, no haya habido situación de emergencia.

(b) La probabilidad de que haya una situación de emergencia y la alarma no funcione.

(c) La probabilidad de que, no habiendo funcionado la alarma, haya una situación de peligro.

Solución. Consideramos los sucesos $A = \{\text{hay una situación de peligro}\}$ y $B = \{\text{la alarma funciona}\}$. Sabemos que $p(A) = 0.1$, $p(B/A) = 0.95$ y $p\left(B/\overline{A}\right) = 0.03$.

(a) Hay que calcular $p\left(\overline{A}/B\right)$ y para ello utilizamos la regla de Bayes:

$$p\left(\overline{A}/B\right) = \frac{p\left(\overline{A}\right)p\left(B/\overline{A}\right)}{p(A)\,p(B/A) + p\left(\overline{A}\right)p\left(B/\overline{A}\right)} = \frac{(0.9)\,(0.03)}{(0.1)\,(0.95) + (0.9)\,(0.03)} = 0.221$$

(b) Hay que calcular $p\left(A\overline{B}\right)$ y para ello utilizamos la regla de la multiplicación:

$$p\left(A\overline{B}\right) = p(A)\,p\left(\overline{B}/A\right) = p(A)\,(1 - p(B/A)) = (0.1)\,(0.05) = 0.005$$

(c) Hay que calcular $p\left(A/\overline{B}\right)$ y para ello usamos la regla de Bayes:

$$p\left(A/\overline{B}\right) = \frac{p(A)\,p\left(\overline{B}/A\right)}{p(A)\,p\left(\overline{B}/A\right) + p\left(\overline{A}\right)p\left(\overline{B}/\overline{A}\right)} = \frac{(0.1)\,(1-0.95)}{(0.1)\,(1-0.95) + (0.9)\,(1-0.03)} = 0.0057$$

42. Las causas por las que puede dejar de funcionar el motor de una automóvil se clasifican en tres categorías, que se suponen independientes, A, B y C. La probabilidad de fallo del motor por la causa A en su primer año de uso es 0.1 y las probabilidades análogas por las causas B y C son 0.2 y 0.3, respectivamente. Hallar la probabilidad de que un motor falle en su primer año de uso.

    Solución. Sabemos que $p(A) = 0.1$, $p(B) = 0.2$, $p(C) = 0.3$ y los sucesos son independientes. La probabilidad de que el motor falle en su primer año de uso será $p(A \cup B \cup C)$. Utilizando la primera ley de De Morgan y la independencia de los sucesos tendremos que:

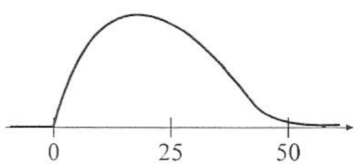
$$\begin{aligned} p(A \cup B \cup C) &= 1 - p\left(\overline{A \cup B \cup C}\right) = 1 - p\left(\overline{A}\ \overline{B}\ \overline{C}\right) = 1 - p\left(\overline{A}\right) p\left(\overline{B}\right) p\left(\overline{C}\right) \\ &= 1 - (0.9)(0.8)(0.7) = 0.496 \end{aligned}$$

1.2. PROBLEMAS RESUELTOS CAPÍTULO 4

1. Una compañía de refrescos anuncia premio en las chapas asegurando que de cada 1000 chapas hay 400 con "inténtalo de nuevo", 300 con premio de 0.5 euros, 250 con premio de 1 euro, 40 con premio de 5 euros y 10 con premio de 10 euros. Un individuo, al que no le gusta el refresco. decide comprar una botella cuyo coste es de 1 euro. Caracterizar su ganancia mediante una variable aleatoria. ¿Cuál es la probabilidad de que el individuo no pierda dinero con la compra?

   Solución. Si consideramos la variable aleatoria X=ganancia obtenida con la compra de la botella expresada en euros, su soporte será $Sop(X) = \{-1, -0.5, 0, 4, 9\}$ y sus probabilidades serán $p_1 = \frac{400}{1000} = 0.4$, $p_2 = \frac{300}{1000} = 0.3$, $p_3 = \frac{250}{1000} = 0.25$, $p_4 = \frac{40}{1000} = 0.04$, y $p_5 = \frac{10}{1000} = 0.01$, respectivamente. La probabilidad de que el individuo no pierda dinero con la compra será $p(X \geq 0) = p_3 + p_4 + p_5 = 0.25 + 0.04 + 0.01 = 0.3$

2. Para determinadas bacterias se ha estudiado la variable aleatoria X=*tiempo de vida de una bacteria* (horas) resultando ser la función de densidad de dicha variable la que se muestra en la figura.

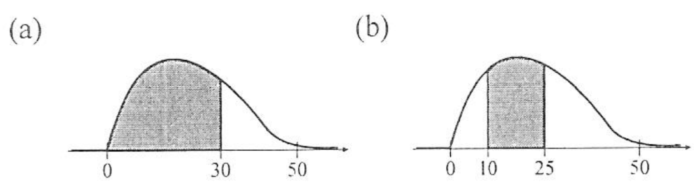
   

   Sombrear la región bajo la curva de densidad que corresponde a cada una de las siguientes probabilidades:

   (a) Probabilidad de que una bacteria viva menos de 30 horas.

   (b) Probabilidad de que una bacteria viva entre 10 y 25 horas

   Solución.

3. Sea X una variable aleatoria con función de densidad:

$$f(x) = \begin{cases} k\left(1 + x^2\right) & \text{si} \quad x \in (0,3) \\ 0 & \text{si} \quad x \notin (0,3) \end{cases}$$

   Se pide:

   (a) Hallar el valor de la constante k para que $f(x)$ sea una función de densidad.

   (b) Obtener la función de distribución de la variable aleatoria X.

   (c) Calcular la probabilidad de que X esté comprendida entre 1 y 2.

   (d) Obtener la mediana de la distribución.

   (e) Sabiendo que X es mayor que 1, obtener la probabilidad de que sea menor que 2.

Solución.

(a) Para que $f(x)$ sea una función de densidad deberá verificar que $\int_{-\infty}^{\infty} f(x)dx = 1$. Entonces:

$$\int_{-\infty}^{\infty} f(x)dx = \int_0^3 k\left(1+x^2\right)dx = k\left(x+\frac{x^3}{3}\right)\Big|_0^3 = 12k = 1$$

y por tanto tendrá que ser $k = 1/12$.

(b) Si $x \in (0,3)$ la función de distribución será:

$$F(x) = p\left(X \leq x\right) = \frac{1}{12}\int_0^x \left(1+t^2\right)dt = \frac{1}{12}\left(t+\frac{t^3}{3}\right)\Big|_0^x = \frac{3x+x^3}{36}$$

Para el resto de valores tendremos, $F(x) = 0$ si $x \leq 0$ y $F(x) = 1$ si $x \geq 3$

(c) $p\left(1 \leq X \leq 2\right) = F\left(2\right) - F\left(1\right) = \frac{14}{36} - \frac{4}{36} = \frac{10}{36} = 0.278$

(d) Si denotamos por Me a la mediana tendrá que ser $F\left(Me\right) = 0.5$. Por tanto tendremos que resolver la ecuación $\frac{3Me+(Me)^3}{36} = 0.5$ o, equivalentemente, $\left(Me\right)^3 + 3Me - 18 = 0$. Resolviendo esta ecuación por métodos numéricos (método de Newton-Fourier) se obtiene $Me \simeq 2.2422$

(e) Habrá que calcular $p\left(X < 2/X > 1\right)$. Utilizando la definición de probabilidad condicionada tendremos:

$$p\left(X < 2/X > 1\right) = \frac{p\left(1 < X < 2\right)}{p\left(X > 1\right)} = \frac{p\left(1 \leq X \leq 2\right)}{1 - p\left(X \leq 1\right)} = \frac{10/36}{1 - F(1)} = \frac{10/36}{32/36} = 0.3125$$

4. La eficacia de las calefacciones que funcionan mediante energía solar depende de la cantidad de radiación de sol. Para un mes de octubre típico, la radiación total diaria en Miami, Florida, es una variable aleatoria cuya función de densidad viene dada por (las unidades son cientos de kilocalorías)

$$f(x) = \begin{cases} \frac{3}{32}\left(x-2\right)\left(6-x\right) & \text{si } 2 \leq x \leq 6 \\ 0 & \text{en el resto} \end{cases}$$

(a) Calcular la probabilidad de que la radiación solar sea mayor que 300 kilocalorías en un día normal de octubre.

(b) Según este modelo, ¿qué cantidad de radiación solar queda rebasada exactamente el 50 % de los días de octubre?

Solución.

(a) Tenemos que calcular $p\left(X > 3\right)$ utilizando la función de densidad $f(x)$.

$$\begin{aligned} p\left(X > 3\right) &= \int_3^{\infty} f(x)dx = \int_3^6 \frac{3}{32}\left(x-2\right)\left(6-x\right)dx = \frac{3}{32}\int_3^6 \left(-x^2+8x-12\right)dx \\ &= \frac{3}{32}\left[-\frac{x^3}{3} + 4x^2 - 12x\right]_3^6 = \frac{3}{32}\left[0 - (-9)\right] = \frac{27}{32} = 0.8438 \end{aligned}$$

(b) Tenemos que calcular la mediana M de la variable aleatoria X, es decir, el valor que verifica:

$$\frac{3}{32}\left[-\frac{x^3}{3} + 4x^2 - 12x\right]_2^M = \frac{1}{32}\left[-M^3 + 12M^2 - 36M - (-32)\right] = \frac{1}{2}$$

Por tanto hay que resolver la ecuación

$$M^3 - 12M^2 + 36M - 16 = 0$$

con $M \in (2,6)$. Teniendo en cuenta que $M^3 - 12M^2 + 36M - 16 = (M-4)\left(M^2 - 8M + 4\right)$ y que la ecuación $M^2 - 8M + 4 = 0$ no tiene raíces en el intervalo $(2,6)$, podemos concluir que $M = 4$. Es decir, el 50 % de los días de octubre la radiación solar supera las 400 kilocalorías.

5. El pH, con el que se mide la acidez del agua, es importante en los estudios sobre lluvia ácida. Para determinado lago se llevan a cabo mediciones testigo de la acidez para que se pueda notar cualquier cambio en la acidez del agua originado por la lluvia ácida. El pH del agua del lago es un variable aleatoria X que, según se ha podido comprobar a partir de las muestras tomadas, tiene una función de densidad:

$$f(x) = \begin{cases} \frac{3}{8}(7-x)^2 & \text{si } 5 \leq x \leq 7 \\ 0 & \text{en el resto} \end{cases}$$

(a) Demostrar que $f(x)$ es una función de densidad.

(b) Dibujar la curva $f(x)$.

(c) Calcular la función de distribución de la variable aleatoria X.

(d) Calcular la probabilidad de que el pH sea menor que 6 en una muestra del agua de este lago.

(e) Calcular $E(X)$ y $Var(X)$.

(f) Calcular la mediana. los cuartiles y el rango intercuartílico de la variable X.

(g) Calcular un intervalo menor que $[5,7]$ en el que estén por lo menos el 75 % de las mediciones.

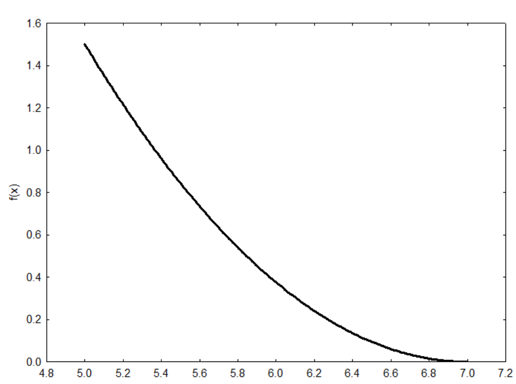
(h) ¿Es de esperar con mucha frecuencia valores del pH menores que 5.5?

Solución.

(a) Como $f(x) \geq 0$ para todo $x \in \mathbb{R}$, bastará demostrar que $\int_{-\infty}^{\infty} f(x)dx = 1$. Veamos:

$$\int_{-\infty}^{\infty} f(x)dx = \int_5^7 \frac{3}{8}(7-x)^2\, dx = \left[-\frac{(7-x)^3}{8}\right]_5^7 = [0-(-1)] = 1$$

(b)



(c) Buscamos primero la integral indefinida de $f(x)$:

$$F(x) = \int f(x)dx = -\frac{(7-x)^3}{8} + C$$

y elegimos ahora la constante C con la condición de que $F(5) = 0$, o también $F(7) = 1$. Entonces, como $F(7) = C$, tendremos que $C = 1$ y la función de distribución es

$$F(x) = \begin{cases} 0 & \text{si } x \leq 5 \\ 1 - \frac{(7-x)^3}{8} & \text{si } 5 < x < 7 \\ 1 & \text{si } x \geq 7 \end{cases}$$

(d) $p\,(X \leq 6) = F(6) = 1 - 1/8 = 7/8 = 0.875$

(e) Calculamos primero EX:

$$\begin{aligned} EX &= \int_{-\infty}^{\infty} x f(x)dx = \int_5^7 x\frac{3}{8}(7-x)^2\,dx = -\frac{x(7-x)^3}{8}\bigg|_5^7 + \int_5^7 \frac{(7-x)^3}{8}dx = \\ &= 5 + \left[\frac{-(7-x)^4}{32}\right]_5^7 = 5 + \frac{1}{2} = 5.5 \end{aligned}$$

Del mismo modo calculamos ahora $E\left(X^2\right)$:

$$\begin{aligned} E\left(X^2\right) &= \int_{-\infty}^{\infty} x^2 f(x)dx = \int_5^7 x^2\frac{3}{8}(7-x)^2\,dx = -\frac{x^2(7-x)^3}{8}\bigg|_5^7 + \int_5^7 \frac{x(7-x)^3}{4}dx \\ &= 25 + -\frac{x(7-x)^4}{16}\bigg|_5^7 + \int_5^7 \frac{(7-x)^4}{16}dx = 25 + 5 + \left[\frac{-(7-x)^5}{80}\right]_5^7 \\ &= 30 + \frac{32}{80} = 30.4 \end{aligned}$$

Por tanto tendremos que:

$$Var(X) = E\left(X^2\right) - [EX]^2 = 30.4 - (5.5)^2 = 0.15$$

(f) Para calcular la mediana M tendremos que resolver la ecuación $F(M) = 0.5$, es decir $1 - (7-M)^3/8 = 1/2$, o equivalentemente $(7-M)^3 = 4$. Por tanto tendremos que $M = 7 - \sqrt[3]{4} = 5.41$. Del mismo modo, para el primer cuartil Q_1 tendremos que resolver la ecuación $1 - (7-Q_1)^3/8 = 1/4$, o equivalentemente $(7-Q_1)^3 = 6$. Por tanto tendremos $Q_1 = 7 - \sqrt[3]{6} = 5.18$. Y para el tercer cuartil Q_3 tendremos $(7-Q_3)^3 = 2$, es decir, $Q_3 = 7 - \sqrt[3]{2} = 5.74$. Por último, el rango intercuartílico será $RI = Q_3 - Q_1 = \sqrt[3]{6} - \sqrt[3]{2} = 0.56$.

(g) Un intervalo menor que $[5,7]$ en el que estén por lo menos el 75 % de las mediciones sería $[Q_1, 7] = [5.18, 7]$, o $[5, Q_3] = [5, 5.74]$, o también, por la regla de Tchebyscheff, $EX \pm 2\sqrt{VarX)} = 5.5 \pm 2\sqrt{0.15} = [4.72, 6.27]$. Como puede comprobarse este último intervalo es mucho más grande que los dos anteriores porque la regla de Tchebyscheff proporciona sólo un intervalo muy amplio válido para cualquier distribución de probabilidad que tenga la misma esperanza y la misma varianza.

(h) $p\,(X \leq 5.5) = F\,(5.5) = 1 - \frac{(7-5.5)^3}{8} = 0.5781$ y por tanto es bastante probable que el pH sea menor que 5.5.

6. La acidez X de cierto compuesto depende de la proporción Y de uno de sus componentes químicos y viene dada por la relación $X = (1+Y)^2$. La proporción Y es una variable aleatoria con función de densidad

$$f(y) = \begin{cases} 2y & \text{si } 0 \leq y \leq 1 \\ 0 & \text{en el resto} \end{cases}$$

Calcular la función de distribución y la función de densidad de la variable aleatoria X. Obtener la esperanza de X de dos formas: (i) utilizando la función de densidad de X, y (ii) utilizando la función de densidad de Y. Calcular también $Var\,(X)$, σ_X, CV_X y la mediana Me_X.

Solución. Calculamos primero la función de distribución de Y, $F(y)$, teniendo en cuenta que será un primitiva de $f(y)$ con $F(0) = 0$ y $F(1) = 1$. Por tanto será

$$F_Y(y) = \begin{cases} 0 & \text{si } y < 0 \\ y^2 & \text{si } 0 \leq y \leq 1 \\ 1 & \text{si } y > 1 \end{cases}$$

Entonces, para la función de distribución de X tendremos:

$$F_X(x) = p\,(X \leq x) = p\left((1+Y)^2 \leq x\right) = p\left(Y \leq \sqrt{x}-1\right) = \begin{cases} 0 & \text{si } x < 1 \\ \left(\sqrt{x}-1\right)^2 & \text{si } 1 \leq x \leq 4 \\ 1 & \text{si } x > 4 \end{cases}$$

Derivando esta función obtendremos la función de densidad de X como:

$$f_X\,(x) = \begin{cases} 1 - \frac{1}{\sqrt{x}} & \text{si } 1 \leq x \leq 4 \\ 0 & \text{en el resto} \end{cases}$$

Teniendo en cuenta que $Y = \sqrt{X} - 1$, también podríamos haber calculado esta funcion de densidad utilizando el Teorema del Cambio Variable del siguiente modo:

$$f_X\,(x) = f_Y\,(y(x)) \left(\frac{dy}{dx}\right) = 2\left(\sqrt{x}-1\right)\left(\frac{1}{2\sqrt{x}}\right) = 1 - \frac{1}{\sqrt{x}}$$

Ahora podemos obtener la esperanza de X como:

$$EX = \int_1^4 x f_X(x)dx = \int_1^4 x\left(1 - \frac{1}{\sqrt{x}}\right)dx = \left.\frac{x^2}{2} - \frac{x^{3/2}}{3/2}\right|_1^4 = \left(8 - \frac{16}{3}\right) - \left(\frac{1}{2} - \frac{2}{3}\right) = \frac{17}{6}$$

También podríamos calcular esta esperanza utilizando la función de densidad $f(y)$, del siguiente modo:

$$EX = E\left((1+Y)^2\right) = \int_0^1 (1+y)^2 f(y)dy = 2\int_0^1 (1+y)^2 y dy = 2\left.\left[\frac{y^2}{2} + \frac{2y^3}{3} + \frac{y^4}{4}\right]\right|_0^1 = \frac{17}{6} = 2.83$$

Del mismo modo para X^2 tendríamos:

$$E\left(X^2\right) = \int_1^4 x^2 f_X(x)dx = \int_1^4 x^2\left(1 - \frac{1}{\sqrt{x}}\right)dx = \left.\frac{x^3}{3} - \frac{x^{5/2}}{5/2}\right|_1^4 = \left(\frac{64}{3} - \frac{64}{5}\right) - \left(\frac{1}{3} - \frac{2}{5}\right) = \frac{129}{15}$$

Por tanto $Var(X) = E\left(X^2\right) - (EX)^2 = \frac{129}{15} - \frac{289}{36} = \frac{1548-1445}{180} = \frac{103}{180} = 0.5722$, $\sigma_X = \sqrt{\frac{103}{180}} = 0.76$, $CV_X = \frac{\sigma_X}{EX}100 = 26.7\,\%$. Por último, para la mediana deberá verificarse que $F_X(Me_X) = 1/2$ y por tanto habrá que resolver la ecuación $\left(\sqrt{Me_X} - 1\right)^2 = 0.5$ y despejando tendremos que $Me_X = \left(1 + \sqrt{0.5}\right)^2 = 2.91$

7. Al examinar los pozos de agua en un zona con respecto a tres impurezas encontradas frecuentemente en el agua potable, resultó que el 20 % de los pozos no revelaba impureza alguna, el 40 % tenía la impureza A, en el 60 % se encontró la impureza B y en el 30 % la impureza C. Además, de los pozos en los que se encontró la impureza A, se vio que el 60 % también contenía la impureza B y que un 30 % contenía la impureza C. Del mismo modo, en un 40 % de los pozos que tenían impureza B también se encontró la impureza C. Encontrar la función de probabilidad y la función de distribución de la variable aleatoria "número de tipos distintos de impurezas encontradas en un pozo de agua".

   Solución. Denotamos por A, B y C a los sucesos definidos por la presencia de la respectiva impureza en el pozo. Por los datos del problema tenemos que $p\left(\overline{A}\;\overline{B}\;\overline{C}\right) = 0.2$, $p(A) = 0.4$, $p(B) = 0.6$, $p(C) = 0.3$, $p(B/A) = 0.6$, $p(C/A) = 0.3$ y $p(C/B) = 0.4$. Además, por la regla de la multiplicación tenemos que $p(AB) = p(A)\,p(B/A) = 0.24$. Del mismo modo $p(AC) = 0.12$ y $p(BC) = 0.24$. Por último tenemos que $p(A \cup B \cup C) = 1 - p\left(\overline{A \cup B \cup C}\right) = 1 - p\left(\overline{A}\;\overline{B}\;\overline{C}\right) = 0.8$. Si ahora consideramos la variable aleatoria X =número de impurezas presente en el pozo tendremos que su soporte es $Sop(X) = \{0, 1, 2, 3\}$ y las probabilidades serán:

   $p(X = 0) = p\left(\overline{ABC}\right) = 0.2$

   $p(X = 3) = p(ABC) = p(A \cup B \cup C) - p(A) - p(B) - p(C) + p(AB) + p(AC) + p(BC) = 0.1$

   $p(X = 2) = p\left(AB\overline{C}\right) + p\left(A\overline{B}C\right) + p\left(\overline{A}BC\right) = p(AB) - p(ABC) + p(AC) - p(ABC) + p(BC) - p(ABC)$

   $$= p(AB) + p(AC) + p(BC) - 3p(ABC) = 0.6 - 0.3 = 0.3$$

   $p(X = 1) = 1 - p(X = 0) - p(X = 2) - p(X = 3) = 0.4$

   Por tanto la función de densidad es:

x_i	0	1	2	3
$p(X = x_i)$	0.2	0.4	0.3	0.1

   La función de distribución será:

   $$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 0.2 & \text{si } 0 \leq x < 1 \\ 0.6 & \text{si } 1 \leq x < 2 \\ 0.9 & \text{si } 2 \leq x < 3 \\ 1 & \text{si } x \geq 3 \end{cases}$$

8. En una población, la cantidad de plomo que tiene una persona en la sangre, medida en ppm, es una

variable aleatoria X con función de densidad

$$f(x)=\begin{cases} \frac{x}{500} & \text{si } 0\leq x<20 \\ \frac{50-x}{750} & \text{si } 20\leq x<50 \\ 0 & \text{en el resto} \end{cases}$$

(a) Representar $f(x)$ gráficamente.

(b) Demostrar que $f(x)$ es una función de densidad.

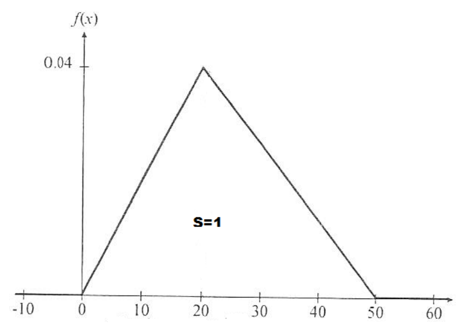
(c) Calcular la cantidad media de plomo en la sangre de los individuos de la población.

(d) Elegimos una persona al azar. Obtener la probabilidad de que la cantidad de plomo en su sangre sea inferior a 15 ppm.

(e) Obtener la probabilidad de que en 40 personas elegidas al azar, haya entre 5 y 12 personas con una cantidad de plomo en la sangre inferior a 15 ppm.

Solución.

(a)



(b) Habrá que demostrar que $\int_0^{50} f(x)dx=1$. Veamos:

$$\begin{aligned}\int_0^{50} f(x)dx &= \int_0^{20}\frac{x}{500}dx+\int_{20}^{50}\frac{50-x}{750}dx=\left.\frac{x^2}{1000}\right|_0^{20}+\left.\frac{50x-0.5x^2}{750}\right|_{20}^{50}\\ &= 0.4+\frac{2500-1250-1000+200}{750}=0.4+\frac{45}{75}=0.4+0.6=1\end{aligned}$$

(c) Hay que calcular EX:

$$\begin{aligned}EX &= \int_0^{50} xf(x)dx=\int_0^{20}\frac{x^2}{500}dx+\int_{20}^{50}\frac{50x-x^2}{750}dx=\left.\frac{x^3}{1500}\right|_0^{20}+\left.\frac{25x^2-\frac{x^3}{3}}{750}\right|_{20}^{50}\\ &= 6+\frac{62500-\frac{125000}{3}-10000+\frac{8000}{3}}{750}=6+\frac{52500-39000}{750}=24\end{aligned}$$

(d) Hay que calcular $p\,(X\leq 15)$:

$$p\,(X\leq 15)=\int_0^{15}\frac{x}{500}dx=\left.\frac{x^2}{1000}\right|_0^{15}=\frac{225}{1000}=0.225$$

(e) Si definimos la variable aleatoria Y=número de personas entre las 40 con cantidad de plomo en la sangre inferior a 15 ppm, por haber sido elegidas las personas de forma aleatoria, tendremos:

$p\left(Y=5\right)=\binom{40}{5}\left(p\left(X\leq 15\right)\right)^{5}\left(1-p\left(X\leq 15\right)\right)^{35}=658008\left(0.225\right)^{5}\left(0.775\right)^{35}$

Del mismo modo, para $k=6,7,...,12$ tendremos

$$p\left(Y=k\right)=\binom{40}{k}\left(p\left(X\leq 15\right)\right)^{5}\left(1-p\left(X\leq 15\right)\right)^{35}$$

y sumando todos los resultados tendremos

$$p\left(5\leq Y\leq 12\right)=\sum_{5}^{12}\binom{40}{k}\left(0.225\right)^{k}\left(0.775\right)^{40-k}=0.8679$$

9. Para una determinada sección de un bosque de pinos, el número de árboles enfermos por hectárea se puede suponer que tiene una distribución cuyo promedio es de 10 árboles enfermos por hectárea. Los árboles con plaga se fumigan con un insecticida a un coste de 3\$ por árbol, además de un coste fijo de administración por renta de equipo igual a 50\$. Calcular el valor esperado y la desviación estándar del costo total C de la desinfección de una hectárea de bosque. ¿Dentro de qué intervalo se puede esperar que quede C con una probabilidad de por lo menos 0.75?

   Solución. Consideramos las variables aleatorias X =número de árboles enfermos en una hectárea y C =costo total de desinfección de una hectárea. Sabemos que $EX=Var\left(X\right)=10$ y que $C=50+3X$. Por tanto tendremos que $EC=50+3EX=80$, es decir, el costo esperado de deseinfección de una hectárea es de 80\$. Del mismo modo, la varianza de C será $Var\left(C\right)=Var\left(3X\right)=9Var\left(X\right)=90$, y por tanto la desviación estándar del costo total de desinfección de una hectárea será $\sigma=\sqrt{90}=9.49$ dólares. Además, por la regla de Tchebyscheff, podemos asegurar que

   $$p\left(80-2\sqrt{90}\leq C\leq 80+2\sqrt{90}\right)=p\left(61.03\leq C\leq 98.97\right)\geq 1-1/4=0.75$$

   Por tanto podemos asegurar que la probabilidad de que el costo esté dentro del intervalo $(61.03,98.97)$ es de, por lo menos, el 75 %.

10. Sean X_1 y X_2 variables aleatorias que representen el alto y el ancho, respectivamente, de una pieza manufacturada. El alto es una variable aleatoria con media 2 cm y desviación típica de 0.1 cm, mientras que el ancho tiene media de 5 cm con varianza de 0.04 cm^2. Se supone que X_1 y X_2 son variables aleatorias independientes.

   (a) Calcular la media y la desviación estándar del perímetro de la pieza.

   (b) Calcular la media y la desviación estándar del área de la pieza.

   (c) Calcular la covarianza del perímetro y el área de la pieza. ¿Son independientes?

   (d) Calcular el coeficiente de correlación entre el perímetro y el área de la pìeza.

   Solución.

   (a) El perímetro de la pieza se calcula como $P=2X_1+2X_2$, por tanto $EP=2EX_1+2EX_2=2\left(2\right)+2\left(5\right)=14\ cm$. Del mismo modo, teniendo en cuenta que las variables aleatorias X_1 y X_2

son independientes tendremos que $Var(P) = 4VarX_1 + 4VarX_2 = 4\left(0.1^2\right) + 4\,(0.04) = 0.2$ y por tanto la desviación estándar del perímetro de la pieza será $\sigma_P = \sqrt{0.2} = 0.4472\ cm$.

(b) Para el área de la pieza tendríamos $S = X_1X_2$ y, por ser X_1 y X_2 independientes, tendremos:

$$E\,(X_1X_2) = E\,(X_1)\,E\,(X_2) = (2)\,(5) = 10\ cm^2$$

Además, para $(X_1X_2)^2$ tendríamos:

$$\begin{aligned} E\left[(X_1X_2)^2\right] &= E\left(X_1^2X_2^2\right) = E\left(X_1^2\right)E\left(X_2^2\right) = \left(Var\,(X_1) + (EX_1)^2\right)\left(Var\,(X_2) + (EX_2)^2\right) \\ &= \left(0.1^2 + 4\right)(0.04 + 25) = 100.4104 \end{aligned}$$

Por tanto tendremos:

$$Var\,(S) = E\left[(X_1X_2)^2\right] - \left[E\,(X_1X_2)\right]^2 = 100.4104 - 100 = 0.4104$$

y la desviación típica buscada será $\sigma_S = \sqrt{0.4104} = 0.64\ cm^2$.

(c) Calculamos primero $E\,(PS)$:

$$\begin{aligned} E\,(PS) &= E\left(2X_1^2X_2 + 2X_1X_2^2\right) = 2E\left(X_1^2X_2\right) + 2E\left(X_1X_2^2\right) = 2E\left(X_1^2\right)E\,(X_2) + 2E\,(X_1)\,E\left(X_2^2\right) \\ &= 2\,(4.01)\,(5) + 2\,(25.04)\,(2) = 40.1 + 100.16 = 140.26 \end{aligned}$$

Ahora calculamos la covarianza:

$$Cov\,(P,S) = E\,(PS) - E\,(P)\,E\,(S) = 140.26 - (14)\,(10) = 0.26$$

Por tanto $Cov\,(P,S) \neq 0$ y las variables no pueden ser independientes.

(d) Calculamos ahora el coeficiente de correlación:

$$\rho_{P,S} = \frac{Cov\,(P,S)}{\sigma_P\sigma_S} = \frac{0.26}{\sqrt{0.2}\sqrt{0.4104}} = 0.9075$$

Por tanto el perímetro y el área de la pieza están fuertemente correlacionadas de manera positiva.

11. Lanzamos 3 veces una moneda equilibrada y consideramos el vector aleatorio (X,Y) definido como $X =$ número de caras e $Y=$ diferencia, en valor absoluto, entre el número de caras y el número de cruces. Se pide: calcular la función de probabilidad conjunta, las funciones de probabilidad marginales, las funciones de probabilidad condicionadas, las esperanzas y varianzas de ambas variables aleatorias y también $Cov(X,Y)$. ¿Son X e Y independientes? ¿Cuál es su coeficiente de correlación?

Solución.

El espacio muestral del experimento es $\Omega = \{CCC, CCX, CXC, CXX, XCC, XCX, XXC, XXX\}$ y la probabilidad asociada vendrá dada por la regla de Laplace por ser Ω un conjunto finito con todos sus elementos igualmente probables (1/8 cada uno). El soporte de la variable aleatoria X

es $Sop(X) = \{0, 1, 2, 3\}$ y para la variable aleatoria Y es $Sop(Y) = \{1, 3\}$. Construimos ahora la función de probabilidad conjunta del vector aleatorio discreto (X, Y):

Y/X	0	1	2	3	
1	0	3/8	3/8	0	3/4
3	1/8	0	0	1/8	1/4
	1/8	3/8	3/8	1/8	

Las distribuciones de probabilidad marginales serán

$Sop(X)$	0	1	2	3
$p(X = x)$	1/8	3/8	3/8	1/8

$Sop(Y)$	1	3
$p(Y = y)$	3/4	1/4

Las funciones de probabilidad de X concicionadas por los valores de Y son:

$Sop(X)$	0	1	2	3
$p(X = x/Y = 1)$	0	1/2	1/2	0
$p(X = x/Y = 3)$	1/2	0	0	1/2

Las funciones de probabilidad de Y concicionadas por los valores de X son:

$Sop(Y)$	1	3
$p(Y = y/X = 0)$	0	1
$p(Y = y/X = 1)$	1	0
$p(Y = y/X = 2)$	1	0
$p(Y = y/X = 3)$	0	1

Para los valores esperados tendremos:

$$E(X) = (0)(1/8) + (1)(3/8) + (2)(3/8) + (3)(1/8) = 12/8 = 1.5$$

$$E(Y) = (1)(3/4) + (3)(1/4) = 6/4 = 1.5$$

De forma similar, para los cuadrados tendremos:

$$E\left(X^2\right) = \left(0^2\right)(1/8) + \left(1^2\right)(3/8) + \left(2^2\right)(3/8) + \left(3^2\right)(1/8) = 24/8 = 3$$

$$E\left(Y^2\right) = \left(1^2\right)(3/4) + \left(3^2\right)(1/4) = 12/4 = 3$$

Por tanto las varianzas serán:

$$Var(X) = E\left(X^2\right) - (EX)^2 = 3 - 1.5^2 = 0.75$$

$$Var(Y) = E\left(Y^2\right) - (EY)^2 = 3 - 1.5^2 = 0.75$$

Ahora calculamos $E(XY)$:

$$E(XY) = (0)(1)(0) + (1)(1)(3/8) + (2)(1)(3/8) + (3)(1)(0) + (0)(3)(1/8) + (1)(3)(0)$$
$$+ (2)(3)(0) + (3)(3)(1/8) = 18/8 = 2.25$$

Por tanto tendremos que $Cov(X, Y) = E(XY) - (EX)(EY) = 2.25 - (1.5)(1.5) = 0$

Aunque $Cov(X, Y) = 0$ las variables X e Y no son independientes porque la distribución de probabilidad conjunta no coincide con el producto de las funciones de probabilidad marginales. El coeficiente de correlación también será 0 por serlo la covarianza $\rho(X, Y) = 0$.

12. Supongamos que (X, Y) es un vector aleatorio continuo con función de densidad conjunta:

$$f(x, y)=\begin{cases} k & \text{si } 0<y<x<1 \\ 0 & \text{en el resto} \end{cases}$$

Se pide: calcular el valor de k, las funciones de densidad marginales y condicionadas, las esperanzas, varianzas y covarianza $Cov(X, Y)$. ¿Son X e Y independientes? ¿Cuál es el coeficiente de correlación? Calcular $p\left((X>0.5) \cap (Y<0.5)\right)$.

Solución. El valor de k se calculará obligando a que $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)dxdy=1$. Entonces

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)dx\, dy=\int_0^1\left(\int_0^x kdy\right)dx=\int_0^1 kxdx=k\left(\left.\frac{x^2}{2}\right|_0^1\right)=\frac{k}{2}=1$$

Por tanto tendrá que ser $k=2$. La función de densidad marginal de X será:

$$f_X(x)=\int_{-\infty}^{\infty} f(x, y)dy=\begin{cases} \int_0^x 2dy=2x & \text{si } 0<x<1 \\ 0 & \text{en el resto} \end{cases}$$

Del mismo modo, la función de densidad marginal de Y será

$$f_Y(y)=\int_{-\infty}^{\infty} f(x, y)dx=\begin{cases} \int_y^1 2dx=2(1-y) & \text{si } 0<y<1 \\ 0 & \text{en el resto} \end{cases}$$

La función densidad de X condicionado por $Y=y$ será

$$f_{X/Y=y}(x)=\frac{f(x, y)}{f_Y(y)}=\begin{cases} \frac{1}{1-y} & \text{si } y<x<1 \\ 0 & \text{en el resto} \end{cases}$$

Del mismo modo, La función densidad de Y condicionado por $X=x$ será

$$f_{Y/X=x}(y)=\frac{f(x, y)}{f_X(x)}=\begin{cases} \frac{1}{x} & \text{si } 0<y<x \\ 0 & \text{en el resto} \end{cases}$$

Calculamos ahora las esperanzas:

$$EX=\int_{-\infty}^{\infty} x f_X(x)=\int_0^1 2x^2 dx=\left.\frac{2x^3}{3}\right|_0^1=\frac{2}{3}$$

$$EY=\int_{-\infty}^{\infty} y f_Y(y)=\int_0^1 2y(1-y)\,dy=\left.y^2-\frac{2y^3}{3}\right|_0^1=1-\frac{2}{3}=\frac{1}{3}$$

Ahora calculamos las esperanzas de X^2 e Y^2:

$$E\left(X^2\right)=\int_{-\infty}^{\infty} x^2 f_X(x)=\int_0^1 2x^3 dx=\left.\frac{2x^4}{4}\right|_0^1=\frac{1}{2}$$

$$E\left(Y^2\right)=\int_{-\infty}^{\infty} y^2 f_Y(y)=\int_0^1 2y^2(1-y)\,dy=\left.\frac{2y^3}{3}-\frac{2y^4}{4}\right|_0^1=\frac{1}{6}$$

Por tanto ya podemos calcular las varianzas:

$$Var(X)=E\left(X^2\right)-\left(E(X)\right)^2=\frac{1}{2}-\frac{4}{9}=\frac{1}{18}$$

$$Var(Y) = E\left(Y^2\right) - \left(E\left(Y\right)\right)^2 = \frac{1}{6} - \frac{1}{9} = \frac{1}{18}$$

Para calcular la covarianza haremos primero $E\left(XY\right)$:

$$E\left(XY\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy = \int_0^1 \left(\int_0^x 2ydy\right) xdx = \int_0^1 x^3 dx = \left.\frac{x^4}{4}\right|_0^1 = \frac{1}{4}$$

Por tanto la covarianza será:

$$Cov\left(X,Y\right) = E\left(XY\right) - \left(EX\right)\left(EY\right) = \frac{1}{4} - \left(\frac{2}{3}\right)\left(\frac{1}{3}\right) = \frac{1}{36}$$

Por tanto las variables X e Y no son independientes porque $Cov\left(X,Y\right) \neq 0$. El coeficiente de correlación entre las variables será:

$$\rho\left(X,Y\right) = \frac{Cov\left(X,Y\right)}{\sqrt{Var(X)}\sqrt{Var(Y)}} = \frac{1/36}{1/18} = \frac{1}{2}$$

Tenemos por tanto una relación positiva entre las variables de magnitud lineal 0.5.

Por último calculamos la probabilidad que nos piden:

$$p\left(\left(X > 0.5\right) \cap \left(Y < 0.5\right)\right) = \int_{0.5}^1 \left(\int_0^{0.5} 2dy\right) dx = \int_{0.5}^1 dx = 1 - 0.5 = 0.5$$

1.3. PROBLEMAS RESUELTOS CAPÍTULO 5

1. Un club de automovilistas comienza una campaña telefónica para la captación de nuevos socios. De acuerdo con la experiencia previa se sabe que una de cada 10 personas que recibe la llamada se une al club. Si en un día 25 personas son contactadas por teléfono, se pide:

   (a) ¿Cuál es la probabilidad de que se unan exactamente 4 personas al club?

   (b) ¿Cuál es la probabilidad de que por lo menos 3 personas se unan al club?

   (c) ¿Cuál es el número esperado de personas que se van a unir al club?

   Solución. La probabilidad de que una persona elegida al azar se una al club es $p = 1/10 = 0.1$. Si preguntamos a 25 personas elegidas aleatoriamente y consideramos la variable aleatoria X =número de personas que se unen al club tendremos que $X \rightsquigarrow B(25, 0.1)$.

   (a) $p(X = 4) = \binom{25}{4}(0.1)^4(0.9)^{21} = 0.1384$

   (b) $p(X \geq 3) = 1 - p(X < 3) = 1 - p(X = 0) - p(X = 1) - p(X = 2)$

   $$= 1 - 0.9^{25} - 25(0.1)(0.9)^{24} - 300(0.1)^2(0.9)^{23} = 0.4629$$

   (c) $EX = 25(0.1) = 2.5$ personas

2. En un taller hay 10 máquinas iguales. Se ha visto que cada máquina se suele averiar un día de cada cinco. Evaluar las probabilidades de que en un cierto día haya r o más máquinas averiadas. Si la pérdida diaria ocasionada por tener una máquina averiada es de 3000 euros, calcular la pérdida media diaria.

   Solución. Para cada máquina la probabilidad de que se averíe en un día concreto es $p = 1/5 = 0.2$ y el número de máquinas averiadas X en un día concreto tendrá una distribución binomial $X \rightsquigarrow B(10, 0.2)$. Por tanto, para cada $r = 0, 1, ..., 10$ se tendrá $p(X \geq r) = \sum_{x=r}^{10}\binom{10}{r}(0.2)^r(0.8)^{10-r}$. Evaluando tendremos la siguiente tabla:

r	0	1	2	3	4	5	6	7	8	9	10
$p(X \geq r)$	1	0.8926	0.6242	0.3222	0.1209	0.0328	0.0064	0.0009	0.0001	0.0000	0.0000

   La pérdida diaria vendrá dada por la variable aleatoria $Y = 3000X$ y por tanto la pérdida media diaria será $EY = 3000EX = 3000(10)(0.2) = 6000$ euros.

3. En una reunión de 300 individuos. calcular la probabilidad de que hayan nacido k de ellos el día de Navidad.

   Solución. Si consideramos que todos los días del año son igualmente probables para nacer, la probabilidad de que un individuo nazca el día de Navidad será $p = 1/365$ (no tenemos en cuenta los años bisiestos). Por tanto, si tenemos 300 individuos y X es el número de ellos que han nacido el día de Navidad tendremos que $X \rightsquigarrow B(300, 1/365)$. Para un k cualquier entre 0 y 300 tendremos que $p(X = k) = \binom{300}{k}(1/365)^k(364/365)^{300-k}$. Por ejemplo, para $k \leq 4$ tendremos $p(X = 0) = 0.4391$, $p(X = 1) = 0.3619$, $p(X = 2) = 0.1486$, $p(X = 3) = 0.0406$ y $p(X = 4) = 0.0083$. Para el resto de valores la probabilidad sería ya muy pequeña porque la suma de esas cinco probabilidades es ya 0.9985.

4. La probabilidad de que un enfermo se recupere de cierta enfermedad es de 0.70. En un hospital hay ingresadas 15 personas con dicha enfermedad. ¿Cuál es la probabilidad de que:

   (a) ¿se recuperen 5 personas o menos?

   (b) ¿sobrevivan exactamente 5 personas?

   (c) ¿al menos 10 se recuperen?

   (d) ¿se recuperen entre 8 y 12 personas?

   (e) Calcular el número de personas que se espera se recuperen.

   Solución. Si consideramos la variable aleatoria X =número de personas que se recuperan tendremos que $X \rightsquigarrow B(15, 0.7)$ Por tanto:

   (a) $p(X \leq 5) = \sum_{k=0}^{5} \binom{15}{k} (0.7)^k (0.3)^{15-k} = 0.0037$

   (b) $p(X = 5) = 0.003$

   (c) $p(X \geq 10) = \sum_{k=10}^{15} \binom{15}{k} (0.7)^k (0.3)^{15-k} = 0.7216$

   (d) $p(8 \leq X \leq 12) = \sum_{k=8}^{12} \binom{15}{k} (0.7)^k (0.3)^{15-k} = 0.8232$

   (e) $EX = 15(0.7) = 10.5$ personas.

5. El número medio de resistencias que produce una fábrica es de 100 por día. De ellas, aproximadamente un 5 % no cumple con las normas de calidad mínimas que exige el mercado.

   (a) Obtener la distribución del número de resistencias diarias que no cumplen con las normas de calidad.

   (b) Calcular la probabilidad de que en un día todas las resistencias fabricadas cumplan con la normativa.

   (c) Calcular la probabilidad de que más de 5 resistencias no cumplan con las especificaciones en un día determinado.

   Solución.

   (a) Si definimos la variable aleatoria X=número de resistencias diarias que no cumplen con las normas de calidad tendremos que $X \rightsquigarrow B(100, 0.05)$. Por ser n grande ($n \geq 30$) y p pequeño ($p \leq 0.1$) podemos aproximar $B(100, 0.05) \approx P(5)$ y utilizar la distribución de Poisson.

   (b) Si usamos la distribución exacta tendremos $p(X = 0) = 0.95^{100} = 0.0059$, y si usamos la distribución de Poisson tendremos $p(X = 0) = e^{-5}\left(5^0\right)/0! = e^{-5} = 0.0067$.

   (c) Si usamos la distribución exacta tendremos

$$p(X > 5) = 1 - p(X \leq 5) = 1 - \sum_{k=0}^{5} \binom{100}{k} (0.05)^k (0.95)^{100-k} = 1 - 0.6160 = 0.384$$

y si usamos la distribución de Poisson tendremos

$$\begin{aligned}
p\left(X>5\right) &= 1-p\left(X\leq 5\right)=1-\sum_{k=0}^{5}\frac{e^{-5}\left(5\right)^{k}}{k!}=1-e^{-5}\left(1+5+\frac{25}{2}+\frac{125}{6}+\frac{625}{24}+\frac{3125}{120}\right)\\
&= 1-e^{-5}\left(\frac{10970}{120}\right)=0.384
\end{aligned}$$

6. Un fabricante de automóviles compra motores a una compañía. El fabricante recibe un lote de 40 motores. Su plan para aceptar el lote consiste en seleccionar 8 motores de entre los 40 de manera aleatoria y someterlos a prueba. Si encuentra que ninguno de los motores presenta serios defectos el fabricante acepta el lote; de otra forma, lo rechaza. Si el lote de 40 contiene cuatro motores con serios defectos, ¿cuál es la probabilidad de que sea aceptado?

   Solución. Si consideramos la variable aleatoria X=número de motores con defecto entre los 8 seleccionados tendremos que $X \rightsquigarrow H\left(40,4,8\right)$. El lote será aceptado si $X=0$, por tanto

$$p\left(X=0\right)=\frac{\binom{4}{0}\binom{36}{8}}{\binom{40}{8}}=0.3935$$

7. Demostrar que la distribución hipergeométrica tiende a la binomial en el límite cuando $N\to\infty$ y $p=\frac{D}{N}$ permanece constante. Es decir, demostrar que, si $p=\frac{D}{N}$, entonces

$$\lim_{N\to\infty}\frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}}=\binom{n}{x}p^{x}\left(1-p\right)^{n-x}$$

   Solución.

$$\begin{aligned}
\frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}} &= \frac{\left(\frac{D(D-1)\cdots(D-x+1)}{x!}\right)\left(\frac{(N-D)(N-D+1)\cdots(N-D-n+x+1)}{(n-x)!}\right)}{\frac{N(N-1)\cdots(N-x+1)(N-x)\cdots(N-n+1)}{n!}}\\
&= \left(\frac{n!}{x!(n-x)!}\right)\left(\frac{D(D-1)\cdots(D-x+1)}{N(N-1)\cdots(N-x+1)}\right)\left(\frac{(N-D)(N-D+1)\cdots(N-D-n+x+1)}{(N-x)\cdots(N-n+1)}\right)\\
&= \binom{n}{x}\left(\frac{D}{N}\right)\left(\frac{\frac{D}{N}-\frac{1}{N}}{1-\frac{1}{N}}\right)\cdots\left(\frac{\frac{D}{N}-\frac{x-1}{N}}{1-\frac{x-1}{N}}\right)\left(\frac{1-\frac{D}{N}}{1-\frac{x}{N}}\right)\cdots\left(\frac{1-\frac{D}{N}-\frac{x-1}{N}}{1-\frac{x-1}{N}}\right)\left(\frac{1-\frac{D}{N}-\frac{n-x-1}{N}}{1-\frac{n-1}{N}}\right)\\
&= \binom{n}{x}p\left(\frac{p-\frac{1}{N}}{1-\frac{1}{N}}\right)\cdots\left(\frac{p-\frac{x-1}{N}}{1-\frac{x-1}{N}}\right)\left(\frac{1-p}{1-\frac{x}{N}}\right)\cdots\left(\frac{1-p-\frac{x-1}{N}}{1-\frac{x-1}{N}}\right)\left(\frac{1-p-\frac{n-x-1}{N}}{1-\frac{n-1}{N}}\right)
\end{aligned}$$

   Ahora, teniendo en cuenta que todos los cocientes con N en el denominador convergen hacia 0 cuando N tiende a infinito tendremos que

$$\lim_{N\to\infty}\frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}}=\binom{n}{x}p^{x}\left(1-p\right)^{n-x}$$

8. Un lote de 200 piezas tiene un 40 % de defectuosas. Calcular la probabilidad de obtener 10 piezas defectuosas al sacar una muestra de 20 piezas del lote si

   (a) si el muestreo se hace sin reemplazamiento.

(b) si el muestreo se hace con reemplazamiento.

(c) ¿Se puede considerar $N = 200$, el tamaño del lote, lo suficientemente grande para que la función de probabilidad binomial sea buena aproximación de la hipergeométrica?

Solución. Consideramos la variable aleatoria X=número de piezas defectuosas en la muestra de tamaño 20.

(a) Si el muestreo se hace sin reemplazamiento tendremos que $X \rightsquigarrow H(200, 80, 20)$ y por tanto

$$p(X = 10) = \frac{\binom{80}{10}\binom{120}{10}}{\binom{200}{20}} = 0.1184$$

(b) Si el muestreo se hace con reemplazamiento tendremos que $X \rightsquigarrow B(20, 0.4)$ y por tanto

$$p(X = 10) = \binom{20}{10}(0.4)^{10}(0.6)^{10} = 0.1171$$

(c) Teniendo en cuenta que $N > 50$ y $n/N \leq 0.1$ podemos considerar que la aproximación de la distribución hipergeométrica por la distribución binomial es satisfactoria.

9. El número de vehículos que pasa por determinado punto de una carretera es 10 por minuto.

(a) Calcular la probabilidad de que pasen por lo menos 15 vehículos por este punto en cualquier minuto.

(b) Calcular la probabilidad de que pasen por lo menos 15 vehículos por este punto en un intervalo de dos minutos.

Solución. Si consideramos la variable aleatoria X=número de vehículos que pasa por un punto de la carretera en un minuto tendremos que $EX = 10$ y además $X \rightsquigarrow P(\lambda)$. Como $EX = \lambda$ tendremos que $X \rightsquigarrow P(10)$.

(a) $p(X \geq 15) = 1 - p(X < 15) = 1 - \sum_{k=0}^{14} \frac{e^{-10}(10)^k}{k!} = 1 - 0.9165 = 0.0835$

También podríamos utilizar las tablas de la distribución de Poisson para ver que

$$p(X \geq 15) = 1 - p(X \leq 14) = 1 - 0.9165 = 0.0835$$

(b) Definimos las variables aleatorias X_1 =número de vehículos que pasan en el primer minuto y X_2 =número de vehículos que pasan en el segundo minuto. Sabemos que $X_1 \rightsquigarrow P(10)$, $X_2 \rightsquigarrow P(10)$ y además son independientes. El número de vehículos que pasan en un intervalo de dos minutos será $X = X_1 + X_2$ y, por las propiedades de la distribución de Poisson, tendremos que $X \rightsquigarrow P(20)$. Por tanto:

$$p(X \geq 15) = 1 - p(X < 15) = 1 - \sum_{k=0}^{14} \frac{e^{-20}(20)^k}{k!} = 1 - 0.1049 = 0.8951$$

Utilizando las tablas de la distribución de Poisson tendríamos

$$p(X \geq 15) = 1 - p(X \leq 14) = 1 - 0.1049 = 0.8951$$

10. Se está desarrollando una nueva variedad de maíz en una estación agrícola experimental. Se espera que germinen el 90 % de las semillas. Para verificar esta tasa se plantan 100 semillas en un suelo de idéntica composición y se les dedican los mismos cuidados. Si la tasa de germinación del 90 % fuera correcta, ¿cuántas semillas habría que esperar que germinaran? Si germinan 84 semillas o menos, ¿habría razones para sospechar de que la tasa del 90 % fuera correcta?

    Solución. Si la tasa de germinación es cierta la probabilidad de que una semilla cualquiera germine será $p = 0.9$. Si ahora sembramos 100 semillas y definimos la variable aleatoria X=número de semillas que germinan entre las 100 sembradas tendremos que $X \rightsquigarrow B(100, 0.9)$. Por tanto $EX = (100)(0.9) = 90$ semillas esperamos que germinen. Ahora calculamos $p(X \leq 84)$:

    $$p(X \leq 84) = 1 - p(X > 84) = 1 - \sum_{k=84}^{100} \binom{100}{k} (0.9)^k (0.1)^{100-k} = 1 - 0.9601 = 0.0399$$

    Por tanto la probabilida es pequeña (< 0.05) y tendríamos razones para sospechar que la tasa del 90 % sea correcta.

11. A fin de aceptar o rechazar lotes de piezas se propone el siguiente plan de inspección por muestreo: los lotes tienen 1000 piezas y se toman 50 piezas al azar. Se aceptará el lote si la muestra da una o menos defectuosas y se rechazará en caso contrario. Calcular la probabilidad de rechazar el lote si p es la proporción de defectuosos en el lote, según que el muestreo se haga con o sin reemplazamiento. Particularizarlo para los casos en que la proporción de defectuosos del lote sea del 4 por mil y del 1 por ciento.

    Solución. Si p es la proporción de piezas defectuosas que contiene el lote y definimos la variable aleatoria X =número de piezas defectuosas en la muestra de tamaño 50 tendremos que $X \rightsquigarrow B(50, p)$ si el muestreo se hace con reemplazamiento, y $X \rightsquigarrow H(1000, 1000p, 50)$ si el muestreo se hace sin reemplazamiento. El lote se rechazará si $X > 1$, por tanto, si el muestreo se hace con reemplazamiento la probabilidad de rechazar el lote será:

    $$p(X > 1) = 1 - p(X = 0) - p(X = 1) = 1 - (1-p)^{50} - 50p(1-p)^{49}$$

    mientras que, si el muestreo se hace sin reemplazamiento será:

    $$p(X > 1) = 1 - p(X = 0) - p(X = 1) = 1 - \frac{\binom{1000(1-p)}{50}}{\binom{1000}{50}} - \frac{1000p\binom{1000(1-p)}{49}}{\binom{1000}{50}}$$

    Evaluando para $p = 0.004$ tendremos

    $$p(X > 1) = 1 - 0.996^{50} - 0.2\left(0.996^{49}\right) = 0.0173$$

    en el primer caso y

    $$p(X > 1) = 1 - \frac{\binom{996}{50}}{\binom{1000}{50}} - \frac{4\binom{996}{49}}{\binom{1000}{50}} = 1 - 0.8142 - 0.1720 = 0.0138$$

en el segundo caso.

Del mismo modo, si $p = 0.01$ tendremos

$$p\left(X > 1\right) = 1 - 0.99^{50} - 0.5\left(0.99^{49}\right) = 0.0894$$

en el primer caso y

$$p\left(X > 1\right) = 1 - \frac{\binom{990}{50}}{\binom{1000}{50}} - \frac{10\binom{990}{49}}{\binom{1000}{50}} = 1 - 0.5973 - 0.3174 = 0.0853$$

en el segundo caso.

Por tanto los dos casos proporcionan prácticamente las mismas probabilidades, justificando que la distribución hipergeométrica podría aproximarse por una distribución binomial.

12. En una línea de montaje operada con robots industriales se pueden instalar cajas de engranajes en un minuto si los agujeros para los tornillos han sido barrenados de forma correcta. y en diez minutos si hay que volver a barrenar. Se dispone de 20 cajas de engranajes, de las cuales dos tienen los agujeros barrenados de forma incorrecta . Se deben seleccionar cinco cajas de las 20 disponibles para que las instalen los cinco robots de la línea.

    (a) Calcular la probabilidad de que las cinco cajas de engranajes se ajusten adecuadamente.

    (b) Calcular el valor esperado y la desviación estándar del tiempo que se necesita para instalar las cinco cajas de engranajes.

    Solución. Si definimos la variable aleatoria X=número de cajas de engranajes que se ajustan adecuadamente entre las cinco elegidas tendremos que $X \rightsquigarrow H\left(20, 18, 5\right)$ y el tiempo que se necesita para instalar las cinco cajas será $T = X + 10\left(5 - X\right) = 50 - 9X$. Por tanto:

    (a) $p\left(X = 5\right) = \dfrac{\binom{18}{5}\binom{2}{0}}{\binom{20}{5}} = \frac{18\cdot17\cdot16\cdot15\cdot14}{20\cdot19\cdot18\cdot17\cdot16} = \frac{210}{380} = 0.5526$

    (b) Por ser $X \rightsquigarrow H\left(20, 18, 5\right)$ tendremos que $EX = 5\left(\frac{18}{20}\right) = 4.5$ y por tanto

$$E\left(T\right) = E\left(50 - 9X\right) = 50 - 9\left(EX\right) = 50 - 40.5 = 9.5 \text{ minutos}$$

Además, $Var\left(X\right) = 5\left(\frac{18}{20}\right)\left(1 - \frac{18}{20}\right)\left(\frac{20-5}{20--1}\right)$,

$$\begin{aligned} Var\left(T\right) &= Var\left(50 - 9X\right) = 9^2 Var\left(X\right) = 81\left(5\right)\left(\frac{18}{20}\right)\left(1 - \frac{18}{20}\right)\left(\frac{15}{19}\right) \\ &= 405\left(0.9\right)\left(0.1\right)\left(\frac{15}{19}\right) = 28.7763 \end{aligned}$$

y por tanto $\sigma_T = \sqrt{28.7763} = 5.4$ minutos.

13. El número esperado de llamadas que entran en una centralita telefónica es de cuatro por minuto. Calcular la probabilidad de que:

(a) no lleguen llamadas en un período de un minuto.

(b) por lo menos lleguen cuatro llamadas en un período de un minuto.

(c) por lo menos lleguen cinco llamadas en un período de dos minutos.

Solución. Si definimos la variable aleatoria X=número de llamadas que entran en la centralita durante un minuto tenemos que $EX = 4$ y $X \rightsquigarrow P(4)$.

(a) $p(X=0) = \frac{e^{-4}4^0}{0!} = e^{-4} = 0.0183$

(b) $p(X \geq 4) = 1 - p(X < 4) = 1 - \sum_{k=0}^{3} \frac{e^{-4}(4)^k}{k!} = 1 - e^{-4}\left(1 + 4 + 8 + \frac{64}{6}\right) = 0.5665$

(c) Si definimos las variables aleatorias X_1=número de llamadas que entran en el primer minuto y X_2=número de llamadas que entran en el segundo minuto, sabemos que $X_1 \rightsquigarrow P(4)$, $X_2 \rightsquigarrow P(4)$ y además son independientes. El número de llamadas que entran en un intervalo de dos minutos será $X = X_1 + X_2$ y, por las propiedades de la distribución de Poisson, tendremos que $X \rightsquigarrow P(8)$. Por tanto:

$$\begin{aligned} p(X \geq 5) &= 1 - p(X \leq 4) = 1 - \sum_{k=0}^{4} \frac{e^{-8}(8)^k}{k!} = 1 - e^{-8}\left(1 + 8 + 32 + \frac{256}{3} + \frac{512}{3}\right) \\ &= 1 - e^{-8}\left(\frac{891}{3}\right) = 0.9004 \end{aligned}$$

14. Se ha observado un telar durante un período de tiempo, anotándose el número de roturas por 10000 pasadas de lanzadera. obteniéndose la siguiente tabla:

Número de roturas	0	1	2	3	4	5	6
Frecuencia	39	48	39	16	6	1	1

Ajustar a una ley de Poisson. Según la ley ajustada, ¿qué frecuencia cabría esperar para cada valor del número de roturas si realizáramos 150 observaciones?

Solución. Si p es la probabilidad de rotura en cada pasada y definimos la variable aleatoria X=número de roturas en 10000 pasadas tendremos que $X \rightsquigarrow B(10000, p)$. Por ser n muy grande y p muy pequeño podríamos aproximar por una distribución de Poisson, es decir, $X \rightsquigarrow P(10000p)$ y $EX = 10000p$. Con los datos observados el número medio de roturas por 10000 pasadas es $\frac{0\cdot39+1\cdot48+2\cdot39+3\cdot16+4\cdot6+5\cdot1+6\cdot1}{150} = \frac{209}{150}$ y por tanto estimaríamos que $EX = \frac{209}{150}$ y $X \rightsquigarrow P\left(\frac{209}{150}\right)$. Utilizando esta distribución tendríamos que:

$p(X=0) = e^{-209/150} = 0.2482$

$p(X=1) = e^{-209/150}(209/150) = 0.3459$

$p(X=2) = e^{-209/150}\left(\frac{(209/150)^2}{2}\right) = 0.2410$

$p(X=3) = e^{-209/150}\left(\frac{(209/150)^3}{6}\right) = 0.1112$

$p(X=4) = e^{-209/150}\left(\frac{(209/150)^4}{24}\right) = 0.0390$

$p(X=5) = e^{-209/150}\left(\frac{(209/150)^5}{120}\right) = 0.0109$

La frecuencia esperada del número de roturas para 150 observaciones se obtendría multiplicando estas probabilidades por 150 y redondeando a números enteros. Por tanto tendríamos:

Número de roturas	0	1	2	3	4	5	6 o más
Frecuencia teórica	37	52	36	17	6	2	0

15. Supongamos que la demanda de televisores de una cierta marca en un mes cualquiera sigue una distribución de Poisson de parámetro 25. ¿Qué existencias ha de tener el comerciante al principio del mes para tener una probabilidad de 0.90 de satisfacer toda la demanda del mes? ¿Y para satisfacerla con una probabilidad de 0.99?

    Solución. Si consideramos la variable aleatoria X=demanda mensual de televisores tendremos que $X \rightsquigarrow P(25)$. Si queremos tener una probabilidad de 0.90 de satisfacer toda la demanda del mes tenemos que buscar x de modo que $p(X \leq x) = 0.90$. Por tanto tenemos que encontrar el percentil 90 % de esta distribución. Utilizando Statgraphics tenemos $P_{90} = 32$ (la tabla de la distribución de Poisson sólo llega a $\lambda = 20$). Del mismo modo, $P_{99} = 37$.

16. El 60 % de los consumidores prefiere la pasta dental marca A. Si se entrevista a un grupo de consumidores, ¿cuál es la probabilidad de que se tenga que entrevistar exactamente a cinco personas para encontrar a una persona que prefiera la marca A? ¿y de entrevistar por lo menos a 5 personas?

    Solución. Con los datos del problema, si consideramos la variable aleatoria X=número de personas entrevistadas antes de encontrar la primera que prefiere la marca A tendremos que $X \rightsquigarrow G(0.6)$. Tenemos que calcular $p(X = 4)$ para tener que entrevistar exactamente a cinco personas. Utilizando la distribución geométrica tendremos que $p(X = 4) = (0.4)^4\, 0.6 = 0.0154$. Para entrevistar por lo menos a 5 necesitaremos calcular $p(X \geq 4)$ del siguiente modo:

    $$p(X \geq 4) = 1 - p(X \leq 3) = 1 - \sum_{k=0}^{3} (0.4)^k (0.6) = 1 - 0.6(1 + 0.4 + 0.16 + 0.064) = 0.0256$$

17. Cierta centralita telefónica es de pequeña capacidad en cuanto al número de llamadas que puede recibir. De hecho, se sabe que en el período de mayor congestión la probabilidad de que la centralita esté libre es de 0.60. Calcular la probabilidad de que:

    (a) Se logre línea a la quinta llamada.

    (b) Hagan falta por lo menos 5 llamadas para conseguir línea.

    (c) ¿Cuál es el número medio de llamadas necesario para conseguir línea?

    Solución. Para cada llamada la probabilidad de conseguir línea es $p = 0.6$. Si consideramos la variable aleatoria X=número de llamadas fallidas antes de conseguir línea tendremos que $X \rightsquigarrow G(0.6)$. Entonces:

    (a) $p(X = 4) = (0.4)^4 (0.6) = 0.0154$

    (b) $p(X \geq 4) = 1 - p(X \leq 3) = 1 - \sum\limits_{k=0}^{3} (0.4)^k (0.6) = 1 - 0.6(1 + 0.4 + 0.16 + 0.064) = 0.0256$

    (c) El número media de llamadas necesario será $EX + 1 = \frac{0.4}{0.6} + 1 = \frac{1}{0.6} = 1.67$ llamadas.

18. Un zoólogo desea capturar un ejemplar de cierta especie de arácnidos en un paraje en el que se sabe que dicha especie supone el 15 % de los arácnidos presentes. ¿Cuál es la probabilidad de que tenga que capturar 10 o más ejemplares para obtener el de la especie deseada?

Solución. Para cada ejemplar capturado la probabilidad de que sea de la especie deseada es $p = 0.15$. Si definimos la variable aleatoria X=número de capturas fallidas antes de encontrar el ejemplar deseado tendremos que $X \rightsquigarrow G(0.15)$. Por tanto:

$$p(X \geq 9) = 1 - p(X \leq 8) = 1 - \sum_{k=0}^{8} (0.85)^k (0.15) = 1 - 0.15 \left(\frac{0.85^9 - 1}{0.85 - 1} \right) = 0.85^9 = 0.2316$$

19. Un científico necesita encontrar 5 monos afectados de cierta enfermedad para realizar un estudio. La incidencia de esta enfermedad en la población de monos de una reserva es del 30 %. El científico examinará uno a uno los monos hasta dar con los 5 afectados por la enfermedad. Calcular:

(a) La probabilidad de que tenga que examinar al menos 20 monos.

(b) El número medio de exámenes requeridos.

Solución. Para cada mono capturado la probabilidad de que esté enfermo es $p = 0.3$. Si consideramos la variable aleatoria X=número de monos no afectados por la enfermedad antes de encontrar el quinto afectado tendremos que $X \rightsquigarrow BN(5, 0.3)$.

(a) Tenemos que cacular $p(X \geq 15)$ del siguiente modo.

$$p(X \geq 15) = 1 - p(X \leq 14) = 1 - \sum_{k=0}^{14} \binom{k+4}{4} (0.3)^5 (0.7)^k = 1 - (0.3)^5 \sum_{k=0}^{14} \binom{k+4}{4} (0.7)^k = 0.2822$$

(b) El número medio de exámenes requeridos será $EX + 5 = 5\left(\frac{0.7}{0.3}\right) + 5 = \frac{5}{0.3} = 16.67$

20. Se estima que sólo 1 de cada 50 loros capturados en la cuenca del Amazonas para su utilización como animales domésticos sobrevive al cambio.

(a) Unos cazadores furtivos han capturado 700 pájaros. ¿Cuál es el número esperado de sobrevivientes? ¿Cuál es la probabilidad de que sobrevivan a lo sumo 10 pájaros?

(b) El dueño de una tienda ilegal de mascotas paga 4$ a los cazadores furtivos por cada loro capturado. El precio de venta de un loro en su tienda es de 300$. ¿Cuántos loros deben capturar los cazadores si el dueño de la tienda espera tener al menos una ganancia de 1000$?

Solución.

(a) Para cada loro capturado la probabilidad de que sobreviva al cambio es $p = 1/50 = 0.02$. Si consideramos la variable aleatoria X=número de loros que sobrevive al cambio entre los 700 capturados tendremos que $X \rightsquigarrow B(700, 0.02)$. Por tanto $EX = (700)(0.02) = 14$ loros. La probabilidad de que sobrevivan a los sumo 10 será:

$$p(X \leq 10) = \sum_{k=0}^{10} \binom{700}{k} (0.02)^k (0.98)^{700-k} = 0.1730$$

También, por ser n muy grande y p muy pequeño podríamos aproximar la distribución binomial por una distribución de Poisson: $B(700, 0.02) \approx P(14)$. En este caso, utilizando las tablas de la distribución de Poisson tendríamos $p(X \leq 10) = 0.1757$.

(b) Si n es el número de loros capturados, la ganancia para el dueño de la tienda será $G = 300X - 4n$ siendo $X \rightsquigarrow B(n, 0.02)$. Por tanto $E(G) = 300(0.02)n - 4n = 2n$. Si queremos que la ganancia esperada sea al menos 1000$ necesitaremos que $2n \geq 1000$, es decir, $n \geq 500$. Por tanto como mínimo necesitaríamos 500 loros.

21. El número de colonias bacteriológicas por mm^2 en un determinado cultivo da la siguiente distribución de frecuencias:

Número de colonias	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Frecuencia	10	50	100	168	190	180	120	80	45	20	8	7	2	1	0

¿Qué distribución teórica podríamos emplear para ajustar los datos? Obtener las frecuencias teóricas utilizando dicha distribución.

Solución. Si consideramos la variable aleatoria X=número de colonias bacteriológicas por mm^2 tendremos que $X \rightsquigarrow P(\lambda)$, donde $\lambda = EX$. Con los datos del problema podemos estimar λ con la media muestral del número de colonias bacteriológicas por mm^2, es decir:

$$\lambda = \frac{0 \cdot 10 + 1 \cdot 50 + ... + 12 \cdot 2 + 13 \cdot 1}{10 + 50 + ... + 2 + 1} = \frac{4428}{981} = 4.5138$$

Por tanto, utilizando un distribución $P(4.5138)$ tendremos que, para cada $x = 0, 1, ..., 14$ tendremos una frecuencia esperada de $981p(X = x) = 981e^{-4.5138}\left(\frac{4.5138^x}{x!}\right)$. Evaluando y redondeando a número enteros tendremos:

Número de colonias	0	1	2	3	4	5	6	7	8	9	10	11	12	13 o más
Frecuencia esperada	11	49	109	165	186	168	126	81	46	23	10	4	2	1

Podemos comprobar que las frecuencias esperadas se parecen mucho a las observadas y por tanto la distribución $P(4.5138)$ funciona adecuadamente.

22. Una máquina produce en serie cierto tipo de tornillos que son recogidos en cajas que contienen 1200. La experiencia ha demostrado que la proporción de estas cajas, según el número de tornillos defectuosos que contienen, es:

% de defectuosos	0	1	2	3	4	5	6
proporción con ese %	0.78	0.17	0.034	0.009	0.005	0.002	0.000

Se considera aceptable una caja que contiene el 2 % o menos de defectuosos. El objeto de la inspección, por lo tanto, es tratar de rechazar aquellas que tengan un porcentaje de defectuosos superior al 2 %. La inspección normal consiste en el examen de 50 tornillos de cada caja.

(a) Calcular la probabilidad de encontrar 6 tornillos defectuosos entre los 50 inspeccionados según sea el porcentaje real de tornillos defectuosos que contenga la caja y según que el muestreo se haga con o sin reemplazamiento.

(b) Si decidimos rechazar toda caja que nos muestre exactamente 6 tornillos defectuosos entre los 50 inspeccionados, ¿qué porcentaje de cajas rechazadas mediante la inspección son realmente aceptables? (dejar indicados los resultados).

Solución. Consideramos la variables aleatorias Y=porcentaje de piezas defectuosas que contiene la caja, y X=número de piezas defectuosas que hay en la muestra de tamaño 50. La distribución de probabilidad de Y viene dada por la tabla del enunciado del problema. Si el muestreo se hace con reemplazamiento la distribución de probabilidad de X condicionada por el valor de Y será una binomial, concretamente, $X/(Y=k) \rightsquigarrow B(50, 0.01k)$. En cambio, si el muestreo se hace sin reeplazamiento tendremos una distribución hipergeométrica, concretamente, $X/(Y=k) \rightsquigarrow H(1200, 12k, 50)$. Por tanto:

(a) Con reemplazamiento tendremos

$$p(X=6/Y=k) = \binom{50}{6}\left(\frac{k}{100}\right)^6\left(1-\frac{k}{100}\right)^{44}$$

y sin reemplazamiento tendremos

$$p(X=6/Y=k) = \frac{\binom{12k}{6}\binom{120-12k}{44}}{\binom{1200}{50}}$$

Evaluando para $k=0,1,...,6$ la probabilidad buscada $p(X=6/Y=k)$ será:

% defectuosas en la caja	0	1	2	3	4	5	6
Con reemplazamiento	0	0.00001	0.00042	0.003	0.011	0.026	0.049
Sin reemplazamiento	0	0.0000029	0.00026	0.0024	0.0096	0.024	0.047

(b) Como las cajas aceptables son aquellas que tienen $Y \leq 2$ y las rechazadas son las que tienen $X=6$, necesitaremos calcular $p(Y \leq 2/X=6)$. Utilizando la regla de Bayes tendremos que

$$p(Y \leq 2/X=6) = \frac{\sum_{k=0}^{2} p(Y=k)\, p(X=6/Y=k)}{\sum_{k=0}^{6} p(Y=k)\, p(X=6/Y=k)}$$

Las probabilidades $p(Y=k)$ vienen dadas por la tabla del enunciado y las probalidades $p(X=6/Y=k)$ vienen dadas por la tabla del apartado anterior, según que el muestreo se haga con reemplazamiento o sin reemplazamiento. Por tanto no tenemos más que sustituir los valores correspondientes para obtener que $p(Y \leq 2/X=6) = 0.107$ si el muestreo se hace con reemplazamiento, y $p(Y \leq 2/X=6) = 0.074$ si el muestreo se hace sin reemplazamiento.

23. Un proceso de fabricación de cierto tipo de piezas es tal que el 2 % de las piezas que se producen son defectuosas. Las piezas se embalan en lotes de 100. A un cierto comprador no le importa aceptar el lote si lleva 4 o menos piezas defectuosas, pero no quiere aceptar lotes con más de 4 piezas defectuosas. El vendedor asegura que en ningún lote hay más de 4 piezas defectuosas, aunque realmente no se ha preocupado de que esto sea cierto y ha empaquetado todas las piezas producidas. El comprador, fiado de lo que le dicen, no lleva a cabo ninguna inspección. ¿Qué proporción de lotes con más de 4 piezas defectuosas aceptará el comprador?

Solución. Para cada pieza fabricada la probabilidad de ser defectuosa es $p=0.02$. Si consideramos la variable aleatoria X=número de piezas defectuosas en un lote de tamaño 100 tendremos que

$X \rightsquigarrow B(100, 0.02)$. Entonces la proporción de lotes con más de 4 piezas defectuosas que aceptará el comprador será:

$$p(X > 4) = 1 - \sum_{k=0}^{3} \binom{100}{k} (0.02)^k (0.98)^{100-k} = 1 - 0.9492 = 0.0508 \simeq 5\,\%$$

24. Una muestra de sangre se examina al microscopio sobre una cuadrícula dividida en 400 cuadrados iguales. Suponiendo que la muestra ha sido diluida y homogeneizada de manera que los hematíes queden distribuidos al azar, y habiéndose observado 25 cuadrados vacíos, hallar el número medio de hematíes por cuadrado. Si la muestra ha sido diluida al 0.25 % y el volumen de disolución distribuido sobre la cuadricula es 0.1 mm^3, ¿cuántos hematíes por mm^3 de sangre se puede esperar que tenga el paciente?

    Solución. Si consideramos la variable aleatoria X=número de hematíes por cuadrado tendremos que $X \rightsquigarrow P(\lambda)$. Además, habiendo observado 25 cuadrados vacíos entre los 400 observados podemos estimar $p(X = 0) = 25/400 = 1/16$. Como $p(X = 0) = e^{-\lambda}$, podemos decir que $e^{-\lambda} \simeq 1/16$ y por tanto $\lambda = -\ln(1/16) = \ln(16)$. Entonces podemos decir que $EX = \ln(16)$. Como tenemos 400 cuadrados, si la muestra ha sido diluida al 0.25 % y el volumen de disolución distribuido sobre la cuadrícula es 0.1 mm^3, el número medio de hematíes por mm^3 de sangre será $\frac{400EX}{(0.0025)(0.1)} = \frac{400\ln(16)}{0.00025} = 4436142$ hematíes por mm^3 de sangre.

25. Un instituto de certificación de semillas desea garantizar que el porcentaje de semillas virosadas en las partidas que certifica, no supera el 5 por mil. Para ello se selecciona al azar de cada partida (que consta de un gran número de semillas) n semillas que tritura formando una pasta con la que inocula a una planta testigo. Se sabe que basta con que una de las n semillas esté virosada para que la planta testigo desarrolle síntomas de contagio, en cuyo caso, la partida correspondiente se rechazará.

    (a) Determinar cuánto debe valer como mínimo n si se desea que la probabilidad de aceptar una partida que contenga el 5 por mil o más de semillas virosadas sea menor o igual que el 1 %.

    (b) Calcular cuál es, con el plan de control establecido en el apartado anterior, la probabilidad de rechazar una partida que sólo tenga un 1 por mil de semillas infectadas.

    Solución. Si la partida contiene un 5 por mil de semillas virosadas y definimos la variable X=número de semillas virosadas en las n semillas trituradas e inoculadas en la planta tendremos que $X \rightsquigarrow B(n, 0.005)$. Para que la partida sea aceptada necesitaremos que $X = 0$ y por tanto la probabilidad será $p(X = 0) = 0.995^n$.

    (a) Si la partida tiene un 5 por mil de semillas infectadas tendremos que $p(X = 0) = 0.995^n$ y necesitaremos que $p(X = 0) \leq 0.01$, es decir, $0.995^n \leq 0.01$. Despejando tendremos $n \geq \frac{\ln(0.01)}{\ln(0.995)} = 918.7$ y por tanto necesitaremos al menos 919 semillas.

    (b) Si la partida sólo tiene un 1 por mil de semillas infectadas la probabilidad de rechazarla será

    $$1 - p(X = 0) = 1 - 0.999^{919} = 0.6013$$

1.4. PROBLEMAS RESUELTOS CAPÍTULO 6

1. Supóngase que la concentración de cierto contaminante se encuentra uniformemente distribuida en el intervalo 4 a 20 ppm. Si se considera como tóxica una concentración de 15 ppm. o más. ¿cuál es la probabilidad de que al tomarse una muestra la concentración de ésta sea tóxica?

   Solución. Sea X=concentración de contaminante en ppm. Sabemos que $X \rightsquigarrow U(4,20)$ y por tanto $f(x) = k$ para $x \in (4,20)$. Necesariamente tiene que ser $k = \frac{1}{16}$ para que $\int_4^{20} f(x)dx = 1$. Por tanto:

$$p(X \geq 15) = \int_{15}^{20} \frac{1}{16} dx = \frac{20-15}{16} = \frac{5}{16} = 0.3125$$

2. De una estación parte un tren cada 20 minutos. Un pasajero llega de improviso. Hallar:

   (a) La función de distribución y de densidad de la variable tiempo de espera.

   (b) Probabilidad de que el pasajero tenga que esperar menos de 7 minutos.

   (e) La esperanza y la desviación típica de la variable aleatoria tiempo de espera.

   Solución. Si X=tiempo de espera del viajero tendremos que $X \rightsquigarrow U(0,20)$. Por tanto:

   (a) La función de densidad será

$$f(x) = \begin{cases} 1/20 & \text{si } 0 < x < 20 \\ 0 & \text{en el resto} \end{cases}$$

   y la función de distribución $F(x) = \int_{-\infty}^{x} f(x)\, dx$ será

$$F(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ x/20 & \text{si } 0 < x < 20 \\ 1 & \text{si } x \geq 20 \end{cases}$$

   (b) $p(X \leq 7) = F(7) = 7/20 = 0.35$

   (c) $EX = \int_{-\infty}^{\infty} xf(x)\, dx = \int_0^{20} x/20 dx = \left.\frac{x^2}{40}\right|_0^{20} = 10$ minutos.

   $E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\, dx = \int_0^{20} x^2/20 dx = \left.\frac{x^3}{60}\right|_0^{20} = \frac{400}{3}$.

   $Var(X) = E(X^2) - (EX)^2 = \frac{400}{3} - 100 = \frac{100}{3}$

   $\sigma_X = \sqrt{\frac{100}{3}} = \frac{10}{\sqrt{3}} = 5.77$ minutos.

3. En un estudio sobre la conducta animal, fueron liberados pájaros, de uno en uno, con determinadas condiciones que hacían muy difícil la orientación. Se vio que los pájaros elegían al azar la dirección en la que comenzaban a volar. La dirección puede definirse mediante el acimut α, es decir, mediante el ángulo formado por el Norte y la dirección elegida por el pájaro. Obtener la función de densidad y la función de distribución de la dirección seguida por el pájaro al ser liberado.

   Solución. Si los pájaros eligen al azar la dirección el acimut α será una variable aleatoria con distribución uniforme $\alpha \rightsquigarrow U(0,360)$. Por tanto la función de densidad será

$$f(x) = \begin{cases} 1/360 & \text{si } 0 < x < 360 \\ 0 & \text{en el resto} \end{cases}$$

y la función de distribución será

$$F(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ x/360 & \text{si } 0 < x < 360 \\ 1 & \text{si } x \geq 360 \end{cases}$$

4. Se ha comprobado que el tiempo de vida de unas determinadas bacterias sigue una distribución exponencial, siendo el tiempo de vida medio de una bacteria de 225 días.

   (a) Calcular la probabilidad de que una bacteria viva más de 8 meses.

   (b) Obtener el 95-percentil de la distribución.

   (c) Calcular la probabilidad de que una bacteria que ha vivido al menos 9 meses viva como poco 3 meses más.

   Solución. Si definimos la variable aleatoria X=tiempo de vida de una bacteria en meses tendremos que $X \rightsquigarrow \exp(\lambda)$ con $EX = 225/30 = 7.5$ meses (consideramos meses de 30 días). Como $EX = 1/\lambda$ tendremos que $X \rightsquigarrow \exp(1/7.5)$. La función de distribución será entonces:

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - e^{-x/7.5} & \text{si } x \geq 0 \end{cases}$$

   (a) $p(X > 8) = 1 - p(X \leq 7) = 1 - F(7) = e^{-7/7.5} = 0.3932$

   (b) Si $x = P_{95}$ tendrá que verificarse que $F(x) = 0.95$. Por tanto hay que resolver la ecuación $1 - e^{-x/7.5} = 0.95$. Despejando tendremos que $x = -7.5 \ln 0.05 = 22.5$ meses.

   (c) Hay que calcular $p(X > 12/X > 9)$. Por tanto

$$p(X > 12/X > 9) = \frac{p(X > 12)}{p(X > 9)} = \frac{1 - F(12)}{1 - F(9)} = \frac{e^{-12/7.5}}{e^{-9/7.5}} = e^{-3/7.5} = 0.6703$$

   Observar que $e^{-3/7.5} = 1 - F(3)$ y por tanto $p(X > 12/X > 9) = p(X > 3)$, como ya sabemos.

5. La concentración de monóxido de carbono durante una hora en una ciudad grande tiene una distribución exponencial cuya media es 3.6 ppm (partes por millón).

   (a) Calcular la probabilidad de que la concentración sea mayor que 9 ppm.

   (b) Una estrategia de control del tránsito de vehículos redujo el promedio a 2.5 ppm. Estimar en este caso la probabilidad de que la concentración sea mayor que 9 ppm.

   Solución. Sea X=concentración de monóxido de carbono durante una hora (ppm). Tenemos que $EX = 3.6$ y por tanto $X \rightsquigarrow \exp(1/3.6)$. La función de distribución será entonces

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - e^{-x/3.6} & \text{si } x \geq 0 \end{cases}$$

   (a) $p(X > 9) = 1 - p(X \leq 9) = 1 - F(9) = e^{-9/3.6} = 0.0821$

   (b) Si el valor esperado se reduce a 2.5 tendremos que $X \rightsquigarrow \exp(1/2.5)$. Por tanto:

$$p(X > 9) = 1 - p(X \leq 9) = 1 - F(9) = e^{-9/2.5} = 0.0273$$

6. Los totales semanales de precipitación para una determinada zona del medio oeste de los Estados Unidos sigue una distribución exponencial con promedio de 40 mm.

   (a) Hallar la probabilidad de que un total semanal de lluvia en esta zona sea mayor que 50 mm.

   (b) Calcular la probabilidad de que los totales semanales no sean mayores que 50 mm.

   Solución. Sea X=total semanal de precipitación (mm). Tenemos que $EX = 40$ y por tanto $X \rightsquigarrow \exp(1/40)$. La función de distribución será entonces

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - e^{-x/40} & \text{si } x \geq 0 \end{cases}$$

   (a) $p(X > 50) = 1 - p(X \leq 50) = 1 - F(50) = e^{-50/40} = 0.2865$

   (b) $p(X \leq 50) = F(50) = 0.7135$

7. (a) Supongamos que el número de sucesos que tiene lugar a lo largo del tiempo constituye un proceso de Poisson con un número esperado de λ ocurrencias por unidad de tiempo. Encontrar la distribución de la variable T =tiempo transcurrido hasta la ocurrencia del primer suceso. (Sugerencia: nótese que $p[T > t]$ es igual a la probabilidad de que en el intervalo $[0, t]$ no haya ocurrido ningún suceso.)

   (b) El número esperado de llamadas telefónicas que entran en una centralita es de 4 por minuto. ¿Cuál es la probabilidad de que pase un minuto sin recibir una llamada telefónica?

   Solución.

   (a) Sea X=número de sucesos que ocurren en t unidades de tiempo. Sabemos que $EX = \lambda t$ y $X \rightsquigarrow P(\lambda t)$. Entonces, si ahora consideramos la variable aleatoria T =tiempo transcurrido hasla la ocurrencia del primer suceso podemos calcular su función de distribución como

$$F(t) = p(T \leq t) = 1 - p(T > t) = 1 - p(X = 0) = 1 - e^{-\lambda t}$$

   Por tanto la variable aleatoria T tiene una distribución exponencial $T \rightsquigarrow \exp(\lambda)$ y $ET = \frac{1}{\lambda}$

   (b) En este caso tendremos $\lambda = 4$, $t = 1$ y $T \rightsquigarrow \exp(4)$. Por tanto

$$p(T > 1) = 1 - p(T \leq 1) = 1 - F(1) = e^{-4} = 0.0183$$

8. Se ha medido el diámetro normal de los 9708 árboles de una hectárea de un bosque encontrándose los siguiente resultados:

Clase diamétrica (cm)	$5-15$	$15-25$	$25-35$	$35-45$	$45-55$	$55-65$
Frecuencia observada (n_i)	4966	2466	1208	616	305	147

   Tratar de ajustar los datos a una distribución exponencial.

   Solución. Definimos la variable aleatoria D =diámetro normal de un árbol del bosque. Utilizando los puntos medios de las clases diamétricas como representantes de cada clase podemos calcular la media muestral de los diametros medidos como

$$\overline{d} = \frac{10 \cdot 4966 + 20 \cdot 2466 + 30 \cdot 1208 + 40 \cdot 616 + 50 \cdot 305 + 60 \cdot 147}{9708} = \frac{183930}{9708} = 18.95$$

Si queremos ajustar una distribución exponencial, como el soporte de una distribución exponencial es el intervalo $(0, \infty)$, deberemos utilizar la variable $X = D - 5$ para que los valores de la variable comienzen en 0 y no en 5. Entonces podemos estimar el valor esperado de X como $\overline{d} - 5$, es decir, $EX = 1/\lambda = 13.95$. Por tanto utilizaremos $X \rightsquigarrow \exp(1/13.95)$. Utilizando esta distribución tenemos:

$p(0 \leq X \leq 10) = F(10) - F(0) = 1 - e^{-10/13.95} = 0.5117$

$p(10 \leq X \leq 20) = F(20) - F(10) = e^{-10/13.95} - e^{-20/13.95} = 0.2499$

$p(20 \leq X \leq 30) = F(30) - F(20) = e^{-20/13.95} - e^{-30/13.95} = 0.1220$

$p(30 \leq X \leq 40) = F(40) - F(30) = e^{-30/13.95} - e^{-40/13.95} = 0.0596$

$p(40 \leq X \leq 50) = F(50) - F(40) = e^{-40/13.95} - e^{-50/13.95} = 0.0291$

$p(50 \leq X \leq 60) = F(60) - F(50) = e^{-50/13.95} - e^{-60/13.95} = 0.0142$

Entonces las frecuencias esperadas para cada clase diamétrica se obtendrían multiplicando esas probabilidades por 9708 y redondeando a números enteros:

Clase diamétrica (cm)	$5-15$	$15-25$	$25-35$	$35-45$	$45-55$	$55-65$
Frecuencia esperada (n_i)	4968	2426	1184	579	283	138

Podemos observar que las frecuencias esperadas se parecen mucho a las observadas y por tanto el ajuste a la distribución de probabilidad exponencial es satisfactorio.

9. Si Z es una variable aleatoria que tiene distribución normal estándar, encontrar la probabilidad de que Z asuma un valor:

   (a) mayor que 1.14 (b) mayor que -0.36 (c) menor que 2.13

   (d) menor que -1.23 (e) entre -0.46 y -0.09 (f) entre -0.58 y 1.12

   Solución. Utilizando siempre las tablas de la distribución normal estándar tenemos:

   (a) $p(Z > 1.14) = 0.1271$

   (b) $p(Z > -0.36) = 1 - p(Z > 0.36) = 1 - 0.3594 = 0.6406$

   (c) $p(Z < 2.13) = 1 - p(Z > 2.13) = 1 - 0.0166 = 0.9834$

   (d) $p(Z < -1.23) = p(Z > 1.23) = 0.1093$

   (e) $p(-0.46 < Z < -0.09) = p(0.09 < Z < 0.46) = p(Z > 0.09) - p(Z > 0.46) = 0.4641 - 0.3228 = 0.1413$

   (f) $p(-0.58 < Z < 1.12) = p(Z < 1.12) - p(Z < -0.58) = 1 - p(Z > 1.12) - p(Z > 0.58)$

   $$= 1 - 0.1314 - 0.2810 = 0.5876$$

10. Si $Z \rightsquigarrow N(0,1)$, encontrar el valor de z_i tal que:

    (a) $p(0 < Z < z_1) = 0.4306$ (b) $p(Z \geq z_2) = 0.7704$

    (c) $p(Z > z_3) = 0.2912$ (d) $p(-z_4 < Z < z_4) = 0.9700$

    Solución. Utilizando las tablas de la distribución normal estándar tenemos:

(a) $p(0<Z<z_1) = p(Z>0) - p(Z>z_1) = 0.5 - p(Z>z_1) = 0.4306$ y por tanto $p(Z>z_1) = 0.0694$. Buscando en el interior de la tabla tendremos que $z_1 = 1.48$

(b) $p(Z \geq z_2) = p(Z \leq -z_2) = 1 - p(Z > -z_2) = 0.7704$ y por tanto $p(Z>-z_2) = 0.2296$. Buscando en el interior de la tabla tendremos que $-z_2 = 0.74$ y por tanto $z_2 = -0.74$

(c) $p(Z > z_3) = 0.2912$ y buscando en el interior de la tabla tenemos que $z_3 = 0.55$

(d) $p(-z_4 < Z < z_4) = 1 - 2p(Z > z_4) = 0.9700$ y por tanto $p(Z > z_4) = 0.0150$. Buscando en el interior de la tabla tendremos que $z_4 = 2.17$

11. Si X es una variable aleatoria con distribución normal de media 10 y desviación típica de 2, calcular la probabilidad de que X tome un valor:

    (a) menor que 13 (b) mayor que 9 (c) entre 6 y 14 (d) entre 2 y 4

    Solución. Sabemos que $X \rightsquigarrow N(10,2)$. Entonces:

    (a) $p(X<13) = p\left(\frac{X-10}{2} < \frac{13-10}{2}\right) = p(Z<1.50) = 1 - p(Z>1.50) = 1 - 0.0668 = 0.9332$

    (b) $p(X>9) = p\left(\frac{X-10}{2} > \frac{9-10}{2}\right) = p(Z>-0.5) = 1 - p(Z<-0.5) = 1 - p(Z>0.5) = 1 - 0.3085 = 0.6915$

    (c) $p(6<X<14) = p\left(\frac{6-10}{2} < Z < \frac{14-10}{2}\right) = p(-2<Z<2) = 1 - 2p(Z>2) = 1 - 2(0.0228) = 0.9544$

    (d) $p(2<X<4) = p\left(\frac{2-10}{2} < Z < \frac{4-10}{2}\right) = p(-4<Z<-3) = p(3<Z<4) = p(Z>3) - p(Z>4) = 0.00135 - 0.0000317 = 0.0013183$

12. Sea X una variable aleatoria con distribución normal de media 5 y varianza 16. Calcular el valor de x_i que hace que:

    (a) $p(X>x_1) = 0.1075$ (b) $p(X>x_2) = 0.95$ (c) $p(5-x_3<X<5+x_3) = 0.7198$

    Solución. Sabemos que $X \rightsquigarrow N(5,4)$. Entonces:

    (a) $p(X>x_1) = 0.1075 \Leftrightarrow p\left(Z > \frac{x_1-5}{4}\right) = 0.1075$. Buscando en el interior de la tabla tendremos que $\frac{x_1-5}{4} = 1.24$ y por tanto $x_1 = 5 + 4(1.24) = 9.96$

    (b) $p(X>x_2) = 0.95 \Leftrightarrow p\left(Z > \frac{x_2-5}{4}\right) = 0.95 \Leftrightarrow p\left(Z > \frac{-x_2+5}{4}\right) = 0.05$. Buscando en el interior de la tabla tendremos que $\frac{-x_2+5}{4} = 1.645$ y por tanto $x_2 = 5 - 4(1.645) = -1.58$

    (c) $p(5-x_3<X<5+x_3) = 0.7198 \Leftrightarrow p\left(\frac{-x_3}{4} < Z < \frac{x_3}{4}\right) = 1-2p\left(Z > \frac{x_3}{4}\right) = 0.7198 \Leftrightarrow p\left(Z > \frac{x_3}{4}\right) = 0.1401$. Buscando en el interior de la tabla tendremos que $\frac{x_3}{4} = 1.08$ y por tanto $x_3 = 4(1.08) = 4.32$. El intervalo buscado alrededor de 5 sería $(0.68, 9.32)$

13. En cierta comunidad de arenques comunes adultos (*Clupea harengus L.*) la longitud de los peces sigue una distribución normal con media 25 cm y desviación típica 4.5 cm.

    (a) ¿Qué porcentaje de los peces tiene una longitud comprendida entre 25 y 30 cm?

    (b) ¿Qué porcentaje de los peces tiene una longitud superior a 18 cm?

    (c) ¿Qué porcentaje de los peces tiene una longitud comprendida entre 32 y 36 cm?

(d) Calcular los percentiles 20 y 70 de la población.

Solución. Sea X=longitud de un arenque común adulto. Sabemos que $X \rightsquigarrow N(25, 4.5)$.

(a) $p(25 < X < 30) = p\left(\frac{25-25}{4.5} < Z < \frac{30-25}{4.5}\right) = p(0 < Z < 1.11) = p(Z > 0) - p(Z > 1.11) = 0.5 - 0.135 = 0.3665$

(b) $p(X > 18) = p\left(Z > \frac{18-25}{4.5}\right) = p(Z > -1.56) = 1 - p(Z > 1.56) = 1 - 0.0594 = 0.9406$

(c) $p(32 < X < 36) = p(1.56 < Z < 2.44) = p(Z > 1.56) - p(Z > 2.44) = 0.0594 - 0.00734 = 0.05206$

(d) Si $x = P_{20}$ tendremos que $p(X < x) = p\left(Z < \frac{x-25}{4.5}\right) = p\left(Z > -\frac{x-25}{4.5}\right) = 0.2$. Por tanto tenemos que buscar en el interior de la tabla 0.2, que está en la zona donde $z_\alpha < 1$ y por consiguiente hay que interpolar. Buscamos los más cercanos: $z_{0.2005} = 0.84$ y $z_{0.1977} = 0.85$. Por tanto, interpolando:

$$\frac{z - 0.84}{0.2 - 0.2005} = \frac{0.85 - 0.84}{0.1977 - 0.2005}$$

Despejando tendremos

$$z = 0.84 + \frac{(0.0005)(0.01)}{0.0028} = 0.8418$$

Por tanto deberá ser $-\frac{x-25}{4.5} = 0.8418$ y despejando tenemos $x = P_{20} = 25 - 4.5(0.8418) = 21.21$

Ahora, Si $x = P_{70}$ tendremos que $p(X < x) = p\left(Z < \frac{x-25}{4.5}\right) = 1 - p\left(Z > \frac{x-25}{4.5}\right) = 0.7$. Por tanto tendrá que ser $p\left(Z > \frac{x-25}{4.5}\right) = 0.3$. Buscamos en el interior de la tabla 0.3 y como está en la zona donde $z_\alpha < 1$ tenemos interpolar. Buscamos los más cercanos: $z_{0.3015} = 0.52$ y $z_{0.2981} = 0.53$. Por tanto, interpolando:

$$\frac{z - 0.52}{0.3 - 0.3015} = \frac{0.53 - 0.52}{0.2981 - 0.3015}$$

Despejando tendremos

$$z = 0.52 + \frac{(0.0015)(0.01)}{0.0034} = 0.5244$$

Por tanto deberá ser $\frac{x-25}{4.5} = 0.5244$ y despejando tenemos $x = P_{70} = 25 + 4.5(0.5244) = 27.36$

14. Un botánico ha observado que la anchura de las hojas de un álamo sigue una distribución normal con una media de 6 cm, y, además, que el 90 % de las hojas tiene una anchura inferior a 7.5 cm. Obtener una estimación de σ y calcular la probabilidad de que una hoja mida más de 8 cm.

    Solución. Sea la variable aleatoria X=anchura de la hoja de un álamo. Sabemos que $X \rightsquigarrow N(6, \sigma)$. Además sabemos que $p(X < 7.5) = 0.9$. Por tanto $p\left(Z < \frac{7.5-6}{\sigma}\right) = 1 - p\left(Z > \frac{7.5-6}{\sigma}\right) = 0.9$ y tendrá que ser $p\left(Z > \frac{7.5-6}{\sigma}\right) = 0.1$. Buscando en el interior de la tabla (no es necesario interpolar) tendremos $\frac{7.5-6}{\sigma} = 1.28$ y despejando $\sigma = \frac{1.5}{1.28} = 1.172$. Ahora podemos calcular la probabilidad de que una hoja mida más de 8 cm. como:

$$p(X > 8) = p\left(Z > \frac{8-6}{1.172}\right) = p(Z > 1.71) = 0.0436$$

15. La anchura, en milímetros, de los coleópteros de una población sigue una distribución normal. Se estima que el 77 % de la población mide menos de 12 mm y que el 84 % mide más de 7 mm. Dar una estimación de la anchura media y de la desviación típica de la población.

Solución. Sea X=anchura de un coléoptero (mm) con $X \rightsquigarrow N(\mu, \sigma)$. Sabemos que $p(X < 12) = p\left(Z < \frac{12-\mu}{\sigma}\right) = 0.77$ y $p(X > 7) = p\left(Z > \frac{7-\mu}{\sigma}\right) = 0.84$. La primera igualdad es equivalente a $p\left(Z > \frac{12-\mu}{\sigma}\right) = 0.23$ y la segunda a $p\left(Z > -\frac{7-\mu}{\sigma}\right) = 0.16$. En los dos casos tenemos que interpolar. Para el primero tendremos: $z_{0.2327} = 0.73$ y $z_{0.2296} = 0.74$ e interpolando:

$$\frac{z - 0.73}{0.23 - 0.2327} = \frac{0.74 - 0.73}{0.2296 - 0.2327}$$

Por tanto $z = 0.73 + \frac{(0.0027)(0.01)}{0.0031} = 0.7387$. Para el segundo tendremos $z_{0.1611} = 0.99$ y $z_{0.1587} = 1.00$ e interpolando:

$$\frac{z - 0.99}{0.16 - 0.1611} = \frac{1.00 - 0.99}{0.1587 - 0.1611}$$

Por tanto $z = 0.99 + \frac{(0.0011)(0.01)}{0.0024} = 0.9946$.

Planteamos un sistema de dos ecuaciones con dos incógnitas

$$\left.\begin{array}{r} \frac{12-\mu}{\sigma} = 0.7387 \\ \frac{7-\mu}{\sigma} = -0.9946 \end{array}\right\}$$

Despejando μ e igualando tendremos $12 - 0.7387\sigma = 7 + 0.9946\sigma$. Por tanto $\sigma = \frac{5}{0.9946+0.7387} = 2.8847$. Finalmente, despejando μ en la primera ecuación tendremos $\mu = 12 - 0.7387\sigma = 12 - (0.7387)(2.8847) = 9.8691$

La anchura media buscada será 9.8691 cm. y la desviación típica será 2.8847 cm.

16. Sean X e Y variables aleatorias normales e independientes. La variable X tiene una media igual a 5 y la varianza igual a 4. La esperanza de la variable Y es de 10 y su varianza de 9. Calcular:

(a) $p(3X + 2Y > 25)$ (b) $p(30 \leq 3X + 2Y \leq 50)$

Solución. Tenemos que $X \rightsquigarrow N(5, 2)$ e $Y \rightsquigarrow N(10, 3)$ y además son independientes. Si consideramos la variable $T = 3X + 2Y$ tendremos que $E(T) = 3EX + 2EY = 35$ y $Var(T) = 9Var(X) + 4Var(Y) = 72$. Además, T también tendrá una distribución normal, es decir, $T \rightsquigarrow N\left(35, 6\sqrt{2}\right)$. Por tanto:

(a) $p(3X + 2Y > 25) = p(T > 25) = p\left(Z > \frac{25-35}{6\sqrt{2}}\right) = p(Z > -1.18) = 1 - p(Z > 1.18) = 1 - 0.1190 = 0.881$

(b) $p(30 \leq 3X + 2Y \leq 50) = p(30 \leq T \leq 50) = p\left(\frac{30-35}{6\sqrt{2}} \leq Z \leq \frac{50-35}{6\sqrt{2}}\right) = p\left(\frac{-5}{6\sqrt{2}} \leq Z \leq 1.77\right)$

$$= p\left(Z > \frac{-5}{6\sqrt{2}}\right) - p(Z > 1.77) = 1 - p\left(Z > \frac{5}{6\sqrt{2}}\right) - p(Z > 1.77) = 1 - 0.2778 - 0.0384 = 0.6838$$

Para la primera probabilidad hemos interpolado porque $\frac{5}{6\sqrt{2}} < 1$:

$$\frac{x - 0.2810}{\frac{5}{6\sqrt{2}} - 0.58} = \frac{0.2776 - 0.2810}{0.59 - 0.58}$$

y por tanto $x = 0.2810 - \frac{(0.0093)(0.0034)}{(0.01)} = 0.2778$.

17. Sean $X_1, X_2, ...X_7$, variables aleatorias independientes tales que $X_i \rightsquigarrow N(3,4)$.

(a) Obtener la distribución de la variable $Y = 7X_1$.

(b) Obtener la distribución de la variable $W = \sum_{i=1}^{7} X_i$.

Solución.

(a) $E(Y) = 7E(X_1) = 21$ y $\sigma_Y = \sqrt{49Var(X_1)} = \sqrt{(49)(16)} = \sqrt{784} = 28$. Por tanto $Y \rightsquigarrow N(21, 28)$.

(b) $E(W) = E\left(\sum_{i=1}^{7} X_i\right) = \sum_{i=1}^{7} E(X_i) = 21$ y $\sigma_W = \sqrt{Var\left(\sum_{i=1}^{7} X_i\right)} = \sqrt{\sum_{i=1}^{7} Var(X_i)} = 4\sqrt{7}$. Por tanto tendremos que $W \rightsquigarrow N(21, 4\sqrt{7})$.

18. El peso de las personas de una población sigue una distribución normal con media 72 kg y desviación típica 10 kg.

(a) Cuatro personas elegidas al azar en esa población entran en un ascensor cuya carga máxima es de 350 kg. ¿Cuál es la probabilidad de que entre los cuatro se supere esta carga máxima?

(b) ¿Cuál es la probabilidad de que dos personas elegidas al azar en esa población puedan jugar en un balancín si sólo pueden hacerlo cuando sus pesos difieran en menos de 5 kg?

Solución. Sea X_i =peso de una persona cualquiera de la población (kg). Sabemos que $X_i \rightsquigarrow N(72, 10)$.

(a) El peso que tiene que soportar el ascensor sería $Y = \sum_{i=1}^{4} X_i$, con $E(Y) = 4(72) = 288$, $\sigma_Y = \sqrt{4(100)} = \sqrt{400} = 20$ y $Y \rightsquigarrow N(288, 20)$. Por tanto la probabilidad buscada será

$$p(Y > 350) = p\left(Z > \frac{350 - 288}{20}\right) = p(Z > 3.1) = 0.000968$$

(b) Para las dos personas elegidas al azar tendremos $E(X_1 - X_2) = E(X_1) - E(X_2) = 0$, $Var(X_1 - X_2) = Var(X_1) + Var(X_2) = 200$ y por tanto $X_1 - X_2 \rightsquigarrow N(0, 10\sqrt{2})$. Entonces la probabilidad buscada será

$$\begin{aligned} p(|X_1 - X_2| < 5) &= p(-5 < X_1 - X_2 < 5) = p\left(\frac{-5}{10\sqrt{2}} < Z < \frac{5}{10\sqrt{2}}\right) = 1 - 2p\left(Z > \frac{5}{10\sqrt{2}}\right) \\ &= 1 - 2p\left(Z > \frac{1}{2\sqrt{2}}\right) \end{aligned}$$

Interpolando tendremos

$$\frac{x - 0.3632}{\frac{1}{2\sqrt{2}} - 0.35} = \frac{0.3594 - 0.3632}{0.36 - 0.35}$$

y por tanto $x = 0.3632 - \frac{(0.0036)(0.0038)}{(0.01)} = 0.3618$. La probabilidad buscada será entonces $1 - 2(0.3618) = 0.2764$

19. El peso de una gacela, en kilogramos, es una variable aleatoria que tiene una distribución normal con media 50 kg y desviación típíca 6 kg. Si capturamos 10 gacelas:

(a) Obtener la distribución de la variable aleatoria Y=“*número de gacelas, de las 10 capturadas, que pesan menos de* 40 *kg*”. Calcular $p(Y = 2)$.

(b) ¿Cuál es la probabilidad de que podamos transportar a las 10 gacelas en un vehículo que admite una carga máxima de 450 kg? ¿Y si admite una carga de 525 kg?

Solución. Sea X=peso de una gacela (kg), con $X \rightsquigarrow N(50,6)$. La probabilidad de que una gacela cualquiera pese menos de 40 kg será $p\left(X<40\right)=p\left(Z<\frac{40-50}{6}\right)=p\left(Z<-1.67\right)=p\left(Z>1.67\right)=$ 0.0475

(a) Como hemos capturado 10 gacelas tendremos que $Y \rightsquigarrow B(10,0.0475)$. Por tanto

$$p\left(Y=2\right)=\binom{10}{2}\left(0.0475\right)^{2}\left(0.9525\right)^{8}=0.0688$$

(b) El peso total de las 10 gacelas será $W=\sum_{i=1}^{10}X_i$ con $X_i \rightsquigarrow N(50,6)$ e independientes. Por tanto $W \rightsquigarrow N(500,6\sqrt{10})$. Si la carga máxima es 450 kg la probabilidad buscada será

$$p\left(W<450\right)=p\left(Z<\frac{450-500}{6\sqrt{10}}\right)=p\left(Z<-2.64\right)=p\left(Z>2.64\right)=0.00415$$

Si la carga máxima fuese 525 kg tendríamos

$$p\left(W<525\right)=p\left(Z<\frac{525-500}{6\sqrt{10}}\right)=p\left(Z<1.32\right)=1-p\left(Z>1.32\right)=1-0.0934=0.9066$$

20. Una línea eléctrica se avería cuando la tensión sobrepasa la capacidad de la línea. Si la tensión es una variable aleatoria con distribución $N(100,20)$ y la capacidad sigue una distribución $N(130,10)$, calcular la probabilidad de que se produzca una avería.

    Solución. Consideramos la variables aleatoria X=tensión de la línea e Y=capacidad de la línea. Por hipótesis sabemos que $X \rightsquigarrow N(100,20)$ e $Y \rightsquigarrow N(130,10)$, con X e Y independientes. Por tanto $X-Y \rightsquigarrow N(-30,10\sqrt{5})$ y la probabilidad de que se produzca una avería será

$$p\left(X-Y>0\right)=p\left(Z>\frac{30}{10\sqrt{5}}\right)=p\left(Z>1.34\right)=0.0901$$

21. Una máquina de empaquetado automático deposita en cada paquete 81.5 gr, por término medio, de cierto producto, con $\sigma=8$ gr. El peso medio del paquete vacío es 14.5 gr con $\sigma=6$ gr. Ambas distribuciones son normales e independientes.

    (a) Obtener la distribución del peso de los paquetes llenos.

    (b) Distribuimos los paquetes llenos, de 40 en 40, en cajas cuyo peso medio, vacías, es de 520 gr con $\sigma=50$ gr (admitimos también normalidad). Hallar la distribución del peso de las cajas llenas medido en kg.

    Solución. Sean X=peso depositado en el paquete (gr) e Y=peso del paquete vacío (gr). Sabemos que $X \rightsquigarrow N(81.5,8)$ e $Y \rightsquigarrow N(14.5,6)$, con X e Y independientes.

    (a) El peso del paquete lleno será $X+Y$ y su distribución de probabilidad será $X+Y \rightsquigarrow N(96,10)$ porque $Var\left(X+Y\right)=Var(x)+Var\left(Y\right)=64+36=100$

    (b) Si consideramos la variable aleatoria C =peso de la caja vacía tendremos $C \rightsquigarrow N(520,50)$. Entonces el peso total de la caja llena con los 40 paquetes expresado en kg será $W=\frac{C+\sum_{i=1}^{40}(X_i+Y_i)}{1000}$

y por tanto

$$E(W) = \frac{E(C) + \sum_{i=1}^{40} E(X_i + Y_i)}{1000} = \frac{520 + 40\,(96)}{1000} = 4.36$$

$$Var(W) = \frac{Var(C) + \sum_{i=1}^{40} Var(X_i + Y_i)}{1000000} = \frac{2500 + 40\,(100)}{1000000} = 0.0065$$

Por tanto la distribución de probabilidad de las cajas llenas, medido en kg, será $W \rightsquigarrow N(4.36, 0.0806)$

22. **Distribución de Weibull**: Se dice que una variable aleatoria X sigue una distribución de Weibull de parámetros a y b ($a > 0$ y $b > 0$) y lo denotaremos diciendo que $X \rightsquigarrow W(a, b)$, si su función de distribución viene dada por: $F(x) = \begin{cases} 0 & \text{si} \quad x < 0 \\ 1 - e^{-ax^b} & \text{si} \quad x \geq 0 \end{cases}$

    La distribución de Weibull aparece a menudo para modelizar la duración (se utiliza para caracterizar el tiempo de vida de sistemas con la particularidad de tener una tasa de fallo que crece, o decrece, de forma potencial a lo largo del tiempo). Debido al gran número de formas que adquieren las funciones de densidad de Weibull, según el valor de sus dos parámetros, la distribución de Weibull se utiliza en numerosas aplicaciones.(Nota, el programa Statgraphics utiliza $F(x) = 1 - e^{-\left(\frac{x}{a}\right)^b}$).

    (a) Demostrar que $F(x)$ es una función de distribución.

    (b) Obtener la función de densidad de una variable de Weibull.

    Solución.

    (a) Es fácil comprobar que $F(x)$ es una función continua y creciente con $F(0) = 0$ y $\lim\limits_{x\to\infty} F(x) = 1$. Por tanto se trata de una función de distribución.Observar que si $b = 1$ tendremos una distribución $\exp(a)$. Es decir, $W(a, 1) = \exp(a)$ y la distribución de Weibull es una generalización de la distribución exponencial.

    (b) Sabemos que la función densidad es la derivada la función distribución. Por tanto será

$$f(x) = \begin{cases} 0 & \text{si } x < 0 \\ abx^{b-1}e^{-ax^b} & \text{si } x \geq 0 \end{cases}$$

23. Los niveles máximo de inundación, en millones de pies cúbicos por segundo, de determinado río siguen una distribución de probabilidad de Weibull con parámetros $a = 1/0.6$ y $b = 1.5$. Estimar la probabilidad de que el nivel máximo de inundación para el año próximo sea:

    (a) mayor que 500000 pies cúbicos por segundo.

    (b) menor que 800000 pies cúbicos por segundo.

    Solución. Sea X=nivel máximo de inundación del río (millones de pies cúbicos). Sabemos que $X \rightsquigarrow W(1/0.6, 1.5)$.

    (a) $p(X > 0.5) = 1 - p(X \leq 0.5) = 1 - F(0.5) = e^{-\frac{0.5^{1.5}}{0.6}} = 0.5547$

    (b) $p(X < 0.8) = F(0.8) = 1 - e^{-\frac{0.8^{1.5}}{0.6}} = 0.6966$

24. La resistencia máxima a la tensión del alambre de acero que se usa para envolver a los tubos de concreto tuvo una distribución de Weibull con $a = 1/270$ y $b = 1.2$ (mediciones hechas en miles de libras). En un determinado punto de un tubo, la presión existente necesita una resistencia máxima a la tensión de cuanto menos 300000 libras. ¿Cuál es la probabilidad de que el alambre tenga esta resistencia necesaria?

    Solución. Sea X=resistencia máxima a la tensión del alambre de acero (miles de libras). Sabemos que $X \rightsquigarrow W(1/270, 1.2)$. La probabilidad de que el alambre tenga la resistencia necesaria será

$$p\left(X > 300\right) = 1 - F\left(300\right) = e^{-\frac{300^{1.2}}{270}} = 0.0309$$

25. A partir de los datos de la velocidad media (medida en km/h) diaria del viento, medidos en el observatorio meteorológico de La Coruña durante varios años, se ha ajustado una ley de Weibull con parámetros $a = 1/117$ y $b = 1.4$.

    (a) ¿Qué velocidad media del viento se espera que se supere el 50 % de los días?

    (b) Calcular el porcentaje de días que se espera superen una velocidad media del viento de 80 km/h.

    Solución. Sea X=velocidad media diaria del viento (km/h). Sabemos que $X \rightsquigarrow W(1/117, 1.4)$.

    (a) Buscamos el valor x para el cual $p\left(X > x\right) = 0.5$, es decir, la mediana. Como $p\left(X > x\right) = e^{-\frac{x^{1.4}}{117}}$ tendremos que resolver la ecuación $e^{-\frac{x^{1.4}}{117}} = 0.5$ Despejando tendremos $x = \left(-117 \ln 0.5\right)^{1/1.4} = 23.1$

    (b) $p\left(X > 80\right) = e^{-\frac{80^{1.4}}{117}} = 0.0193$

26. Se supone que las plantas de una determinada especie se distribuyen aleatoriamente en una región. con una densidad promedio de λ plantas por unidad de área. Para una planta seleccionada al azar en la zona, sea R la distancia a la planta vecina más próxima de la misma especie. Determinar la función de densidad de la variable R. (Sugerencia: nótese que $p\left(R > r\right)$ es igual a la probabilidad de que en el círculo de radio r no haya ninguna planta de la especie examinada).

    Solución. Si las plantas se distribuyen aleatoriamente y consideramos un círculo de radio r sabemos que X_r =número de plantas en el círculo tiene una distribución de Poisson cuyo parámetro será el valor esperado de X_r. Si el promedio de plantas por unidad de área es λ tendremos que el promedio de plantas en el círculo de radio r será $\lambda\pi r^2$. Por tanto $X_r \rightsquigarrow P(\lambda\pi r^2)$. Entonces, para un $r > 0$ cualquiera, la probabilidad de que R sea más grande que r será la probabilidad de que no haya ninguna planta en el círculo de radio r alrededor de dicha planta. Ahora podemos calcular la función de distribución de R como:

$$F\left(r\right) = p(R \leq r) = 1 - p(R > r) = 1 - p\left(X_r = 0\right) = 1 - e^{-\lambda\pi r^2}$$

    Por tanto la función de densidad de R se obtendrá derivando esta función de distribución:

$$f(r) = \begin{cases} 2\lambda\pi r e^{-\lambda\pi r^2} & \text{si } r > 0 \\ 0 & \text{en el resto} \end{cases}$$

    Observar que tendremos entonces una distribución de Weibull, es decir, $R \rightsquigarrow W(\lambda\pi, 2)$.

27. A partir de una muestra de los diámetros de los árboles de un bosque (medidos en cm.), se ha ajustado satisfactoriamente la variable diámetro de un árbol a una Weibull de parámetros $a = 1.5581 \cdot 10^{-10}$ y $b = 5.45$. Si observamos una muestra de 1000 árboles, ¿qué frecuencias podríamos esperar bajo la distribución ajustada? Formar una tabla de frecuencias esperadas para clases diamétricas de 10 cm de amplitud.

Solución. Sea X=diámetro de un árbol (cm). Sabemos que $X \rightsquigarrow W(1.5581 \cdot 10^{-10}, 5.45)$. Para clases diamétricas de 10 cm tendríamos:

$p\,(0 < X < 10) = F\,(10) = 1 - e^{-1.5581\left(\frac{10^{5.45}}{10^{10}}\right)} = 0.000044$

$p\,(10 < X < 20) = F\,(20) - F\,(10) = e^{-1.5581\left(\frac{10^{5.45}}{10^{10}}\right)} - e^{-1.5581\left(\frac{20^{5.45}}{10^{10}}\right)} = 0.001874$

$p\,(20 < X < 30) = F\,(30) - F\,(20) = e^{-1.5581\left(\frac{20^{5.45}}{10^{10}}\right)} - e^{-1.5581\left(\frac{30^{5.45}}{10^{10}}\right)} = 0.015425$

$p\,(30 < X < 40) = F\,(40) - F\,(30) = e^{-1.5581\left(\frac{30^{5.45}}{10^{10}}\right)} - e^{-1.5581\left(\frac{40^{5.45}}{10^{10}}\right)} = 0.063145$

$p\,(40 < X < 50) = F\,(50) - F\,(40) = e^{-1.5581\left(\frac{40^{5.45}}{10^{10}}\right)} - e^{-1.5581\left(\frac{50^{5.45}}{10^{10}}\right)} = 0.166089$

$p\,(50 < X < 60) = F\,(60) - F\,(50) = e^{-1.5581\left(\frac{50^{5.45}}{10^{10}}\right)} - e^{-1.5581\left(\frac{60^{5.45}}{10^{10}}\right)} = 0.287974$

$p\,(60 < X < 70) = F\,(70) - F\,(60) = e^{-1.5581\left(\frac{60^{5.45}}{10^{10}}\right)} - e^{-1.5581\left(\frac{70^{5.45}}{10^{10}}\right)} = 0.295399$

$p\,(70 < X < 80) = F\,(80) - F\,(70) = e^{-1.5581\left(\frac{70^{5.45}}{10^{10}}\right)} - e^{-1.5581\left(\frac{80^{5.45}}{10^{10}}\right)} = 0.144525$

$p\,(80 < X < 90) = F\,(90) - F\,(80) = e^{-1.5581\left(\frac{80^{5.45}}{10^{10}}\right)} - e^{-1.5581\left(\frac{90^{5.45}}{10^{10}}\right)} = 0.024586$

$p\,(90 < X < 100) = F\,(100) - F\,(90) = e^{-1.5581\left(\frac{90^{5.45}}{10^{10}}\right)} - e^{-1.5581\left(\frac{100^{5.45}}{10^{10}}\right)} = 0.000936$

Las frecuencias esperadas en una muestra de tamaño 1000 se obtendrían multiplicando las probabilidades anteriores por 1000 y redondeando a númeos enteros. Entonces obtenemos la siguiente tabla:

Clase diamétrica (cm)	$0-10$	$10-20$	$20-30$	$30-40$	$40-50$
Frecuencia esperada	0	2	15	63	166

Clase diamétrica (cm)	$50-60$	$60-70$	$70-80$	$80-90$	90 o más
Frecuencia esperada	288	295	145	25	1

1.5. PROBLEMAS RESUELTOS CAPÍTULO 7

1. La probabilidad de que un paciente se recupere de cierta enfermedad es 0.8 . Si 50 personas han contraído la enfermedad, ¿cuál es la probabilidad de que se recuperen 40? ¿ Y la probabilidad de que se recuperen más del 80 %?

   Solución. Si consideramos la variable aleatoria X=número de personas que se recuperan entre las 50 tendremos que $X \rightsquigarrow B(50, 0.8)$. Como $n > 30$ y $0.1 < p < 0.9$ podemos aproximar por una distribución normal con $EX = np = 40$ y $\sigma_X = \sqrt{np(1-p)} = \sqrt{8} = 2\sqrt{2}$.

   Utilizando el cálculo exacto tendríamos $p(X = 40) = \binom{50}{40} 0.8^{40} 0.2^{10} = 0.1398$

   Utilizando la aproximación por la normal con corrección por continuidad tendríamos

   $$p(X = 40) \simeq p\left(\frac{39.5 - 40}{2\sqrt{2}} < Z < \frac{40.5 - 40}{2\sqrt{2}}\right) = p\left(\frac{-1}{4\sqrt{2}} < Z < \frac{1}{4\sqrt{2}}\right) = 1 - 2p\left(Z > \frac{1}{4\sqrt{2}}\right)$$

   Interpolando obtenemos:

   $$\frac{x - 0.4325}{\frac{1}{4\sqrt{2}} - 0.17} = \frac{0.4286 - 0.4325}{0.18 - 0.17}$$

   y por tanto $x = 0.4325 - \frac{(0.0068)(0.0039)}{(0.01)} = 0.4298$. La probabilidad será entonces $1 - 2(0.4298) = 0.1403$

   El 80 % de 50 personas serían 40 personas. Por tanto el cáculo exacto sería

   $$p(X > 40) = \sum_{k=41}^{50} \binom{50}{k} (0.8)^k (0.2)^{50-k} = 0.44374$$

   El cálculo aproximado será

   $$p(X > 40.5) = p\left(Z > \frac{1}{4\sqrt{2}}\right) = 0.4298$$

   Podemos comprobar que la aproximación de la binomial por la normal funciona satisfactoriamente.

2. En un proceso de fabricación de papel aparece, por término medio, un defecto cada 20 metros de papel. Si se puede considerar que el número de defectos sigue una distribución de Poisson, calcular la probabilidad de que haya 30 o más defectos en un rollo de 500 metros de papel.

   Solución. Si consideramos la variable aleatoria X=número de defectos en 20 metros de papel sabemos que $X \rightsquigarrow P(1)$. Entonces el número de defectos en un rollo de 500 metros será $Y = \sum_{k=1}^{25} X_k$ con $X_k \rightsquigarrow P(1)$ e independientes. Por tanto, podemos asegurar que $Y \rightsquigarrow P(25)$. Como el parámetro de la distribución es mayor que 20 podemos aproximar por una distribución normal con $\mu = 25$ y $\sigma = \sqrt{25} = 5$, es decir, $Y \rightsquigarrow N(25, 5)$. Por tanto, utilizando la corrección por continuidad, tendremos que:

   $$p(Y \geq 30) = p(Y \geq 29.5) \simeq p\left(Z > \frac{29.5 - 25}{5}\right) = p(Z > 0.9) = 0.1841$$

   El cáculo exacto sería

   $$p(Y \geq 30) = 1 - p(Y \leq 29) = 1 - \sum_{k=1}^{29} e^{-25}\left(\frac{25^k}{k!}\right) = 0.1821$$

Podemos observar que la aproximación funciona satisfactoriamente.

3. Un distribuidor ha determinado a partir de numerosos ensayos que el 5 % de un grupo grande de semillas no germina. Para asegurarse, vende las semillas en paquetes de 200, garantizando al cliente la germinación del 90 %. ¿Cuál es la probabilidad de que un paquete no cumpla la garantía?

Solución. Para cada semillae podemos asegurar que la probabilidad de que no germine es $p = 0.05$. Sea X=Número de semillas que no germina en un paquete de 200 semillas. Entonces tenemos que $X \rightsquigarrow B(200, 0.05)$. Como $n > 30$ y $p < 0.1$ podemos aproximar esta distribución por una distribución de Poisson con $\lambda = EX = (200)(0.05) = 10$, es decir, $X \rightsquigarrow P(10)$. Para que un paquete no cumpla la garantía necesitamos que $X \geq 20$ y por tanto la probabilidad será:

$$p(X \geq 20) = 1 - p(X < 19) \simeq 1 - \sum_{k=1}^{19} e^{-10}\left(\frac{10^k}{k!}\right) = 1 - 0.9965 = 0.0035$$

El cáculo exacto sería

$$p(X \geq 20) = 1 - p(X < 19) = 1 - \sum_{k=1}^{19} \binom{200}{k} 0.05^k 0.95^{200-k} = 0.0027$$

Por tanto la aproximación es satisfactoria.

4. Tiramos 400 veces una moneda equilibrada.

(a) Hallar la probabilidad de que el número de caras esté comprendido entre 160 y 190.

(b) Obtener un intervalo centrado en 200, tal que la probabilidad de que el número de caras esté en dicho intervalo sea de 0.95.

Solución. Si consideramos la variable aleatoria X=número de caras al tirar una moneda equilibrada 400 veces tendremos que $X \rightsquigarrow B(400, 0.5)$. Como $n > 30$ y $0.1 < p < 0.9$ podemos aproximara esta distribución por una normal con $EX = (400)(0.5) = 200$ y $\sigma = \sqrt{(400)(0.5)(0.5)} = 10$, es decir $X \rightsquigarrow N(200, 10)$.

(a) $p(160 \leq X \leq 190) = p(159.5 \leq X \leq 190.5) \simeq p\left(\frac{159.5-200}{10} < Z < \frac{190.5-200}{10}\right) = p(-4.1 < Z < -0.95)$

$= p(0.95 < Z < 4.1) = p(Z > 0.95) - p(Z > 4.1) = 0.1711 - 0.0000207 = 0.1711$

Utilizando Statgraphics el cálculo exacto sería $p(160 \leq X \leq 190) = 0.1710$. Por tanto la aproximación funciona satisfactoriamente.

(b) Tenemos que buscar x de modo que $p(200 - x \leq X \leq 200 + x) = 0.95$. Utilizando la aproximación por la distribución normal tendremos que

$$p\left(\frac{200 - x - 200}{10} < Z < \frac{200 + x - 200}{10}\right) = 1 - 2p(Z > 0.1x) = 0.95$$

Por tanto debería ser $p(Z > 0.1x) = 0.025$. Buscando en el interior de la tabla tendremos que $0.1x = 1.96$ y por tanto $x = 19.6 \simeq 20$. El intervalo buscado será entonces $(180, 220)$, es decir, entre 181 y 219 caras.

5. Se supone que el número de bacterias por mm^3 de agua en un estanque, es una variable aleatoria X con distribución de Poisson de parámetro $\lambda = 0.13$.

(a) ¿Cuál es la probabilidad de que en un mm^3 de agua del estanque no haya ninguna bacteria?

(b) En 40 tubos de ensayo, se toman muestras del agua del estanque (10 mm3 de agua en cada tubo). ¿Qué distribución sigue la variable *número de tubos de ensayo entre los* 40 *que no contienen bacterias*? Calcular la probabilidad de que haya el 30 % o más de tubos en los que no hay bacterias.

(c) De los tubos que contienen bacterias, ¿qué porcentaje contendrá al menos 4 bacterias?

Solución.

(a) Sea X=número de bacterias en un mm^3 de agua, con $X \rightsquigarrow P(0.13)$. Entonces $p\,(X = 0) = e^{-0.13}\frac{0.13^0}{0!} = e^{-0.13} = 0.8781$

(b) Si cada tubo de ensayo contiene 10 mm3 de agua y consideramos la variable aleatoria Y=número de bacterias en un tubo de ensayo tendremos que $Y \rightsquigarrow P(1.3)$. Por tanto $p\,(Y = 0) = e^{-1.3} = 0.2725$. Si ahora consideramos la variable aleatoria T =*número de tubos de ensayo entre los* 40 *que no contienen bacterias* tendremos que $T \rightsquigarrow B(40, 0.2725)$. Como el 30 % de 40 tubos son 12 tubos tenemos que calcular $p\,(T \geq 12)$. Entonces el cáculo exacto sería:

$$p\,(T \geq 12) = 1 - p\,(T < 12) = 1 - \sum_{k=0}^{11} \binom{40}{k} (0.2725)^k\,(0.7275)^{40-k} = 1 - 0.5943 = 0.4057$$

También podríamos aproximar por una distribución normal $B(40, 0.2725) \approx N\,(10.9, 2.8160)$. En este caso:

$$p\,(T \geq 12) = p\,(T > 11.5) \simeq p\left(Z > \frac{11.5 - 10.9}{2.8160}\right) = p\,(Z > 0.2131)$$

Entonces, interpolando en la tablas de la normal tendremos

$$\frac{x - 0.4168}{0.2131 - 0.21} = \frac{0.4129 - 0.4168}{0.22 - 0.21}$$

y por tanto $x = 0.4168 - \frac{(0.0031)(0.0039)}{(0.01)} = 0.4156$.

(c) Tendremos que calcular $p\,(Y \geq 4/Y > 0)$. Entonces:

$$p\,(Y \geq 4/Y > 0) = \frac{p\,(Y \geq 4)}{p\,(Y > 0)} = \frac{1 - \sum\limits_{k=0}^{3} e^{-1.3}\left(\frac{1.3^k}{k!}\right)}{1 - e^{-1.3}} = \frac{1 - (3.5112)\,e^{-1.3}}{0.7275} = 0.0592$$

Por tanto aproximadamente el 6 % de los tubos que contienen bacterias contendrán por lo menos 4 bacterias.

6. Sean $X_1, X_2, ..., X_{100}$ variables aleatorias independientes tales que $E(X_i) = 2$ y $Var(X_i) = 9$. Calcular, aproximadamente, $p\left(225 \leq \sum\limits_{i=1}^{100} X_i \leq 250\right)$.

Solución. Por ser las variables aleatorias independientes el Teorema Central del Límite nos permitirá asegurar que, aproximadamente, $\sum\limits_{i=1}^{100} X_i \rightsquigarrow N(200, 30)$, porque $E\left(\sum\limits_{i=1}^{100} X_i\right) = \sum\limits_{i=1}^{100} E\,(X_i) = 200$

y $Var\left(\sum_{i=1}^{100} X_i\right) = \sum_{i=1}^{100} Var\left(X_i\right) = 900$. Por tanto tendremos que

$$\begin{aligned} p\left(225 \leq \sum_{i=1}^{100} X_i \leq 250\right) &\simeq p\left(\frac{225-200}{30} < Z < \frac{250-200}{30}\right) = p\left(\frac{5}{6} < Z < \frac{5}{3}\right) = p\left(Z > \frac{5}{6}\right) - p\left(Z > \frac{5}{3}\right) \\ &= p\left(Z > \frac{5}{6}\right) - p\left(Z > 1.67\right) \end{aligned}$$

Interpolando en la tablas de la normal para la primera probabilidad tendremos

$$\frac{x - 0.2033}{\frac{5}{6} - 0.83} = \frac{0.2005 - 0.2033}{0.84 - 0.83}$$

y por tanto $x = 0.2033 - \frac{(0.0033)(0.0028)}{(0.01)} = 0.2024$. La probabilidad buscada será entonces

$$p\left(225 \leq \sum_{i=1}^{100} X_i \leq 250\right) \simeq 0.2024 - 0.0475 = 0.1549$$

7. En una asignatura del colegio la probabilidad de que te saquen a la pizarra en cada clase es del 10 %. A lo largo del año tienes 100 clases de esa asignatura. ¿Cuál es la probabilidad de tener que salir a la pizarra más de 15 veces?

   Solución. Sea X=número de veces que sale a la pizarra en las 100 clases. Tendremos que $X \rightsquigarrow B(100, 0.1)$. Por ser $n > 30$ y $0.1 \leq p \leq 0.9$ podremos aproximar por una distribución normal $B(100, 0.1) \approx N\left(10, 3\right)$ .Entonces:

$$p\left(X > 15\right) = p\left(X > 15.5\right) \simeq p\left(Z > \frac{15.5 - 10}{3}\right) = p\left(Z > 1.83\right) = 0.0336$$

8. Se considera que el número diario de ejemplares vendidos de una determinada publicación en un quiosco es una variable aleatoria que se distribuye según una Poisson de media 30. El precio de venta de cada ejemplar es de 1,3€ euros. Obtener, razonadamente, la probabilidad de que en los próximos 40 días los ingresos por la venta de ejemplares superen los 1625€.

   Solución. Sea X=número de ejemplares vendidos en un día. Sabemos que $EX = 30$ y por tanto $X \rightsquigarrow P(30)$. El número de ejemplares vendidos en los próximos 40 días será $N = \sum_{i=1}^{40} X_i$ y por tanto $N \rightsquigarrow P(1200)$. Como $\lambda > 25$ podemos aproximar esta distribución por un distribución normal $P(1200) \approx N\left(1200, 20\sqrt{3}\right)$.Los ingresos por la venta de ejemplares serán entonces $I = 1.3N$ y tendremos que calcular $p\left(I > 1625\right)$. Entonces:

$$p\left(I > 1625\right) = p\left(N > \frac{1625}{1.3}\right) = p\left(N > 1250\right) \simeq p\left(Z > \frac{1250.5 - 1200}{20\sqrt{3}}\right) = p\left(Z > 1.46\right) = 0.0721$$

9. El dinero que se gastan los adolescentes de entre 16 y 18 años durante un fin de semana sigue una distribución desconocida de media 6.20 € y desviación típica 1,90 €. Se toma al azar una muestra de 60 de esos adolescentes.

   (a) ¿Qué distribución sigue la media del gasto en dicha muestra?

   (b) ¿Cuál es la probabilidad de que la media del gasto en esa muestra sea superior a 7€?

   (c) ¿Qué hubiese ocurrido si tomamos una muestra de solo 28 estudiantes?

Solución.

(a) Sea X_i =gasto de un adolescente durante un fin de semana, con $i = 1, 2, ..., 60$. Por el Teorema Central del Límite podemos asegurar que la distribución aproximada del gasto medio de los 60 adolescentes es $\overline{X} = \frac{\sum_{i=1}^{60} X_i}{60} \rightsquigarrow N\left(6.2, \frac{1.9}{\sqrt{60}}\right)$ porque $E\left(\overline{X}\right) = E\left(X_i\right) = 6.2$ y $\sigma_{\overline{X}} = \frac{\sigma_{X_i}}{\sqrt{60}} = \frac{1.9}{\sqrt{60}}$.

(b) $p\left(\overline{X} > 7\right) = p\left(Z > \frac{7-6.2}{1.9/\sqrt{60}}\right) = p\left(Z > 3.3\right) = 0.000483$

(c) Si sólo hubiese 28 estudiantes la muestra no sería lo suficientemente grande para aplicar el Teorema Central de Límite para una distribución de probabilidad cualquiera (en general $n > 30$).

10. Ante una medida adoptada por el Gobierno, se sabe que el 35 % de la población está a favor de dicha medida. Se toma una muestra al azar de 200 personas.

(a) ¿Qué número de personas se espera que estén a favor?

(b) Halla la distribución en el muestreo de la proporción de personas que están a favor de esta medida.

(c) Halla la probabilidad de que en la muestra elegida, más del 40 % de los integrantes que la forman estén a favor.

Solución. Sea X=número de personas que están a favor en la muestra de 200 personas. Tendremos que $X \rightsquigarrow B\left(200, 0.35\right)$. Entonces:

(a) $EX = (200)(0.35) = 70$

(b) La proporción de personas que están a favor sería $\overline{X} = \left(\sum_{i=1}^{60} X_i\right)/200$ siendo X_i variables aleatorias independientes con distribución de Bernouilli $X_i \rightsquigarrow B\left(0.35\right)$. Por el Teorema Central del Límite podemos decir que la distribución de probabilidad de $\overline{X}$ es aproximadamente normal con $E\left(\overline{X}\right)$ y $Var\left(\overline{X}\right) = (0.35)(0.65)/200$ $\overline{X} \rightsquigarrow N\left(0.35, \sqrt{0.011375}\right) = N\left(0.35, 0.034\right)$.

(c) $p\left(\overline{X} > 0.4\right) \simeq p\left(Z > \frac{0.4-0.35}{0.034}\right) = p\left(Z > 1.47\right) = 0.0708$

1.6. PROBLEMAS RESUELTOS CAPÍTULO 8

1. Según el Environmental News (Septiembre de 1975), la “lluvia ácida” causada por la reacción de ciertos contaminantes en el aire con el agua de la lluvia es un problema creciente en ciertas regiones. La lluvia pura que se precipita a través del aire limpio tiene un pH de 5.7 (el pH es una medida de la acidez: 0 ácido y 14 alcalino). Se analizan 40 muestras de agua de diferentes lluvias con respecto a su pH dando una media igual a 3.7 y desviación típica de 0.5. Determinar un intervalo de confianza del 99 % para el pH medio de las lluvias e interpretar el resultado. ¿Que supuestos deben hacerse para que el intervalo sea válido?.

   Si se desea estimar el promedio de PH de las lluvias en un área que experimenta una gran contaminación debido a la presencia de una fábrica que emite gran cantidad de humos, y se sabe que $\sigma = 0.5$, ¿cuántas lluvias deben tomarse para que la estimación difiera de la media teórica como mucho en 0.1 con una probabilidad de al menos 0.95?

   Solución. Sea X=ph del agua de lluvia y $X_1, X_2, ..., X_{40}$ una muestra aleatoria simple (m.a.s) de la variable X. Si suponemos que $X \rightsquigarrow N(\mu, \sigma)$ entonces tendremos que la media muestral $\overline{X} = \frac{1}{40}\sum_{i=1}^{40} X_i$ también tendrá una distribución normal, concretamente, $\overline{X} \rightsquigarrow N\left(\mu, \frac{\sigma}{\sqrt{40}}\right)$. Además, si definimos $S_c^2 = \left(\frac{40}{39}\right) S^2$ con $S^2 = \left(\frac{1}{40}\right)\sum_{i=1}^{40} X_i^2 - \left(\overline{X}\right)^2$, sabemos que $\frac{\overline{X}-\mu}{S_c/\sqrt{40}} \rightsquigarrow t_{39}$. Si queremos un grado de confianza de 99 % tendremos que $1-\alpha = 0.99$ y por tanto $\alpha/2 = 0.005$. Buscando en la tabla de la distribución t-Student tenemos que $t_{39;0.005} = 2.704$ (hemos utilizado $n = 40$ que es el más próximo) y podemos asegurar que

$$p\left(-2.704 \leq \frac{\overline{X}-\mu}{S_c/\sqrt{40}} \leq 2.704\right) = 0.99$$

   Por tanto, despejando el parámetro μ tendremos:

$$p\left(\overline{X} - 2.704\left(\frac{S_c}{\sqrt{40}}\right) \leq \mu \leq \overline{X} + 2.704\left(\frac{S_c}{\sqrt{40}}\right)\right) = 0.99$$

   De este modo podemos asegurar que $\left(\overline{X} - 2.704S_c/\sqrt{40}, \overline{X} + 2.704S_c/\sqrt{40}\right)$ es un intervalo de confianza al 99 % para el pH medio de las lluvias, es decir, el valor esperado de X. Si en la muestra analizada hemos obtenido $\bar{x} = 3.7$ y $s_c = 0.5$ entonces el intervalo obtenido será

$$\left(3.7 - (2.704)(0.5)/\sqrt{40}, 3.7 + (2.704)(0.5)/\sqrt{40}\right) = (3.48, 3.92)$$

   Entonces, podemos asegurar con una confianza del 99 % que el pH medio de las lluvias está entre 3.48 y 3.92 y por tanto este valor medio está por debajo del valor de la lluvia pura (5.7) con un confianza del 99 %. Hemos necesitado suponer que la distribución de probabilidad del pH de la lluvia tiene una distribución normal.

   Si suponemos que la desviación típica σ es conocida con $\sigma = 0.5$ y consideramos un tamaño muestral n, entonces podríamos asegurar que $\overline{X} \rightsquigarrow N\left(\mu, \frac{0.5}{\sqrt{n}}\right)$. Si queremos que la estimación $\overline{X}$ difiera de la media teórica como mucho en 0.1 con una probabilidad de al menos 0.95 necesitamos que se verifique $p\left(\left|\overline{X}-\mu\right|\leq 0.1\right)\geq 0.95$, es decir, $p\left(-0.1\leq \overline{X}-\mu\leq 0.1\right)\geq 0.95$. Esta condición sería equivalente a $p\left(\frac{-0.1}{0.5/\sqrt{n}}\leq Z\leq \frac{0.1}{0.5/\sqrt{n}}\right)\geq 0.95$ y, como $p\left(\frac{-0.1}{0.5/\sqrt{n}}\leq Z\leq \frac{0.1}{0.5/\sqrt{n}}\right)=1-2p\left(Z\geq \frac{0.1}{0.5/\sqrt{n}}\right)$, esto sería equivalente a $p\left(Z\geq \frac{0.1}{0.5/\sqrt{n}}\right)\leq 0.025$. Buscando en la tabla de la distribución normal tenemos que $z_{0.025}=1.96$ y por tanto deberá ser $\frac{0.1}{0.5/\sqrt{n}}\geq 1.96$. Finalmente, despejando el valor de n tendremos que $n\geq\left(\frac{(1.96)(0.5)}{0.1}\right)^2=96.04$, es decir, deberíamos tomar al menos 97 lluvias para garantizar la estimación deseada.

2. Un periódico señala en su número del 23 de mayo de 1979 que "los estudiantes de derecho se oponen a la pena de muerte". Esta declaración se hizo en base a una encuesta para la cual se escogieron al azar y se entrevistaron 86 estudiantes de derecho. El 52 % de los entrevistados declararon que se oponían a la pena de muerte. A partir de esta información obtenga un intervalo de confianza del 95 % para la proporción real de estudiantes de derecho que se oponen a la pena de muerte. ¿Se justifica la afirmación de este periódico? ¿Cuántos estudiantes se tendrán que entrevistar por parte del periódico para estimar la proporción de estudiantes en contra de la pena de muerte, con un error máximo de 0.01 con una probabilidad de al menos 0.95?

   Solución. Para cada entrevistado, $i=1,2,...,86$, consideramos la variable aleatoria

$$X_i=\begin{cases}1 & \text{si el estudiante se opone a la pena de muerte}\\ 0 & \text{en caso contrario}\end{cases}$$

   Sabemos que $X_i\rightsquigarrow B(p)$ siendo p la verdadera proporción de estudiantes que se oponen a la pena de muerte. Queremos calcular un intervalo de confianza para p con un grado de confianza del 95 %, es decir, $1-\alpha=0.95$ y por tanto $\alpha/2=0.025$. El estimador puntual para p es la media muestral de las variables aleatorias X_i, que en este caso representa la proporción muestral que estudiantes que se oponen a la pena de muerte, es decir, $\widehat{p}=\overline{X}=\frac{1}{86}\sum_{i=1}^{86}X_i$. Por el Teorema Central de Límite sabemos que la distribución de probabilidad de $\widehat{p}$ es aproximadamente normal con $\widehat{p}\widetilde{\rightsquigarrow}N\left(p,p\left(1-p\right)/\sqrt{86}\right)$ y también

$$\widehat{p}\widetilde{\rightsquigarrow}N\left(p,\frac{\widehat{p}\left(1-\widehat{p}\right)}{86}\right)$$

   Entonces, aproximadamente, tenemos que

$$\frac{\widehat{p}-p}{\sqrt{\widehat{p}\left(1-\widehat{p}\right)/86}}\widetilde{\rightsquigarrow}N\left(0,1\right)$$

   Utilizando este estadístico pivote, y teniendo en cuenta que $z_{0.025}=1.96$, tendremos que

$$p\left(-1.96\leq\frac{\widehat{p}-p}{\sqrt{\widehat{p}\left(1-\widehat{p}\right)/86}}\leq 1.96\right)\simeq 0.95$$

   y despejando el parámetro p tendremos que

$$p\left(\widehat{p}-1.96\sqrt{\frac{\widehat{p}\left(1-\widehat{p}\right)}{86}}\leq p\leq\widehat{p}+1.96\sqrt{\frac{\widehat{p}\left(1-\widehat{p}\right)}{86}}\right)\simeq 0.95$$

Por tanto un intervalo de confianza asintótico para p con 95 % de confianza será $\widehat{p} \pm 1.96\sqrt{\frac{\widehat{p}(1-\widehat{p})}{86}}$. Con la vuestra observada tenemos que $\widehat{p} = 0.52$ y el intervalo de confianza será

$$0.52 \pm 1.96\sqrt{\frac{(0.52)(0.48)}{86}} = 0.52 \pm 0.106 = (0.414, 0.626)$$

Ahora, como $0.5 \in (0.414, 0.626)$ no podemos asegurar con un confianza del 95 % que los estudiantes que se oponen a la pena de muerte sean mayoría y no se justifica la afirmación del periódico.

Si n fuese el nuevo tamaño muestral para que la afirmación fuese justificada necesitaríamos que

$$0.52 - 1.96\sqrt{\frac{(0.52)(0.48)}{n}} > 0.5$$

lo cual equivale a que $n > (0.52)(0.48)\left(\frac{1.96}{0.02}\right)^2 = 2397.2$. Por tanto el tamaño muestral debería ser por lo menos $n = 2398$ personas entrevistadas.

3. La dirección médica de una clínica desea estimar el número promedio de días necesarios para el tratamiento de pacientes con edades entre 25 y 34 años. Una muestra aleatoria de 500 pacientes con estas características proporcionó una media y una desviación estándar (utilizando la cuasi-varianza) de 5.4 y 3.1 días, respectivamente. Obtener un intervalo de confianza del 95 % para el promedio de estancia en el hospital de la población de pacientes de la que se obtuvo la muestra.

   Solución. Consideramos la variable aleatoria X=número de días necesarios para el tratamiento. Teniendo en cuenta que el tamaño muestral, $n = 500$, es lo suficientemente grande como para poder aplicar el Teorema Central del Límite, sea cual sea la distribución de probabilidad de X, podemos asegurar que aproximadamente

$$\frac{\overline{X} - \mu}{S_c/\sqrt{500}} \widetilde{\rightsquigarrow} N(0, 1)$$

   Entonces, como queremos una confianza del 95 %, tendremos que $\alpha/2 = 0.025$, $z_{0.025} = 1.96$ y podemos asegurar que

$$p\left(-1.96 \leq \frac{\overline{X} - \mu}{S_c/\sqrt{500}} \leq 1.96\right) \simeq 0.95$$

   o equivalentemente

$$p\left(\overline{X} - 1.96\frac{S_c}{\sqrt{500}} \leq \mu \leq \overline{X} + 1.96\frac{S_c}{\sqrt{500}}\right) \simeq 0.95$$

   Por tanto el intervalo de confianza asintótico será $\overline{X} \pm 1.96\frac{S_c}{\sqrt{500}}$. Si en nuestro caso hemos obtenido $\bar{x} = 5.4$ y $s_c = 3.1$ entonces el intervalo de confianza buscado será

$$5.4 \pm (1.96)(3.1)/\sqrt{500} = 5.4 \pm 0.271 = (5.12, 5.68)$$

4. Una encuesta intenta estudiar la relación entre la participación en los deportes y la destreza manual. De una muestra aleatoria de 37 alumnos de Enseñanza Secundaria que participaban regularmente en actividades deportivas se obtuvo una calificación media en asignaturas que medían la destreza manual de 32.19 con una desviación estándar de 4.34. De otra muestra aleatoria independiente de 37 alumnos de Enseñanza Secundaria que no participaban en deportes se calculó una calificación media de 31.68 y una desviación estándar de 4.56. Estimar la diferencia de los verdaderos promedios

de ambos grupos con un intervalo de confianza del 90 %. ¿Se desprende de los resultados que la calificación promedio de ambos grupos es diferente?

Si debe efectuarse otro estudio similar, ¿cuántas observaciones deben incluirse en cada grupo (igual número en ambos) para producir un intervalo de confianza del 90 % con una amplitud de 2 unidades?

Solución. Consideramos las variables aleatorias X=destreza manual de un alumno que participa en actividades deportivas e Y=destreza manual de un alumno que no participa en actividades deportivas. Sean μ_X, μ_Y los respectivos valores esperados, y σ_X, σ_Y las respectivas desviaciones típicas de las dos variables aleatorias. Para saber si la calificación promedio de ambos grupos es diferente tenemos que calcular un intervalo de confianza para $\mu_X - \mu_y$. Sean $X_1, X_2, ..., X_{37}$ e $Y_1, Y_2, ..., Y_{37}$ las respectivas muestras aleatorias simples de alumnos, que consideramos independientes entre si. Como no nos dice qué distribuciones de probabilidad tienen estas variables aleatorias vamos a suponer que el tamaño muestral ($n = 37$) es suficientemente grande para poder aplicar el Teorema Central del Límite para las medias muestrales de ambas muestras $\overline{X} = \frac{1}{37}\sum_{i=1}^{37} X_i$ e $\overline{Y} = \frac{1}{37}\sum_{i=1}^{37} Y_i$. Podemos entonces calcular un intervalo asintótico a partir del siguiente estadístico pivote:

$$\frac{\left(\overline{X} - \overline{Y}\right) - \left(\mu_X - \mu_y\right)}{\sqrt{\frac{S_X^2}{37} + \frac{S_Y^2}{37}}} \underset{\sim}{\approx} N(0,1)$$

siendo S_X^2, S_Y^2 las respectivas varianzas muestrales corregidas de ambas muestras. Como queremos una confianza del 90 % tendremos que $\alpha/2 = 0.05$ y $z_{0.05} = 1.645$. Entonces podemos asegurar que

$$p\left(-1.645 \leq \frac{\left(\overline{X} - \overline{Y}\right) - \left(\mu_X - \mu_y\right)}{\sqrt{\frac{S_X^2}{37} + \frac{S_Y^2}{37}}} \leq 1.645\right) \simeq 0.90$$

o equivalentemente

$$p\left(\left(\overline{X} - \overline{Y}\right) - 1.645\sqrt{\frac{S_X^2}{37} + \frac{S_Y^2}{37}} \leq \mu_X - \mu_y \leq \left(\overline{X} - \overline{Y}\right) + 1.645\sqrt{\frac{S_X^2}{37} + \frac{S_Y^2}{37}}\right) \simeq 0.90$$

Entonces un intervalo de confianza asintótico al 90 % para $\mu_X - \mu_y$ será

$$\left(\overline{X} - \overline{Y}\right) \pm 1.645\sqrt{\frac{S_X^2 + S_Y^2}{37}}$$

Como nuestras muestras han proporcionado unos valores $\overline{x} = 32.19$, $\overline{y} = 31.68$, $s_X = 4.34$, $s_Y = 4.56$ el intervalo será

$$(32.19 - 31.68) \pm 1.645\sqrt{\frac{4.34^2 + 4.56^2}{37}} = 0.51 \pm 1.71 = (-1.19, 2.22)$$

Por tanto, como $0 \in (-1.19, 2.22)$ no podemos asegurar que la calificación promedio de ambos grupos sea diferente con una confianza de 90 %.

Para que el intervalo de confianza tuviese una amplitud máxima de 2 unidades necesitaríamos que

$$2\,(1.645)\sqrt{\frac{4.34^2 + 4.56^2}{n}} < 2$$

y por tanto $n > \left(4.34^2 + 4.56^2\right)(1.645)^2 = 107.2$. Entonces necesitaríamos una muestra de 108 alumnos en cada grupo.

5. Una encuesta realizada en Otoño de 1979 con respecto a la política de jubilaciones, reveló el pesimismo de la población respecto a sus perspectivas cuando lleguen a jubilarse. El 62.9 % de los 6100 trabajadores entrevistados, indicaron que creían que sus ingresos al jubilarse no serían suficientes. Calcular un intervalo de confianza del 95 % para esta proporción e interpretarlo.

   Solución. Para cada entrevistado, $i = 1, 2, ..., 6100$, consideramos la variable aleatoria

$$X_i = \begin{cases} 1 & \text{si considera que sus ingresos al jubilarse no serán suficientes} \\ 0 & \text{en caso contrario} \end{cases}$$

   Sabemos que $X_i \rightsquigarrow B(p)$ siendo p la verdadera proporción de personas que opinan que sus ingresos al jubilarse no serán suficientes. Queremos calcular un intervalo de confianza para p con un grado de confianza del 95 %, es decir, $1-\alpha = 0.95$ y por tanto $\alpha/2 = 0.025$. El estimador puntual para p es la media muestral de las variables aleatorias X_i, que en este caso representa la proporción muestral de personas que opinan que sus ingresos al jubilarse no serán suficientes, es decir, $\widehat{p} = \overline{X} = \frac{1}{86}\sum_{i=1}^{86} X_i$. Entonces el estadístico pivote para encontrar un intervalo de confianza asintótico en este caso es

$$\frac{\widehat{p}-p}{\sqrt{\widehat{p}\left(1-\widehat{p}\right)/6100}} \widetilde{\rightsquigarrow} N\left(0,1\right)$$

   Utilizando este estadístico pivote, y teniendo en cuenta que $z_{0.025} = 1.96$, tendremos que

$$p\left(-1.96 \leq \frac{\widehat{p}-p}{\sqrt{\widehat{p}\left(1-\widehat{p}\right)/6100}} \leq 1.96\right) \simeq 0.95$$

   y despejando el parámetro p tendremos que

$$p\left(\widehat{p}-1.96\sqrt{\frac{\widehat{p}\left(1-\widehat{p}\right)}{6100}} \leq p \leq \widehat{p}+1.96\sqrt{\frac{\widehat{p}\left(1-\widehat{p}\right)}{6100}}\right) \simeq 0.95$$

   Por tanto un intervalo de confianza asintótico para p con 95 % de confianza será $\widehat{p} \pm 1.96\sqrt{\frac{\widehat{p}(1-\widehat{p})}{6100}}$. Con la muestra observada tenemos que $\widehat{p} = 0.629$ y el intervalo de confianza será

$$0.629 \pm 1.96\sqrt{\frac{(0.629)\,(0.371)}{6100}} = 0.629 \pm 0.0122 = (0.616, 0.642)$$

   Podemos asegurar entonces con un confianza del 95 % que la proporción de personas que opinan que sus ingresos al jubilarse no serán suficientes se encuentra entre 0.616 y 0.642.

6. Se realizó una encuesta preguntando si el sexo de un candidato político era decisivo a la hora de recibir el voto. El 62 % por ciento de 241 hombres y el 49 % de 256 mujeres opinaron que el sexo del candidato no importaba. Construir un intervalo de confianza para la diferencia de proporciones de hombres y mujeres que opinan que el sexo del candidato no tiene importancia en la votación.

   Solución. Sea $X_1, X_2, ..., X_{241}$ la muestra aleatoria simple de hombres, con $X_i \rightsquigarrow B(p_X)$ siendo p_X la verdadera proporción de hombres que opinan que el sexo del candidato no importa. Del mismo modo sea $Y_1, Y_2, ..., Y_{256}$ la muestra aleatoria simple de mujeres, con $Y_i \rightsquigarrow B(p_Y)$ siendo p_Y la verdadera proporción de mujeres que opinan que el sexo del candidato no importa. Vamos a utilizar un grado de confianza de 95 % y un intervalo de confianza asintótico para la diferencia $p_X - p_Y$. El estadístico pivote en este caso será

$$\frac{(\widehat{p_X} - \widehat{p_Y}) - (p_X - p_Y)}{\sqrt{\frac{\widehat{p_X}(1-\widehat{p_X})}{241} + \frac{\widehat{p_Y}(1-\widehat{p_Y})}{256}}} \widetilde{\rightsquigarrow} N(0,1)$$

y, como $z_{0.025} = 1.96$, tenemos que

$$p\left(-1.96 \leq \frac{(\widehat{p_X} - \widehat{p_Y}) - (p_X - p_Y)}{\sqrt{\frac{\widehat{p_X}(1-\widehat{p_X})}{241} + \frac{\widehat{p_Y}(1-\widehat{p_Y})}{256}}} \leq 1.96\right) \simeq 0.95$$

o, equivalentemente

$$p\left((\widehat{p_X} - \widehat{p_Y}) - 1.96\sqrt{\frac{\widehat{p_X}(1-\widehat{p_X})}{241} + \frac{\widehat{p_Y}(1-\widehat{p_Y})}{256}} \leq (p_X - p_Y) \leq (\widehat{p_X} - \widehat{p_Y}) - 1.96\sqrt{\frac{\widehat{p_X}(1-\widehat{p_X})}{241} + \frac{\widehat{p_Y}(1-\widehat{p_Y})}{256}}\right) \simeq 0.95$$

Por tanto el intervalo de confianza asintótico buscado será

$$(\widehat{p_X} - \widehat{p_Y}) \pm 1.96\sqrt{\frac{\widehat{p_X}(1-\widehat{p_X})}{241} + \frac{\widehat{p_Y}(1-\widehat{p_Y})}{256}}$$

Como en nuestro caso hemos obtenido $\widehat{p_X} = 0.62$ y $\widehat{p_Y} = 0.49$ el intervalo será

$$0.13 \pm 1.96\sqrt{\frac{(0.62)(0.38)}{241} + \frac{(0.49)(0.51)}{256}} = 0.13 \pm (1.96)(0.044) = 0.13 \pm 0.087 = (0.043, 0.217)$$

Observar que $0 < 0.043$ y por tanto podemos afirmar, con una confianza del 95 %, que la proporción de hombres que opina que el sexo del candidato no importa es mayor que para las mujeres.

7. Los resultados de un estudio sobre la concentración de plomo en el agua indicaron que el 20 % de los 248 hogares que se analizaron tenían una concentración de plomo en el agua que superaba el nivel permitido. En otro punto geográfico se observó que solo el 5 % de los 110 hogares analizados superaban este nivel. ¿Existe diferencia entre las dos zonas? Utilizar para responder a esta pregunta un intervalo de confianza del 95 %.

   Solución. Sea $X_1, X_2, ..., X_{248}$ la muestra aleatoria simple de hogares en el primer punto geográfico donde la concentración de plomo en el agua supera el nivel permitido, con $X_i \rightsquigarrow B(p_X)$ siendo p_X la verdadera proporción de hogares en el primer punto geográfico donde la concentración de plomo en el agua supera el nivel permitido. Del mismo modo sea $Y_1, Y_2, ..., Y_{110}$ la muestra aleatoria simple de hogares en el segundo punto geográfico donde la concentración de plomo en el agua supera el nivel permitido, con $Y_i \rightsquigarrow B(p_Y)$ siendo p_Y la verdadera proporción de hogares en el segundo punto geográfico donde la concentración de plomo en el agua supera el nivel permitido. Como queremos un grado de confianza de 95 %, tenemos que $1 - \alpha = 0.95$, $\alpha/2 = 0.025$ y $z_{0.025} = 1.96$. Vamos a utilizar un intervalo de confianza asintótico para las diferencia $p_X - p_Y$. El estadístico pivote en este caso será

$$\frac{(\widehat{p_X} - \widehat{p_Y}) - (p_X - p_Y)}{\sqrt{\frac{\widehat{p_X}(1-\widehat{p_X})}{248} + \frac{\widehat{p_Y}(1-\widehat{p_Y})}{110}}} \widetilde{\rightsquigarrow} N(0,1)$$

   y, como $z_{0.025} = 1.96$, tenemos que

$$p\left(-1.96 \leq \frac{(\widehat{p_X} - \widehat{p_Y}) - (p_X - p_Y)}{\sqrt{\frac{\widehat{p_X}(1-\widehat{p_X})}{248} + \frac{\widehat{p_Y}(1-\widehat{p_Y})}{110}}} \leq 1.96\right) \simeq 0.95$$

o, equivalentemente

$$p\left((\widehat{p_X}-\widehat{p_Y})-1.96\sqrt{\frac{\widehat{p_X}(1-\widehat{p_X})}{248}+\frac{\widehat{p_Y}(1-\widehat{p_Y})}{110}}\leq(p_X-p_Y)\leq(\widehat{p_X}-\widehat{p_Y})-1.96\sqrt{\frac{\widehat{p_X}(1-\widehat{p_X})}{248}+\frac{\widehat{p_Y}(1-\widehat{p_Y})}{110}}\right)\simeq 0.95$$

Por tanto el intervalo de confianza asintótico buscado será

$$(\widehat{p_X}-\widehat{p_Y})\pm 1.96\sqrt{\frac{\widehat{p_X}(1-\widehat{p_X})}{248}+\frac{\widehat{p_Y}(1-\widehat{p_Y})}{110}}$$

Como en nuestro caso hemos obtenido $\widehat{p_X}=0.20$ y $\widehat{p_Y}=0.05$ el intervalo será

$$0.15\pm 1.96\sqrt{\frac{(0.2)(0.8)}{248}+\frac{(0.05)(0.95)}{110}}=0.15\pm(1.96)(0.033)=0.15\pm 0.064=(0.085,0.215)$$

Observar que $0<0.085$ y por tanto podemos afirmar, con una confianza del 95 %, que la proporción de hogares donde la concentración de plomo en el agua supera el nivel permitido es mayor en el primer punto geográfico que en el segundo.

8. La tasa de consumo de oxígeno es una medida importante a la hora de estudiar la actividad fisiológica de los atletas. Se realizó un estudio para intentar observar las diferencias entre el consumo de oxígeno en dos grupos de varones universitarios entrenados con dos métodos diferentes, en uno el entrenamiento se realizaba de forma continua y en otro de forma intermitente. El consumo de oxígeno se mide en mililitros por kilogramo-minuto, y los resultados obtenidos vienen dados en la siguiente tabla:

	Entrenamiento I	Entrenamiento II
Tamaño muestral	9	7
Media muestral	43.71	39.63
Desviación típica corregida	5.88	7.68

Si se supone que las mediciones provienen de poblaciones normales con varianzas iguales, estimar la diferencia entre las medias poblacionales de ambos tipos de entrenamiento con una confianza del 95 %. ¿Como podría comprobar que las varianzas se pueden considerar iguales?

Responder a las mismas preguntas sin suponer que las mediciones provienen de dos poblaciones normales, esto es, provienen de dos distribuciones arbitrarias.

Solución. Consideramos las variables aleatorias X=tasa de consumo de oxígeno en atletas con entrenamiento I e Y=tasa de consumo de oxígeno en atletas con entrenamiento II. Suponemos que $X \rightsquigarrow N(\mu_X,\sigma)$ e $Y \rightsquigarrow N(\mu_Y,\sigma)$ porque el problema supone poblaciones normales con la misma varianza. Para estimar la diferencia $\mu_X-\mu_y$ entre las medias poblacionales utilizaremos un intervalo de confianza al 95 % para esa diferencia. Sean $X_1,X_2,...,X_9$ e $Y_1,Y_2,...,Y_7$ las respectivas muestras aleatorias simples de alumnos, que consideramos independientes entre si. En este caso sabemos que los estimadores puntuales de μ_X y μ_Y son, respectivamente, $\widehat{\mu_X}=\overline{X}=\frac{1}{9}\sum\limits_{i=1}^{9}X_i$ e $\widehat{\mu_Y}=\overline{Y}=\frac{1}{7}\sum\limits_{i=1}^{7}Y_i$, y también que el estimador de la varianza común es

$$\widehat{\sigma^2}=S_p^2=\frac{8S_X^2+6S_Y^2}{14}$$

siendo $S_X^2 = \left(\frac{9}{8}\right)\left[\left(\frac{1}{9}\right)\sum_{i=1}^{9} X_i^2 - \overline{X}^2\right]$ y $S_Y^2 = \left(\frac{7}{6}\right)\left[\left(\frac{1}{7}\right)\sum_{i=1}^{7} Y_i^2 - \overline{Y}^2\right]$ las respectivas varianzas muestrales corregidas de ambas muestras. En estas circunstancia sabemos que el estadístico pivote para el intervalo de confianza buscado es

$$\frac{\left(\overline{X}-\overline{Y}\right)-\left(\mu_X-\mu_y\right)}{S_p\sqrt{\frac{1}{9}+\frac{1}{7}}} \rightsquigarrow t_{14}$$

y, buscando en la tablas de la distribución $t-$Student, tenemos que $t_{14;0.025} = 2.145$. Por tanto podemos asegurar que

$$p\left(-2.145 \le \frac{\left(\overline{X}-\overline{Y}\right)-\left(\mu_X-\mu_y\right)}{S_p\sqrt{\frac{1}{9}+\frac{1}{7}}} \le 2.145\right) = 0.95$$

o, equivalentemente, que

$$p\left(\left(\overline{X}-\overline{Y}\right)-2.145S_p\sqrt{\frac{1}{9}+\frac{1}{7}} \le \mu_X-\mu_y \le \left(\overline{X}-\overline{Y}\right)+2.145S_p\sqrt{\frac{1}{9}+\frac{1}{7}}\right) = 0.95$$

Por tanto el intervalo de confianza buscado será

$$\left(\overline{X}-\overline{Y}\right) \pm 2.145S_p\sqrt{\frac{1}{9}+\frac{1}{7}}$$

Con los datos del problema los valores obtenidos son $\overline{x} = 43.71$, $\overline{y} = 39.63$, $s_X^2 = 5.88^2$ y $s_Y^2 = 7.68^2$. Entonces, para S_p obtenemos

$$s_p = \sqrt{\frac{8\,(5.88^2)+6\,(7.68^2)}{14}} = 6.711$$

y el intervalo de confianza buscado será

$$4.08 \pm (2.145)\,(6.711)\sqrt{\frac{1}{9}+\frac{1}{7}} = 4.08 \pm 7.254 = (-3.174, 11.334)$$

Por tanto, como $-3.174 < 0 < 11.334$ no podemos asegurar con un 95 % de confianza que las dos medias poblacionales sean distintas.

Para comprobar si las varianzas poblacionales σ_X^2 y σ_Y^2 pueden considerarse iguales podríamos calcular un intervalo de confianza del 90 % para el cociente $\left(\frac{\sigma_X^2}{\sigma_Y^2}\right)$ y ver si el valor 1 esta dentro del intervalo. Utilizando como estimador puntual de este cociente el cociente de las varianzas muestrales corregidas, es decir

$$\widehat{\left(\frac{\sigma_X^2}{\sigma_Y^2}\right)} = \frac{S_X^2}{S_y^2}$$

sabemos que el estadístico pivote en este caso es

$$\frac{S_X^2/S_Y^2}{\sigma_X^2/\sigma_Y^2} \rightsquigarrow F_{8,6}$$

Utilizando las tablas de la distribución $F-$Fisher con $\alpha = 0.05$ tenemos que $F_{8,6;0.05} = 4.1468$ y además

$$F_{8,6;0.95} = \frac{1}{F_{6,8;0.05}} = \frac{1}{3.5806}$$

Por tanto podemos asegurar que

$$p\left(\frac{1}{3.5806} \leq \frac{S_X^2/S_Y^2}{\sigma_X^2/\sigma_Y^2} \leq 4.1468\right) = 0.90$$

o equivalentemente

$$p\left(\frac{S_X^2/S_Y^2}{4.1468} \leq \sigma_X^2/\sigma_Y^2 \leq (3.5806)\left(S_X^2/S_Y^2\right)\right) = 0.90$$

Entonces el intervalo de confianza buscado es

$$\left(\frac{S_X^2/S_Y^2}{4.1468}, (3.5806)\left(S_X^2/S_Y^2\right)\right)$$

Como en nuestro caso hemos obtenido $s_X^2 = 5.88^2$ y $s_Y^2 = 7.68^2$, sustituyendo tendremos

$$\left(\frac{5.88^2}{(4.1468)\left(7.68^2\right)}, \frac{(3.5806)\left(5.88^2\right)}{7.68^2}\right) = (0.1413, 2.0989)$$

Podemos observar que $1 \in (0.1413, 2.0989)$ y entonces no podemos asegurar con un 90 % de confianza que las varianzas sean distintas. Por tanto podemos considerar que son iguales.

Si suponemos que las distribuciones de las variables aleatorias son arbitrarias (no necesariamente normales) no podríamos responder a las anteriores preguntas porque los tamaños muestrales son demasiado pequeños para calcular un intervalo de confianza asintótico, que sería la única alternativa.

9. La Agencia para la Protección Ambiental reunió información respecto a las mediciones de $CL50$ (concentración letal de una sustancia que mata el 50 % de los animales utilizados en la experimentación) para ciertos productos que se encuentran en ocasiones en ríos y lagos de agua dulce. Para cierta especie de peces las mediciones de $CL50$ para el DDT en 12 experimentos fueron las siguientes (en partes por millón): 16, 5, 21, 19, 10, 5, 8, 2, 7, 2, 4, 9 Estimar el verdadero promedio de $CL50$ para el DDT utilizando un intervalo de confianza del 90 % suponiendo que las mediciones de $CL50$ siguen aproximadamente una distribución normal.

   Otro insecticida común, Diazinón, dio las siguientes mediciones de $CL50$ en tres experimentos: 7.8, 1.6 y 1.3. Calcular en este caso un intervalo de confianza del 90 % para la media poblacional. Estimar la diferencia entre los promedios de $CL50$ para el DDT y el Diazinón utilizando un intervalo de confianza del 90 %. ¿Qué supuestos hay que realizar para que el intervalo construido sea válido?

   Solución. Sea la variable aleatoria X=concentración de $CL50$ para el DDT, con $X \rightsquigarrow N(\mu_X, \sigma_X)$, y $X_1, X_2, ..., X_{12}$ la muestra aleatoria simple de X. Tenemos que estimar el verdadero promedio de $CL50$, es decir μ_X, utilizando un intervalo de confianza con $1-\alpha = 0.9$ y por tanto $\alpha/2 = 0.05$. Sabemos que el estimador puntual es $\widehat{\mu_X} = \overline{X} = \frac{1}{12}\sum_{i=1}^{12} X_i$ y el estadístico pivote para el intervalo buscado es

$$\frac{\overline{X}-\mu_X}{S_c/\sqrt{12}} \rightsquigarrow t_{11}$$

   siendo S_c^2 la varianza muestral corregida, con $S_c^2 = \left(\frac{12}{11}\right)\left[\left(\frac{1}{12}\right)\sum_{i=1}^{12} X_i^2 - \overline{X}^2\right]$. Buscando en las tablas de $t-$Student obtenemos $t_{11;0.05} = 1.796$ y por tanto podemos asegurar que

$$p\left(-1.796 \leq \frac{\overline{X}-\mu_X}{S_c/\sqrt{12}} \leq 1.796\right) = 0.90$$

o equivalentemente

$$p\left(\overline{X}-1.796\left(\frac{S_c}{\sqrt{12}}\right)\leq \mu_X\leq \overline{X}+1.796\left(\frac{S_c}{\sqrt{12}}\right)\right)=0.90$$

Por tanto el intervalo de confianza buscado es $\overline{X}\pm 1.796\left(\frac{S_c}{\sqrt{12}}\right)$. En nuestro caso tenemos que $\sum_{i=1}^{12} x_i = 108$, $\sum_{i=1}^{12} x_i^2 = 1426$ y por tanto obtenemos $\overline{x}=9$ y $s_c^2=\left(\frac{12}{11}\right)\left[\frac{1426}{12}-9^2\right]=41.27$. Evaluando el intervalo tendremos

$$9\pm 1.796\left(\sqrt{\frac{41.27}{12}}\right)=9\pm 3.34=(5.66,12.33)$$

Ahora consideramos la variable aleatoria Y=concentración de $CL50$ para el Diazinón, y su muestra aleatoria simple Y_1,Y_2,Y_3. Suponemos que $Y\rightsquigarrow N(\mu_Y,\sigma_Y)$ porque la muestra es muy pequeña y no podemos hacer un intervalo asintótico. Razonando del mismo modo que antes obtenemos una media muestral $\overline{y}=3.57$ y una desviación típica muestral corregida de $s_Y=3.67$. Además, en este caso necesitamos $t_{2;0.05}=2.920$. Entonces, evaluando el intervalo de confianza del 90 % para μ_Y tenemos:

$$3.57\pm 2.920\left(\frac{3.67}{\sqrt{3}}\right)=3.57\pm 6.19=(-2.62,9.76)$$

Para estimar la diferencia entre los promedios de $CL50$ para el DDT y el Diazinón utilizando un intervalo de confianza del 90 % supondremos que las dos variables aleatorias tienen una distribución normal con la misma varianza σ^2. En este caso sabemos que el estadístico pivote es

$$\frac{\left(\overline{X}-\overline{Y}\right)-\left(\mu_X-\mu_y\right)}{S_p\sqrt{\frac{1}{12}+\frac{1}{3}}}\rightsquigarrow t_{13}$$

siendo S_p la desviación típica corregida promedio calculada a partir de las varianzas muestrales corregidas de ambas muestras como

$$S_p=\sqrt{\frac{11S_X^2+2S_Y^2}{13}}$$

Buscando en la tablas de la distribución $t-$Student, tenemos que $t_{13;0.05}=1.771$, y por tanto podemos asegurar que

$$p\left(-1.771\leq\frac{\left(\overline{X}-\overline{Y}\right)-\left(\mu_X-\mu_y\right)}{S_p\sqrt{\frac{1}{12}+\frac{1}{3}}}\leq 1.771\right)=0.90$$

o, equivalentemente, que

$$p\left(\left(\overline{X}-\overline{Y}\right)-1.771S_p\sqrt{\frac{1}{12}+\frac{1}{3}}\leq\mu_X-\mu_y\leq\left(\overline{X}-\overline{Y}\right)+1.771S_p\sqrt{\frac{1}{12}+\frac{1}{3}}\right)=0.90$$

El intervalo de confianza buscado será entonces $\left(\overline{X}-\overline{Y}\right)\pm 1.771S_p\sqrt{\frac{1}{12}+\frac{1}{3}}$. Con los datos obtenidos ya sabemos que $\overline{x}=9$, $x_X=\sqrt{41.27}$, $\overline{y}=3.57$, $s_Y=3.67$ y por tanto $s_p=\sqrt{\frac{(11)(41.27)+2(13.46)}{13}}=6.08$. Por tanto el intervalo buscado es

$$5.43\pm(1.771)(6.08)\sqrt{\frac{1}{12}+\frac{1}{3}}=5.43\pm 6.96=(-1.52,12.39)$$

No podemos garantizar entonces, con una confianza del 90 %, que la concentración media de $CL50$ sea distinta para el DDT que para el Diazinón, porque $0 \in (-1.52, 12.39)$ y por tanto puede ser que $\mu_X - \mu_y = 0$.

10. Se registraron las áreas (en hectáreas) que ocupan los caimanes, según las estaciones, en un lago de las cercanías de Gainesville (Florida) por la Comisión de Caza y Pesca de Florida. Cinco caimanes registrados en la primavera mostraron áreas de 8.0, 12.1, 8.1, 18.2 y 31.7. Cuatro caimanes diferentes registrados en el verano mostraron áreas de 102.0, 81.7, 54.7 y 50.7. Estimar la diferencia en el promedio de las áreas ocupadas por los caimanes en la primavera y en el verano utilizando un intervalo de confianza del 95 %. ¿Qué supuestos deben establecerse?

    Solución. Consideramos las variables aleatorias X=superficie ocupada por caimanes en primavera (ha) e Y=superficie ocupara por caimanes en verano (ha). Como las muestras son muy pequeñas necesitamos suponer que ambas variables tienen distribución normal y con varianzas iguales, es decir, que $X \rightsquigarrow N(\mu_X, \sigma)$ e $Y \rightsquigarrow N(\mu_Y, \sigma)$. Sean X_1, X_2, X_3, X_4, X_5 e Y_1, Y_2, Y_3, Y_4 las respectivas muestras aleatorias simples, que consideramos independientes entre si. Los estimadores puntuales de μ_X y μ_Y son, respectivamente, $\widehat{\mu_X} = \overline{X} = \frac{1}{5}\sum_{i=1}^{5} X_i$ e $\widehat{\mu_Y} = \overline{Y} = \frac{1}{4}\sum_{i=1}^{4} Y_i$, y el estimador de la varianza común es

$$\widehat{\sigma^2} = S_p^2 = \frac{4S_X^2 + 3S_Y^2}{7}$$

    siendo $S_X^2 = \left(\frac{5}{4}\right)\left[\left(\frac{1}{5}\right)\sum_{i=1}^{5} X_i^2 - \overline{X}^2\right]$ y $S_Y^2 = \left(\frac{4}{3}\right)\left[\left(\frac{1}{4}\right)\sum_{i=1}^{4} Y_i^2 - \overline{Y}^2\right]$ las respectivas varianzas muestrales corregidas de ambas muestras. En estas circunstancias sabemos que el estadístico pivote para el intervalo de confianza buscado es

$$\frac{\left(\overline{X} - \overline{Y}\right) - \left(\mu_X - \mu_y\right)}{S_p\sqrt{\frac{1}{5} + \frac{1}{4}}} \rightsquigarrow t_7$$

    Además, como el grado de confianza deseado es 95 %, tenemos que $1 - \alpha = 0.95$, $\alpha/2 = 0.025$ y buscando en la tablas de la distribución $t-$Student obtenemos que $t_{7;0.025} = 2.365$. Por tanto podemos asegurar que

$$p\left(-2.365 \leq \frac{\left(\overline{X} - \overline{Y}\right) - \left(\mu_X - \mu_y\right)}{S_p\sqrt{\frac{1}{5} + \frac{1}{4}}} \leq 2.365\right) = 0.95$$

    o, equivalentemente, que

$$p\left(\left(\overline{X} - \overline{Y}\right) - 2.365 S_p\sqrt{\frac{1}{5} + \frac{1}{4}} \leq \mu_X - \mu_y \leq \left(\overline{X} - \overline{Y}\right) + 2.365 S_p\sqrt{\frac{1}{5} + \frac{1}{4}}\right) = 0.95$$

    Por tanto el intervalo de confianza buscado será

$$\left(\overline{X} - \overline{Y}\right) \pm 2.365 S_p\sqrt{\frac{1}{5} + \frac{1}{4}}$$

    Con los datos del problema tenemos que $\sum_{i=1}^{5} x_i = 78.1$, $\sum_{i=1}^{5} x_i^2 = 1612.5$, $\sum_{i=1}^{4} y_i = 289.1$ y $\sum_{i=1}^{4} y_i^2 = 22641.47$. Podemos evaluar entonces todos los estadísticos necesarios: $\overline{x} = 15.62$, $\overline{y} = 72.275$, $s_x^2 =$

$\left(\frac{5}{4}\right)\left[\frac{1612.5}{5} - 15.62^2\right] = 98.1445$ y $s_y^2 = \left(\frac{4}{3}\right)\left[\frac{22641.47}{4} - 72.275^2\right] = 582.2558$. Entonces, para s_p obtenemos

$$s_p = \sqrt{\frac{4\,(98.1445) + 3\,(582.2558)}{7}} = 17.48$$

y el intervalo de confianza buscado será

$$-56.655 \pm (2.365)\,(17.48)\sqrt{\frac{1}{5} + \frac{1}{4}} = -56.655 \pm 27.73 = (-84.38, -28.93)$$

Por tanto, podemos asegurar con un 95 % de confianza que $\mu_X - \mu_y < 0$ y la superficie media ocupada por los caimanes es mayor en verano que en primavera.

11. En el trabajo de un laboratorio es deseable verificar cuidadosamente la variabilidad de las lecturas obtenidas de muestras estándar. En un estudio de la concentración de calcio en agua potable como parte de la valoración de la calidad del agua se pasó el mismo patrón de medida seis veces por el laboratorio en intervalos aleatorios. Las lecturas en partes por millón fueron 9.54, 9.61, 9.32, 9.48, 9.70 y 9.26. Estimar la varianza de las lecturas y dar para ella un intervalo de confianza del 90 %.

Solución. Sea X=concentración de calcio en agua potable (ppm) y supongamos que $X \rightsquigarrow N(\mu, \sigma)$. Sea $X_1, X_2, ..., X_6$ la muestra aleatoria simple de X. Queremos calcular un intervalo de confianza para σ^2 con $1 - \alpha = 0.9$ y por tanto $\alpha/2 = 0.05$. El estimador puntual para σ^2 es

$$\widehat{\sigma^2} = S_c^2 = \left(\frac{6}{5}\right)\left(\frac{1}{6}\sum_{i=1}^{6} X_i^2 - \left(\frac{1}{6}\sum_{i=1}^{6} X_i\right)^2\right)$$

y el estadístico pivote para el intervalo de confianza es

$$\frac{5S_c^2}{\sigma^2} \rightsquigarrow \chi_5^2$$

Utilizando las tabla de la distribución χ^2-Pearson obtenemos que $\chi^2_{5;0.05} = 11.070$ y $\chi^2_{5;0.95} = 1.145$. Podemos asegurar entonces que:

$$p\left(1.145 \leq \frac{5S_c^2}{\sigma^2} \leq 11.070\right) = 0.90$$

o, equivalentemente,

$$p\left(\frac{5S_c^2}{11.070} \leq \sigma^2 \leq \frac{5S_c^2}{1.145}\right) = 0.90$$

Por tanto el intervalo de confianza buscado será

$$\left(\frac{5S_c^2}{11.070}, \frac{5S_c^2}{1.145}\right)$$

Con los datos obtenidos tenemos que $\sum_{i=1}^{6} x_i = 56.91$ y $\sum_{i=1}^{6} x_i^2 = 539.9341$. Por tanto el valor para la estimación puntual de la varianza de las lecturas será

$$s_c^2 = \left(\frac{6}{5}\right)\left(\frac{539.9341}{6} - \left(\frac{56.91}{6}\right)^2\right) = 0.02855$$

y el intervalo de confianza para la varianza de las lecturas será $(0.0129, 0.1247)$.

12. Se seleccionó una muestra aleatoria de 21 ingenieros de un grupo mayor que trabaja para una determinada empresa. La desviación estándar de la muestra de las horas de trabajo por semana fue de 7 horas. Determinar un intervalo de confianza del 95 % para la varianza de la población suponiendo que las medidas siguen una distribución normal.

    Solución. Sea X=horas de trabajo por semana con $X \rightsquigarrow N(\mu, \sigma)$. Buscamos un intervalo de confianza para σ^2 con $1 - \alpha = 0.95$ y por tanto $\alpha/2 = 0.025$. El estadístico pivote para este intervalo de confianza es

$$\frac{20S_c^2}{\sigma^2} \rightsquigarrow \chi^2_{20}$$

    Utilizando las tabla de la distribución χ^2-Pearson obtenemos que $\chi^2_{20;0.025} = 34.17$ y $\chi^2_{20;0.975} = 9.591$. Podemos asegurar entonces que:

$$p\left(9.591 \leq \frac{20S_c^2}{\sigma^2} \leq 34.17\right) = 0.95$$

    o, equivalentemente,

$$p\left(\frac{20S_c^2}{34.17} \leq \sigma^2 \leq \frac{20S_c^2}{9.591}\right) = 0.95$$

    Por tanto el intervalo de confianza buscado será

$$\left(\frac{20S_c^2}{34.17}, \frac{20S_c^2}{9.591}\right)$$

    Como el problema nos dice que la desviación estándar (entendemos que corregida) de la muestra es 7 tenemos que $s_c^2 = 49$ y el intervalo de confianza para la varianza de las lecturas será

$$\left(\frac{(20)(49)}{34.17}, \frac{(20)(49)}{9.591}\right) = (28.68, 102.18)$$

13. El crecimiento del tronco principal para una muestra de 17 pinos rojos de 4 años tiene una media de 11.3 pulgadas y una desviación estándar de 3.4 pulgadas. Obtener un intervalo de confianza del 90 % para la media del crecimiento del tronco principal para una población de pinos rojos de 4 años sujeta a condiciones ambientales similares. Supóngase que el crecimiento tiene un distribución normal.

    Solución. Sea la variable aleatoria X=crecimiento del tronco principal de un pino rojo de 4 años (pulgadas), con $X \rightsquigarrow N(\mu, \sigma)$, y $X_1, X_2, ..., X_{17}$ la muestra aleatoria simple de X. Tenemos que estimar el crecimiento medio, es decir μ, utilizando un intervalo de confianza con $1 - \alpha = 0.9$ y por tanto $\alpha/2 = 0.05$. Sabemos que el estimador puntual es $\widehat{\mu_X} = \overline{X} = \frac{1}{17}\sum_{i=1}^{17} X_i$ y el estadístico pivote para el intervalo buscado es

$$\frac{\overline{X} - \mu_X}{S_c/\sqrt{17}} \rightsquigarrow t_{16}$$

    siendo S_c^2 la varianza muestral corregida, con $S_c^2 = \left(\frac{17}{16}\right)\left[\left(\frac{1}{17}\right)\sum_{i=1}^{17} X_i^2 - \overline{X}^2\right]$. Buscando en las tablas de $t-$Student obtenemos $t_{16;0.05} = 1.746$ y por tanto podemos asegurar que

$$p\left(-1.746 \leq \frac{\overline{X} - \mu_X}{S_c/\sqrt{17}} \leq 1.746\right) = 0.90$$

o equivalentemente

$$p\left(\overline{X}-1.746\left(\frac{S_c}{\sqrt{17}}\right)\leq\mu_X\leq\overline{X}+1.746\left(\frac{S_c}{\sqrt{17}}\right)\right)=0.90$$

Por tanto el intervalo de confianza buscado es $\overline{X}\pm1.746\left(\frac{S_c}{\sqrt{17}}\right)$. En nuestro caso el problema nos dice que $\overline{x}=11.3$ y $s_c=3.4$ (entendemos que la desviación estándar que nos da es ya corregida). Evaluando el intervalo tendremos

$$11.3\pm1.746\left(\frac{3.4}{\sqrt{17}}\right)=11.3\pm1.44=(9.86,12.74)$$

Podemos asegurar entonces que el crecimiento medio del tronco principal de los pinos está entre 9.68 y 12.74 pulgadas.

14. Se aplicaron dos métodos para enseñar la lectura a dos grupos de niños de la escuela primaria, comparándose los resultados mediante una prueba de lectura y comprensión al final del periodo de aprendizaje. Obtener un intervalo de confianza del 95 % para la diferencia de promedios. ¿Qué supuestos hay que hacer?

	Método I	Método II
Tamaño muestral	11	14
Media muestral	64	69
Varianza corregida	52	71

Solución. Consideramos las variables aleatorias X=puntuación en la prueba con método I e Y=puntuación en la prueba con método II. Como los tamaños muestrales no son grandes necesitamos suponer que ambas variables tienen distribución normal y con varianzas iguales, es decir, que $X\rightsquigarrow N(\mu_X,\sigma)$ e $Y\rightsquigarrow N(\mu_Y,\sigma)$. Sean $X_1,X_2,...,X_{11}$ e $Y_1,Y_2,...,Y_{14}$ las respectivas muestras aleatorias simples, que consideramos independientes entre si. Los estimadores puntuales de μ_X y μ_Y son, respectivamente, $\widehat{\mu_X}=\overline{X}=\frac{1}{11}\sum_{i=1}^{11}X_i$ e $\widehat{\mu_Y}=\overline{Y}=\frac{1}{14}\sum_{i=1}^{14}Y_i$, y el estimador de la varianza común es

$$\widehat{\sigma^2}=S_p^2=\frac{10S_X^2+13S_Y^2}{23}$$

siendo $S_X^2=\left(\frac{11}{10}\right)\left[\left(\frac{1}{11}\right)\sum_{i=1}^{11}X_i^2-\overline{X}^2\right]$ y $S_Y^2=\left(\frac{14}{13}\right)\left[\left(\frac{1}{14}\right)\sum_{i=1}^{14}Y_i^2-\overline{Y}^2\right]$ las respectivas varianzas muestrales corregidas de ambas muestras. En estas circunstancias sabemos que el estadístico pivote para el intervalo de confianza buscado es

$$\frac{\left(\overline{X}-\overline{Y}\right)-\left(\mu_X-\mu_y\right)}{S_p\sqrt{\frac{1}{11}+\frac{1}{14}}}\rightsquigarrow t_{23}$$

Además, como el grado de confianza deseado es 95 %, tenemos que $1-\alpha=0.95$, $\alpha/2=0.025$ y buscando en la tablas de la distribución $t-$Student obtenemos que $t_{23;0.025}=2.069$. Por tanto podemos asegurar que

$$p\left(-2.069\leq\frac{\left(\overline{X}-\overline{Y}\right)-\left(\mu_X-\mu_y\right)}{S_p\sqrt{\frac{1}{11}+\frac{1}{14}}}\leq2.069\right)=0.95$$

o, equivalentemente, que

$$p\left(\left(\overline{X}-\overline{Y}\right)-2.069_p\sqrt{\frac{1}{11}+\frac{1}{14}}\leq \mu_X-\mu_y\leq \left(\overline{X}-\overline{Y}\right)+2.069S_p\sqrt{\frac{1}{11}+\frac{1}{14}}\right)=0.95$$

Por tanto el intervalo de confianza buscado será

$$\left(\overline{X}-\overline{Y}\right)\pm 2.069S_p\sqrt{\frac{1}{11}+\frac{1}{14}}$$

Con los datos del problema tenemos que $\overline{x}=64$, $\overline{y}=69$, $s_x^2=52$ y $s_y^2=71$. Entonces, para s_p obtenemos

$$s_p=\sqrt{\frac{10\,(52)+13\,(71)}{23}}=7.92$$

y el intervalo de confianza buscado será

$$-5\pm(2.069)\,(7.92)\sqrt{\frac{1}{11}+\frac{1}{14}}=-5\pm 6.61=(-11.60,1.61)$$

Por tanto, no podemos asegurar con un 95 % de confianza que $\mu_X-\mu_y\neq 0$ y no hay diferencias significativas entre los dos métodos de aprendizaje.

15. Una comparación de los tiempos de reacción a dos estímulos diferentes en un experimento con 16 animales produjo los resultados (en segundos) dados por la siguiente tabla:

Estímulo 1:	1	3	2	1	2	1	3	3
Estímulo 2:	4	3	3	3	1	2	3	3

Obtener un intervalo de confianza del 90 % para la diferencia de medias de los tiempos de reacción. ¿Qué supuestos hay que hacer?

Solución. Consideramos las variables aleatorias X=tiempo de reacción al estímulo 1 e Y=tiempo de reacción al estímulo 2. Como los tamaños muestrales no son suficientemente grandes para hacer un intervalo asintótico necesitamos suponer que ambas variables tienen distribución normal y con varianzas iguales, es decir, que $X\rightsquigarrow N(\mu_X,\sigma)$ e $Y\rightsquigarrow N(\mu_Y,\sigma)$. Sean $X_1,X_2,...,X_8$ e $Y_1,Y_2,...,Y_8$ las respectivas muestras aleatorias simples, que consideramos independientes entre si. Los estimadores puntuales de μ_X y μ_Y son, respectivamente, $\widehat{\mu_X}=\overline{X}=\frac{1}{8}\sum\limits_{i=1}^{8}X_i$ e $\widehat{\mu_Y}=\overline{Y}=\frac{1}{8}\sum\limits_{i=1}^{8}Y_i$, y el estimador de la varianza común es

$$\widehat{\sigma^2}=S_p^2=\frac{7S_X^2+7S_Y^2}{14}=\frac{S_X^2+S_Y^2}{2}$$

siendo $S_X^2=\left(\frac{8}{7}\right)\left[\left(\frac{1}{8}\right)\sum\limits_{i=1}^{8}X_i^2-\overline{X}^2\right]$ y $S_Y^2=\left(\frac{8}{7}\right)\left[\left(\frac{1}{8}\right)\sum\limits_{i=1}^{8}Y_i^2-\overline{Y}^2\right]$ las respectivas varianzas muestrales corregidas de ambas muestras. En estas circunstancias sabemos que el estadístico pivote para el intervalo de confianza buscado es

$$\frac{\left(\overline{X}-\overline{Y}\right)-\left(\mu_X-\mu_y\right)}{S_p\sqrt{\frac{1}{8}+\frac{1}{8}}}\rightsquigarrow t_{14}$$

Además, como el grado de confianza deseado es 90 %, tenemos que $1-\alpha=0.90$, $\alpha/2=0.05$ y buscando en la tablas de la distribución $t-$Student obtenemos que $t_{14;0.05}=1.761$. Por tanto podemos asegurar que

$$p\left(-1.761 \leq \frac{\left(\overline{X}-\overline{Y}\right)-\left(\mu_X-\mu_y\right)}{S_p\sqrt{\frac{1}{4}}} \leq 1.761\right) = 0.95$$

o, equivalentemente, que

$$p\left(\left(\overline{X}-\overline{Y}\right)-1.761\left(\frac{S_p}{2}\right) \leq \mu_X-\mu_y \leq \left(\overline{X}-\overline{Y}\right)+1.761\left(\frac{S_p}{2}\right)\right) = 0.95$$

Por tanto el intervalo de confianza buscado será

$$\left(\overline{X}-\overline{Y}\right) \pm 0.8805 S_p$$

Con los datos del problema tenemos que $\sum_{i=1}^{8} x_i = 16$, $\sum_{i=1}^{8} x_i^2 = 38$, $\sum_{i=1}^{8} y_i = 22$ y $\sum_{i=1}^{8} y_i^2 = 66$. Podemos evaluar entonces todos los estadísticos necesarios: $\overline{x} = 2$, $\overline{y} = 2.75$, $s_x^2 = \left(\frac{8}{7}\right)\left[\frac{38}{8} - 2^2\right] = 0.8571$ y $s_y^2 = \left(\frac{8}{7}\right)\left[\frac{66}{8} - 2.75^2\right] = 0.7857$. Entonces, para s_p obtenemos

$$s_p = \sqrt{\frac{0.8571+0.7857}{2}} = 0.9063$$

y el intervalo de confianza buscado será

$$-0.75 \pm (0.8805)(0.9063) = -5 \pm 0.7980 = (-5.79, -4.21)$$

Por tanto, podemos asegurar con un 90 % de confianza que $\mu_X - \mu_y < 0$ y el tiempo medio de reacción al estímulo 2 es mayor que para el estímulo 1.

16. El tiempo transcurrido entre la facturación y el pago recibido se registró para una muestra aleatoria de 100 clientes de una empresa de contadores públicos. La media y la desviación estándar de las 100 cuentas fueron respectivamente 39.1 días y 17.3 días. Obtener un intervalo de confianza del 90 % para el tiempo medio que transcurre entre la facturación y el pago recibido para todas las cuentas de la firma de contadores públicos. Interpretar el resultado.

    Solución. Sea X=tiempo transcurrido entre la facturación y el pago (días) y sean μ y σ respectivamente el valor esperado y la desviación típica de X. Como no conocemos la distribución de probabilidad de X pero tenemos una muestra suficientemente grande vamos a calcular un intervalo asintótico para el tiempo medio μ. Sabemos que el estimador puntual es $\widehat{\mu} = \overline{X}$ y el estadístico pivote es

$$\frac{\overline{X}-\mu_X}{S_c/\sqrt{100}} \widetilde{\rightsquigarrow} N(0,1)$$

siendo S_c^2 la varianza muestral corregida, con $S_c^2 = \left(\frac{100}{99}\right)\left[\left(\frac{1}{100}\right)\sum_{i=1}^{100} X_i^2 - \overline{X}^2\right]$. Como $1-\alpha = 0.9$ y $\alpha/2 = 0.05$, buscando en las tablas de la distribución normal obtenemos $z_{0.05} = 1.645$ y por tanto podemos asegurar que

$$p\left(-1.645 \leq \frac{\overline{X}-\mu_X}{S_c/\sqrt{100}} \leq 1.645\right) \simeq 0.90$$

o equivalentemente

$$p\left(\overline{X}-1.645\left(\frac{S_c}{\sqrt{100}}\right)\leq\mu_X\leq\overline{X}+1.645\left(\frac{S_c}{\sqrt{100}}\right)\right)\simeq 0.90$$

Por tanto el intervalo de confianza asintótico buscado es $\overline{X}\pm 0.1645S_c$. En nuestro caso el problema nos dice que $\overline{x}=39.1$ y $s_c=17.3$ (entendemos que la desviación estándar que nos da es ya corregida). Evaluando el intervalo tendremos

$$39.1\pm(0.1645)\,(17.3)=39.1\pm 2.85=(36.25,41.95)$$

Podemos asegurar entonces con una confianza del 90 % que el tiempo medio que transcurre entre la facturación y el pago se encuentra entre 36 y 42 días.

17. Se supone que los diámetros normales de los pies de una determinada masa forestal siguen una distribución normal. Se eligen 10 pies al azar obteniéndose las siguientes medidas de sus diámetros:

25.5 26.8 24.2 25.0 27.3 26.1 23.2 28.4 27.8 25.7 ($\sum x_i=260$ y $\sum x_i^2=6783.76$)

a) Obtener las estimaciones puntuales para la media y la varianza del diámetro de la masa.

b) Determinar un intervalo de confianza al 90 % para el diámetro medio de la masa.

c) Determinar un intervalo de confianza al 90 % para la varianza y la desviación típica de la población.

Solución. Sea la variable aleatoria X=diámetro normal de un pino (cm), con $X\rightsquigarrow N(\mu,\sigma)$, y $X_1,X_2,...,X_{10}$ la muestra aleatoria simple de X.

a) El estimador puntual del diámetro medio de los pinos μ será $\widehat{\mu_X}=\overline{X}=\frac{1}{10}\sum\limits_{i=1}^{10}X_i$ y el estimador puntual de la varianza del diámetro de la masa será $\widehat{\sigma^2}=S_c^2=\left(\frac{10}{9}\right)\left[\left(\frac{1}{10}\right)\sum\limits_{i=1}^{10}X_i^2-\overline{X}^2\right]$. Con los datos obtenidos en la muestra dada tenemos que

$$\widehat{\mu_X}=\overline{x}=\frac{1}{10}\sum_{i=1}^{10}x_i=\frac{260}{10}=26\ cm$$

y

$$\widehat{\sigma^2}=s_c^2=\left(\frac{10}{9}\right)\left[\left(\frac{1}{10}\right)\sum_{i=1}^{10}x_i^2-\overline{x}^2\right]=\left(\frac{10}{9}\right)\left(\frac{6783.76}{10}-26^2\right)=2.64\ cm^2$$

b) Queremos calcular un intervalo de confianza con $1-\alpha=0.9$ y por tanto $\alpha/2=0.05$. Sabemos que el estimador puntual es $\widehat{\mu_X}=\overline{X}=\frac{1}{17}\sum\limits_{i=1}^{17}X_i$ y el estadístico pivote para el intervalo buscado es

$$\frac{\overline{X}-\mu_X}{S_c/\sqrt{10}}\rightsquigarrow t_9$$

siendo S_c^2 la varianza muestral corregida, con $S_c^2=\left(\frac{10}{9}\right)\left[\left(\frac{1}{10}\right)\sum\limits_{i=1}^{10}X_i^2-\overline{X}^2\right]$. Buscando en las tablas de $t-$Student obtenemos $t_{9;0.05}=1.833$ y por tanto podemos asegurar que

$$p\left(-1.833\leq\frac{\overline{X}-\mu_X}{S_c/\sqrt{10}}\leq 1.833\right)=0.90$$

o equivalentemente

$$p\left(\overline{X} - 1.833\left(\frac{S_c}{\sqrt{10}}\right) \leq \mu_X \leq \overline{X} + 1.833\left(\frac{S_c}{\sqrt{10}}\right)\right) = 0.90$$

Por tanto el intervalo de confianza buscado es $\overline{X} \pm 1.833\left(\frac{S_c}{\sqrt{10}}\right)$. En nuestro caso, como hemos obtenido en el apartado anterior, tenemos que $\overline{x} = 26$ y $s_c = \sqrt{2.64}$. Entonces, evaluando el intervalo tendremos

$$26 \pm 1.833\left(\frac{\sqrt{2.64}}{\sqrt{10}}\right) = 26 \pm 0.95 = (25.0, 27.0)$$

Por tanto el diámetro medio de la masa, con un 90 % de confianza, está entre 25 y 27 centímetros.

c) Calculamos primero el intervalo de confianza del 90 % para la varianza. Como ya hemos dicho en el primer apartado el estimador puntual de la varianza es $\widehat{\sigma^2} = S_c^2 = \left(\frac{10}{9}\right)\left[\left(\frac{1}{10}\right)\sum_{i=1}^{10} X_i^2 - \overline{X}^2\right]$. Además, el estadístico pivote para el intervalo buscado es

$$\frac{9S_c^2}{\sigma^2} \rightsquigarrow \chi_9^2$$

Utilizando la tabla de la distribución χ^2-Pearson obtenemos que $\chi^2_{9;0.05} = 16.919$ y $\chi^2_{9;0.95} = 3.325$. Podemos asegurar entonces que:

$$p\left(3.325 \leq \frac{9S_c^2}{\sigma^2} \leq 16.919\right) = 0.90$$

o, equivalentemente,

$$p\left(\frac{9S_c^2}{16.919} \leq \sigma^2 \leq \frac{9S_c^2}{3.325}\right) = 0.90$$

Por tanto el intervalo de confianza buscado será

$$\left(\frac{9S_c^2}{16.919}, \frac{9S_c^2}{3.325}\right)$$

Con la muestra dada por el problema tenemos $s_c^2 = 2.64$ y el intervalo de confianza para la varianza de los diámetros será

$$\left(\frac{(9)(2.64)}{16.919}, \frac{(9)(2.64)}{3.325}\right) = (1.40, 7.15)$$

El intervalo de confianza para la desviación típica de los diámetros se obtendrá con las raíces cuadradas del intervalo de confianza para la varianza, es decir,

$$\left(\sqrt{1.40}, \sqrt{7.15}\right) = (1.1, 2.7)$$

Por tanto la desviación típica de los diámetros, con un 90 % de confianza, está entre 1.1 y 2.7 centímetros.

18. En un estudio sobre el crecimiento de una determinada especie vegetal se plantaron 10 semillas, 5 en una solución de nutrientes estándar y otras 5 en una solución de nutrientes que contenía una

cantidad extra de nitrógeno. Después de 22 días, las plantas fueron recogidas y pesadas en seco, obteniéndose los siguientes resultados:

	Estándar	Nitrógeno extra
Tamaño muestral	5	5
Media muestral	3.62	4.17
Desviación típica corregida	0.54	0.67

a) Construir un intervalo de confianza del 90 % para la razón de varianzas entre ambas poblaciones. ¿A qué conclusión podemos llegar sobre las varianzas de estas poblaciones? ¿Por qué?

b) A partir de la conclusión obtenida en el apartado (a), construir un intervalo de confianza para la diferencia de medias entre las dos poblaciones a un nivel de confianza del 95 %. ¿Podemos asegurar que los tratamientos producen diferentes rendimientos en la producción? ¿Por qué?

(En ambos apartados suponer que las dos poblaciones siguen una distribución normal)

Solución.

a) Consideramos las variables aleatorias X=crecimiento de una planta con la solución estándar e Y=crecimiento de una planta con solución extra de nitrógeno. Como nos dice el problema, suponemos que ambas variables tienen distribución normal, es decir, que $X \rightsquigarrow N(\mu_X, \sigma_X)$ e $Y \rightsquigarrow N(\mu_Y, \sigma_Y)$. Sean $X_1, X_2, ..., X_5$ e $Y_1, Y_2, ..., Y_5$ las respectivas muestras aleatorias simples, que consideramos independientes entre si. El estimador puntual para la razón de varianzas es $\widehat{\left(\frac{\sigma_X^2}{\sigma_Y^2}\right)} = \frac{S_X^2}{S_Y^2}$ siendo $S_X^2 = \left(\frac{5}{4}\right)\left[\left(\frac{1}{5}\right)\sum_{i=1}^{5} X_i^2 - \overline{X}^2\right]$ y $S_Y^2 = \left(\frac{5}{4}\right)\left[\left(\frac{1}{5}\right)\sum_{i=1}^{5} Y_i^2 - \overline{Y}^2\right]$ las respectivas varianzas muestrales corregidas de ambas muestras. Además el estadístico pivote para el intervalo de confianza de la razón de varianzas es

$$\frac{S_X^2/S_Y^2}{\sigma_X^2/\sigma_Y^2} \rightsquigarrow F_{4,4}$$

Como queremos utilizar una confianza de 90 % tenemos que $1-\alpha = 0.9$, $\alpha/2 = 0.05$ y necesitamos encontrar los valores $F_{4,4;0.05}$ y $F_{4,4;0.95}$. El primero de ellos podemos encontrarlo en la tabla de la distribución F-Fisher con $\alpha = 0.05$ y tenemos que $F_{4,4;0.05} = 6.3882$. Para el segundo sabemos que $F_{4,4;0.95} = 1/F_{4,4;0.05}$ y por tanto $F_{4,4;0.95} = 1/6.3882$. Entonces podemos asegurar que

$$p\left(\frac{1}{6.3882} \leq \frac{S_X^2/S_Y^2}{\sigma_X^2/\sigma_Y^2} \leq 6.3882\right) = 0.90$$

o equivalentemente

$$p\left(\frac{S_X^2/S_Y^2}{6.3882} \leq \sigma_X^2/\sigma_Y^2 \leq (6.3882)\left(S_X^2/S_Y^2\right)\right) = 0.90$$

Entonces el intervalo de confianza buscado es

$$\left(\frac{S_X^2/S_Y^2}{6.3882}, (6.3882)\left(S_X^2/S_Y^2\right)\right)$$

Con los datos del problema, en nuestra muestra hemos obtenido $s_X^2 = 0.54^2$ y $s_Y^2 = 0.67^2$, y sustituyendo tendremos

$$\left(\frac{0.54^2/0.67^2}{6.3882}, (6.3882)\left(0.54^2/0.67^2\right)\right) = (0.10, 4.15)$$

Por tanto no podemos decir, con una confianza de 90 % que el cociente σ_X^2/σ_Y^2 sea distinto 1, porque $1 \in (0.10, 4.15)$. Es decir no podemos decir que las varianzas sean distintas. Supondremos entonces que son iguales para el apartado siguiente.

b) Ahora queremos calcular un intervalo de confianza del 95 % para la diferencia de medias $\mu_X - \mu_Y$. Sabemos que el estimador puntual para esta diferencia es $\widehat{\mu_X - \mu_Y} = \overline{X} - \overline{Y}$ y, como hemos supuesto que las varianzas son iguales, el estadístico pivote para el intervalo buscado es

$$\frac{\left(\overline{X} - \overline{Y}\right) - \left(\mu_X - \mu_y\right)}{S_p\sqrt{\frac{1}{5} + \frac{1}{5}}} \rightsquigarrow t_8$$

siendo

$$S_p^2 = \frac{4S_X^2 + 4S_Y^2}{8}$$

el estimador de la varianza común y $S_X^2 = \left(\frac{5}{4}\right)\left[\left(\frac{1}{5}\right)\sum_{i=1}^{5} X_i^2 - \overline{X}^2\right]$, $S_Y^2 = \left(\frac{5}{4}\right)\left[\left(\frac{1}{5}\right)\sum_{i=1}^{5} Y_i^2 - \overline{Y}^2\right]$ las respectivas varianzas muestrales corregidas de ambas muestras. Además, como el grado de confianza deseado es 95 %, tenemos que $1 - \alpha = 0.95$, $\alpha/2 = 0.025$, y buscando en la tablas de la distribución $t-$Student obtenemos que $t_{8;0.025} = 2.308$. Por tanto podemos asegurar que

$$p\left(-2.308 \leq \frac{\left(\overline{X} - \overline{Y}\right) - \left(\mu_X - \mu_y\right)}{S_p\sqrt{\frac{1}{5} + \frac{1}{5}}} \leq 2.308\right) = 0.95$$

o, equivalentemente, que

$$p\left(\left(\overline{X} - \overline{Y}\right) - 2.308\left(S_p\sqrt{\frac{2}{5}}\right) \leq \mu_X - \mu_y \leq \left(\overline{X} - \overline{Y}\right) + 2.308\left(S_p\sqrt{\frac{2}{5}}\right)\right) = 0.95$$

Por tanto el intervalo de confianza buscado será

$$\left(\overline{X} - \overline{Y}\right) \pm 2.308\left(S_p\sqrt{\frac{2}{5}}\right)$$

Con los datos del problema tenemos que $\overline{x} = 3.62$, $\overline{y} = 4.17$, $s_X^2 = 0.54^2$ y $s_Y^2 = 0.67^2$. Entonces, para s_p obtenemos

$$s_p = \sqrt{\frac{4\left(0.54^2\right) + 4\left(0.67^2\right)}{8}} = \sqrt{\frac{0.54^2 + 0.67^2}{2}} = 0.6085$$

y el intervalo de confianza buscado será

$$-0.55 \pm (2.308)\left(0.6085\sqrt{\frac{2}{5}}\right) = -0.55 \pm 0.8882 = (-1.43, 0.34)$$

No podemos entonces asegurar, con un 95 % de confianza, que $\mu_X - \mu_y \neq 0$, y por tanto no podemos asegurar que los tratamientos producen diferentes rendimientos en la producción.

19. Se desea comparar la altura de dos poblaciones de *Pinus sylvestris* ubicadas en dos localizaciones de características a priori diferentes. Se midieron, en cada localización, las alturas totales de 51 ejemplares. El resumen de los resultados aparece en la siguiente tabla:

	Localidad 1	Localidad 2
Tamaño muestral	51	51
Media muestral	10.576	8.259
Varianza corregida	6.952	8.479

Considerando que ambas poblaciones siguen una distribución normal:

a) Construir un intervalo de confianza del 95 % para la razón de varianzas entre ambas poblaciones. ¿A qué conclusión podemos llegar sobre las varianzas de estas poblaciones? ¿Por qué?

b) A partir de la conclusión obtenida en el apartado anterior, construir un intervalo de confianza del 95 % para la diferencia de medias entre las dos poblaciones. ¿A qué conclusiones podemos llegar sobre la altura de las dos poblaciones? ¿Por qué?

c) Además se observó que de los 51 pinos observados en la primera localización 16 estaban atacados por la procesionaria, mientras que en los 51 de la segunda población tan sólo 8 lo estaban. ¿Podemos concluir con una confianza del 99 % que las proporciones de pinos atacados en ambas poblaciones es diferente? ¿Y con una confianza del 90 %?

Solución. Consideramos las variables aleatorias X=altura de *Pinus sylvestris* en la localidad 1 e Y=altura de *Pinus sylvestris* en la localidad 2. Como nos dice el problema, suponemos que ambas variables tienen distribución normal, es decir, que $X \rightsquigarrow N(\mu_X, \sigma_X)$ e $Y \rightsquigarrow N(\mu_Y, \sigma_Y)$. Sean $X_1, X_2, ..., X_{51}$ e $Y_1, Y_2, ..., Y_{51}$ las respectivas muestras aleatorias simples, que consideramos independientes entre si.

a) El estimador puntual para la razón de varianzas es $\widehat{\left(\frac{\sigma_X^2}{\sigma_Y^2}\right)} = \frac{S_X^2}{S_Y^2}$ siendo $S_X^2 = \left(\frac{51}{50}\right)\left[\left(\frac{1}{51}\right)\sum_{i=1}^{51} X_i^2 - \overline{X}^2\right]$ y $S_Y^2 = \left(\frac{51}{50}\right)\left[\left(\frac{1}{51}\right)\sum_{i=1}^{51} Y_i^2 - \overline{Y}^2\right]$ las respectivas varianzas muestrales corregidas de ambas muestras. Además el estadístico pivote para el intervalo de confianza de la razón de varianzas es

$$\frac{S_X^2/S_Y^2}{\sigma_X^2/\sigma_Y^2} \rightsquigarrow F_{50,50}$$

Como queremos utilizar una confianza de 95 % tenemos que $1-\alpha = 0.95$, $\alpha/2 = 0.025$ y necesitamos encontrar los valores $F_{50,50;0.025}$ y $F_{50,50;0.975}$. Como no disponemos de tablas de la distribución F-Fisher con $\alpha = 0.025$ calculamos esos valores utilizando el programa Statgraphics: $F_{50,50;0.025} = 1.7520$ y $F_{50,50;0.975} = 0.5708$. Entonces podemos asegurar que

$$p\left(0.5708 \leq \frac{S_X^2/S_Y^2}{\sigma_X^2/\sigma_Y^2} \leq 1.7520\right) = 0.95$$

o equivalentemente

$$p\left(\frac{S_X^2/S_Y^2}{1.7520} \leq \sigma_X^2/\sigma_Y^2 \leq \frac{S_X^2/S_Y^2}{0.5708}\right) = 0.95$$

Entonces el intervalo de confianza buscado es

$$\left(\frac{S_X^2/S_Y^2}{1.7520}, \frac{S_X^2/S_Y^2}{0.5708}\right)$$

Con los datos del problema, en nuestra muestra hemos obtenido $s_X^2 = 6.952$ y $s_Y^2 = 8.479$, y sustituyendo tendremos

$$\left(\frac{6.952/8.479}{1.7520}, \frac{6.952/8.479}{0.5708}\right) = (0.46, 1.44)$$

Por tanto no podemos decir, con una confianza de 95 % que el cociente σ_X^2/σ_Y^2 sea distinto de 1, porque $1 \in (0.46, 1.44)$. Es decir no podemos decir que las varianzas sean distintas. Supondremos entonces que son iguales para el apartado siguiente.

b) Ahora queremos calcular un intervalo de confianza del 95 % para la diferencia de medias $\mu_X - \mu_Y$. Sabemos que el estimador puntual para esta diferencia es $\widehat{\mu_X - \mu_Y} = \overline{X} - \overline{Y}$ y, como hemos supuesto que las varianzas son iguales, el estadístico pivote para el intervalo buscado es

$$\frac{\left(\overline{X} - \overline{Y}\right) - \left(\mu_X - \mu_y\right)}{S_p\sqrt{\frac{1}{51} + \frac{1}{51}}} \rightsquigarrow t_{100}$$

siendo

$$S_p^2 = \frac{50S_X^2 + 50S_Y^2}{100} = \frac{S_X^2 + S_Y^2}{2}$$

el estimador de la varianza común y $S_X^2 = \left(\frac{51}{50}\right)\left[\left(\frac{1}{51}\right)\sum_{i=1}^{51} X_i^2 - \overline{X}^2\right]$, $S_Y^2 = \left(\frac{51}{50}\right)\left[\left(\frac{1}{51}\right)\sum_{i=1}^{51} Y_i^2 - \overline{Y}^2\right]$ las respectivas varianzas muestrales corregidas de ambas muestras. Además, como el grado de confianza deseado es 95 %, tenemos que $1-\alpha = 0.95$, $\alpha/2 = 0.025$, y buscando en la tablas de la distribución $t-$Student obtenemos que $t_{100;0.025} = 1.984$. Por tanto podemos asegurar que

$$p\left(-1.984 \leq \frac{\left(\overline{X} - \overline{Y}\right) - \left(\mu_X - \mu_y\right)}{S_p\sqrt{\frac{1}{51} + \frac{1}{51}}} \leq 1.984\right) = 0.95$$

o, equivalentemente, que

$$p\left(\left(\overline{X} - \overline{Y}\right) - 1.984\left(S_p\sqrt{\frac{2}{51}}\right) \leq \mu_X - \mu_y \leq \left(\overline{X} - \overline{Y}\right) + 1.984\left(S_p\sqrt{\frac{2}{51}}\right)\right) = 0.95$$

Por tanto el intervalo de confianza buscado será

$$\left(\overline{X} - \overline{Y}\right) \pm 1.984\left(S_p\sqrt{\frac{2}{51}}\right)$$

Con los datos del problema tenemos que $\overline{x} = 10.576$ $\overline{y} = 8.259$, $s_X^2 = 6.952$ y $s_Y^2 = 8.479$. Entonces, para s_p obtenemos

$$s_p = \sqrt{\frac{6.952 + 8.479}{2}} = 2.7777$$

y el intervalo de confianza buscado será

$$2.317 \pm (1.984)\left(2.7777\sqrt{\frac{2}{51}}\right) = 2.317 \pm 1.0914 = (1.22, 3.41)$$

Podemos entonces asegurar, con un 95 % de confianza, que $\mu_X - \mu_y > 0$, y por tanto podemos asegurar que la altura es mayor en la primera localidad que en la segunda.

c) Sea ahora $X_1, X_2, ..., X_{51}$ la muestra de pinos de la primera localidad, con $X_i \rightsquigarrow B(p_X)$, siendo p_X la verdadera proporción de pinos atacados por la procesionaria en la primera localidad. Del mismo modo sea $Y_1, Y_2, ..., Y_{51}$ la muestra de pinos de la segunda localidad, con $Y_i \rightsquigarrow B(p_Y)$, siendo p_Y la verdadera proporción de pinos atacados por la procesionaria en la segunda localidad. Como queremos un grado de confianza de 99 %, tenemos que $1-\alpha = 0.99$, $\alpha/2 = 0.005$ y $z_{0.005} = 2.58$. Vamos a utilizar un intervalo de confianza asintótico para la diferencia $p_X - p_Y$. El estadístico pivote en este caso será

$$\frac{(\widehat{p_X} - \widehat{p_Y}) - (p_X - p_Y)}{\sqrt{\frac{\widehat{p_X}(1-\widehat{p_X})}{51} + \frac{\widehat{p_Y}(1-\widehat{p_Y})}{51}}} \widetilde{\rightsquigarrow} N(0,1)$$

y, como $z_{0.025} = 2.58$, tenemos que

$$p\left(-2.58 \leq \frac{(\widehat{p_X} - \widehat{p_Y}) - (p_X - p_Y)}{\sqrt{\frac{\widehat{p_X}(1-\widehat{p_X})}{51} + \frac{\widehat{p_Y}(1-\widehat{p_Y})}{51}}} \leq 2.58\right) \simeq 0.99$$

o, equivalentemente

$$p\left((\widehat{p_X} - \widehat{p_Y}) - 2.58\sqrt{\frac{\widehat{p_X}(1-\widehat{p_X})}{51} + \frac{\widehat{p_Y}(1-\widehat{p_Y})}{51}} \leq (p_X - p_Y) \leq (\widehat{p_X} - \widehat{p_Y}) + 2.58\sqrt{\frac{\widehat{p_X}(1-\widehat{p_X})}{51} + \frac{\widehat{p_Y}(1-\widehat{p_Y})}{51}}\right) \simeq 0.99$$

Por tanto el intervalo de confianza asintótico buscado será

$$(\widehat{p_X} - \widehat{p_Y}) \pm 2.58\sqrt{\frac{\widehat{p_X}(1-\widehat{p_X})}{51} + \frac{\widehat{p_Y}(1-\widehat{p_Y})}{51}}$$

Como en nuestro caso hemos obtenido $\widehat{p_X} = 16/51 =$ y $\widehat{p_Y} = 8/51$ el intervalo será

$$8/51 \pm 2.58\sqrt{\frac{(16/51)(35/51)}{51} + \frac{(8/51)(43/51)}{51}} = 0.1569 \pm (2.58)\,(0.083) = 0.1569 \pm 0.2130 = (-0.06, 0.37)$$

Observar que $0 \in (-0.05, 0.37)$ y por tanto no podemos afirmar, con una confianza del 99 %, que la proporción de pinos atacados por la procesionaria sea distinta en las dos localidades.

Si utilizásemos una confianza del 90 % todo el razonamiento sería el mismo pero utilizando $z_{0.05} = 1.645$ y el intervalo sería

$$8/51 \pm 1.645\sqrt{\frac{(16/51)(35/51)}{51} + \frac{(8/51)(43/51)}{51}} = 0.1569 \pm (1.645)\,(0.083) = 0.1569 \pm 0.1366 = (0.02, 0.30)$$

Por tanto ahora podríamos asegurar, con un 90 % de confianza, que $p_X - p_Y > 0$, es decir, que la proporción de pinos atacados por la procesionaria es mayor en la primera localidad que en la segunda.

20. En un estudio sobre la evolución de una plantación de *Pinus pinaster* fueron medidas las alturas del fuste de 20 ejemplares, obteniéndose los siguientes resultados:

5.3, 7.6, 7.7, 8.2, 8.3, 9.2, 9.3, 9.4, 9.7, 10.1, 10.5, 10.7, 10.8, 11.0, 11.3, 12.1, 12.2, 12.5, 13.3, 16.3

($\sum x_i = 205.5$ y $\sum x_i^2 = 2221.01$)

a) Suponiendo que los datos siguen una distribución normal calcular intervalos de confianza, con una confianza del 95 %, para la altura media y variabilidad de la población.

b) Calcular una intervalo de confianza al 90 %, para la proporción de árboles de la plantación con altura superior a 10 metros.

Solución.

a) Sea la variable aleatoria X=altura del fuste de un pino (m), con $X \rightsquigarrow N(\mu, \sigma)$, y $X_1, X_2, ..., X_{20}$ la muestra aleatoria simple de X. Queremos calcular intervalos de confianza para μ y σ con $1-\alpha = 0.95$ y por tanto $\alpha/2 = 0.025$. Sabemos que el estimador puntual de μ es $\widehat{\mu_X} = \overline{X} = \frac{1}{20}\sum_{i=1}^{20} X_i$ y el estadístico pivote para el intervalo de confianza de μ es

$$\frac{\overline{X}-\mu_X}{S_X/\sqrt{20}} \rightsquigarrow t_{19}$$

siendo S_c^2 la varianza muestral corregida, con $S_X^2 = \left(\frac{20}{19}\right)\left[\left(\frac{1}{20}\right)\sum_{i=1}^{20} X_i^2 - \overline{X}^2\right]$. Buscando en las tablas de $t-$Student obtenemos $t_{19;0.025} = 2.093$ y por tanto podemos asegurar que

$$p\left(-2.093 \le \frac{\overline{X}-\mu_X}{S_X/\sqrt{20}} \le 2.093\right) = 0.95$$

o equivalentemente

$$p\left(\overline{X} - 2.093\left(\frac{S_X}{\sqrt{20}}\right) \le \mu_X \le \overline{X} + 2.093\left(\frac{S_X}{\sqrt{20}}\right)\right) = 0.95$$

Por tanto el intervalo de confianza buscado es $\overline{X} \pm 2.093\left(\frac{S_X}{\sqrt{20}}\right)$. En nuestro caso, tendremos que

$$\overline{x} = \frac{\sum x_i}{20} = \frac{205.5}{20} = 10.275$$

y

$$s_X^2 = \left(\frac{20}{19}\right)\left[\left(\frac{1}{20}\right)\sum x_i^2 - \overline{x}^2\right] = \left(\frac{20}{19}\right)\left(\frac{2221.01}{20} - 10.275^2\right) = 5.7630$$

Entonces, evaluando el intervalo tendremos

$$10.275 \pm 2.093\left(\frac{\sqrt{5.7630}}{\sqrt{20}}\right) = 10.275 \pm 1.124 = (9.151, 11.399)$$

Por tanto la altura media del fuste, con un 95 % de confianza, está entre 9.151 y 11.399 metros.

En cuanto a la variabilidad de la población vamos a calcular un intervalo de confianza para σ. Para ello construimos primero un intervalo de confianza del 95 % para σ^2. El estimador puntual de la varianza es

$$\widehat{\sigma^2} = S_X^2 = \left(\frac{20}{19}\right)\left[\left(\frac{1}{20}\right)\sum_{i=1}^{20} X_i^2 - \overline{X}^2\right]$$

y el estadístico pivote para el intervalo buscado es

$$\frac{19 S_X^2}{\sigma^2} \rightsquigarrow \chi_{19}^2$$

Utilizando las tabla de la distribución χ^2-Pearson obtenemos que $\chi^2_{19;0.025} = 32.852$ y $\chi^2_{19;0.975} = 8.907$. Podemos asegurar entonces que:

$$p\left(8.907 \leq \frac{19S_X^2}{\sigma^2} \leq 32.852\right) = 0.95$$

o, equivalentemente,

$$p\left(\frac{19S_X^2}{32.852} \leq \sigma^2 \leq \frac{9S_X^2}{8.907}\right) = 0.95$$

Por tanto el intervalo de confianza buscado será

$$\left(\frac{19S_X^2}{32.852}, \frac{9S_X^2}{8.907}\right)$$

Con la muestra dada por el problema, como hemos calculado antes, tenemos $s_X^2 = 5.7630$ y el intervalo de confianza para la varianza de las alturas será

$$\left(\frac{(19)\,(5.7630)}{32.852}, \frac{(19)\,(5.7630)}{8.907}\right) = (3.3330, 12.2934)$$

Entonces el intervalo de confianza para la desviación típica de las alturas se obtendrá con las raíces cuadradas del intervalo de confianza para la varianza, es decir,

$$\left(\sqrt{3.3330}, \sqrt{12.2934}\right) = (1.825, 3.506)$$

Por tanto la desviación típica de la altura del fuste está, con un 95 % de confianza, entre 1.825 y 3.506 metros.

b) Para cada pino $i = 1, 2, ..., 20$, consideramos la variable aleatoria

$$X_i = \begin{cases} 1 & \text{si la altura es superior a 10 metros} \\ 0 & \text{en caso contrario} \end{cases}$$

Sabemos que $X_i \rightsquigarrow B(p)$ siendo p la verdadera proporción de pinos con una altura superior a 10 metros. Queremos calcular un intervalo de confianza para p con un grado de confianza del 90 %, es decir, $1-\alpha = 0.90$ y por tanto $\alpha/2 = 0.05$. El estimador puntual para p es la media muestral de las variables aleatorias X_i, que en este caso representa la proporción muestral de pinos con una altura superior a 10 metros, es decir, $\widehat{p} = \overline{X} = \frac{1}{20}\sum_{i=1}^{20} X_i$. Entonces el estadístico pivote para encontrar un intervalo de confianza asintótico en este caso es

$$\frac{\widehat{p} - p}{\sqrt{\widehat{p}\,(1-\widehat{p})\,/20}} \,\widetilde{\rightsquigarrow}\, N\,(0,1)$$

Utilizando este estadístico pivote, y teniendo en cuenta que $z_{0.05} = 1.645$, tendremos que

$$p\left(-1.645 \leq \frac{\widehat{p} - p}{\sqrt{\widehat{p}\,(1-\widehat{p})\,/20}} \leq 1.645\right) \simeq 0.90$$

y despejando el parámetro p tendremos que

$$p\left(\widehat{p} - 1.645\sqrt{\frac{\widehat{p}\,(1-\widehat{p})}{20}} \leq p \leq \widehat{p} + 1.645\sqrt{\frac{\widehat{p}\,(1-\widehat{p})}{20}}\right) \simeq 0.90$$

Por tanto un intervalo de confianza asintótico para p con 95 % de confianza será $\widehat{p} \pm 1.645\sqrt{\frac{\widehat{p}(1-\widehat{p})}{20}}$. Observando los datos de la muestra nosotros tenemos 11 árboles con altura superior a 10 metros, es decir $\widehat{p} = 11/20 = 0.55$ y el intervalo de confianza será

$$0.55 \pm 1.645\sqrt{\frac{(0.55)\,(0.45)}{20}} = 0.55 \pm 0.0182 = (0.367, 0.733)$$

Podemos asegurar entonces con un confianza del 90 % que el porcentaje de pínos con altura superior a 10 metros se encuentra entre 36.7 % y 73.3 %.

21. Se disponen de los siguientes resultados sobre el punto de ebullición (^{o}K) de un aldehído: 376, 351, 346, 363, 380, 372, 365, 356, 393, 367, 358, 342, 370, 385, 360, 365 (suma de valores= 5849 y suma de valores al cuadrado= 2141023).

    Se asegura que, añadiendo al aldehído un determinado aditivo, aumenta su temperatura de ebullición. Para verificar esta afirmación se realizan 10 pruebas con el aldehído más el aditivo y se obtienen los siguientes puntos de ebullición (también en $^{\circ}K$): 355, 378, 390, 400, 379, 381, 411, 374, 396, 366 (suma de valores= 3830 y suma de valores al cuadrado= 1469380). Suponiendo que en ambos casos el punto de ebullición sigue una distribución normal, se pide:

    a) Encontrar un intervalo de confianza al 95 % para la temperatura media de ebullición en ambos casos. En base a estos intervalos, ¿podríamos decir que el aditivo aumenta la temperatura de ebullición? Razona la respuesta.

    b) Encontrar un intervalo de confianza al 90 % para el cociente de las varianzas teóricas del punto de ebullición del aldehído según se use o no aditivo. En vista del resultado, ¿podríamos decir, con un 90 % de confianza, que dichas varianzas son distintas?

    c) Teniendo en cuenta el apartado anterior, encontrar un intervalo de confianza al 95 % para la diferencia entre los puntos medios de ebullición en ambos casos. ¿Podríamos afirmar, con un 95 % de confianza, que el aditivo aumenta la temperatura de ebullición?

Solución. Consideramos las variables aleatorias X=punto de ebullición del aldehído sin aditivo (^{o}K) e Y=punto de ebullición del aldehído con aditivo (^{o}K). Como nos dice el problema, suponemos que ambas variables tienen distribución normal, es decir, que $X \rightsquigarrow N(\mu_X, \sigma_X)$ e $Y \rightsquigarrow N(\mu_Y, \sigma_Y)$. Sean $X_1, X_2, ..., X_{16}$ e $Y_1, Y_2, ..., Y_{10}$ las respectivas muestras aleatorias simples, que consideramos independientes entre si.

a) Calculamos primero un intervalo de confianza del 95 % para μ_X. El estimador puntual es $\widehat{\mu_X} = \overline{X} = \frac{1}{16}\sum_{i=1}^{16} X_i$ y el estadístico pivote para el intervalo buscado es

$$\frac{\overline{X} - \mu_X}{S_X/\sqrt{16}} \rightsquigarrow t_{15}$$

siendo S_X^2 la varianza muestral corregida, con $S_X^2 = \left(\frac{16}{15}\right)\left[\left(\frac{1}{16}\right)\sum_{i=1}^{16} X_i^2 - \overline{X}^2\right]$. Buscando en las tablas de $t-$Student obtenemos $t_{15;0.025} = 2.131$ y por tanto podemos asegurar que

$$p\left(-2.131 \leq \frac{\overline{X} - \mu_X}{S_X/\sqrt{16}} \leq 2.131\right) = 0.95$$

o equivalentemente

$$p\left(\overline{X} - 2.131\left(\frac{S_X}{\sqrt{16}}\right) \leq \mu_X \leq \overline{X} + 2.131\left(\frac{S_X}{\sqrt{16}}\right)\right) = 0.95$$

Por tanto el intervalo de confianza buscado es $\overline{X} \pm 2.131\left(\frac{S_X}{\sqrt{16}}\right)$. En nuestro caso, tendremos que

$$\overline{x} = \frac{\sum x_i}{16} = \frac{5849}{16} = 365.56255$$

y

$$s_X^2 = \left(\frac{16}{15}\right)\left[\left(\frac{1}{16}\right)\sum x_i^2 - \overline{x}^2\right] = \left(\frac{16}{15}\right)\left(\frac{2141023}{16} - 365.56255^2\right) = 189.8235$$

Entonces, evaluando el intervalo tendremos

$$365.56255 \pm 2.131\left(\frac{\sqrt{189.8235}}{4}\right) = 365.56255 \pm 7.34 = (358.19, 372.91)$$

Por tanto la temperatura media de ebullición sin aditivo, con un 95 % de confianza, está entre 358.19 y 372.91 ^oK.

Razonando del mismo modo para el caso del aditivo tendremos $\widehat{\mu_Y} = \overline{Y} = \frac{1}{10}\sum_{i=1}^{10} Y_i$ y el estadístico pivote para el intervalo buscado es

$$\frac{\overline{Y} - \mu_Y}{S_Y/\sqrt{11}} \leadsto t_9$$

siendo S_Y^2 la varianza muestral corregida, con $S_Y^2 = \left(\frac{10}{9}\right)\left[\left(\frac{1}{10}\right)\sum_{i=1}^{10} Y_i^2 - \overline{Y}^2\right]$. Buscando en las tablas de $t-$Student obtenemos $t_{9;0.025} = 2.262$ y por tanto podemos asegurar que

$$p\left(-2.262 \leq \frac{\overline{Y} - \mu_Y}{S_Y/\sqrt{10}} \leq 2.262\right) = 0.95$$

o equivalentemente

$$p\left(\overline{Y} - 2.262\left(\frac{S_Y}{\sqrt{10}}\right) \leq \mu_Y \leq \overline{Y} + 2.262\left(\frac{S_Y}{\sqrt{10}}\right)\right) = 0.95$$

Por tanto el intervalo de confianza buscado es $\overline{Y} \pm 2.262\left(\frac{S_Y}{\sqrt{10}}\right)$. En nuestro caso, tendremos que

$$\overline{y} = \frac{\sum y_i}{10} = \frac{3830}{10} = 383$$

y

$$s_Y^2 = \left(\frac{10}{9}\right)\left[\left(\frac{1}{10}\right)\sum y_i^2 - \overline{y}^2\right] = \left(\frac{10}{9}\right)\left(\frac{1469380}{10} - 383^2\right) = 276.6667$$

Entonces, evaluando el intervalo tendremos

$$383 \pm 2.262\left(\frac{\sqrt{276.6667}}{\sqrt{10}}\right) = 383 \pm 11.90 = (371.10, 394.90)$$

Por tanto la temperatura media de ebullición con aditivo, con un 95 % de confianza, está entre 371.10 y 394.90 ^oK.

En base a estos intervalos no podemos decir que la temperatura media de ebullición sea diferente en los dos casos puesto que hay puntos comunes a ambos intervalos, justamente todos los comprendidos en el intervalo $(371.10, 372.91)$.

b) Buscamos ahora un intervalo de confianza del 90 % para el cociente de varianzas σ_X^2/σ_Y^2. El estimador puntual de este cociente es $\widehat{\left(\frac{\sigma_X^2}{\sigma_Y^2}\right)} = \frac{S_X^2}{S_Y^2}$ siendo S_X^2 y S_Y^2 las respectivas varianzas muestrales corregidas de ambas muestras utilizadas en el apartado anterior. Además el estadístico pivote para el intervalo de confianza de la razón de varianzas es

$$\frac{S_X^2/S_Y^2}{\sigma_X^2/\sigma_Y^2} \rightsquigarrow F_{15,9}$$

Como queremos utilizar una confianza de 90 % tenemos que $1-\alpha = 0.90$, $\alpha/2 = 0.05$ y necesitamos encontrar los valores $F_{15,9;0.05}$ y $F_{15,9;0.95}$. Utilizando las tablas de la distribución F-Fisher con $\alpha = 0.05$ obtenemos el primer valor $F_{15,9;0.05} = 3.0061$. Para el segundo utilizaremos la propiedad de que $F_{15,9;0.95} = 1/F_{9,15;0.05}$ y $F_{9,15;0.05} = 2.5876$. Entonces podemos asegurar que

$$p\left(\frac{1}{2.5876} \leq \frac{S_X^2/S_Y^2}{\sigma_X^2/\sigma_Y^2} \leq 3.0061\right) = 0.90$$

o equivalentemente

$$p\left(\frac{S_X^2/S_Y^2}{3.0061} \leq \sigma_X^2/\sigma_Y^2 \leq (2.5876)\left(S_X^2/S_Y^2\right)\right) = 0.90$$

Entonces el intervalo de confianza buscado es

$$\left(\frac{S_X^2/S_Y^2}{3.0061}, (2.5876)\left(S_X^2/S_Y^2\right)\right)$$

En nuestro caso, como ya hemos calculado en el apartado anterior, tenemos $s_X^2 = 189.8235$ y $s_Y^2 = 276.6667$. Entonces, sustituyendo en el intervalo obtenemos

$$\left(\frac{189.8235/276.6667}{3.0061}, (2.5876)\,(189.8235/276.6667)\right) = (0.22, 1.78)$$

Por tanto no podemos decir, con una confianza de 90 % que el cociente σ_X^2/σ_Y^2 sea distinto 1, porque $1 \in (0.22, 1.78)$. Es decir no podemos decir que las varianzas sean distintas. Supondremos entonces que son iguales para el apartado siguiente.

c) Ahora queremos calcular un intervalo de confianza del 95 % para la diferencia de medias $\mu_X - \mu_Y$ suponiendo que ambas varianzas son iguales ($\sigma_X^2 = \sigma_Y^2 = \sigma^2$). Sabemos que el estimador puntual para esta diferencia es $\widehat{\mu_X - \mu_Y} = \overline{X} - \overline{Y}$ y, como hemos supuesto que las varianzas son iguales, el estadístico pivote para el intervalo buscado es

$$\frac{\left(\overline{X}-\overline{Y}\right) - \left(\mu_X - \mu_y\right)}{S_p\sqrt{\frac{1}{16}+\frac{1}{10}}} \rightsquigarrow t_{24}$$

siendo

$$S_p^2 = \frac{15S_X^2 + 9S_Y^2}{24}$$

el estimador de la varianza común. Además, como el grado de confianza deseado es 95 %, tenemos que $1-\alpha = 0.95$, $\alpha/2 = 0.025$, y buscando en la tablas de la distribución $t-$Student obtenemos que $t_{24;0.025} = 2.064$. Por tanto podemos asegurar que

$$p\left(-2.064 \leq \frac{\left(\overline{X}-\overline{Y}\right)-\left(\mu_X-\mu_y\right)}{S_p\sqrt{\frac{1}{16}+\frac{1}{10}}} \leq 2.064\right) = 0.95$$

o, equivalentemente, que

$$p\left(\left(\overline{X}-\overline{Y}\right) - 2.064\left(S_p\sqrt{\frac{1}{16}+\frac{1}{10}}\right) \leq \mu_X - \mu_y \leq \left(\overline{X}-\overline{Y}\right) + 2.064\left(S_p\sqrt{\frac{1}{16}+\frac{1}{10}}\right)\right) = 0.95$$

Por tanto el intervalo de confianza buscado será

$$\left(\overline{X}-\overline{Y}\right) \pm 2.064\left(S_p\sqrt{\frac{1}{16}+\frac{1}{10}}\right)$$

Con los datos ya obtenidos antes tenemos que $\overline{x} = 365.56255$, $\overline{y} = 383$, $s_X^2 = 189.8235$ y $s_Y^2 = 276.6667$. Entonces, para s_p obtenemos

$$s_p = \sqrt{\frac{(15)\,(189.8235)+(9)\,(276.6667)}{24}} = 14.9127$$

y el intervalo de confianza buscado será

$$-17.4375 \pm (2.064)\,(14.9127)\sqrt{\frac{1}{16}+\frac{1}{10}} = -17.4375 \pm 12.4078 = (-29.85, -5.03)$$

Podemos entonces asegurar, con un 95 % de confianza, que $\mu_X - \mu_y < 0$, y por tanto la temperatura media de ebullición es mayor con aditivo que sin él.

22. Para comparar el pH del suelo de dos explotaciones forestales (A y B) se tomaron 8 muestras de tierra de cada una de ellas en puntos elegidos al azar sobre su superficie. Se analizaron las muestras para determinar el pH, obteniéndose los siguientes datos:

Muestra	1	2	3	4	5	6	7	8		
Explotación A	6.55	5.98	5.59	6.17	5.92	6.18	6.43	5.68	$\sum x_i = 48.5$	$\sum x_i^2 = 294.826$
Explotación B	6.78	6.14	6.80	6.91	6.10	6.01	6.18	6.88	$\sum y_i = 51.8$	$\sum y_i^2 = 336.513$

Suponiendo que en ambas explotaciones el pH tiene una distribución normal, se pide:

a) Encontrar un intervalo de confianza al 90 % para el cociente de las varianzas teóricas del pH de ambas explotaciones. En base a este intervalo, ¿podríamos decir, con un 90 % de confianza, que dichas varianzas son distintas?

b) Teniendo en cuenta el resultado obtenido en el apartado anterior, encontrar un intervalo de confianza al 95 % para la diferencia de los valores medios teóricos del pH del suelo de ambas explotaciones. ¿Podríamos afirmar, con un 95 % de confianza, que dichos valores son distintos? ¿Y con una confianza del 99 %?

c) Si en la explotación A de 40 pies se observaron 23 atacados por un determinado hongo, mientras que en la explotación B se detectaron 18 pies infectados de otros 40 analizados ¿Cuál es el mayor nivel de confianza con el que podemos afirmar que las proporciones de pies infectados es diferente en ambas poblaciones?

Solución. Consideramos las variables aleatorias X=pH del suelo en la explotación A e Y=pH del suelo en la explotación B. Como nos dice el problema, suponemos que ambas variables tienen distribución normal, es decir, que $X \rightsquigarrow N(\mu_X, \sigma_X)$ e $Y \rightsquigarrow N(\mu_Y, \sigma_Y)$. Sean $X_1, X_2, ..., X_8$ e $Y_1, Y_2, ..., Y_8$ las respectivas muestras aleatorias simples, que consideramos independientes entre si.

a) Buscamos un intervalo de confianza del 90 % para el cociente de varianzas σ_X^2/σ_Y^2. El estimador puntual de este cociente es $\widehat{\left(\frac{\sigma_X^2}{\sigma_Y^2}\right)} = \frac{S_X^2}{S_Y^2}$ siendo S_X^2 y S_Y^2 las respectivas varianzas muestrales corregidas de ambas muestras definidas como $S_X^2 = \left(\frac{8}{7}\right)\left[\left(\frac{1}{8}\right)\sum_{i=1}^{8} X_i^2 - \overline{X}^2\right]$, $S_Y^2 = \left(\frac{8}{7}\right)\left[\left(\frac{1}{8}\right)\sum_{i=1}^{8} Y_i^2 - \overline{Y}^2\right]$. Además el estadístico pivote para el intervalo de confianza de la razón de varianzas es

$$\frac{S_X^2/S_Y^2}{\sigma_X^2/\sigma_Y^2} \rightsquigarrow F_{7,7}$$

Como queremos utilizar una confianza de 90 % tenemos que $1-\alpha = 0.90$, $\alpha/2 = 0.05$ y necesitamos encontrar los valores $F_{7,7;0.05}$ y $F_{7,7;0.95}$. Utilizando las tablas de la distribución F-Fisher con $\alpha = 0.05$ obtenemos el primer valor $F_{7,7;0.05} = 3.7871$. Para el segundo utilizaremos la propiedad de que $F_{7,7;0.95} = 1/F_{7,7;0.05} = 1/3.7871$. Entonces podemos asegurar que

$$p\left(1/3.7871 \leq \frac{S_X^2/S_Y^2}{\sigma_X^2/\sigma_Y^2} \leq 3.7871\right) = 0.90$$

o equivalentemente

$$p\left(\frac{S_X^2/S_Y^2}{3.7871} \leq \sigma_X^2/\sigma_Y^2 \leq (3.7871)\left(S_X^2/S_Y^2\right)\right) = 0.90$$

Entonces el intervalo de confianza buscado es

$$\left(\frac{S_X^2/S_Y^2}{3.7871}, (3.7871)\left(S_X^2/S_Y^2\right)\right)$$

Con los datos del problema tenemos que $\sum_{i=1}^{8} x_i = 48.5$, $\sum_{i=1}^{8} x_i^2 = 294.826$, $\sum_{i=1}^{8} y_i = 51.8$ y $\sum_{i=1}^{8} y_i^2 = 336.513$. Podemos evaluar entonces todos los estadísticos necesarios: $\overline{x} = 6.0625$, $\overline{y} = 6.475$, $s_x^2 = \left(\frac{8}{7}\right)\left[\frac{294.826}{8} - 6.0625^2\right] = 0.1135$ y $s_y^2 = \left(\frac{8}{7}\right)\left[\frac{336.513}{8} - 6.475^2\right] = 0.1583$. Entonces, sustituyendo en el intervalo obtenemos

$$\left(\frac{0.1135}{(0.1583)(3.7871)}, \frac{(3.7871)(0.1135)}{0.1583}\right) = (0.1893, 2.7153)$$

Por tanto no podemos decir, con una confianza del 90 % que el cociente σ_X^2/σ_Y^2 sea distinto 1, porque $1 \in (0.1893, 2.7153)$. Es decir no podemos decir que las varianzas sean distintas. Supondremos entonces que son iguales para el apartado siguiente.

b) Ahora queremos calcular un intervalo de confianza del 95 % para la diferencia de medias $\mu_X - \mu_Y$. Sabemos que el estimador puntual para esta diferencia es $\widehat{\mu_X - \mu_Y} = \overline{X} - \overline{Y}$ y, como hemos supuesto que las varianzas son iguales, el estadístico pivote para el intervalo buscado es

$$\frac{\left(\overline{X} - \overline{Y}\right) - \left(\mu_X - \mu_y\right)}{S_p\sqrt{\frac{1}{8} + \frac{1}{8}}} \rightsquigarrow t_{14}$$

siendo

$$S_p^2 = \frac{7S_X^2 + 7S_Y^2}{14} = \frac{S_X^2 + S_Y^2}{2}$$

y S_X^2, S_Y^2 las respectivas varianzas muestrales corregidas de ambas muestras utilizadas en el apartado anterior. Además, como el grado de confianza deseado es 95 %, tenemos que $1 - \alpha = 0.95$, $\alpha/2 = 0.025$, y buscando en la tablas de la distribución $t-$Student obtenemos que $t_{14;0.025} = 2.145$. Por tanto podemos asegurar que

$$p\left(-2.145 \leq \frac{\left(\overline{X} - \overline{Y}\right) - \left(\mu_X - \mu_y\right)}{S_p\sqrt{\frac{1}{8} + \frac{1}{8}}} \leq 2.145\right) = 0.95$$

o, equivalentemente, que

$$p\left(\left(\overline{X} - \overline{Y}\right) - 2.145\left(S_p\sqrt{\frac{2}{8}}\right) \leq \mu_X - \mu_y \leq \left(\overline{X} - \overline{Y}\right) + 2.145\left(S_p\sqrt{\frac{2}{8}}\right)\right) = 0.95$$

Por tanto el intervalo de confianza buscado será

$$\left(\overline{X} - \overline{Y}\right) \pm 1.0725 S_p$$

Con los datos del problema tenemos que $\overline{x} = 6.0625$, $\overline{y} = 6.475$, $s_X^2 = 0.1135$ y $s_Y^2 = 0.1583$. Entonces, para s_p obtenemos

$$s_p = \sqrt{\frac{0.1135 + 0.1583}{2}} = 0.3686$$

y el intervalo de confianza buscado será

$$-0.4125 \pm (1.0725)(0.3686) = -0.4125 \pm 0.3954 = (-0.8078, -0.0171)$$

Podemos entonces asegurar, con un 95 % de confianza, que $\mu_X - \mu_y < 0$, y por tanto podemos asegurar que la pH medio es mayor en la explotación B que en la explotación A. Si quisiésemos utilizar una confianza de 99 % todo se haría igual pero ahora tendríamos $t_{14;0.005} = 2.977$ y el intervalo sería

$$-0.4125 \pm (2.977)(0.3686/2) = -0.4125 \pm 0.5487 = (-0.9611, 0.1362)$$

Por tanto ahora tendríamos que $0 \in (-0.9611, 0.1362)$ y no podríamos asegurar, con una confianza de 99 %, que el pH medio en las dos explotaciones fuese distinto.

c) Sea ahora $X_1, X_2, ..., X_{40}$ la muestra de pies de la explotación A, con $X_i \rightsquigarrow B(p_X)$, siendo p_X la verdadera proporción de pies atacados por el hongo en la explotación A. Del mismo modo sea $Y_1, Y_2, ..., Y_{40}$ la muestra de pies de la explotación B, con $Y_i \rightsquigarrow B(p_Y)$, siendo p_Y la verdadera proporción de pies atacados por el hongo en la explotación B. Para un grado de confianza cualquiera $1-\alpha$ vamos a calcular un intervalo de confianza asintótico para la diferencia $p_X - p_Y$. El estadístico pivote en este caso será

$$\frac{(\widehat{p_X} - \widehat{p_Y}) - (p_X - p_Y)}{\sqrt{\frac{\widehat{p_X}(1-\widehat{p_X})}{40} + \frac{\widehat{p_Y}(1-\widehat{p_Y})}{40}}} \widetilde{\rightsquigarrow} N(0,1)$$

y, considerando el valor $z_{\alpha/2}$, tenemos que

$$p\left(-z_{\alpha/2} \leq \frac{(\widehat{p_X} - \widehat{p_Y}) - (p_X - p_Y)}{\sqrt{\frac{\widehat{p_X}(1-\widehat{p_X})}{40} + \frac{\widehat{p_Y}(1-\widehat{p_Y})}{40}}} \leq z_{\alpha/2}\right) \simeq 1-\alpha$$

o, equivalentemente

$$p\left((\widehat{p_X} - \widehat{p_Y}) - z_{\alpha/2}\sqrt{\frac{\widehat{p_X}(1-\widehat{p_X})}{40} + \frac{\widehat{p_Y}(1-\widehat{p_Y})}{40}} \leq (p_X - p_Y) \leq (\widehat{p_X} - \widehat{p_Y}) + z_{\alpha/2}\sqrt{\frac{\widehat{p_X}(1-\widehat{p_X})}{40} + \frac{\widehat{p_Y}(1-\widehat{p_Y})}{40}}\right) \simeq 1-\alpha$$

Por tanto el intervalo de confianza asintótico buscado será

$$(\widehat{p_X} - \widehat{p_Y}) \pm z_{\alpha/2}\sqrt{\frac{\widehat{p_X}(1-\widehat{p_X})}{40} + \frac{\widehat{p_Y}(1-\widehat{p_Y})}{40}}$$

Con los datos del problema nosotros tenemos $\widehat{p_X} = 23/40 =$ y $\widehat{p_Y} = 18/40$ y el intervalo será

$$5/40 \pm z_{\alpha/2}\sqrt{\frac{(23/40)(17/40)}{40} + \frac{(18/40)(22/40)}{40}} = 0.125 \pm (0.1109)\, z_{\alpha/2}$$

Para que las proporciones de pies infectados en ambas explotaciones fuesen significativamente diferentes necesitaríamos que $0.125 - (0.1109)\, z_{\alpha/2} > 0$, es decir, que $z_{\alpha/2} > 1.12$. Buscando en las tablas de la distribución normal observamos que $p\,(Z > 1.12) = 0.1314$ y por tanto necesitaríamos que $\alpha/2 = 0.1314$, es decir, $\alpha = 0.2628$. Por tanto el mayor nivel de confianza con el que podemos afirmar que la proporción es diferente en las dos explotaciones es $1-\alpha = 1 - 0.2628 = 0.7372$, es decir, 73.72 %.

23. Un grupo de investigadores afirma haber descubierto un tipo de alimentación para las gallinas, bajo la cual éstas producen huevos que no aumentan el colesterol en las personas que los consumen. Para comprobar dicha teoría se seleccionaron al azar 10 personas a las que se les midió el colesterol antes (X) y después (Y) de ser sometidos a una dieta a base de dichos huevos. Suponiendo normalidad, encontrar un intervalo de confianza al 95 % para la diferencia entre los valores medios del colesterol antes y después de la dieta, si los datos obtenidos son los siguientes:

X (antes)	120	312	243	161	314	234	143	287	423	155
Y (después)	130	306	255	168	310	250	158	290	440	140

¿Podemos considerar jusficada la afirmación de los investigadores? En base a los datos obtenidos, ¿cuántas personas necesitaríamos añadir a la muestra para rebatir la afirmación de los investigadores?

Solución. Consideramos las variables aleatorias X=nivel de colesterol antes de la dieta e Y=nivel de colesterol después de la dieta. Suponemos que ambas variables tienen distribución normal, es decir, que $X \rightsquigarrow N(\mu_X, \sigma_X)$ e $Y \rightsquigarrow N(\mu_Y, \sigma_Y)$. Sean $X_1, X_2, ..., X_{10}$ e $Y_1, Y_2, ..., Y_{10}$ las respectivas muestras aleatorias simples, que ahora no son independientes entre si por tratarse de dos muestras pareadas medidas sobre las mismas 10 gallinas. Buscamos un intervalo de confianza para la diferencia $\mu_X - \mu_Y$ con $1-\alpha = 0.95$ y por tanto $\alpha/2 = 0.025$. Consideramos ahora la variable aleatoria $D = X - Y$=diferencia en el nivel de colesterol antes y despues de la dieta. Tenemos entonces una muestra aleatoria simple $D_1, D_2, ..., D_{10}$ para esta variable aleatoria que también tendrá distribución normal con valor esperado $\mu_D = \mu_X - \mu_Y$ y una cierta varianza σ_D^2 que no pòdemos obtener directamente con σ_X^2 y σ_Y^2 por no ser las variables independientes (necesitaríamos tambien el valor de la covarianza). El estimador puntual para $\mu_X - \mu_Y$ podría calcularse restando las medias muestrales $\overline{X}$ e $\overline{Y}$ de las muestras iniciales, que coincidiría con la media muestral de la variable D, es decir, $\overline{D} = \overline{X} - \overline{Y}$. Entonces para encontrar un intervalo de confianza para $\mu_D = \mu_X - \mu_Y$ utilizaremos el estadístico pivote

$$\frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{S_D/\sqrt{10}} \rightsquigarrow t_9$$

siendo S_D^2 la varianza muestral corregida de la variable D, es decir $S_D^2 = \left(\frac{10}{9}\right)\left[\left(\frac{1}{10}\right)\sum_{i=1}^{10} D_i^2 - \overline{D}^2\right]$. Buscando en las tablas de $t-$Student obtenemos $t_{9;0.025} = 2.262$ y por tanto podemos asegurar que

$$p\left(-2.262 \leq \frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{S_D/\sqrt{10}} \leq 2.262\right) = 0.95$$

o equivalentemente

$$p\left(\overline{X} - \overline{Y} - 2.262\left(\frac{S_D}{\sqrt{10}}\right) \leq \mu_X \leq \overline{X} - \overline{Y} + 2.262\left(\frac{S_D}{\sqrt{10}}\right)\right) = 0.95$$

Por tanto el intervalo de confianza buscado es $\overline{X} - \overline{Y} \pm 2.262\left(\frac{S_D}{\sqrt{10}}\right)$. En nuestro caso, la muestra de las diferencias es

$D = X - Y$	-10	6	-12	-7	4	-16	-15	-3	-17	15

y por tanto endremos que

$$\overline{d} = \overline{x} - \overline{y} = \frac{\sum(x_i - y_i)}{10} = \frac{-55}{10} = -5.5$$

y

$$s_D^2 = \left(\frac{10}{9}\right)\left[\left(\frac{1}{10}\right)\sum(x_i - y_i)^2 - \overline{d}^2\right] = \left(\frac{10}{9}\right)\left(\frac{1349}{10} - \left(-5.5^2\right)\right) = 116.278$$

Entonces, evaluando el intervalo tendremos

$$-5.5 \pm 2.262\sqrt{\frac{116.278}{10}} = -5.5 \pm 7.71 = (-13.21, 2.22)$$

Como $0 \in (-13.21, 2.22)$ no podemos asegurar, con una confianza de 95 %, que el nivel medio de colesterol sea distinto antes de la dieta que después de la dieta. Para rebatir la afirmación de los investigadores necesitaríamos un nivel de signficación α para el cual

$$2.262\sqrt{\frac{116.278}{n}} < 5.5$$

y por tanto, despejando obtenemos

$$n > 116.278 \left(\frac{2.262}{5} \right)^2 = 23.8$$

El tamaño muestral necesario sería, por lo menos, $n = 24$.

1.7. PROBLEMAS RESUELTOS CAPÍTULO 9

1. Un investigador ha preparado el nivel de dosificación de un fármaco que afirma provocará el sueño en por lo menos el 80 % de las personas que padecen insomnio. Después de examinar la dosificación, se considera que su afirmación acerca de la efectividad del fármaco es exagerada. En un intento de refutar su afirmación se administra la dosificación prescrita a 20 personas que padecen insomnio, y se observa Y="Número de personas que se adormecen debido al fármaco". Se desea realizar el contraste de hipótesis $H_o : p = 0.8$ frente a $H_1 : p < 0.8$, utilizando la región crítica $C = \{Y \leq 12\}$.

   a) Calcular el nivel de significación del test utilizado.

   b) Encontrar la potencia para $p = 0.6$ y $p = 0.4$.

   c) ¿Cuál es la probabilidad de rechazar la hipótesis nula cuando es cierta? ¿Cuál es la probabilidad de no rechazar la hipótesis cuando p toma los valores del apartado anterior?

   d) Se desea contrastar la hipótesis a un nivel de significación $\alpha = 0.01$. ¿Cuál es la región crítica que se debería utilizar? Responder a los dos apartados anteriores utilizando este nuevo test.

   Solución. Suponiendo que la hipótesis nula es cierta la distribución de probabilidad de Y es una binomial $B(20, 0.8)$.

   a) El nivel de significación α del test se calcula como la probabilidad de rechazar la hipótesis nula (región crítica) suponiendo que sea cierta. Es decir, tendremos que calcular $p(Y \leq 12)$ suponiendo $Y \rightsquigarrow B(20, 0.8)$. Por tanto el nivel de significación del test es

   $$\alpha = p(Y \leq 12) = \sum_{k=0}^{12} \binom{20}{k} (0.8)^k (0.2)^{20-k} = 0.032$$

   b) Para cada posible p la potencia del test se calcula como la probabilidad de la región crítica suponiendo que p es el verdadero valor. En este caso tendríamos $Y \rightsquigarrow B(20, p)$. Por tanto, la potencia del test para $p = 0.6$ se calcularía suponiendo que $Y \rightsquigarrow B(20, 0.6)$ como

   $$1 - \beta = p(Y \leq 12) = \sum_{k=0}^{12} \binom{20}{k} (0.6)^k (0.4)^{20-k} = 0.584$$

   La probailidad de cometer un error de tipo II es entonces $\beta = 1 - 0.584 = 0.416$, de modo que la potencia es $1 - \beta$ como ya sabemos. Del mismo modo, para $p = 0.4$ la potencia sería

   $$p(Y \leq 12) = \sum_{k=0}^{12} \binom{20}{k} (0.4)^k (0.6)^{20-k} = 0.979$$

   y la probabilidad de error de tipo II sería $\beta = 1 - 0.979 = 0.021$,

   c) La probabilidad de rechazar la hipótesis nula cuando es cierta es justamente el nivel de significación, ya calculado en el primer apartado, es decir, $\alpha = 0.032$. La probabilidad de no rechazar la hipótesis nula cuando $p = 0.6$ es la probabilidad de error de tipo II $\beta = 0.416$. Del mismo modo, para $p = 0.4$ la probabilidad de no rechazar la hipótesis nula es la probabilidad de error de tipo II $\beta = 0.021$.

d) Si deseasemos utilizar un nivel de significación $\alpha = 0.01$ la región crítica debería ser $C = \{Y \leq y\}$ siendo y tal que $p(Y \leq y) = 0.01$ con $Y \rightsquigarrow B(20, 0.8)$. Es decir, necesitaremos calcular el cuantil 0.01 de esta distribución binomial. Teniendo en cuenta que

$$p(Y \leq 11) = \sum_{k=0}^{11} \binom{20}{k} (0.8)^k (0.2)^{20-k} = 0.00999 \simeq 0.01$$

obtenemos que la región crítica seria $C = \{Y \leq 11\}$. Con este test, la potencia para $p = 0.6$ sería

$$p(Y \leq 11) = \sum_{k=0}^{11} \binom{20}{k} (0.6)^k (0.4)^{20-k} = 0.404$$

y la probabilidad de error tipo II es $\beta = 1 - 0.404 = 0.596$. Del mismo modo, para $p = 0.4$ la potencia sería

$$p(Y \leq 11) = \sum_{k=0}^{11} \binom{20}{k} (0.4)^k (0.6)^{20-k} = 0.943$$

y la probabilidad de error tipo II sería $\beta = 1 - 0.943 = 0.057$.

2. El voltaje de salida de un cierto circuito eléctrico debe ser igual a 130, segun las especificaciones. Una muestra de 40 lecturas independientes para este circuito dio una media muestral $\bar{x} = 128.6$. Suponiendo que los datos provienen de una población normal con desviación típica $\sigma = 2.1$, contrastar la hipótesis de que el voltaje de salida promedio es de 130 frente a la alternativa de que es menor, con un nivel de significación $\alpha = 0.05$. ¿Cuál es el p-valor que nos da la muestra obtenida para este test?

Solución. Sea X=voltaje de salida de un circuito eléctrico con $X \rightsquigarrow N(\mu, 2.1)$. Sea $X_1, X_2, ..., X_{40}$ la muestra aleatoria simple de X. Queremos contrastar la hipótesis nula $H_o : \mu = 130$ frente a $H_1 : \mu < 130$ con un nivel de significación $\alpha = 0.05$. Si consideramos la media muestral $\overline{X} = \sum_{i=1}^{40} X_i$. El estadístico test en este caso sería

$$\frac{\overline{X} - 130}{2.1/\sqrt{40}} \rightsquigarrow N(0, 1)$$

y. como $z_{0.95} = -z_{0.05} = -1.645$, la región crítica sería

$$C = \left\{ \frac{\overline{X} - 130}{2.1/\sqrt{40}} \leq -1.645 \right\}$$

o, equivalentemente, $C = \left\{\overline{X} \leq 129.45\right\}$. Teniendo en cuenta que nosotros hemos obtenido $\bar{x} = 128.6 < 129.45$ se debe rechazar la hipótesis nula con un nivel de significación $\alpha = 0.05$. El p-valor que nos da la muestra obtenida es

$$p\left(Z \leq \frac{128.6 - 130}{2.1/\sqrt{40}}\right) = p(Z \leq -4.2) = p(Z > 4.2) = 0.0000133$$

Por tanto el p-valor que nos da la muestra obtenida para este test es $\alpha = 0.0000133$.

3. Para determinar el índice de dureza Rockwell del acero se oprime una punta de diamante en acero y se mide la profundidad de penetración. Para 50 muestras de un determinado acero se obtuvo un

índice de dureza promedio $\bar{x} = 62$. El fabricante afirma que el acero tiene un índice promedio de dureza de por lo menos 64. Suponiendo que los datos provienen de una distribución normal con desviación típica $\sigma = 8$, contrastar la afirmación con un nivel de significación $\alpha = 0.01$. Calcular el p-valor para la muestra obtenida.

El acero tiene la dureza suficiente cuando la media del índice de dureza Rockwell no baja de 60. En el test construido anteriormente, calcular la probabilidad de cometer un error de tipo II en la alternativa $\mu = 60$. Para contrastar $H_o : \mu = 64$ frente a $H_1 : \mu = 60$, determinar el tamaño de muestra necesario para que $\alpha = 0.01$ y $\beta = 0.05$.

Solución. Sea X=índice de dureza del acero con $X \rightsquigarrow N(\mu, 8)$. Sea $X_1, X_2, ..., X_{50}$ la muestra aleatoria simple de X. Queremos contrastar la hipótesis nula $H_o : \mu = 64$ frente a $H_1 : \mu > 64$ con un nivel de significación $\alpha = 0.01$. Si consideramos la media muestral $\overline{X} = \left(\frac{1}{50}\right) \sum_{i=1}^{50} X_i$, el estadístico test sería

$$\frac{\overline{X} - 64}{8/\sqrt{50}} \rightsquigarrow N(0,1)$$

y. como $z_{0.01} = 2.33$, la región crítica sería

$$C = \left\{ \frac{\overline{X} - 64}{8/\sqrt{50}} \geq 2.33 \right\}$$

o, equivalentemente, $C = \left\{ \overline{X} \geq 66.7 \right\}$. Teniendo en cuenta que nosotros hemos obtenido $\bar{x} = 62 < 66.7$ no podemos rechazar la hipótesis nula con un nivel de significación $\alpha = 0.05$. El p-valor que nos da la muestra obtenida es

$$p\left(Z \geq \frac{62 - 64}{8/\sqrt{50}}\right) = p(Z \geq -1.77) = 1 - p(Z > 1.77) = 1 - 0.0384 = 0.9616$$

Por tanto el p-valor que nos da la muestra obtenida para este test es $\alpha = 0.9616$. Para este test, la probabilidad de cometer un error de tipo II en la alternativa $\mu = 60$ sería

$$\beta = p\left(\frac{\overline{X} - 60}{8/\sqrt{50}} < \frac{66.7 - 60}{8/\sqrt{50}}\right) = p(Z < 5.9) = 1 - p(Z > 5.9) \approx 1$$

Si queremos contrastar $H_o : \mu = 64$ frente a $H_1 : \mu = 60$ con un tamaño muestral n, la región crítica sería $C = \left\{ \overline{X} \leq x \right\}$para un cierto x que dependerá del nivel de significación del test. Si queremos que el nivel de significación sea $\alpha = 0.01$ y que la probabilidad de error de tipo II sea $\beta = 0.05$, tendrá que verificarse:

$$\left.\begin{array}{r} p\left(Z \leq \frac{x-64}{8/\sqrt{n}}\right) = 0.01 \\ p\left(Z > \frac{x-60}{8/\sqrt{n}}\right) = 0.05 \end{array}\right\}$$

o, equivalentemente,

$$\left.\begin{array}{r} p\left(Z > -\frac{x-64}{8/\sqrt{n}}\right) = 0.01 \\ p\left(Z > \frac{x-60}{8/\sqrt{n}}\right) = 0.05 \end{array}\right\}$$

Buscando en la tabla de la distribución normal estándar tenemos $z_{0.01} = 2.33$ y $z_{0.05} = 1.645$ y por tanto el sistema que tenemos que resolver sería:

$$\left.\begin{array}{r} \frac{x-64}{8/\sqrt{n}} = -2.33 \\ \frac{x-60}{8/\sqrt{n}} = 1.645 \end{array}\right\}$$

o equivalentemente

$$\left.\begin{array}{c} x = 64 - 2.33\left(\frac{8}{\sqrt{n}}\right) \\ x = 60 + 1.645\left(\frac{8}{\sqrt{n}}\right) \end{array}\right\}$$

Por tanto debería ser

$$\left(\frac{8}{\sqrt{n}}\right) = \frac{4}{2.33 + 1.645} = 1.0063$$

y despejando n tenemos que $n = (8/1.0063)^2 = 63.2$. Es decir, el tamaño de muestra necesario sería $n = 64$. La región crítica para este test sería $C = \left\{\overline{X} \leq x\right\}$ siendo $x = 64 - 2.33\,(1.0063) = 61.66$.

4. El pH del agua que sale de una planta depuradora debe ser 7.0 por especificación. Se tomaron de forma independiente 30 muestras de agua de esa planta y se obtuvo un promedio de pH de $\bar{x} = 6.8$ y un desviación típica muestral corregida $s_c = 0.9$. Suponiendo que el pH del agua tiene una distribución normal, ¿hay razón para dudar del cumplimiento de la especificación de la planta? Ver la significatividad de los datos a un nivel $\alpha = 0.05$. Calcular el p-valor del test.

   Solución. Sea X=pH del agua $X \rightsquigarrow N(\mu, \sigma)$. Sea $X_1, X_2, ..., X_{30}$ la muestra aleatoria simple de X. Queremos contrastar la hipótesis nula $H_o : \mu = 7.0$ frente a $H_1 : \mu \neq 7.0$ con un nivel de significación $\alpha = 0.05$. Si consideramos la media muestral $\overline{X} = \left(\frac{1}{30}\right) \sum\limits_{i=1}^{30} X_i$ y la desviación típica muestral corregida

$$S_c = \sqrt{\left(\frac{30}{29}\right)\left[\left(\frac{1}{30}\right)\sum_{i=1}^{30} X_i^2 - \overline{X}^2\right]}$$

   el estadístico test sería

$$T = \frac{\overline{X} - 7}{S_c/\sqrt{30}} \rightsquigarrow t_{29}$$

   Entonces, como estamos haciendo el test de dos lados con $\alpha = 0.05$, necesitamos calcular $t_{29;0.025} = 2.045$ y la región crítica para el test sería $C = (-\infty, -2.045) \cup (2.045, \infty)$. Con los datos del problema tenemos $\overline{x} = 6.8$, $s_c = 0.9$ y por tanto el valor del estadístico test sería $T_0 = \frac{6.8-7}{0.9/\sqrt{30}} = -1.22$. Teniendo en cuenta que $-1.22 \notin C$ no podríamos rechazar la hipótesis nula y por tanto, con un nivel de significación $\alpha = 0.05$, no hay razón para dudar del cumplimiento de la especificación de la planta. El p-valor del test sería

$$p\left(|T| > 1.22\right) = 2p\left(t_{29} > 1.22\right)$$

   Utilizando el programa Statgraphics podemos evaluar esa probabilidad: $p\left(t_{29} > 1.22\right) = 0.116$ y por tanto el p-valor del test sería $2\,(0.116) = 0.232$.

5. Se afirma que la dureza, en grados Shore, de un determinado caucho debe ser 65. Se tomaron 14 pruebas independientes y se obtuvo $\bar{x} = 63.1$ con $s_c = 1.4$. ¿Hay evidencia suficiente como para rechazar la afirmación anterior a un nivel de significación $\alpha = 0.05$? ¿Qué hipótesis se necesita asumir para que la respuesta sea válida?

   Solución. Sea X=dureza del caucho (grados Shore) y sea $X_1, X_2, ..., X_{14}$ la muestra aleatoria simple de X. Como el tamaño muestral no es grande necesitamos suponer que $X \rightsquigarrow N(\mu, \sigma)$. Queremos contrastar la hipótesis nula $H_o: \mu = 65$ frente a $H_1: \mu \neq 65$ con un nivel de significación $\alpha = 0.05$. Si consideramos la media muestral $\overline{X} = \left(\frac{1}{30}\right)\sum_{i=1}^{30} X_i$ y la desviación típica muestral corregida

$$S_c = \sqrt{\left(\frac{14}{13}\right)\left[\left(\frac{1}{14}\right)\sum_{i=1}^{14} X_i^2 - \overline{X}^2\right]}$$

el estadístico test sería

$$T = \frac{\overline{X} - 65}{S_c/\sqrt{14}} \rightsquigarrow t_{13}$$

Entonces, como estamos haciendo el test de dos lados con $\alpha = 0.05$, necesitamos calcular $t_{13;0.025} = 2.16$ y la región crítica para el test sería $C = (-\infty, -2.16) \cup (2.16, \infty)$. Con los datos del problema tenemos $\overline{x} = 63.1$, $s_c = 1.4$ y por tanto el valor del estadístico test sería $T_0 = \frac{63.1-65}{1.4/\sqrt{14}} = -5.08$. Teniendo en cuenta que $-5.08 \in C$ podemos rechazar la hipótesis nula con un nivel de significación $\alpha = 0.05$ y concluir que hay evidencia suficiente para concluir que la dureza media no es 65. Como ya hemos dicho hemos necesitado asumir que la distribución de probabilidad de la dureza es normal.

6. La dispersión de los tiempos de descarga de material de un proyecto de construcción es de gran importancia, ya que si estos tiempos son muy variables es muy difícil organizar las tareas de forma eficiente. El encargado del transporte de una obra afirma que la diferencia entre los tiempos máximo y mínimo de descarga no debe ser mayor de 40 minutos. Si se supone que estos tiempos están distribuidos de forma aproximadamente normal, el encargado de la obra cree que esta afirmación quiere decir que la desviación típica σ debe ser aproximadamente 10 minutos. Se midieron, en minutos, 15 tiempos en los que se obtuvo $\bar{x} = 142$ con $s_c = 12$. ¿Podría rechazarse la hipótesis de que la desviación típica es igual a 10 con un nivel de significación $\alpha = 0.05$? Calcular el p-valor del test.

Solución. Sea X=tiempo de descarga de material (minutos), con $X \rightsquigarrow N(\mu, \sigma)$. Sea $X_1, X_2, ..., X_{15}$ la muestra aleatoria simple de X. Queremos contrastar la hipótesis nula $H_o: \sigma = 10$ frente a $H_1: \sigma > 10$ con un nivel de significación $\alpha = 0.05$. Si consideramos la media muestral $\overline{X} = \left(\frac{1}{15}\right)\sum_{i=1}^{15} X_i$ y la desviación típica muestral corregida

$$S_c = \sqrt{\left(\frac{15}{14}\right)\left[\left(\frac{1}{15}\right)\sum_{i=1}^{15} X_i^2 - \overline{X}^2\right]}$$

el estadístico test sería

$$T = \frac{14 S_c^2}{10^2} \rightsquigarrow \chi^2_{14}$$

Como estamos haciendo el test de un lado con $\alpha = 0.05$, necesitamos calcular $\chi^2_{14;0.05} = 23.685$ y la región crítica para el test sería $C = (23.685, \infty)$. Con los datos del problema tenemos $s_c = 12$ y por tanto el valor del estadístico test sería $T_0 = \frac{(14)(12^2)}{100} = 20.16$. Teniendo en cuenta que $20.16 \notin C$, no podemos rechazar la hipótesis nula de que $\sigma = 10$ frente $\sigma > 10$ con un nivel de significación $\alpha = 0.05$. Utilizando el programa Statgraphics podemos calcular el p-valor del test como

$$p(T > 20.16) = p\left(\chi^2_{14} > 20.16\right) = 0.1252$$

7. Las pruebas de aptitud deben producir calificaciones con una gran cantidad de variación para que se pueda distinguir a las personas con pocas aptitudes de aquellas que tienen muchas. La prueba normal que se emplea en una determinada empresa ha dado calificaciones con una desviación típica $\sigma = 5$ puntos. Se ensaya una prueba en 20 aspirantes, obteniéndose una muestra con $s_c = 8$ puntos. ¿Son las calificaciones de la prueba apreciablemente más variables que la prueba normal? Usar un nivel de significación $\alpha = 0.05$ y calcular el p-valor de la prueba.

   Solución. Sea X=calificación en la prueba de aptitud, con $X \rightsquigarrow N(\mu, \sigma)$. Sea $X_1, X_2, ..., X_{20}$ la muestra aleatoria simple de X. Queremos contrastar la hipótesis nula $H_o : \sigma = 5$ frente a $H_1 : \sigma > 5$ con un nivel de significación $\alpha = 0.05$. Si consideramos la media muestral $\overline{X} = \left(\frac{1}{20}\right) \sum_{i=1}^{20} X_i$ y la desviación típica muestral corregida

$$S_c = \sqrt{\left(\frac{20}{19}\right)\left[\left(\frac{1}{20}\right)\sum_{i=1}^{20} X_i^2 - \overline{X}^2\right]}$$

   el estadístico test sería

$$T = \frac{19S_c^2}{5^2} \rightsquigarrow \chi^2_{19}$$

   Como estamos haciendo el test de un lado con $\alpha = 0.05$, necesitamos calcular $\chi^2_{19;0.05} = 30.144$ y la región crítica para el test sería $C = (30.114, \infty)$. Con los datos del problema tenemos $s_c = 8$ y por tanto el valor del estadístico test sería $T_0 = \frac{(19)(8^2)}{25} = 48.64$. Teniendo en cuenta que $48.64 \in C$, podemos rechazar la hipótesis nula de que $\sigma = 5$ frente $\sigma > 5$ con un nivel de significación $\alpha = 0.05$. Por tanto la calificaciones de la prueba son más variables que la prueba normal. Utilizando el programa Statgraphics podemos calcular el p-valor del test como

$$p\,(T > 48.64) = p\left(\chi^2_{19} > 48.64\right) = 0.0002$$

8. En una encuesta realizada entre 2207 agricultores, el 54 % respondieron que el sistema de ayudas a la producción tiene unos trámites demasiado complicados. ¿Se puede concluir que la mayoría de los agricultores piensan que el sistema de ayudas es muy complicado? Utilizar un nivel de significación $\alpha = 0.05$ y calcular el p-valor. Calcular la potencia del test en la alternativa dada por que una tercera parte de los agricultores piensan que los trámites a seguir son demasiado complicados.

   Solución. Para cada agricultor encuestado $i = 1, 2, ..., 2207$, consideramos la variable aleatoria

$$X_i = \begin{cases} 1 & \text{si responde que el sistema es complicado} \\ 0 & \text{en caso contrario} \end{cases}$$

   Sabemos que $X_i \rightsquigarrow B(p)$ siendo p la verdadera proporción de agricultores que piensa que el sistema de ayudas a la producción tiene unos trámites demasiado complicados. Vamos a construir un test estadístico para contrastar la hipótesis nula $H_o : p = 0.5$ frente a $H_1 : p > 0.5$ con un nivel de significación $\alpha = 0.05$. Si $\widehat{p}$ es la proporción muestral de agricultores que piensan que el sistema es complicado, como el tamaño muestral es muy grande, podemos construir un test asintótico utilizando el estadístico test

$$T = \frac{\widehat{p} - 0.5}{\sqrt{0.5\,(1-0.5)\,/2207}} \widetilde{\rightsquigarrow} N\,(0,1)$$

Con la hipótesis alternativa elegida y el nivel de significación dado necesitamos $z_{0.05} = 1.645$. La región crítica será entonces $C = \left\{ \frac{\widehat{p}-0.5}{\sqrt{0.25/2207}} > 1.645 \right\}$ o, equivalentemente, $C = \{\widehat{p} > 0.518\}$. Con los datos del problema tenemos $\widehat{p} = 0.54 \in C$ y por tanto, con un nivel de significación $\alpha = 0.05$ debemos de rechazar la hipótesis nula y concluir que los agricultores que piensan que el sistema de ayudas es muy complicado son mayoría. El p-valor del test se calcularía como

$$p\left(Z > \frac{0.54 - 0.5}{\sqrt{0.25/2207}}\right) = p\,(Z > 3.8) = 0.0000723$$

Además, la potencia del test en la alternativa $p = 1/3$ se calcularía como

$$p\left(Z > \frac{0.518 - 1/3}{\sqrt{(1/3)\,(2/3)\,/2207}}\right) = p\,(Z > 18.4) \approx 0$$

9. Mediciones respecto del esfuerzo cortante obtenidas a partir de pruebas de compresión independientes para dos tipos de suelos dieron los siguientes resultados (en toneladas por pie cuadrado):

	Suelo I	Suelo II
Tamaño muestral	30	35
Media muestral	1.65	1.43
Desviación típica corregida	0.26	0.22

¿Son los datos significativamente diferentes a un nivel $\alpha = 0.01$. Calcular el p-valor del test.

Solución. Consideramos las variables aleatorias X=esfuerzo cortante del suelo I Y=esfuerzo cortante del suelo II. Como los tamaños muestrales no son muy grandes vamos a suponer que ambas variables tienen distribución normal con la misma varianza, es decir, que $X \rightsquigarrow N\,(\mu_X, \sigma)$ e $Y \rightsquigarrow N\,(\mu_Y, \sigma)$. Sean $X_1, X_2, ..., X_{30}$ e $Y_1, Y_2, ..., Y_{35}$ las respectivas muestras aleatorias simples, que consideramos independientes entre si. Queremos contrastar la hipótesis nula $H_o : \mu_X - \mu_Y = 0$ frente a $H_1 : \mu_X - \mu_Y \neq 0$ con un nivel de significación $\alpha = 0.01$. El estadístico test para este caso es

$$T = \frac{\overline{X} - \overline{Y}}{S_p\sqrt{\frac{1}{30} + \frac{1}{35}}} \rightsquigarrow t_{63}$$

siendo $\overline{X} = \frac{1}{30}\sum\limits_{i=1}^{30} X_i$, $\overline{Y} = \frac{1}{35}\sum\limits_{i=1}^{35} Y_i$ y

$$S_p = \sqrt{\frac{29S_X^2 + 34S_Y^2}{63}}$$

el estimador de la varianza común, con $S_X^2 = \left(\frac{30}{29}\right)\left[\left(\frac{1}{30}\right)\sum\limits_{i=1}^{30} X_i^2 - \overline{X}^2\right]$ y $S_Y^2 = \left(\frac{35}{34}\right)\left[\left(\frac{1}{35}\right)\sum\limits_{i=1}^{35} Y_i^2 - \overline{Y}^2\right]$. Como estamos haciendo el test de dos lados y queremos un nivel de significación $\alpha = 0.01$, necesitamos calcular $t_{63;0.005} = 2.660$ (hemos utilizado las tabla de la distribución $t-$Student con $n = 60$ que es el más cercano). La región crítica para el test sería entonces $T \in (-\infty, -2.66) \cup (2.66, \infty)$, es decir, $|T| > 2.66$. Con los datos que proporciona el problema nosotros tenemos $\overline{x} = 1.65$, $\overline{y} = 1.43$, $s_X^2 = 0.26^2$ y $s_Y^2 = 0.22^2$ y

$$s_p = \sqrt{\frac{(29)\left(0.26^2\right) + (34)\left(0.22^2\right)}{63}} = 0.2392$$

y el valor observado para el estadístico test sería

$$T_0 = \frac{1.65 - 1.43}{(0.2392)\sqrt{\frac{1}{30} + \frac{1}{35}}} = 3.6966$$

Por tanto, como $|T_0| = 3.70 > 2.66$ debemos rechazar la hipótesis nula y concluir que el esfuerzo medio cortante es significativamente distinto en ambos tipos de suelo con un nivel de significación $\alpha = 0.01$. Calculamos ahora el p-valor del test utilizando el programa Statgraphics:

$$p\left(T > T_0\right) = p\left(t_{63} > 3.6966\right) = 0.0002$$

Por tanto rechazaríamos la hipótesis nula para cualquier nivel de significación $\alpha > 0.0002$.

10. Unos cohetes se fabrican con un supuesto alcance de 2500 metros. Teóricamente se supone que el alcance se reduce cuando se almacenan durante algún tiempo. Seis de esos cohetes se almacenaron durante un determinado periodo y a continuación se probaron. Los alcances obtenidos fueron los siguientes: 2490, 2510, 2360, 2410, 2300 y 2440. ¿Es más corto el alcance después del almacenamiento? Suponiendo distribución normal, contrastar la hipótesis con un nivel de significación del 1 % y calcular el p-valor de la prueba para los datos obtenidos.

    En este problema, la variabilidad de los alcances es también importante. Los cohetes nuevos tienen una desviación típica $\sigma = 50$ metros. ¿Podemos decir que el almacenamiento aumenta la variabilidad de los alcances? Usar un nivel de significación $\alpha = 0.05$ y calcular el p-valor del test.

    Solución. Sea X=alcance de un cohete (m), con $X \rightsquigarrow N(\mu, \sigma)$. Sea $X_1, X_2, ..., X_6$ la muestra aleatoria simple de X. Queremos contrastar la hipótesis nula $H_o : \mu = 2500$ frente a $H_1 : \mu < 2500$ con un nivel de significación $\alpha = 0.01$. Si consideramos la media muestral $\overline{X} = \left(\frac{1}{16}\right) \sum_{i=1}^{6} X_i$ y la desviación típica muestral corregida

$$S_c = \sqrt{\left(\frac{6}{5}\right)\left[\left(\frac{1}{6}\right)\sum_{i=1}^{6} X_i^2 - \overline{X}^2\right]}$$

    el estadístico test sería

$$T = \frac{\overline{X} - 2500}{S_c/\sqrt{6}} \rightsquigarrow t_5$$

    Como estamos haciendo el test de un lado con $\alpha = 0.01$, necesitamos calcular $t_{5;0.01} = 3.365$ y la región crítica para el test sería $C = (-\infty, -3.365)$ o, equivalentemente, $T < -3.365$. Con los datos del problema tenemos que $\sum x_i = 14510$ y $\sum x_i^2 = 35121500$ y por tanto $\overline{x} = 14510/6 = 2418.33$ y

$$s_c = \sqrt{(6/5)\left[65121500/6\right] - \left(14510/6\right)^2} = 79.35$$

El valor del estadístico test será entonces

$$T_0 = \frac{\overline{X} - 2500}{S_c/\sqrt{6}} = \frac{-81.67}{32.40} = -2.52$$

Teniendo en cuenta que $-2.52 \notin C$, no podemos rechazar la hipótesis nula de que $\mu = 2500$ frente a $\mu < 2500$ con un nivel de significación $\alpha = 0.01$. Utilizando el programa Statgraphics podemos calcular el p-valor del test como

$$p\,(T < -2.52) = p\,(t_5 < -2.52) = p\,(t_5 > 2.52) = 0.027$$

Observar que, como $\alpha = 0.01 < 0.027$, no hemos podido rechazar la hipótesis nula.

Para la variabilidad de los alcances necesitamos contrastar la hipótesis nula $H_o : \sigma = 50$ frente a $H_1 : \sigma > 50$ con un nivel de significación $\alpha = 0.05$. En este caso el estadístico test sería

$$T = \frac{5S_c^2}{50^2} \rightsquigarrow \chi_5^2$$

Como estamos haciendo el test de un lado con $\alpha = 0.05$, necesitamos calcular $\chi^2_{5;0.05} = 11.070$ y la región crítica para el test sería $C = (11.070, \infty)$ o, equivalentemente, $T > 11.070$. Con los datos del problema tenemos $s_c = 79.35$ y por tanto el valor del estadístico test sería $T_0 = \frac{(5)\left(79.35^2\right)}{2500} = 12.59$. Teniendo en cuenta que $12.59 \in C$, podemos rechazar la hipótesis nula de que $\sigma = 50$ frente $\sigma > 50$ con un nivel de significación $\alpha = 0.05$. Por tanto el almacenamiento aumenta la variabilidad de los alcances. Utilizando el programa Statgraphics podemos calcular el p-valor del test como

$$p\,(T > 12.59) = p\left(\chi_5^2 > 12.59\right) = 0.0275$$

Observar que, como el nivel de significación $\alpha = 0.05 > 0.0275$ hemos rechazado la hipótesis nula.

11. Se efectuó un estudio por parte de una Comisión de Caza y Pesca para estimar las cantidades de residuos químicos encontrados en los tejidos cerebrales de pelícanos. En una prueba sobre mediciones de DDT, muestras aleatorias de 10 pelícanos jóvenes y 13 polluelos dieron los siguientes resultados:

	Jóvenes	Polluelos
Tamaño muestral	10	13
Media muestral	0.041	0.026
Desviación típica corregida	0.017	0.006

Suponiendo que las mediciones de DDT provienen de distribuciones normales con la misma desviación típica, contrastar la hipótesis de que no existe diferencia en las cantidades promedio de DDT encontradas en los pelícanos jóvenes y en los polluelos, frente a la alternativa de que los pelícanos jóvenes presentan un mayor promedio. Utilizar un nivel de significación $\alpha = 0.05$ y calcular el p-valor. ¿Son los datos estadísticamente significativos a un nivel del 1 %?

¿Hay suficiente evidencia, a un nivel de significación del 5 %, para concluir que la varianza en las mediciones de los niveles de DDT es mayor en los pelícanos jóvenes que en los polluelos? ¿Qué implicaciones tiene en el contraste realizado con anterioridad? De acuerdo con esto, volver a realizar el test anterior sin suponer que las varianzas son iguales.

Solución. Consideramos las variables aleatorias X=residuos químicos cerebrales en pelícanos jóvenes e Y=residuos químicos cerebrales en polluelos. Como nos dice el problema, suponemos que ambas variables tienen distribución normal con la misma varianza, es decir, que $X \rightsquigarrow N(\mu_X, \sigma)$ e $Y \rightsquigarrow N(\mu_Y, \sigma)$. Sean $X_1, X_2, ..., X_{10}$ e $Y_1, Y_2, ..., Y_{13}$ las respectivas muestras aleatorias simples, que consideramos independientes entre si. Queremos contrastar la hipótesis nula $H_o : \mu_X - \mu_Y = 0$ frente a $H_1 : \mu_X - \mu_Y > 0$ con un nivel de significación $\alpha = 0.05$. El estadístico test para este caso es

$$T = \frac{\overline{X} - \overline{Y}}{S_p \sqrt{\frac{1}{10} + \frac{1}{13}}} \rightsquigarrow t_{21}$$

siendo $\overline{X} = \frac{1}{10} \sum_{i=1}^{10} X_i$, $\overline{Y} = \frac{1}{13} \sum_{i=1}^{13} Y_i$ y

$$S_p = \sqrt{\frac{9S_X^2 + 12S_Y^2}{21}}$$

el estimador de la varianza común, con $S_X^2 = \left(\frac{10}{9}\right) \left[\left(\frac{1}{10}\right) \sum_{i=1}^{10} X_i^2 - \overline{X}^2\right]$ y $S_Y^2 = \left(\frac{13}{12}\right) \left[\left(\frac{1}{13}\right) \sum_{i=1}^{13} Y_i^2 - \overline{Y}^2\right]$. Como estamos haciendo el test de un lado y queremos un nivel de significación $\alpha = 0.05$, necesitamos calcular $t_{21;0.05} = 1.721$. La región crítica para el test sería entonces $T \in (1.721, \infty)$, es decir, $T > 1.721$. Con los datos que proporciona el problema nosotros tenemos $\overline{x} = 0.041$, $\overline{y} = 0.026$, $s_X^2 = 0.017^2$ y $s_Y^2 = 0.006^2$ y

$$s_p = \sqrt{\frac{(9)(0.017^2) + (12)(0.006^2)}{21}} = 0.0120$$

y el valor observado para el estadístico test sería

$$T_0 = \frac{0.041 - 0.026}{(0.0120)\sqrt{\frac{1}{10} + \frac{1}{13}}} = 2.9674$$

Por tanto, como $T_0 = 2.9674 > 1.721$ debemos rechazar la hipótesis nula y concluir que los pelícanos jóvenes presentan un mayor promedio de residuos químicos. Utilizando el programa Statgraphics el p-valor del test sería:

$$p(T > T_0) = p(t_{21} > 2.9674) = 0.0037$$

Por tanto rechazaríamos la hipótesis nula para cualquier nivel de significación $\alpha > 0.0037$, en particular para $\alpha = 0.05$ como ya hemos hecho. Si ahora quisiésemos utilizar un nivel de signifcación del 1 % tendríamos que $\alpha = 0.01 > .0037$ y por tanto también deberíamos rechazar la hipótesis nula. Es decir, los datos son estadísticamente significativos a un nivel del 1 %.

Para comparar las varianzas en las mediciones de los niveles de DDT necesitamos contrastar la hipótesis nula $H_0 : \sigma_X^2/\sigma_Y^2 = 1$ frente a $H_1 : \sigma_X^2/\sigma_Y^2 > 1$ siendo σ_X^2 y σ_Y^2, las dos varianzas poblacionales en pelícanos jóvenes y en polluelos respectivamente. En este caso, el estadístico test sería

$$T = \frac{S_X^2}{S_Y^2} \rightsquigarrow F_{9,12}$$

siendo S_X^2 y S_Y^2 las varianzas muestrales corregidas utilizadas anteriormente. Como estamos haciendo el test de un lado y queremos un nivel de significación $\alpha = 0.05$, necesitamos calcular $F_{9,12;0.05} = 2.7964$. La región crítica para el test sería entonces $T \in (2.7964, \infty)$, es decir, $T > 2.7964$. Con los datos que proporciona el problema nosotros tenemos $s_X^2 = 0.017^2$ y $s_Y^2 = 0.006^2$ y por tanto el valor del estadístico test sería

$$T_0 = \frac{0.017^2}{0.006^2} = 8.0278$$

Por tanto, como $T_0 = 8.0278 > 2.7964$ debemos rechazar la hipótesis nula y concluir que los pelícanos jóvenes presentan un mayor varianza en los valores de DDT. Observar que esto implica que el test realizado anteriormente para los valores promedio de residuos químicos no sería correcto. Vamos entonces a rehacer ese test sin suponer que las varianzas son iguales. Para este caso el estadístico test sería

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_X^2}{10} + \frac{S_Y^2}{13}}} \rightsquigarrow t_d$$

siendo

$$d = \frac{(\Delta_1 + \Delta_2)^2}{\frac{\Delta_1^2}{11} + \frac{\Delta_2^2}{14}}$$

con $\Delta_1 = \frac{S_X^2}{10}$ y $\Delta_1 = \frac{S_Y^2}{13}$. Con los datos del problema tenemos $\Delta_1 = \frac{0.017^2}{10} = 0.000289$, $\Delta_2 = \frac{0.006^2}{13} = 0.0000277$ y

$$d = \frac{(0.000289 + 0.0000277)^2}{0.000289/11 + 0.0000277/14} = 11.21 \simeq 11$$

Como estamos haciendo el test de un lado y queremos un nivel de significación $\alpha = 0.05$, necesitamos calcular $t_{11;0.05} = 1.796$. La región crítica para el test sería entonces $T \in (1.796, \infty)$, es decir, $T > 1.796$. Con los valores ya utilizados antes tenemos que el valor del estadístico pivote es

$$T_0 = \frac{0.041 - 0.026}{\sqrt{\frac{0.017^2}{10} + \frac{0.006^2}{13}}} = 2.6655$$

Por tanto, como $T_0 = 2.6655 > 1.796$ también debemos rechazar la hipótesis nula y concluir que los pelícanos jóvenes presentan un mayor promedio de residuos químicos. Utilizando el programa Statgraphics el p-valor del test sería:

$$p\,(T > T_0) = p\,(t_{11} > 2.6655) = 0.011$$

Observar que en este caso, con un nivel de significación $\alpha = 0.01$ no podríamos rechazar la hipótesis nula, mientras que antes sí.

12. Se compararon dos proyectos para la construcción de un laboratorio con respecto a la cantidad media de luz que se tiene en la superficies de las mesas. Se tomaron 40 mediciones independientes en cada laboratorio obteniéndose los siguientes resultados:

	Diseño I	Diseño II
Tamaño muestral	40	40
Media muestral	28.8	32.6
Desviación típica corregida	15.1	15.8

Suponiendo que los datos provienen de dos poblaciones con distribución normal de varianzas desconocidas (pero iguales), contrastar la igualdad de los promedios con un nivel de significación $\alpha = 0.05$ y calcular el p-valor del test.

Solución. Consideramos las variables aleatorias X=cantidad de luz en superficie de una mesa con diseño I e Y=cantidad de luz en superficie de una mesa con diseño II. Como nos dice el problema, suponemos que ambas variables tienen distribución normal con la misma varianza, es decir, que $X \rightsquigarrow N(\mu_X, \sigma)$ e $Y \rightsquigarrow N(\mu_Y, \sigma)$. Sean $X_1, X_2, ..., X_{40}$ e $Y_1, Y_2, ..., Y_{43}$ las respectivas muestras aleatorias simples, que consideramos independientes entre si. Queremos contrastar la hipótesis nula $H_o : \mu_X - \mu_Y = 0$ frente a $H_1 : \mu_X - \mu_Y \neq 0$ con un nivel de significación $\alpha = 0.05$. El estadístico test para este caso es

$$T = \frac{\overline{X} - \overline{Y}}{S_p\sqrt{\frac{1}{40} + \frac{1}{40}}} \rightsquigarrow t_{78}$$

siendo $\overline{X} = \frac{1}{40}\sum_{i=1}^{40} X_i$, $\overline{Y} = \frac{1}{40}\sum_{i=1}^{40} Y_i$ y

$$S_p = \sqrt{\frac{39S_X^2 + 39S_Y^2}{78}} = \sqrt{\frac{S_X^2 + S_Y^2}{2}}$$

el estimador de la varianza común, con $S_X^2 = \left(\frac{40}{39}\right)\left[\left(\frac{1}{40}\right)\sum_{i=1}^{40} X_i^2 - \overline{X}^2\right]$ y $S_Y^2 = \left(\frac{40}{39}\right)\left[\left(\frac{1}{40}\right)\sum_{i=1}^{40} Y_i^2 - \overline{Y}^2\right]$. Como estamos haciendo el test de dos lados y queremos un nivel de significación $\alpha = 0.05$, necesitamos calcular $t_{78;0.025} = 1.99$ (hemos buscado en la tabla $t-$Student con $n = 80$ que es el más cercano). La región crítica para el test sería entonces $T \in (-\infty, -1.99) \cup (1.99, \infty)$, es decir, $|T| > 1.99$. Con los datos que proporciona el problema nosotros tenemos $\overline{x} = 28.8$, $\overline{y} = 32.6$, $s_X^2 = 15.1^2$ y $s_Y^2 = 15.8^2$ y

$$s_p = \sqrt{\frac{15.1^2 + 15.8^2}{2}} = 15.4540$$

y el valor observado para el estadístico test sería

$$T_0 = \frac{28.8 - 32.6}{(15.4540)\sqrt{\frac{1}{40} + \frac{1}{40}}} = -1.0997$$

Por tanto, como $|T_0| = 1.10 < 1.99$ no podemos rechazar la hipótesis nula y no podemos concluir, con un nivel de significación $\alpha = 0.05$, que la cantidad media de luz sea diferente en ambos tipos de mesas. Utilizando el programa Statgraphics el p-valor del test sería:

$$p(|T| > |T_0|) = p(t_{78} > -1.0997) + p(t_{78} > 1.0997) = 2p(t_{78} > 1.0997) = 0.275$$

Por tanto rechazaríamos la hipótesis nula si el nivel de significación $\alpha > 0.275$, en particular para $\alpha = 0.05$ no podríamos rechazar la hipótesis como hemos calculado antes.

13. Los datos obtenidos en una prospección geológica del Departamento de Interior de los Estados Unidos presentan los caudales de un río pequeño en el norte de Florida. El interés está en comparar los caudales de marzo y abril, puesto que son dos meses relativamente secos. Tomadas 31 observaciones para el mes de marzo dieron como resultado $\bar{x} = 6.85$ con $s_1 = 1.2$ pies por segundo. Las 30

mediciones de abril dieron $\bar{y} = 7.47$ con $s_2 = 2.3$ pies por segundo. ¿Hay evidencia suficiente para decir que esos dos meses tuvieron diferentes caudales a un nivel de significación del 5 %? Calcular el p-valor que determinan los datos. ¿Te parece adecuada la hipótesis de igualdad de varianzas con estos datos? (responder utilizando un test de hipótesis con $\alpha = 0.05$ y calcular el p-valor de la prueba).

Solución. Consideramos las variables aleatorias X=caudal del río en el mes de marzo e Y=caudal del río en el mes de abril. En principio, suponemos que ambas variables tienen distribución normal con la misma varianza, es decir, que $X \rightsquigarrow N(\mu_X, \sigma)$ e $Y \rightsquigarrow N(\mu_Y, \sigma)$. Sean $X_1, X_2, ..., X_{31}$ e $Y_1, Y_2, ..., Y_{30}$ las respectivas muestras aleatorias simples, que consideramos independientes entre si. Queremos contrastar la hipótesis nula $H_o : \mu_X - \mu_Y = 0$ frente a $H_1 : \mu_X - \mu_Y \neq 0$ con un nivel de significación $\alpha = 0.05$. El estadístico test para este caso es

$$T = \frac{\overline{X} - \overline{Y}}{S_p \sqrt{\frac{1}{31} + \frac{1}{30}}} \rightsquigarrow t_{59}$$

siendo $\overline{X} = \frac{1}{31} \sum_{i=1}^{31} X_i$, $\overline{Y} = \frac{1}{30} \sum_{i=1}^{30} Y_i$ y

$$S_p = \sqrt{\frac{30S_X^2 + 29S_Y^2}{59}}$$

el estimador de la varianza común, con $S_X^2 = \left(\frac{31}{30}\right) \left[\left(\frac{1}{31}\right) \sum_{i=1}^{31} X_i^2 - \overline{X}^2\right]$ y $S_Y^2 = \left(\frac{30}{29}\right) \left[\left(\frac{1}{30}\right) \sum_{i=1}^{30} Y_i^2 - \overline{Y}^2\right]$. Como estamos haciendo el test de dos lados y queremos un nivel de significación $\alpha = 0.05$, necesitamos calcular $t_{59;0.025} = 2.000$ (hemos buscado en la tabla $t-$Student con $n = 60$ que es el más cercano). La región crítica para el test sería entonces $T \in (-\infty, -2) \cup (2, \infty)$, es decir, $|T| > 2$. Con los datos que proporciona el problema nosotros tenemos $\overline{x} = 6.85$, $\overline{y} = 7.47$, $s_X^2 = 1.2^2$ y $s_Y^2 = 2.3^2$ y

$$s_p = \sqrt{\frac{(30)(1.2^2) + (29)(2.3^2)}{59}} = 1.8255$$

y el valor observado para el estadístico test sería

$$T_0 = \frac{6.85 - 7.47}{(1.8255)\sqrt{\frac{1}{31} + \frac{1}{30}}} = -1.3261$$

Por tanto, como $|T_0| = 1.3261 < 2.99$ no podemos rechazar la hipótesis nula y no podemos concluir, con un nivel de significación $\alpha = 0.05$, que los caudales medios de los meses de marzo y abril sean significativamente distintos con un nivel de significación $\alpha = 0.05$. Utilizando el programa Statgraphics el p-valor del test sería:

$$p(|T| > |T_0|) = p(t_{59} > -1.3261) + p(t_{59} > 1.3261) = 2p(t_{59} > 1.3261) = 0.190$$

Para saber si la hipótesis de igualdad de varianzas es adecuada tendremos que contrastar la hipótesis nula $H_0 : \sigma_X^2/\sigma_Y^2 = 1$ frente a $H_1 : \sigma_X^2/\sigma_Y^2 \neq 1$ siendo σ_X^2 y σ_Y^2, las dos varianzas poblacionales en los meses de marzo y abril respectivamente. En este caso, el estadístico test sería

$$T = \frac{S_X^2}{S_Y^2} \rightsquigarrow F_{30,29}$$

siendo S_X^2 y S_Y^2 las varianzas muestrales corregidas utilizadas anteriormente. Como estamos haciendo el test de dos lados y queremos un nivel de significación $\alpha = 0.05$, necesitamos calcular los valores $F_{30,29;0.025}$ y $F_{30,29;0.975}$. Como no tenemos tablas de la distribución $F-$Fisher con $\alpha = 0.025$ calcularemos esos valores utilizando el programa Statgraphics, obteniendo los valores $F_{30,29;0.025} = 0.4802$ y $F_{30,29;0.975} = 2.0923$. La región crítica para el test sería entonces $C \in (0, 0.4802) \cup (2.0923, \infty)$. Con los datos que proporciona el problema nosotros tenemos $s_X^2 = 1.2^2$ y $s_Y^2 = 2.3^2$ y por tanto el valor del estadístico test sería

$$T_0 = \frac{1.2^2}{2.3^2} = 0.2722$$

Por tanto, como $T_0 = 0.2722 \in C$ debemos rechazar la hipótesis nula y concluir que las varianzas son distintas con un nivel de significación $\alpha = 0.05$. Por tanto no es adecuada la hipótesis de igualdad de varianzas. El p-valor del test se calcularía como

$$p\left(T < T_0\right) + p\left(T > 1/T_0\right)$$

Utilizando el programa Statgraphics tendremos

$$p\left(F_{30,29} < 0.2722\right) + p\left(F_{30,29} > 3.6738\right) = 0.0003 + 0.0004 = 0.0007$$

Por tanto rechazaríamos la hipótesis de igualdad de varianzas para cualquier nivel de significación por encima de 0.0007, en particular para $\alpha = 0.05$ que es el que nosotros hemos usado antes.

14. La retención del nitrógeno por el suelo es un aspecto importante en los métodos de cultivo, incluyendo el cultivo de bosques. Se compararon dos métodos de preparación de cepas para plantar pinos después de limpiar el terreno, analizando el porcentaje de nitrógeno marcado que se recupera. El método A deja intacto gran parte del piso del bosque, mientras que el método B remueve la mayor parte de la materia orgánica. Está claro que con el método B se recuperará una cantidad menor de nitrógeno del piso del bosque. El asunto de interés es si el método B hará que la retención de nitrógeno sea mayor en la biomasa microbiana en compensación por tener menos material orgánico disponible. El porcentaje de nitrógeno recuperado en la biomasa microbiana se midio en 6 cepas de prueba para cada método. Las cepas del método A mostraron un promedio de 12 y una desviación típica corregida de 1. Las cepas del método B mostraron un promedio 15 y una desviación típica corregida de 2. A un nivel de significación del 10 %, ¿se podría decir que el porcentaje promedio de nitrógeno recuperado es mayor para el método B? Calcular el p-valor del contraste utilizado para los datos obtenidos.

    Solución. Consideramos las variables aleatorias X=nitrógeno recuperado con el método A e Y=nitrógeno recuperado con el método B. Como los tamaños muestrales son pequeños y no podemos hacer un test asintótico, vamos a suponer que ambas variables tienen distribución normal con la misma varianza, es decir, que $X \rightsquigarrow N\left(\mu_X, \sigma\right)$ e $Y \rightsquigarrow N\left(\mu_Y, \sigma\right)$. Sean $X_1, X_2, ..., X_6$ e $Y_1, Y_2, ..., Y_6$ las respectivas muestras aleatorias simples, que consideramos independientes entre si. Necesitamos contrastar la hipótesis nula $H_o : \mu_X - \mu_Y = 0$ frente a $H_1 : \mu_X - \mu_Y < 0$ con un nivel de significación $\alpha = 0.10$.

El estadístico test para este caso es

$$T = \frac{\overline{X} - \overline{Y}}{S_p \sqrt{\frac{1}{6} + \frac{1}{6}}} \rightsquigarrow t_{10}$$

siendo $\overline{X} = \frac{1}{6} \sum_{i=1}^{6} X_i$, $\overline{Y} = \frac{1}{6} \sum_{i=1}^{6} Y_i$ y

$$S_p = \sqrt{\frac{5S_X^2 + 5S_Y^2}{10}} = \sqrt{\frac{S_X^2 + S_Y^2}{2}}$$

el estimador de la varianza común, con $S_X^2 = \left(\frac{6}{5}\right) \left[\left(\frac{1}{6}\right) \sum_{i=1}^{6} X_i^2 - \overline{X}^2\right]$ y $S_Y^2 = \left(\frac{6}{5}\right) \left[\left(\frac{1}{6}\right) \sum_{i=1}^{6} Y_i^2 - \overline{Y}^2\right]$. Como estamos haciendo el test de un lado y con nivel de significación $\alpha = 0.10$, necesitamos calcular $t_{10;0.10} = 1.372$. La región crítica para el test sería entonces $C \in (-\infty, -1.372)$, es decir, $T < -1.372$. Con los datos que proporciona el problema nosotros tenemos $\overline{x} = 15$, $\overline{y} = 12$, $s_X^2 = 1^2 = 1$ y $s_Y^2 = 2^2 = 4$ y

$$s_p = \sqrt{\frac{1+4}{2}} = 1.5811$$

y el valor observado para el estadístico test sería

$$T_0 = \frac{12 - 15}{(1.5811) \sqrt{\frac{1}{6} + \frac{1}{6}}} = -3.2863$$

Por tanto, como $T_0 = -3.2863 < -1.372$ podemos rechazar la hipótesis nula con un nivel de significación $\alpha = 0.05$, y concluir que el porcentaje promedio de nitrógeno recuperado es mayor para el método B. Utilizando el programa Statgraphics el p-valor del test sería:

$$p\,(T < T_0) = p\,(t_{10} < -3.2863) = 0.0041$$

15. En una revista se afirma que "las técnicas y los dispositivos de biorretroalimentación para controlar las funciones fisiológicas ayudan a los astronautas a controlar la tensión". Seis personas se sometieron a una situación de tensión (mediante videojuegos) seguidos por un periodo de biorretroalimentación y adaptación. Otro grupo de seis personas fueron sometidas a la misma tensión, y después se les dijo que simplemente se relajaran. El primer grupo obtuvo 70.4 latidos por minuto de promedio, con una desviación típica corregida de 15.3, mientras el segundo grupo tuvo un promedio de 74.9 latidos por minuto, con una desviación típica corregida de 16.0. ¿Se puede decir que los latidos cardíacos promedio con biorretroalimentación son menos rápidos que sin la biorretroalimentación? ¿Qué hipótesis se necesita asumir para que la respuesta sea válida? Calcular el p-valor del contraste.

    Solución. Consideramos las variables aleatorias X=latidos por minuto con periodo de biorretroalimentación y adaptación e Y=latidos por minuto sin periodo de biorretroalimentación y adaptación. Como los tamaños muestrales son muy pequeños y no podemos hacer un test asintótico, vamos a suponer que ambas variables tienen distribución normal con la misma varianza, es decir, que $X \rightsquigarrow N(\mu_X, \sigma)$ e $Y \rightsquigarrow N(\mu_Y, \sigma)$. Sean $X_1, X_2, ..., X_6$ e $Y_1, Y_2, ..., Y_6$ las respectivas muestras aleatorias simples, que consideramos independientes entre si. Necesitamos contrastar la hipótesis nula $H_o: \mu_X - \mu_Y = 0$ frente a $H_1: \mu_X - \mu_Y < 0$ para un nivel de significación cualquiera α. El estadístico test para este caso es

$$T = \frac{\overline{X} - \overline{Y}}{S_p\sqrt{\frac{1}{6} + \frac{1}{6}}} \rightsquigarrow t_{10}$$

siendo $\overline{X} = \frac{1}{6}\sum_{i=1}^{6} X_i$, $\overline{Y} = \frac{1}{6}\sum_{i=1}^{6} Y_i$ y

$$S_p = \sqrt{\frac{5S_X^2 + 5S_Y^2}{10}} = \sqrt{\frac{S_X^2 + S_Y^2}{2}}$$

el estimador de la varianza común, con $S_X^2 = \left(\frac{6}{5}\right)\left[\left(\frac{1}{6}\right)\sum_{i=1}^{6} X_i^2 - \overline{X}^2\right]$ y $S_Y^2 = \left(\frac{6}{5}\right)\left[\left(\frac{1}{6}\right)\sum_{i=1}^{6} Y_i^2 - \overline{Y}^2\right]$. Como estamos haciendo el test de un lado y con nivel de significación α, necesitamos calcular $t_{10;\alpha}$. La región crítica para el test sería entonces $C \in (-\infty, -t_{10;\alpha})$, es decir, $T < -t_{10;\alpha}$. Con los datos que proporciona el problema nosotros tenemos $\overline{x} = 70.4$, $\overline{y} = 74.9$, $s_X^2 = 15.3^2$ y $s_Y^2 = 16^2$ y

$$s_p = \sqrt{\frac{15.3^2 + 16^2}{2}} = 15.6539$$

y el valor observado para el estadístico test sería

$$T_0 = \frac{70.4 - 74.9}{(15.6539)\sqrt{\frac{1}{6} + \frac{1}{6}}} = -0.4979$$

Utilizando el programa Statgraphics podemos calcular el p-valor del test como:

$$p\,(T < T_0) = p\,(t_{10} < -0.4979) = p\,(t_{10} > 0.4979) = 0.3147$$

Por tanto podríamos concluir que los latidos cardíacos promedio con biorretroalimentación son menos rápidos que sin biorretroalimentación siempre y cuando el nivel de significación α deseado fuese mayor o igual que 0.3147. A los niveles de significación más habituales $(0.1, 0.05, 0.01)$ no podríamos afirmarlo.

16. Se probaron dos máquinas A y B para estudiar la resistencia a la torsión de un alambre de acero en 12 pares distintos de alambre. En cada máquina se probó un alambre de cada uno de los pares. Los resultados de la resistencia a la torsión (punto de ruptura) en cada par fueron los siguientes:

Alambre	1	2	3	4	5	6	7	8	9	10	11	12
Máquina A	32	35	38	28	40	42	36	29	33	37	22	42
Máquina B	30	34	39	26	37	42	35	30	30	32	20	41

a) ¿Hay significatividad a un nivel del 5 % de que las máquinas A y B dan lecturas diferentes?

b) ¿Hay significatividad a un nivel del 5 % de que la máquina B de una lectura menor que la máquina A?

En ambos casos calcular el p-valor para los datos obtenidos.

Solución. Consideramos las variables aleatorias X=resistencia a la torsión del alambre medida con máquina A e Y=resistencia a la torsión del alambre medida con máquina B. Como los tamaños muestrales son muy pequeños y no podemos hacer un test asintótico, vamos a suponer que ambas variables tienen distribución normal con la misma varianza, es decir, que $X \rightsquigarrow N(\mu_X, \sigma)$ e $Y \rightsquigarrow N(\mu_Y, \sigma)$. Sean $X_1, X_2, ..., X_{12}$ e $Y_1, Y_2, ..., Y_{12}$ las respectivas muestras aleatorias simples, que consideramos independientes entre si.

a) Necesitamos contrastar la hipótesis nula $H_o : \mu_X - \mu_Y = 0$ frente a $H_1 : \mu_X - \mu_Y \neq 0$ con un nivel de significación $\alpha = 0.05$. El estadístico test para este caso es

$$T = \frac{\overline{X} - \overline{Y}}{S_p\sqrt{\frac{1}{12} + \frac{1}{12}}} \rightsquigarrow t_{22}$$

siendo $\overline{X} = \frac{1}{12}\sum_{i=1}^{12} X_i$, $\overline{Y} = \frac{1}{12}\sum_{i=1}^{12} Y_i$ y

$$S_p = \sqrt{\frac{11S_X^2 + 11S_Y^2}{22}} = \sqrt{\frac{S_X^2 + S_Y^2}{2}}$$

el estimador de la varianza común, con $S_X^2 = \left(\frac{12}{11}\right)\left[\left(\frac{1}{12}\right)\sum_{i=1}^{12} X_i^2 - \overline{X}^2\right]$ y $S_Y^2 = \left(\frac{12}{11}\right)\left[\left(\frac{1}{12}\right)\sum_{i=1}^{12} Y_i^2 - \overline{Y}^2\right]$. Como estamos haciendo el test de dos lados con nivel de significación $\alpha = 0.05$, necesitamos calcular $t_{22;0.025} = 2.074$. La región crítica para el test sería entonces $C \in (-\infty, -2.074) \cup (2.074, \infty)$, es decir, $|T| > 2.074$. Con los datos que proporciona el problema nosotros tenemos $\sum x_i = 414$, $\sum x_i^2 = 14684$, $\sum y_i = 396$, $\sum y_i^2 = 13516$ y por tanto $\overline{x} = 34.5$, $\overline{y} = 33$, $s_X^2 = \left(\frac{12}{11}\right)\left[\frac{14684}{12} - 34.5^2\right] = 36.4545$,y $s_Y^2 = \left(\frac{12}{11}\right)\left[\frac{13516}{12} - 33^2\right] = 40.7273$ y

$$s_p = \sqrt{\frac{36.4545 + 40.7273}{2}} = 6.2122$$

y el valor observado para el estadístico test sería

$$T_0 = \frac{34.5 - 33}{(6.2122)\sqrt{\frac{1}{12} + \frac{1}{12}}} = 0.5915$$

Por tanto, como $|T_0| = 0.5915 < 2.074$ no podemos rechazar la hipótesis nula con un nivel de significación $\alpha = 0.05$. Es decir, no hay significatividad a un nivel del 5 %. Utilizando el programa Statgraphics el p-valor del test sería:

$$p(|T| > |T_0|) = p(t_{22} < -0.5915) + p(t_{22} > 0.5915) = 2p(t_{22} > 0.5915) = 0.5602$$

b) Ahora necesitamos contrastar la hipótesis nula $H_o : \mu_X - \mu_Y = 0$ frente a $H_1 : \mu_X - \mu_Y > 0$ con un nivel de significación $\alpha = 0.05$. El estadístico test sería el mismo que antes, pero ahora necesitaríamos calcular $t_{22;0.05} = 1.717$ y la región crítica sería $C \in (1.717, \infty)$, es decir, $T > 1.717$. Como en nuestro caso tenemos $T_0 = 0.5915 < 1.717$ tampoco ahora podemos rechazar la hipótesis nula y por tanto no hay significatividad, a un nivel de 5 %, de que la máquina B proporcione una lectura media menor que la máquina A. El p-valor del test sería ahora

$$p(T > T_0) = p(t_{22} > 0.5915) = 0.2801$$

Observar que el p-valor para el test de dos lados es el doble del que obtenemos para el test de un lado.

17. Supongamos que se está planeando un experimento en un invernadero para estudiar el crecimiento de las plantas de pimiento. Se plantan n de estas plantas en tierra normal, y otras n en tierra tratada. Después de 21 días se mide la longitud del tallo de cada una de las plantas. Si el efecto del tratamiento de la tierra es que se incrementa la longitud media de los tallos en 2 cm., en el correspondiente contraste se debería tener una probabilidad del 90 % de rechazar la hipótesis nula con el test de la t de Student de un lado. Los datos de un estudio previo en 15 plantas crecidas en tierra normal muestran una media muestral de 12.5 cm. y una desviación típica muestral corregida de 0.8 cm.

    a) Si se quiere utilizar un test a un nivel de significación $\alpha = 0.05$, ¿cuál es el valor de n que se debe tomar?

    b) ¿Qué condiciones deben cumplir los datos para que los cálculos del apartado anterior sean válidos?

    c) Si se decide adoptar una postura más conservadora y utilizar un nivel de significación $\alpha = 0.01$, ¿qué valor de n se debe utilizar?

    Solución. Consideramos las variables aleatorias X=longitud del tallo de una planta crecida en tierra normal (cm) e Y=longitud del tallo de una planta crecida en tierra tratada (cm). Necesitaremos suponer que ambas variables tienen distribución normal con la misma varianza, es decir, que $X \rightsquigarrow N(\mu_X, \sigma)$ e $Y \rightsquigarrow N(\mu_Y, \sigma)$. Sean $X_1, X_2, ..., X_n$ e $Y_1, Y_2, ..., Y_{15}$ las respectivas muestras aleatorias simples, que consideramos independientes entre si. Sabemos que para la muestra de Y de tamaño 15 hemos obtenido $\overline{y} = 12.5$ y $s_Y = 0.8$. Para demostrar que el tratamiento de la tierra incrementa la longitud media de los tallos en 2 cm. necesitaremos hacer un contraste de hipótesis con $H_o : \mu_X - \mu_Y = 0$ frente a $H_1 : \mu_X - \mu_Y > 0$. Con las hipótesis que hemos supuesto, sabemos que el estadístico test es

$$T = \frac{\overline{X} - \overline{Y}}{S_p\sqrt{\frac{1}{n} + \frac{1}{15}}} \rightsquigarrow t_{n+13}$$

siendo $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\overline{Y} = \frac{1}{15}\sum_{i=1}^{15} Y_i$ y

$$S_p = \sqrt{\frac{(n-1)S_X^2 + 14S_Y^2}{n+13}}$$

el estimador de la varianza común, con $S_X^2 = \left(\frac{n}{n-1}\right)\left[\left(\frac{1}{n}\right)\sum_{i=1}^{n} X_i^2 - \overline{X}^2\right]$ y $S_Y^2 = \left(\frac{15}{14}\right)\left[\left(\frac{1}{15}\right)\sum_{i=1}^{15} Y_i^2 - \overline{Y}^2\right]$. Como estamos haciendo el test de un lado con un nivel de significación α necesitamos calcular $t_{n+13;\alpha}$ y la región crítica será $T > t_{n+13;\alpha}$. Para cumplir la condición que nos piden necesitaremos que la probabilidad de la región crítica cuando $\mu_X - \mu_Y = 2$ sea de 0.90.

    a) Para responder a esta pregunta con un nivel de significación $\alpha = 0.05$ supondremos que el verdadero valor σ es el estimador obtenido con la muestra de tamaño 15 (es decir $\sigma = 0.8$) y que el tamaño muestral n es lo suficientemente grande para que $t_{n+13} \approx N(0, 1)$ (recordar que con $n > 30$ ya se puede aproximar $t_{n+13} \approx N(0,1)$). Entonces tendremos que $t_{n+13;0.05} \simeq z_{0.05} = 1.645$ y, si $\mu_X - \mu_Y = 2$, tendremos que

$$\frac{\overline{X} - \overline{Y} - 2}{0.8\sqrt{\frac{1}{n} + \frac{1}{15}}} \rightsquigarrow t_{n+13} \approx N(0,1)$$

Por tanto, para cumplir que la probabilidad de la región crítica cuando $\mu_X - \mu_Y = 2$ sea de 0.90 necesitaremos que

$$\begin{aligned} 0.90 &= p\left(\frac{\overline{X} - \overline{Y}}{0.8\sqrt{\frac{1}{n} + \frac{1}{15}}} > 1.645\right) = p\left(\frac{\overline{X} - \overline{Y} - 2}{0.8\sqrt{\frac{1}{n} + \frac{1}{15}}} > \frac{1.645 - 2}{0.8\sqrt{\frac{1}{n} + \frac{1}{15}}}\right) \\ &\simeq p\left(Z > \frac{1.645 - 2}{0.8\sqrt{\frac{1}{n} + \frac{1}{15}}}\right) \end{aligned}$$

o, equivalentemente,

$$p\left(Z > -\frac{1.645 - 2}{0.8\sqrt{\frac{1}{n} + \frac{1}{15}}}\right) = 0.10$$

Entonces, utilizando las tablas de la distribución normal estándar deberá verificarse que

$$-\frac{1.645 - 2}{0.8\sqrt{\frac{1}{n} + \frac{1}{15}}} = 1.28$$

y por tanto, despejando n tendremos:

$$\frac{1}{n} = \left(\frac{0.355}{(0.8)\,(1.28)}\right)^2 - \frac{1}{15} = 0.0535$$

es decir $n = 1/0.0535 = 18.68$. Por tanto necesitaríamos una muestra de tamaño al menos $n = 19$.

b) Como acabamos de decir necesitaríamos $X \rightsquigarrow N(\mu_Y, 0.8)$ y admitir que $t_{32} \approx N(0,1)$.

c) Si ahora queremos usar $\alpha = 0.01$ en lugar de $\alpha = 0.05$ tendríamos que $t_{n+13;0.01} \simeq z_{0.01} = 2.32$ y la condición sería

$$-\frac{2.32 - 2}{0.8\sqrt{\frac{1}{n} + \frac{1}{15}}} = 1.28$$

Entonces, despejando n tendríamos que

$$\frac{1}{n} = \left(\frac{-0.32}{(0.8)\,(1.28)}\right)^2 - \frac{1}{15} = 0.03099$$

es decir $n = 1/0.03099 = 32.27$. Por tanto necesitaríamos una muestra de tamaño al menos $n = 33$.

18. El siguiente problema trata sobre la eficiencia en la combustión que cabe esperar de un calentador de petróleo. El *Environmental News* (enero de 1977) afirma que el 80 % o más es excelente, del 75 % al 79 % bueno, del 70 % al 74 % mediano y menos del 70 % malo. Un contratista de calefacción que vende dos tipos de calentadores (A y B) decidió comparar las eficiencias de estos dos tipos de calentadores. Los datos obtenidos en una muestra de cada tipo son los siguientes:

Tipo A:	72	78	73	69	75	74	69	75
Tipo B:	78	76	81	74	82	75		

Contrastar la diferencia en la eficiencia promedio de ambos tipos de calentador, suponiendo que en ambos casos se tiene una distribución normal y utilizando un nivel de significación $\alpha = 0.05$.

Solución. Consideramos las variables aleatorias X=eficiencia en la combustión de un calentador tipo A e Y=eficiencia en la combustión de un calentador tipo B. Suponemos que ambas variables tienen distribución normal con la misma varianza, es decir, que $X \rightsquigarrow N(\mu_X, \sigma)$ e $Y \rightsquigarrow N(\mu_Y, \sigma)$. Sean $X_1, X_2, ..., X_8$ e $Y_1, Y_2, ..., Y_6$ las respectivas muestras aleatorias simples, que consideramos independientes entre si. Necesitamos contrastar la hipótesis nula $H_o : \mu_X - \mu_Y = 0$ frente a $H_1 : \mu_X - \mu_Y \neq 0$ con un nivel de significación $\alpha = 0.05$. El estadístico test para este caso es

$$T = \frac{\overline{X} - \overline{Y}}{S_p\sqrt{\frac{1}{8} + \frac{1}{6}}} \rightsquigarrow t_{12}$$

siendo $\overline{X} = \frac{1}{8}\sum_{i=1}^{8} X_i$, $\overline{Y} = \frac{1}{6}\sum_{i=1}^{6} Y_i$ y

$$S_p = \sqrt{\frac{7S_X^2 + 5S_Y^2}{12}}$$

el estimador de la varianza común, con $S_X^2 = \left(\frac{8}{7}\right)\left[\left(\frac{1}{8}\right)\sum_{i=1}^{8} X_i^2 - \overline{X}^2\right]$ y $S_Y^2 = \left(\frac{6}{5}\right)\left[\left(\frac{1}{6}\right)\sum_{i=1}^{6} Y_i^2 - \overline{Y}^2\right]$. Como estamos haciendo el test de dos lados con nivel de significación $\alpha = 0.05$, necesitamos calcular $t_{12;0.025} = 2.179$. La región crítica para el test sería entonces $C \in (-\infty, -2.179) \cup (2.179, \infty)$, es decir, $|T| > 2.179$. Con los datos que proporciona el problema nosotros tenemos $\sum x_i = 585$, $\sum x_i^2 = 42845$, $\sum y_i = 466$, $\sum y_i^2 = 36246$ y por tanto $\overline{x} = 73.125$, $\overline{y} = 77.667$, $s_X^2 = \left(\frac{8}{7}\right)\left[\frac{42845}{8} - 73.125^2\right] = 9.5536$, $s_Y^2 = \left(\frac{6}{5}\right)\left[\frac{36246}{6} - 77.667^2\right] = 10.667$ y

$$s_p = \sqrt{\frac{(7)(9.5536) + (5)(10.667)}{12}} = 3.1650$$

y el valor observado para el estadístico test sería

$$T_0 = \frac{73.125 - 77.667}{(3.1650)\sqrt{\frac{1}{8} + \frac{1}{6}}} = -2.6570$$

Por tanto, como $|T_0| = 2.6570 > 2.179$ podemos rechazar la hipótesis nula con un nivel de significación $\alpha = 0.05$. y concluir que la eficiencia media es distinta en los dos tipos de calentadores. Utilizando el programa Statgraphics el p-valor del test sería:

$$p(|T| > |T_0|) = p(t_{12} < -2.6570) + p(t_{12} > 2.6570) = 2p(t_{12} > 0.6570) = 0.0209$$

19. Para comprobar la potencia en cuanto a vitamina D de un aceite de hígado de pescado, se alimentó a 20 polluelos de un día con un alimento al que le fue añadido un 1 % de este aceite como única fuente de vitamina D. Otros 10 polluelos fueron alimentados con el mismo alimento base, pero añadiendo un 1 % de un aceite de hígado de bacalao (estándar) del que se conoce su potencia en cuanto a la vitamina D. Después de 8 días los pesos de los polluelos (en gramos) fueron los siguientes:

Aceite estándar: 69, 66, 71, 69, 68, 71, 74, 73, 70, 69

Nuevo aceite: 68, 69, 66, 64, 67, 69, 65, 68, 70, 65, 63, 66, 68, 66, 69, 67, 66, 61, 71, 72

Contrastar la hipótesis de que el nuevo aceite es mejor utilizando un nivel de significación $\alpha = 0.05$ y obtener el p-valor que proporcionan los datos (suponer distribuciones normales).

Solución. Consideramos las variables aleatorias X=peso de un polluelo alimentado con el aceite estándar (gr) e Y=peso de un polluelo alimentado con el nuevo aceite (gr). Suponemos que ambas variables tienen distribución normal con la misma varianza, es decir, que $X \rightsquigarrow N(\mu_X, \sigma)$ e $Y \rightsquigarrow N(\mu_Y, \sigma)$. Sean $X_1, X_2, ..., X_{10}$ e $Y_1, Y_2, ..., Y_{20}$ las respectivas muestras aleatorias simples, que consideramos independientes entre si. Necesitamos contrastar la hipótesis nula $H_o : \mu_X - \mu_Y = 0$ frente a $H_1 : \mu_X - \mu_Y < 0$ con un nivel de significación $\alpha = 0.05$. El estadístico test para este caso es

$$T = \frac{\overline{X} - \overline{Y}}{S_p\sqrt{\frac{1}{10} + \frac{1}{20}}} \rightsquigarrow t_{28}$$

siendo $\overline{X} = \frac{1}{10}\sum_{i=1}^{10} X_i$, $\overline{Y} = \frac{1}{20}\sum_{i=1}^{20} Y_i$ y

$$S_p = \sqrt{\frac{9S_X^2 + 19S_Y^2}{28}}$$

el estimador de la varianza común, con $S_X^2 = \left(\frac{10}{9}\right)\left[\left(\frac{1}{10}\right)\sum_{i=1}^{10} X_i^2 - \overline{X}^2\right]$ y $S_Y^2 = \left(\frac{20}{19}\right)\left[\left(\frac{1}{20}\right)\sum_{i=1}^{20} Y_i^2 - \overline{Y}^2\right]$. Como estamos haciendo el test de un lado con nivel de significación $\alpha = 0.05$, necesitamos calcular $t_{28;0.05} = 1.701$. La región crítica para el test sería entonces $C \in (-\infty, -1.701)$, es decir, $T < -1.701$. Con los datos que proporciona el problema nosotros tenemos $\sum x_i = 700$, $\sum x_i^2 = 49050$, $\sum y_i = 1340$, $\sum y_i^2 = 89918$ y por tanto $\overline{x} = 70$, $\overline{y} = 67$, $s_X^2 = \left(\frac{10}{9}\right)\left[\frac{49050}{10} - 70^2\right] = 5.5556$, $s_Y^2 = \left(\frac{20}{19}\right)\left[\frac{89918}{20} - 67^2\right] = 7.2632$ y

$$s_p = \sqrt{\frac{(9)(5.5556) + (19)(7.2632)}{28}} = 2.5912$$

y el valor observado para el estadístico test sería

$$T_0 = \frac{70 - 67}{(2.5912)\sqrt{\frac{1}{10} + \frac{1}{20}}} = 2.9893$$

Por tanto, como $T_0 = 2.9893 > -1.701$ no podemos rechazar la hipótesis nula con un nivel de significación $\alpha = 0.05$. No hay evidencia de que el nuevo aceite produzca un mayor peso medio en los polluelos (observar que de hecho el peso medio es menor con el nuevo aceite). Utilizando el programa Statgraphics el p-valor del test sería:

$$p(T < T_0) = p(t_{28} < 2.9873) = 0.9971$$

20. La cantidad de oxígeno que puede inspirar una persona representa la mayor cantidad de oxígeno que puede consumir dicha persona. Se llevó a cabo un test de fatiga para determinar la capacidad

pulmonar de 9 niñas en edad escolar antes y después de un programa de 10 semanas de ejercicio físico. Los resultados expresados en ml. de oxígeno por minuto y por kg. de peso fueron:

Individuo	1	2	3	4	5	6	7	8	9
Antes	48.6	38.0	31.2	45.5	41.7	41.8	37.9	39.2	47.2
Después	38.8	40.7	32.0	45.4	43.2	45.3	38.9	43.5	45.0

Contrastar la hipótesis de que existe diferencia entre las cantidades de oxígeno consumida “Antes” y “Después” del test de fatiga con un nivel de significación $\alpha = 0.05$ y obtener el p-valor que proporcionan los datos (suponer distribuciones normales).

Solución. Consideramos las variables aleatorias X=cantidad de oxígeno inspirado antes del programa (ml por minuto y kg) e Y=cantidad de oxígeno inspirado después del programa (ml por minuto y kg). Suponemos que ambas variables tienen distribución normal, es decir, que $X \rightsquigarrow N(\mu_X, \sigma_X)$ e $Y \rightsquigarrow N(\mu_Y, \sigma_Y)$. Sean $X_1, X_2, ..., X_9$ e $Y_1, Y_2, ..., Y_9$ las respectivas muestras aleatorias simples, que ahora no son independientes entre si por tratarse de dos muestras pareadas medidas sobre las mismas 9 niñas. Necesitamos contrastar la hipótesis nula $H_o : \mu_X - \mu_Y = 0$ frente a $H_1 : \mu_X - \mu_Y \neq 0$ con un nivel de significación $\alpha = 0.05$. El estadístico test para este caso es

$$T = \frac{\overline{X} - \overline{Y}}{S_D\sqrt{\frac{1}{9}}} \rightsquigarrow t_8$$

siendo $\overline{X} = \frac{1}{9}\sum_{i=1}^{9} X_i$, $\overline{Y} = \frac{1}{9}\sum_{i=1}^{9} Y_i$ y S_D la desviación típica muestral corregida de las diferencias $X_i - Y_i$. Como estamos haciendo el test de dos lados con nivel de significación $\alpha = 0.05$, necesitamos calcular $t_{8;0.025} = 2.306$. La región crítica para el test sería entonces $C \in (-\infty, -2.306) \cup (2.306, \infty)$, es decir, $|T| > 2.306$. Con los datos que proporciona el problema nosotros tenemos $\sum x_i = 371.1$, $\sum y_i = 372.8$, y por tanto $\overline{x} = 41.23$, $\overline{y} = 41.42$. Los valores de las diferencias $X_i - Y_i$ en nuestra muestra son

Individuo	1	2	3	4	5	6	7	8	9
Diferencia	9.8	-2.7	-0.8	0.1	-1.5	-3.5	-1	-4.3	2.2

y entonces tenemos $\sum d_i = -1.7$, $\sum d_i^2 = 142.81$, $\overline{d} = \overline{x} - \overline{y} = -0.1889$ y

$$s_D = \sqrt{\left(\frac{9}{8}\right)\left[\frac{142.81}{9} - 0.1889^2\right]} = 4.2203$$

Entonces el valor observado para el estadístico test será

$$T_0 = \frac{-0.1889}{4.2203/3} = -0.1343$$

Por tanto, como $|T_0| = 0.1343 < 2.306$ no podemos rechazar la hipótesis nula con un nivel de significación $\alpha = 0.05$. No hay evidencia de que exista diferencia entre las cantidades medias de oxígeno consumidas “Antes” y “Después” del test de fatiga con un nivel de significación $\alpha = 0.05$. Utilizando el programa Statgraphics el p-valor del test sería:

$$p(|T| > 0.1343) = 2p(t_8 > 0.1343) = 0.8965$$

21. Los siguientes datos muestran 30 medidas del voltaje eléctrico tomadas durante un cierto proceso industrial. El proceso se considera satisfactorio cuando el voltaje es mayor que 9.2 voltios (las lecturas más grandes son mejores que las más pequeñas).

Vieja localización			Nueva localización		
9.98	10.12	9.84	9.19	10.01	8.82
10.26	10.05	10.15	9.63	8.82	8.65
10.05	9.80	10.02	10.10	9.43	8.51
10.29	10.15	9.80	9.70	10.03	9.14
10.03	10.00	9.73	10.09	9.85	9.75
8.05	9.87	10.01	9.60	9.27	8.78
10.55	9.55	9.98	10.05	8.83	9.35
10.26	9.95	8.72	10.12	9.39	9.54
9.97	9.70	8.80	9.49	9.48	9.36
9.87	8.72	9.84	9.37	9.64	8.68

Contrastar la hipotesis de si hay diferencia entre la nueva y la vieja localización con un nivel de significación $\alpha = 0.01$ y calcular el p-valor que proporcionan los datos. ¿Es mejor la nueva localización que la vieja?

Solución. Consideramos las variables aleatorias X=medida del voltaje eléctrico en la vieja localización e Y=medida del voltaje eléctrico en la nueva localización. Como el problema no dice nada acerca de la distribución de probabilidad de estas variables pero el tamaño muestral es grande ($n = 30$) vamos a plantear un test de hipótesis asintótico que no necesita suponer que las variables tienen distribución normal. Consideramos $\mu_X = EX$, $\mu_Y = EY$, $\sigma_X^2 = Var(X)$ y $\sigma_Y^2 = Var(Y)$. Sean $X_1, X_2, ..., X_{30}$ e $Y_1, Y_2, ..., Y_{30}$ las respectivas muestras aleatorias simples, que consideramos independientes entre si. Necesitamos contrastar la hipótesis nula $H_o : \mu_X - \mu_Y = 0$ frente a $H_1 : \mu_X - \mu_Y \neq 0$ con un nivel de significación $\alpha = 0.01$. El estadístico test (asintótico) para este caso es

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_X^2}{30} + \frac{S_Y^2}{30}}} \widetilde{\rightsquigarrow} N(0,1)$$

siendo $\overline{X} = \frac{1}{30}\sum_{i=1}^{30} X_i$, $\overline{Y} = \frac{1}{30}\sum_{i=1}^{30} Y_i$, $S_X^2 = \left(\frac{30}{29}\right)\left[\left(\frac{1}{30}\right)\sum_{i=1}^{30} X_i^2 - \overline{X}^2\right]$ y $S_Y^2 = \left(\frac{30}{29}\right)\left[\left(\frac{1}{30}\right)\sum_{i=1}^{30} Y_i^2 - \overline{Y}^2\right]$. Como estamos haciendo el test de dos lados con nivel de significación $\alpha = 0.01$, necesitamos calcular $z_{0.005} = 2.58$. La región crítica para el test sería entonces $C \in (-\infty, -2.58) \cup (2.58, \infty)$, es decir, $|T| > 2.58$. Con los datos que proporciona el problema nosotros tenemos $\sum x_i = 294.11$, $\sum x_i^2 = 2891.84$, $\sum y_i = 282.67$, $\sum y_i^2 = 2670.06$ y por tanto $\overline{x} = 9.8037$, $\overline{y} = 9.4223$, $s_X^2 = \left(\frac{30}{29}\right)\left[\frac{2891.84}{30} - 9.8037^2\right] = 0.2926$, $s_Y^2 = \left(\frac{30}{29}\right)\left[\frac{2670.06}{30} - 9.4223^2\right] = 0.2293$. Entonces el valor observado para el estadístico test sería

$$T_0 = \frac{9.8037 - 9.4223}{\sqrt{\frac{0.2926+0.2293}{30}}} = 2.8912$$

Por tanto, como $|T_0| = 2.8912 > 2.58$ podemos rechazar la hipótesis nula con un nivel de significación $\alpha = 0.01$ y concluir que el voltaje medio es distinto en las dos localizaciones. Además, el p-valor del test sería:

$$p\left(|T| > |T_0|\right) = p\left(Z < -2.8912\right) + p\left(Z > 2.8912\right) = 2p\left(Z > 2.8912\right) = 2\left(0.00193\right) = 0.0039$$

No parece ser que la nueva localización sea mejor que la vieja porque la media muestral es más grande en la vieja que en la nueva, y el objetivo es un mayor voltaje. De hecho, si hiciésemos el test de hipótesis $H_o : \mu_X - \mu_Y = 0$ frente a $H_1 : \mu_X - \mu_Y > 0$ con un nivel de significación $\alpha = 0.01$, tendríamos que utilizar $z_{0.01} = 2.33$ y la región crítica sería $T > 2.33$. Como en nuestro caso tenemos $T_0 = 2.8912 > 2.33$, también podríamos rechazar la hipótesis nula en este test con un nivel de significación $\alpha = 0.01$, y por tanto concluir que el voltaje medio es significativamente más alto en la vieja localización que en la nueva, que es justo lo contrario de lo que se planteaba.